2

Vectors and Vector Spaces

There are several conceptual levels to the understanding of vector spaces and linear algebra. First, there is the mechanical interpretation of the term *vector* as it is often taught in physics and mechanics courses. This implies a magnitude and direction, usually with clear physical meaning, such as the magnitude and direction of the velocity of a particle. This is often an idea restricted to the two and three dimensions of physical space perceptible to humans. In this case, the familiar operations of dot product and cross product have physical implications, and can be easily understood through the mental imagery of projections and right-hand rules.

Then there is the idea of the vector as an *n*-tuple of numbers, usually arranged as a column, such as often presented in lower-level mathematics classes. Such columns are given the same properties of magnitude and direction but are not constrained to three dimensions. Further, they are subject to more generalized operators in the form of a matrix. Matrix-vector multiplication can also be treated very mechanically, as when students are first taught to find solutions to simultaneous algebraic equations through gaussian elimination or by using matrix inverses.[M] `inv(A)`

These simplified approaches to vector spaces and linear algebra are valuable tools that serve their purposes well. In this chapter we will be concerned with more generalized treatments of linear spaces. As such, we wish to present the properties of vector spaces with somewhat more abstraction, so that the applications of the theory can be applied to broad classes of problems. However, because the simplified interpretations are familiar, and they are certainly consistent with the generalized concept of a linear space, we will not discard them. Instead, we will use simple mechanical examples of vector spaces to launch the discussions of more general settings. In this way, one can transfer existing intuition into the abstract domain for which intuition is often difficult.

2.1 Vectors

To begin with, we appeal to the intuitive sense of a vector, i.e., as a magnitude and direction, as in the two-dimensional case shown below.

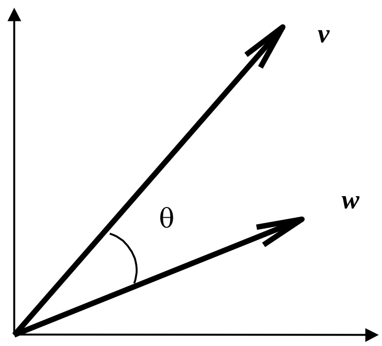

Figure 2.1 The concept of a vector with magnitude and direction. Such vectors can be easily pictured in two and three dimensions only.

In Figure 2.1, vectors ***v*** and ***w*** are depicted as arrows, the length of which indicates their magnitude, with directions implied by the existence of the reference arrows lying horizontally and vertically. It is clear that the two vectors have an angle between them (_), but in order to uniquely fix them in “space,” they should also be oriented with respect to these references. The discussion of these reference directions is as important as the vectors themselves.

2.1.1 Familiar Euclidean Vectors

We begin the discussion of the most basic properties of such vectors with a series of definitions, which we first present in familiar terms but which may be generalized later. We give these as a reminder of some of the simpler operations we perform on vectors that we visualize as “arrows” in space.

`dot(x,y)`

Inner (dot) productM***:*** The inner product of two vectors, ***v*** and ***w***, is denoted as $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$. Although there may exist many definitions for the computation of an inner product, when vectors are interpreted as columns of n real numbers, the inner product is customarily computed as:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \boldsymbol{v}^T \boldsymbol{w} = \boldsymbol{w}^T \boldsymbol{v} = \sum_{i=1}^{n} v_i w_i \tag{2.1}$$

where v_i and w_i are the individual elements of the vectors. In terms of the magnitudes of the vectors (defined below), this is also sometimes given the definition:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \|\boldsymbol{v}\| \, \|\boldsymbol{w}\| \cos\theta \tag{2.2}$$

This gives a geometric relationship between the size and direction of the vectors and their inner product.

Norm (magnitude): The norm[M] of a vector $\boldsymbol{v}$, physically interpreted as its magnitude, can be "induced" from this definition of the inner product above as:

`norm(x)`

$$\|\boldsymbol{v}\| = \langle v, v \rangle^{1/2} = \sqrt{\sum_{i=1}^{n} v_i^2} \tag{2.3}$$

As we will see, there may be many different norms. Each different inner product may produce a different norm. When applied to *n*-tuples, the most common is the euclidean, as given in (2.3) above.

Outer (tensor) product: The outer product of two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ is defined as:

$$\boldsymbol{v}\rangle\langle\boldsymbol{w} = \boldsymbol{v}\boldsymbol{w}^{\mathrm{T}} = -\boldsymbol{w}\boldsymbol{v}^{\mathrm{T}} \tag{2.4}$$

We will not use this definition much in this text.

Cross (vector) product[M]***:*** This product produces a vector quantity from two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$.

`cross(x,y)`

$$\|\boldsymbol{v} \times \boldsymbol{w}\| = \|\boldsymbol{v}\| \, \|\boldsymbol{w}\| \sin\theta \tag{2.5}$$

where the resulting vector has a new direction that is perpendicular to the plane of the two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ and that is generated by the so-called "right-hand rule." As such, this operation is generally used in three dimensions only. Again, it will have limited usefulness for us.

2.2 Vector Spaces

Given these familiar definitions, we turn now to more abstract definitions of vectors. Our concept of vectors will retain the properties of inner products and norms as defined above, but we will consider vectors consisting of quantities that appear to be very different from *n*-tuples of numbers, in perhaps infinite dimensions. Nevertheless, analogies to the intuitive concepts above will survive.

2.2.1 Fields

First, we cannot introduce the idea of a vector space without first giving the definition of a *field*. Conceptually, a field is the set from which we select elements that we call *scalars*. As the name implies, scalars will be used to scale vectors.

Field: A field consists of a set of two or more elements, or members, which must include the following:

1. There must exist in the field a unique element called 0 (zero). For $0 \in F$ and any other element $a \in F$, $0(a) = 0$ and $a + 0 = a$.
2. There must exist in the field another unique element called 1 (one). For $1 \in F$ and $a \in F$, $1(a) = a(1) = (a/1) = a$.
3. For every $a \in F$, there is a unique element called its negative, $-a \in F$, such that $a + (-a) = 0$.

A field must also provide definitions of the operations of addition, multiplication, and division. There is considerable flexibility in the definition of these operations, but in any case the following properties must hold:

1. If $a \in F$ and $b \in F$, then $(a+b) = (b+a) \in F$. That is, the sum of any two vectors in a field is also in the same field. This is known as *closure under addition*.
2. If $a \in F$ and $b \in F$, then $(ab) = (ba) \in F$. That is, their product remains in the field. This is known as *closure under multiplication*.
3. If $a \in F$ and $b \in F$, and if $b \neq 0$, then $a/b \in F$.

Finally, for the addition and multiplication operations defined, the usual associative, commutative, and distributive laws apply. (2.6)

Example 2.1: Candidate Fields

Determine which of the following sets of elements constitute fields, using elementary arithmetic notions of addition, multiplication, and division (as we are familiar with them).*

1. The set of real numbers $\{0, 1\}$
2. The set of all real matrices of the form $\begin{bmatrix} x & -y \\ y & x \end{bmatrix}$ where x and y are real numbers
3. The set of all polynomials in s
4. The set of all real numbers
5. The set of all integers

Solutions:

1. No. This cannot be a field, because $1+1=2$, and the element 2 is not a member of the set.
2. Yes. This set has the identity matrix and the zero matrix, and inverse matrices of the set have the same form. However, this set has the special property of being commutative under multiplication, which is not true of matrices in general. Therefore, the set of all 2×2 matrices is not a field.
3. No. The inverse of a polynomial is not usually a polynomial.
4. Yes. This is the most common field that we encounter.
5. No. Like polynomials, the inverse of an integer is not usually an integer.

Now we can proceed to define vector spaces. The definition of a vector space is dependent on the field over which we specify it. We therefore must refer to *vector spaces over fields*.

---

* Such arithmetic operations need not generally follow our familiar usage. For instance, in Example 1, the answer will vary if the set is interpreted as binary digits with binary operators rather than the two real numbers 0 and 1.

Linear Vector Space: A *linear vector space* **X** is a collection of elements called *vectors*, defined over a field *F*. Among these vectors must be included:

1. A vector $\mathbf{0} \in \mathbf{X}$ such that $\boldsymbol{x} + \mathbf{0} = \mathbf{0} + \boldsymbol{x} = \boldsymbol{x}$.
2. For every vector $\boldsymbol{x} \in \mathbf{X}$, there must be a unique vector $\boldsymbol{y} \in \mathbf{X}$ such that $\boldsymbol{x} + \boldsymbol{y} = \mathbf{0}$. This condition is equivalent to the existence of a negative element for each vector in the vector space, so that $\boldsymbol{y} = -\boldsymbol{x}$.

As with a field, operations on these elements must satisfy certain requirements. (In the rules that follow, the symbols $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are elements of the space **X**, and symbols *a* and *b* are elements of the field *F*.) The requirements are:

1. *Closure* under addition: If $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{v}$, then $\boldsymbol{v} \in \mathbf{X}$.
2. *Commutativity* of addition: $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$.
3. *Associativity* of addition: $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$.
4. *Closure* under scalar multiplication: For every $\boldsymbol{x} \in \mathbf{X}$ and scalar $a \in F$, the product $a\boldsymbol{x}$ gives another vector $\boldsymbol{y} \in \mathbf{X}$. Scalar *a* may be the *unit* scalar, so that $a\boldsymbol{x} = 1 \cdot \boldsymbol{x} = \boldsymbol{x} \cdot 1 = \boldsymbol{x}$.
5. *Associativity* of scalar multiplication: For any scalars $a, b \in F$, and for any vector $\boldsymbol{x} \in \mathbf{X}$, $a(b\boldsymbol{x}) = (ab)\boldsymbol{x}$.
6. *Distributivity* of scalar multiplication over vector addition:

$$\begin{aligned} (a+b)\boldsymbol{x} &= a\boldsymbol{x} + b\boldsymbol{x} \\ a(\boldsymbol{x} + \boldsymbol{y}) &= a\boldsymbol{x} + a\boldsymbol{y} \end{aligned} \tag{2.7}$$

Example 2.2: Candidate Vector Spaces

Determine whether the following sets constitute vector spaces when defined over the associated fields.

1. The set of all *n*-tuples of scalars from any field *F*, defined over *F*. For example, the set of *n*-tuples $\Re^n$ over the field of reals $\Re$, or the set of complex *n*-tuples $\mathbf{C}^n$ over the field of complex numbers $\mathbf{C}$.
2. The set of complex numbers over the reals.

3. The set of real numbers over the complex numbers.
4. The set of all $m \times n$ matrices, over the reals. The same set over complex numbers.
5. The set of all piecewise continuous functions of time, over the reals.
6. The set of all polynomials in s of order less than n, with real coefficients, over $\Re$.
7. The set of all symmetric matrices over the real numbers.
8. The set of all nonsingular matrices.
9. The set of all solutions to a particular linear, constant-coefficient, finite-dimensional homogeneous differential equation.
10. The set of all solutions to a particular linear, constant-coefficient, finite-dimensional *non*-homogeneous differential equation.

Solutions:

1. Yes. These are the so-called "euclidean spaces," which are most familiar. The intuitive sense of vectors as necessary for simple mechanics fits within this category.
2. Yes. Note that a complex number, when multiplied by a real number, is still a complex number.
3. No. A real number, when multiplied by a complex number, is not generally real. This violates the condition of closure under scalar multiplication.
4. Yes in both cases. Remember that such matrices are not generally fields, as they have no unique inverse, but they do, as a collection, form a space.
5. Yes.
6. Yes. Again, such a set would not form a field, but it does form a vector space. There is no requirement for division in a vector space.
7. Yes. The sum of symmetric matrices is again a symmetric matrix.
8. No. One can easily find two nonsingular matrices that add to form a singular matrix.
9. Yes. This is an important property of solutions to differential equations.
10. No. If one were to add the particular solution of a nonhomogeneous differential equation to itself, the result would not be a solution to the same differential equation, and thus the set would not be closed under addition. Such a set is not closed under scalar multiplication either.

With the above definitions of linear vector spaces and their vectors, we are ready for some definitions that relate vectors to each other. An important concept is the *independence* of vectors.

2.2.2 Linear Dependence and Independence

It may be clear from the definitions and examples above that a given vector space will have an infinite number of vectors in it. This is because of the closure rules, which specify that any multiple or sum of vectors in a vector space must also be in the space. However, this fact does not imply that we are doomed to manipulating large numbers of vectors whenever we use vector spaces as descriptive tools. Instead, we are often as interested in the directions of the vectors as we are in the magnitudes, and we can gather vectors together that share a common set of directions.

In future sections, we will learn to decompose vectors into their exact directional components. These components will have to be independent of one another so that we can create categories of vectors that are in some way similar. This will lead us to the concepts of bases and dimension. For now, we start with the concepts of *linear dependence* and *independence*.

Linear Dependence: Consider a set of n vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset \mathbf{X}$. Such a set is said to be *linearly dependent* if there exists a set of scalars $\{a_i\}$, $i = 1, \ldots, n$, *not all of which are zero*, such that

$$a_1 \boldsymbol{x}_1 + a_2 \boldsymbol{x}_2 + \cdots + a_n \boldsymbol{x}_n = \sum_{i=1}^{n} a_i \boldsymbol{x}_i = 0 \tag{2.8}$$

The sum on the left side in this equation is referred to as a *linear combination*.

Linear Independence: If the linear combination shown above, $\sum_{i=1}^{n} a_i \boldsymbol{x}_i = 0$, requires that *all* of the coefficients $\{a_i\}$, $i = 1, \ldots, n$ be zero, then the set of vectors $\{\boldsymbol{x}_i\}$ is linearly independent. (2.9)

If we consider the vectors $\{\boldsymbol{x}_i\}$ to be columns of numbers, we can use a more compact notation. Stacking such vectors side-by-side, we can form a matrix X as follows:

$$X = \left[\boldsymbol{x}_1 \;\vdots\; \boldsymbol{x}_2 \;\vdots\; \cdots \;\vdots\; \boldsymbol{x}_3 \right]$$

Now considering the set of scalars a_i to be similarly stacked into a column called a (this does not mean they constitute a vector), then the condition for linear independence of the vectors $\{\boldsymbol{x}_i\}$ is that the equation $Xa = \mathbf{0}$ has only the trivial solution $a = 0$.

Linear Independence of Functions

When the vectors themselves are functions, such as $\boldsymbol{x}_i(t)$ for $i = 1, \ldots, n$, then we can define linear independence on intervals of *t*. That is, if we can find scalars a_i, *not all zero*, such that the linear combination

$$a_1 \boldsymbol{x}_1(t) + \cdots + a_n \boldsymbol{x}_n(t) = 0$$

for all $t \in [t_0, t_1]$, then the functions $\{\boldsymbol{x}_1(t), \ldots, \boldsymbol{x}_n(t)\}$ are linearly dependent in that interval. Otherwise they are independent. Independence outside that interval, or for that matter, in a subinterval, is not guaranteed. Furthermore, if $\boldsymbol{X}(t) = [\boldsymbol{x}_1(t) \quad \cdots \quad \boldsymbol{x}_n(t)]^{\mathrm{T}}$ is an $n \times 1$ vector of functions, then we can define the *Gram matrix* (or *grammian*) of $\boldsymbol{X}(t)$ as the $n \times n$ matrix:

$$G_{\boldsymbol{X}}(t_1, t_2) = \int_{t_1}^{t_2} \boldsymbol{X}(t) \boldsymbol{X}^*(t)\, dt$$

It can be shown (see Problem 2.26) that the functions $\boldsymbol{x}_i(t)$ are linearly independent if and only if the *Gram determinant* is nonzero, i.e., $|G_{\boldsymbol{X}}(t_1, t_2)| \neq 0$.

Example 2.3: Linear Dependence of Vectors

Consider the set of three vectors from the space of real *n*-tuples defined over the field of reals:

$$\boldsymbol{x}_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \qquad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \qquad \boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 7 \\ 8 \end{bmatrix}$$

This is a linearly dependent set because we can choose the set of *a*-coefficients as

$a_1 = -1$, $a_2 = 2$, and $a_3 = -1$. Clearly not all (indeed, none) of these scalars is zero, yet

$$\sum_{i=1}^{3} a_i \mathbf{x}_i = -\mathbf{x}_1 + 2\mathbf{x}_2 - \mathbf{x}_3 = \mathbf{0}$$

det(A)

Some readers may be familiar with the test from matrix theory that allows one to determine the linear independence of a collection of vectors from the determinant[M] formed by the matrix X as we constructed it above. That is, the dependence of the vectors in the previous example could be ascertained by the test

$$\det(X) = \begin{vmatrix} 2 & 1 & 0 \\ -1 & 3 & 7 \\ 0 & 4 & 8 \end{vmatrix} = 0$$

If this determinant were nonzero, then the vectors would have been found linearly independent.

However, some caution should be used when relying on this test, because the number of components in the vector may not be equal to the number of vectors in the set under test, so that a nonsquare matrix would result. Nonsquare matrices have no determinant defined for them.* We will also be examining vector spaces that are not such simple n-tuples defined over the real numbers. They therefore do not form such a neat determinant. Furthermore, this determinant test does not reveal the underlying geometry that is revealed by applying the definition of linear dependence. The following example illustrates the concept of linear independence with a different kind of vector space.

Example 2.4: Vectors of Rational Polynomials

The set $R(s)$ of all rational polynomial functions in s is a vector space over the field of real numbers $\Re$. It is also known to be a vector space over the field of rational polynomials themselves. Consider two such vectors of the space of *ordered pairs* of such rational polynomials:

---

* In such a situation, one can examine all the square submatrices that can be formed from subsets of the rows (or columns) of the nonsquare matrix. See the problems at the end of the chapter for examples.

$$\mathbf{x}_1 = \begin{bmatrix} \dfrac{1}{s+1} \\ \\ \dfrac{1}{s+2} \end{bmatrix} \qquad \mathbf{x}_2 = \begin{bmatrix} \dfrac{s+2}{(s+1)(s+3)} \\ \\ \dfrac{1}{s+3} \end{bmatrix}$$

If the chosen field is the set of all real numbers, then this set of two vectors is found to be linearly independent. One can verify that if a_1 and a_2 are real numbers, then setting

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = a_1 \begin{bmatrix} \dfrac{1}{s+1} \\ \\ \dfrac{1}{s+2} \end{bmatrix} + a_2 \begin{bmatrix} \dfrac{s+2}{(s+1)(s+3)} \\ \\ \dfrac{1}{s+3} \end{bmatrix}$$

will imply that $a_1 = a_2 = 0$. See Problem 2.3 at the end of this chapter.

If instead the field is chosen as the set of rational polynomials, then we have an entirely different result. Through careful inspection of the vectors, it can be found that if the scalars are chosen as

$$a_1 = 1 \qquad \text{and} \qquad a_2 = -\frac{s+3}{s+2}$$

then

$$\sum_{i=1}^{2} a_i \mathbf{x}_i = \begin{bmatrix} \dfrac{1}{s+1} \\ \\ \dfrac{1}{s+2} \end{bmatrix} + \begin{bmatrix} -\dfrac{1}{s+1} \\ \\ -\dfrac{1}{s+2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The vectors are now seen to be linearly dependent. Note that this is not a matter of the *same* two vectors being dependent where they once were independent. We prefer to think of them as entirely *different vectors*, because they were selected from entirely different spaces, as specified over two different fields.

This particular example also raises another issue. When testing for linear independence, we should check that equality to zero of the linear combinations is true *identically*. The vectors shown are themselves functions of the variable *s*. Given a set of linearly independent vectors, there may therefore be isolated values for s that might produce the condition $\sum a_i \boldsymbol{x}_i = 0$ even though the vectors are independent. We should not hastily mistake this for linear dependence, as this condition is not true identically, i.e., for *all s*.

Geometrically, the concept of linear dependence is a familiar one. If we think of two-dimensional euclidean spaces that we can picture as a plane, then any two collinear vectors will be linearly dependent. To show this, we can always scale one vector to have the negative magnitude of the other. This scaled sum would of course equal zero. Two linearly independent vectors will have to be noncollinear. Scaling with nonzero scalars and adding them will never produce zero, because scaling them cannot change their directions.

In three dimensions, what we think of as a volume, we can have three linearly independent vectors if they are noncoplanar. This is because if we add two scaled vectors, then the result will of course lie in the plane that is formed by the vectors. If the third vector is not in that same plane, then no amount of nonzero scaling will put it there. Thus there is no possibility of scaling it so that it is the negative sum of the first two. This geometric interpretation results in a pair of lemmas that are intuitively obvious and can be easily proven.

> LEMMA: If we have a set of linearly dependent vectors, and we add another vector to this set, the resulting set will also have to be linearly dependent. (2.10)

This is obvious from the geometric description we have given. If the original set of vectors is dependent, it adds to zero with at least one nonzero scaling coefficient. We can then include the new vector into the linear combination with a zero coefficient, as in Figure 2.2. At least one of the other scalars is already known to be nonzero, and the new sum will remain zero. Thus, the augmented set is still dependent.

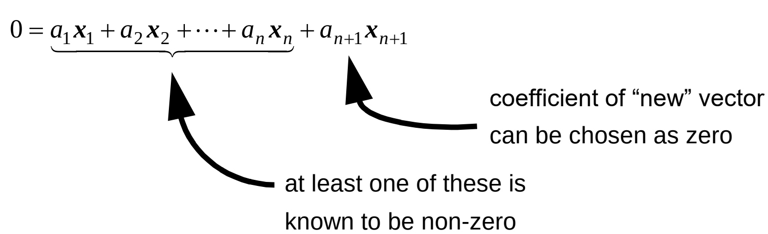

Figure 2.2 Manipulations necessary to show linear dependence of an augmented set of already dependent vectors.

The other lemma is:

LEMMA: If a given set of vectors is linearly dependent, then one of the vectors can be written as a linear combination of the other vectors. (2.11)

This lemma is sometimes given as the definition of a linearly dependent set of vectors. In fact, it can be derived as a result of our definition.

PROOF: If the set $\{\mathbf{x}_i\}$ is linearly dependent, then $\sum_{i=1}^{n} a_i \mathbf{x}_i = 0$ with at least one *a* coefficient not equal to zero. Suppose that $a_j \neq 0$. Then without risking a divide-by-zero, we can perform the following operation:

$$\begin{aligned} \mathbf{x}_j &= \frac{-a_1\mathbf{x}_1 - a_2\mathbf{x}_2 - \cdots - a_{j-1}\mathbf{x}_{j-1} - a_{j+1}\mathbf{x}_{j+1} - \cdots - a_n\mathbf{x}_n}{a_j} \\ &= -\frac{a_1}{a_j}\mathbf{x}_1 - \frac{a_2}{a_j}\mathbf{x}_2 - \cdots - \frac{a_{j-1}}{a_j}\mathbf{x}_{j-1} - \frac{a_{j+1}}{a_j}\mathbf{x}_{j+1} - \cdots - \frac{a_n}{a_j}\mathbf{x}_n \end{aligned} \quad (2.12)$$

By defining $b_i \stackrel{\Delta}{=} -a_i / a_j$,

$$\mathbf{x}_j = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \cdots + b_n\mathbf{x}_n \quad (2.13)$$

thus proving that one vector can be written as a linear combination of the others.

Already, our intuitive sense is that we can have (at most) as many independent vectors in a set as we have “dimensions” in the space. This is strictly true, although we will need to rigorously show this after discussion of the mathematical notion of *dimension*.

Dimension: The *dimension* of a linear vector space is the largest possible number of linearly *in*dependent vectors that can be taken from that space. (2.14)

2.2.3 Bases

If we are working within a particular vector space and we select the maximum number of linearly independent vectors, the set we will have created is known as a *basis*. Officially, we have the definition:

> ***Basis:*** A set of linearly independent vectors in vector space **X** is a *basis* of **X** if and only if every vector in **X** can be written as a *unique* linear combination of vectors from this set. (2.15)

One must be careful to note that in an *n*-dimensional vector space, there must be exactly *n* vectors in any basis set, but there are an infinite number of such sets that qualify as a basis. The main qualification on these vectors is that they be linearly independent and that there be a maximal number of them, equal to the dimension of the space. The uniqueness condition in the definition above results from the fact that if the basis set is linearly independent, then each vector contains some unique "direction" that none of the others contain. In geometric terms, we usually think of the basis as being the set of coordinate axes. Although coordinate axes, for our convenience, are usually orthonormal, i.e., mutually orthogonal and of unit length, basis vectors need not be.

THEOREM: In an *n*-dimensional linear vector space, *any* set of *n* linearly *in*dependent vectors qualifies as a basis. (2.16)

PROOF: This statement implies that a vector $\boldsymbol{x}$ should be described uniquely by any *n* linearly independent vectors, say

$$\{\boldsymbol{e}_i\} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$$

That is, for every vector $\boldsymbol{x}$,

$$\begin{aligned}\boldsymbol{x} &= \text{linear combination of } \boldsymbol{e}_i\text{'s} \\ &= \sum_{i=1}^{n} \alpha_i \boldsymbol{e}_i\end{aligned} \tag{2.17}$$

Because the space is *n*-dimensional, the set of $n+1$ vectors $\{\boldsymbol{x}, \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ must be linearly dependent. Therefore, there exists a set of scalars $\{\alpha_i\}$, not all of which are zero, such that

$$\alpha_0 \boldsymbol{x} + \alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n = \boldsymbol{0} \tag{2.18}$$

Suppose $\alpha_0 = 0$. If this were the case and if the set $\{\boldsymbol{e}_i\}$ is, as specified, linearly independent, then we must have

$\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. To avoid this trivial situation, we assume that $\alpha_0 \neq 0$. This would allow us to write the equation:

$$\begin{aligned} \boldsymbol{x} &= -\frac{\alpha_1}{\alpha_0}\boldsymbol{e}_1 - \frac{\alpha_2}{\alpha_0}\boldsymbol{e}_2 - \cdots - \frac{\alpha_n}{\alpha_0}\boldsymbol{e}_n \\ &\equiv \beta_1\boldsymbol{e}_1 + \beta_2\boldsymbol{e}_2 + \cdots + \beta_n\boldsymbol{e}_n \\ &= \sum_{i=1}^{n} \beta_i \boldsymbol{e}_i \end{aligned} \tag{2.19}$$

So we have written vector $\boldsymbol{x}$ as a linear combination of the $\boldsymbol{e}_i$'s.

Now we must show that this expression is *unique*. To do this, suppose there were *another* set of scalars $\{\bar{\beta}_i\}$ such that $\boldsymbol{x} = \sum_{i=1}^{n} \bar{\beta}_i \boldsymbol{e}_i$. We already have $\boldsymbol{x} = \sum_{i=1}^{n} \beta_i \boldsymbol{e}_i$, so

$$\begin{aligned} \boldsymbol{x} - \boldsymbol{x} &= \sum_{i=1}^{n} \bar{\beta}_i \boldsymbol{e}_i - \sum_{i=1}^{n} \beta_i \boldsymbol{e}_i \\ &= \sum_{i=1}^{n} \left(\bar{\beta}_i - \beta_i\right)\boldsymbol{e}_i \\ &= 0 \end{aligned} \tag{2.20}$$

But the set $\{\boldsymbol{e}_i\}$ is known to be a basis; therefore, for the above equality to hold, we must have $\bar{\beta}_i - \beta_i = 0$ for all $i = 1, \ldots, n$. So $\bar{\beta}_i = \beta_i$ and the uniqueness of the representation is proven.

Once the basis $\{\boldsymbol{e}_i\}$ is chosen, the set of numbers $\{\beta_i\}$ is called the *representation* of *x* in $\{\boldsymbol{e}_i\}$. We can then refer to vector $\boldsymbol{x}$ by this representation, which is simply an *n*-tuple of numbers $\beta = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_n \end{bmatrix}^{\mathrm{T}}$. This is a very important by-product of the theorem, which implies that any finite dimensional vector space is *isomorphic** to the space of n-tuples. This allows us to always use the n-tuple operations such as dot-products and the induced norm on arbitrary spaces. It also allows us to rely on the intuitive picture of a vector space as a set of coordinate axes. For these reasons, it is often much more convenient to use the representation of a vector than to use the vector itself. However, it should be remembered that the same vector will have different representations in different bases.

There is weaker terminology to describe the expansion of a vector in terms of vectors from a particular set.

> ***Span:*** A set of vectors X is *spanned* by a set of vectors $\{\boldsymbol{x}_i\}$ if every $\boldsymbol{x} \in X$ can be written as a linear combination of the $\boldsymbol{x}_i$'s. Equivalently, the $\boldsymbol{x}_i$'s *span* X. The notation is $X = \text{sp}\{\boldsymbol{x}_i\}$. (2.21)

Note that in this terminology, the set $\{\boldsymbol{x}_i\}$ is not necessarily a basis. It may be linearly dependent. For example, five noncollinear vectors in two dimensions suffice to span the two dimensional euclidean space, but as a set they are not a basis. We sometimes use this definition when defining subspaces, which will be discussed later.

We will see after the following examples how one can take a single vector and convert its representation from one given basis to another.

Example 2.5: Spaces and Their Bases

1. Let **X** be the space of all vectors written $\boldsymbol{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^{\mathrm{T}}$ such that $x_1 = x_2 = \cdots = x_n$. One legal basis for this space is the single vector $\boldsymbol{v} = [1 \quad 1 \quad \cdots \quad 1]^{\mathrm{T}}$. This is therefore a one-dimensional space, despite it consisting of vectors containing n “components.”
2. Consider the space of all polynomials in s of degree less than 4, with real coefficients, defined over the field of reals. One basis for this space is the set

$$\boldsymbol{e}_1 = 1 \qquad \boldsymbol{e}_2 = s \qquad \boldsymbol{e}_3 = s^2 \qquad \boldsymbol{e}_4 = s^3$$

In this basis, the vector

* Spaces are said to be isomorphic if there exists a bijective mapping (i.e., one-to-one and onto) from one to the other.

$$\boldsymbol{x} = 3s^3 + 2s^2 - 2s + 10$$

can obviously be expanded as

$$\boldsymbol{x} = 3\boldsymbol{e}_4 + 2\boldsymbol{e}_3 + (-2)\boldsymbol{e}_2 + 10\boldsymbol{e}_1$$

So the vector $\boldsymbol{x}$ has the representation $\boldsymbol{x} = \begin{bmatrix}10 & -2 & 2 & 3\end{bmatrix}^{\mathrm{T}}$ in the $\{\boldsymbol{e}_i\}$ basis. We can write another basis for this same space:

$$\{\boldsymbol{f}_i\} = \{\boldsymbol{f}_1,\ \boldsymbol{f}_2, \boldsymbol{f}_3,\ \boldsymbol{f}_4\} = \left\{s^3 - s^2,\ s^2 - s,\ s - 1, 1\right\}$$

In this different basis, the same vector has a different representation:

$$\boldsymbol{x} = 3(s^3 - s^2) + 5(s^2 - s) + 3(s - 1) + 13$$

or

$$\boldsymbol{x} = \begin{bmatrix} 3 \\ 5 \\ 3 \\ 13 \end{bmatrix}$$

This can be verified by direct expansion of this expression.

Example 2.6: Common Infinite-Dimensional Spaces

While at first the idea of infinite-dimensional spaces may seem unfamiliar, most engineers have had considerable experience with function spaces of infinite dimension. For the following spaces, give one or more examples of a valid basis set.

1. The space of all functions of time that are analytic in the interval $a < t < b$.
2. The space of all bounded periodic functions of period T with at most a finite number of finite discontinuities and a finite number of extrema.

Solution:

The first function space can be described with power series expansions because of the use of the key word *analytic*. Therefore, for example, we can write a Taylor series expansion

$$\boldsymbol{f}(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{d^n \boldsymbol{f}(t)}{dt^n}\bigg|_{t=t_0} (t-t_0)^n \tag{2.22}$$

which suggests the bases $\boldsymbol{e}_i(t) = \{(t-t_0)^i\}$ for $i = 0, 1, \cdots$, or $\boldsymbol{e}_i(t) = \{1, (t-t_0)^1, (t-t_0)^2, \cdots\}$. There are, as usual, an infinite number of such valid bases (consider the MacLaurin series).

The second set of functions is given by the conditions that guarantee a convergent Fourier series. The Fourier expansion of any such function can be written either as

$$\boldsymbol{f}(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left[a_n \cos\left(\tfrac{2n\pi}{T} t\right) + b_n \sin\left(\tfrac{2n\pi}{T} t\right)\right] \tag{2.23}$$

or

$$\boldsymbol{f}(t) = \frac{1}{2} \sum_{n=-\infty}^{+\infty} c_n e^{j(2n\pi/T)t} \tag{2.24}$$

for appropriately calculated coefficients a_0, a_n, b_n, and c_n (where $j = \sqrt{-1}$). Therefore, two suitable bases for these function spaces are

$$\boldsymbol{e}_i = \left\{1, \cos\left(\tfrac{2\pi}{T}\right), \sin\left(\tfrac{2\pi}{T}\right), \cos\left(\tfrac{4\pi}{T}\right), \sin\left(\tfrac{4\pi}{T}\right), \cdots\right\} \tag{2.25}$$

or

$$\boldsymbol{f}_i = \left\{1, e^{j(2\pi/T)t}, e^{j(4\pi/T)t}, \cdots\right\} \tag{2.26}$$

2.2.4 Change of Basis

Suppose a vector $\boldsymbol{x}$ has been given in a basis $\{\boldsymbol{v}_j\}$, $j = 1, \ldots, n$. We are then given a different basis, consisting of vectors $\{\hat{\boldsymbol{v}}_i\}$, $i = 1, \ldots, n$, coming from the same space. Then the same vector $\boldsymbol{x}$ can be expanded into the two bases with different representations:

$$\boldsymbol{x} = \sum_{j=1}^{n} x_j \boldsymbol{v}_j = \sum_{i=1}^{n} \hat{x}_i \hat{\boldsymbol{v}}_i \tag{2.27}$$

But since the basis $\{\boldsymbol{v}_j\}$ consists of vectors in a common space, these vectors themselves can be expanded in the new basis $\{\hat{\boldsymbol{v}}_i\}$ as:

$$\boldsymbol{v}_j = \sum_{i=1}^{n} b_{ij}\hat{\boldsymbol{v}}_i \tag{2.28}$$

By gathering the vectors on the right side of this equation side-by-side into a matrix, this expression can also be expressed as

$$\boldsymbol{v}_j = \begin{bmatrix} \hat{\boldsymbol{v}}_1 & \hat{\boldsymbol{v}}_2 & \cdots & \hat{\boldsymbol{v}}_n \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}$$

or, when all *j*-values are gathered in a single notation,

$$\begin{aligned} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} &= \begin{bmatrix} \hat{\boldsymbol{v}}_1 & \hat{\boldsymbol{v}}_2 & \cdots & \hat{\boldsymbol{v}}_n \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \hat{\boldsymbol{v}}_1 & \hat{\boldsymbol{v}}_2 & \cdots & \hat{\boldsymbol{v}}_n \end{bmatrix} B \end{aligned} \tag{2.29}$$

where the matrix *B* so defined in this equation will hereafter be referred to as the "change of basis matrix."

Substituting the expression (2.28) into Equation (2.27),

$$\sum_{j=1}^{n} x_j \left[\sum_{i=1}^{n} b_{ij}\hat{\boldsymbol{v}}_i \right] = \sum_{i=1}^{n} \hat{x}_i \hat{\boldsymbol{v}}_i \tag{2.30}$$

By changing the order of summation on the left side of this equality, moving both terms to the left of the equal sign, and factoring,

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{n} b_{ij} x_j - \hat{x}_i \right) \hat{\boldsymbol{v}}_i = 0 \tag{2.31}$$

For this to be true given that $\{\hat{\boldsymbol{v}}_i\}$ is an independent set, we must have each constant in the above expansion being zero, implying

$$\hat{x}_i = \sum_{j=1}^{n} b_{ij} x_j \tag{2.32}$$

This is how we get the components of a vector in a new basis from the components of the old basis. The coefficients b_{ij} in the expansion come from our knowledge of how the two bases are related, as determined from Equation (2.29). Notice that this relationship can be written in vector-matrix form by expressing the two vectors as n-tuples, $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$. Denoting by B the $n \times n$ matrix with coefficient b_{ij} in the i^{th} row and j^{th} column, we can write

$$\hat{\boldsymbol{x}} = B\boldsymbol{x} \tag{2.33}$$

Example 2.7: Change of Basis

Consider the space $\Re^2$ and the two bases:

$$\{\boldsymbol{e}_1, \boldsymbol{e}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2\} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix} \right\}$$

Let a vector $\boldsymbol{x}$ be represented by $\boldsymbol{x} = [2 \quad 2]^{\mathrm{T}}$ in the $\{\boldsymbol{e}_j\}$ basis. To find the representation in the $\{\hat{\boldsymbol{e}}_i\}$ basis, we must first write down the relationship between the two basis sets. We do this by expressing the vectors $\boldsymbol{e}_j$ in terms of vectors $\hat{\boldsymbol{e}}_i$:

$$\boldsymbol{e}_1 = 0\hat{\boldsymbol{e}}_1 + (-1)\hat{\boldsymbol{e}}_2 = [\hat{\boldsymbol{e}}_1 \quad \hat{\boldsymbol{e}}_2] \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$\boldsymbol{e}_2 = \tfrac{1}{2}\hat{\boldsymbol{e}}_1 + \tfrac{1}{2}\hat{\boldsymbol{e}}_2 = [\hat{\boldsymbol{e}}_1 \quad \hat{\boldsymbol{e}}_2] \begin{bmatrix} \tfrac{1}{2} \\ \tfrac{1}{2} \end{bmatrix}$$

From this expansion we can extract the matrix

$$B = \begin{bmatrix} 0 & \tfrac{1}{2} \\ -1 & \tfrac{1}{2} \end{bmatrix}$$

giving, in the $\{\hat{\boldsymbol{e}}_i\}$ basis,

$$\hat{\boldsymbol{x}} = B\boldsymbol{x} = \begin{bmatrix} 0 & \tfrac{1}{2} \\ -1 & \tfrac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

REMARK: The reader might verify that the expansion of the $\hat{e}_i$ vectors in terms of the $\boldsymbol{e}_j$ vectors is considerably easier to write by inspection:

$$\hat{\boldsymbol{e}}_1 = 1\boldsymbol{e}_1 + 2\boldsymbol{e}_2 = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\hat{\boldsymbol{e}}_2 = (-1)\boldsymbol{e}_1 + 0\boldsymbol{e}_2 = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Therefore, the inverse of the change-of-basis matrix is more apparent:

$$B^{-1} = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}$$

Note that the columns of this matrix are simply equal to the columns of coefficients of the basis vectors $\{\hat{\boldsymbol{e}}_i\}$ because the $\{\boldsymbol{e}_j\}$ basis happens to be the same as the standard basis, but this is not always the case. We can display the relationship between the two basis sets and the vector $\boldsymbol{x}$ as in Figure 2.3.

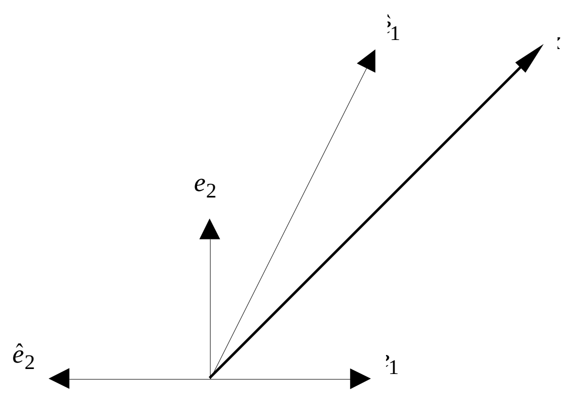

Figure 2.3 Relative positions of vectors in Example 2.7. It can be seen that $\boldsymbol{x} = 2\boldsymbol{e}_1 + 2\boldsymbol{e}_2$ and that $\boldsymbol{x} = 1 \cdot \hat{\boldsymbol{e}}_1 - 1 \cdot \hat{\boldsymbol{e}}_2$.

2.2.5 Rank and Degeneracy

rank(A)

In matrix theory, the *rank*ᴹ of a matrix A, denoted $r(A)$, is defined as the size of the largest nonzero determinant that can be formed by taking square subsets of the rows and columns of A. Given our knowledge of bases and representations, we can construct matrices from collections of vectors by considering each column to be the representation of a different vector (in the same basis):

$$A = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix} = \left[\begin{array}{c:c:c:c} x_{11} & x_{12} & & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & & x_{mn} \end{array}\right] \tag{2.34}$$

With this interpretation of a matrix, the concept of rank can be reconsidered.

> ***Rank:*** The *rank* of a matrix A, $r(A)$, is the maximum number of linearly independent columns in the matrix. This will be the same number as the maximum number of linearly independent rows in the matrix. That is, the *row* and *column* rank of a matrix are identical. (2.35)

Note that this definition for rank is consistent with the test for linear independence of a collection of vectors. Therefore, it is apparent also that the rank is the dimension of the space formed by the columns of the matrix (the *range* space, see Chapter 3). If the collection of columns forms a matrix with a nonzero determinant, then the set of vectors is linearly independent. Of course, this test only works with a number of vectors equal to the number of coefficients in the representation (which is equal to the dimension of the space). If the number of vectors is different from the dimension of their space, then all square subsets of the rows and columns of the matrix must be separately tested with the determinant test. If matrix A is of dimension $m \times n$, then of course, $r(A) \le \min(m, n)$. Furthermore, if an $m \times n$ matrix has $r(A) = \min(m, n)$, it is said to be of "full rank." If an $m \times n$ matrix has $r(A) < \min(m, n)$, it is said to be "rank deficient," or "degenerate." If the number of vectors considered is the same as the dimension of the space, so that the matrix A is square, then the matrix is also called "singular." A square matrix that is of *full* rank is "nonsingular."

> ***Nullity***: The *nullity* of an $m \times n$ matrix A is denoted by $q(A)$ and is defined as
>
> $$q(A) = n - r(A) \tag{2.36}$$

> NOTE: This is also sometimes referred to as the *degeneracy*, but this is not accurate terminology. It is possible for a full-rank matrix (i.e., not degenerate) to have a nonzero nullity, even though its degeneracy is zero. For example, consider a 3×5 matrix with rank 3. The degeneracy of this matrix is zero [$rank = \min(m,n)$], but its nullity is nonzero [$q(A) = n - r(A) = 5 - 3 = 2$]. For a *square* matrix, the degeneracy and nullity are equivalent.

In addition, there are some useful properties for the rank and nullity of products of matrices, given by the following theorem.

> THEOREM: If matrix A is $m\times n$ and matrix B is $n\times p$, and we form the $m\times p$ matrix product $AB = C$, then the following properties hold:
>
> $$\begin{gathered} r(A) + r(B) - n \le r(C) \le \min(r(A), r(B)) \\ q(C) \le q(A) + q(B) \end{gathered} \tag{2.37}$$
>
> [The second line can be derived from the first, using the definition of nullity (2.36)]. Furthermore, if D is an $m\times m$ nonsingular matrix and E is an $n\times n$ nonsingular matrix, then
>
> $$r(DA) = r(A) \qquad \text{and} \qquad r(AE) = r(A) \tag{2.38}$$

2.2.6 Inner Products

In the beginning of this chapter we presented the definitions of some inner products and norms that are commonly used in euclidean n-space, $\Re^n$. Now that more general vector spaces have been introduced, we are prepared for more general definitions of these operators on vectors.

> ***Inner Product:*** An *inner product* is an operation on two vectors, producing a scalar result. The inner product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is denoted $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ and must have the following properties:
>
> 1. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}$ (where the overbar indicates complex conjugate). (2.39)
> 2. $\langle \boldsymbol{x}, \alpha\boldsymbol{y}_1 + \beta\boldsymbol{y}_2 \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{y}_1 \rangle + \beta\langle \boldsymbol{x}, \boldsymbol{y}_2 \rangle$. (2.40)

3. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ for all $\boldsymbol{x}$, and $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = 0$. (2.41)

THEOREM: The following properties are true for inner products and can be derived from the three properties in the definition of the inner product:

4. $\langle \boldsymbol{x}, \alpha \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. (2.42)

5. $\langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle = \bar{\alpha} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. (2.43)

Vector spaces with inner products defined for them are referred to as *inner product spaces*. One can verify that the inner products given at the beginning of this chapter for $\Re^n$ are valid inner products as defined above. Note, however, that if the vectors are defined over a complex field, then the conjugate operator as shown in the conditions must be observed. For example, the inner product of two complex-valued vectors can be written $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^* \boldsymbol{y}$, where $\boldsymbol{x}^*$ is the complex-conjugate transpose.

Example 2.8: Inner Product for a Function Space

Consider the space of complex-valued functions of time defined on the interval $t \in [a, b]$, consisting of vectors of the form $\boldsymbol{x}(t)$. An inner product for this space might be:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \int_a^b \boldsymbol{x}^*(t) \boldsymbol{y}(t)\, dt \qquad (2.44)$$

2.2.7 Norms

A norm generalizes our concept of "length" or "magnitude" of a vector.

`norm(x)`

Norm: A *norm*[M] is a function of a single vector that produces a scalar result. For a vector $\boldsymbol{x}$, it is denoted $\|\boldsymbol{x}\|$ and must satisfy the following rules:

1. $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = 0$. (2.45)

2. $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$ for any scalar α. (2.46)

3. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ ("triangle inequality"). (2.47)

4. $\left|\langle \boldsymbol{x}, \boldsymbol{y} \rangle\right| \le \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|$ ("Cauchy-Schwarz inequality"). (2.48)

Vector spaces with such norms defined for them are referred to as *normed linear spaces*.

Norms are sometimes induced from the given inner product. For example, the euclidean norm given earlier in this chapter is induced by whatever inner product is defined:

$$\|\boldsymbol{x}\| = \langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2} \tag{2.49}$$

A *unit vector* is one that has unit length, i.e., whose norm is 1. Of course, any vector can be converted to a unit vector ("normalized") by dividing it by its own norm. This will not change its direction, only scale its magnitude.

Example 2.9: Norms

One can use the properties listed above to verify that each of the following is a valid norm for vectors (considered as *n*-tuples, $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$). These are known as the "ℓ_k norms":

$$\|\boldsymbol{x}\|_k = \sqrt[k]{\sum_{i=1}^{n} x_i^{\,k}} \tag{2.50}$$

of which the following common norms are special cases:

1. ℓ_1 norm: $$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i| \tag{2.51}$$

2. ℓ_2 norm: $$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \tag{2.52}$$

   (i.e., the euclidean norm). If the vector is a signal vector, $\boldsymbol{x}(t)$, the ℓ_2 norm becomes:

$$\|\boldsymbol{x}(t)\|_2 = \sqrt{\int_{-\infty}^{\infty} \boldsymbol{x}^*(t) \cdot \boldsymbol{x}(t)\, dt} \tag{2.53}$$

3. ℓ_∞ norm: $$\|\boldsymbol{x}\|_\infty = \max_i |x_i| \tag{2.54}$$

2.2.8 Some Other Terms

Until now we have referred to spaces consisting of vectors of *n*-tuples as *euclidean n-spaces*. This is actually a general term that refers to any real finite-dimensional linear space that has an inner product defined for it. A similar space consisting of complex vectors with an inner product satisfying the rules given above is known as a *unitary space*. In this section we will give definitions for other terms, some of which are not critical for our purposes, but which are nevertheless mathematically very important and quite common.

> ***Metric:*** A *metric* is a function of two vectors and gives a scalar measure of the distance between two vectors. For two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, it is denoted as $\rho(\boldsymbol{x}, \boldsymbol{y})$ and is defined as:
>
> $$\rho(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| \tag{2.55}$$
>
> A space with such a defined metric is referred to as *metric space*.

Sometimes we wish to define the distance between vectors independently of the lengths of the vectors involved. Such a concept is available in the form of the *angle* between two vectors.

> ***Angle:*** The *angle* between two vectors is denoted by $\theta(\boldsymbol{x}, \boldsymbol{y})$ and satisfies the following equation:
>
> $$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \cos\theta \tag{2.56}$$

In low dimensions (two- and three-dimensional spaces) and using the euclidean norm, this concept of angle is exactly what one might imagine, yet it is entirely valid for higher dimensions and other norms as well. In addition, this concept of angle extends to sets and subspaces[M] (see Section 2.4) of vectors as well, taking as the angle between two sets X and Y the *minimum* of all angles between pairs of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, where $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$.

`subspace(A,B)`

> ***Hilbert Space:*** A Hilbert space is an inner-product euclidean space with the following additional conditions:
>
> 1. The space is *infinite dimensional*.
> 2. The space is *complete* with respect to the metric $\rho(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$.

> A space is said to be *complete* if all Cauchy sequences converge to an element within the space. A *Cauchy sequence* is a sequence of elements $\{\boldsymbol{x}_n\}$ that satisfies the following criterion: Given any $\varepsilon > 0$, there exists an integer $N(\varepsilon)$ such that $\rho(\boldsymbol{x}_m, \boldsymbol{x}_n) < \varepsilon$ for all $m, n > N(\varepsilon)$ (see [4]). (2.57)

All euclidean spaces are complete, and we will not generally be concerned with the completeness of the spaces we discuss in this book. We will therefore not elaborate on this definition further.

> ***Banach Space:*** A *Banach space* is a normed linear space that is complete according to the definition of completeness given for Hilbert spaces above. (2.58)

> ***Orthogonality and Orthonormality:*** Two vectors are said to be *orthogonal* if their inner product is zero, i.e., $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. A pair of vectors is *orthonormal* if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ and, in addition, $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$. (2.59)

A *set* of vectors $\{\boldsymbol{x}_i\}$ is *orthogonal* if

$$\begin{aligned} \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle &= 0 \text{ if } i \neq j \text{ and} \\ \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle &\neq 0 \text{ if } i = \text{j} \end{aligned} \tag{2.60}$$

A set of vectors is *orthonormal* if

$$\begin{aligned} \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle &= 0 \text{ if } i \neq j \text{ and} \\ \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle &= 1 \text{ if } i = \text{j} \end{aligned} \tag{2.61}$$

2.3 Gram-Schmidt Orthonormalization

It is computationally convenient to use basis sets that are orthonormal. However, if we are provided with a basis set that is not orthonormal, the Gram-Schmidt orthonormalization procedure provides us with a technique for orthonormalizing[M] it. By this we mean that if we are provided with the nonorthonormal set of basis vectors $\{\boldsymbol{y}_i\}$, $i = 1, \ldots, n$, we generate a new set of vectors $\{\boldsymbol{v}_i\}$, $i = 1, \ldots, n$,

`orth(A)`

which is orthonormal. In addition, each subset of vectors, $\{\boldsymbol{v}_j\}$, $j = 1, \ldots, m$, spans $\{\boldsymbol{y}_j\}$, $j = 1, \ldots, m$, for every $m < n$. We build the orthonormal set by starting with one vector from the original nonorthonormal set, and successively include additional vectors, one at a time, while subtracting from each new vector any component that it shares with the previously collected vectors. Therefore, as we add more vectors to the orthonormal set, each will contain only components along directions that were not included in any previously found vectors.

The technique will be presented as the following algorithm:

Let $\boldsymbol{v}_1 = \boldsymbol{y}_1$.

Choose $\boldsymbol{v}_2$ to be $\boldsymbol{y}_2$ with all components along the previous vector, subtracted out, i.e.,

$$\boldsymbol{v}_2 = \boldsymbol{y}_2 - a\boldsymbol{v}_1$$

Here, the coefficient *a* represents the magnitude of the component of $\boldsymbol{y}_2$ in the $\boldsymbol{v}_1$ direction. It is not initially known, so it will be computed. Because the $\boldsymbol{v}_i$'s are orthogonal,

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = 0 = \langle \boldsymbol{v}_1, \boldsymbol{y}_2 \rangle - a\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle$$

so

$$a = \frac{\langle \boldsymbol{v}_1, \boldsymbol{y}_2 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle}$$

Choose $\boldsymbol{v}_3$ as $\boldsymbol{y}_3$ with the components along *both* previous $\boldsymbol{v}_i$ vectors subtracted out:

$$\boldsymbol{v}_3 = \boldsymbol{y}_3 - a_1\boldsymbol{v}_1 - a_2\boldsymbol{v}_2 \tag{2.62}$$

Again, coefficients a_1 and a_2, which represent the components of $\boldsymbol{y}_3$ along the previously computed $\boldsymbol{v}_i$'s, must be calculated with the knowledge that the $\boldsymbol{v}_i$'s are orthogonal. As before, we find the inner product of Equation (2.62) with both $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$:

$$\begin{aligned}\langle \boldsymbol{v}_1, \boldsymbol{v}_3 \rangle = 0 &= \langle \boldsymbol{v}_1, \boldsymbol{y}_3 \rangle - a_1 \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle - a_2 \overbrace{\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle}^{0} \\ \langle \boldsymbol{v}_2, \boldsymbol{v}_3 \rangle = 0 &= \langle \boldsymbol{v}_2, \boldsymbol{y}_3 \rangle - a_1 \underbrace{\langle \boldsymbol{v}_2, \boldsymbol{v}_1 \rangle}_{0} - a_2 \langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle\end{aligned} \tag{2.63}$$

These two equations can be solved independently of one another to give

$$a_1 = \frac{\langle \boldsymbol{v}_1, \boldsymbol{y}_3 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \qquad \text{and} \qquad a_2 = \frac{\langle \boldsymbol{v}_2, \boldsymbol{y}_3 \rangle}{\langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle}$$

Continue using this same process to give

$$\boldsymbol{v}_i = \boldsymbol{y}_i - \sum_{k=1}^{i-1} a_k \boldsymbol{v}_k = \boldsymbol{y}_i - \sum_{k=1}^{i-1} \frac{\langle \boldsymbol{v}_k, \boldsymbol{y}_i \rangle}{\langle \boldsymbol{v}_k, \boldsymbol{v}_k \rangle} \boldsymbol{v}_k \tag{2.64}$$

Finish by normalizing each of the new basis vectors using:

$$\hat{\boldsymbol{v}}_i = \frac{\boldsymbol{v}_i}{\|\boldsymbol{v}_i\|} \tag{2.65}$$

Alternatively, this normalization can be done at the end of each stage, after which the terms $\langle \boldsymbol{v}_k, \boldsymbol{v}_k \rangle = 1$ in (2.64).

It should be noted that there are a number of variations on this procedure. For example, the normalization step (step 5) could be executed each time a new vector $\boldsymbol{v}$ is computed. This will affect the numerical results in subsequent steps, and make the end result look somewhat different, though still entirely correct. It has the benefit, though, that all terms $\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle = 1$ for subsequent iterations. Also, the results will depend on which basis vector is chosen at the start. There is no special reason that we chose $\boldsymbol{y}_1$ to start the process; another vector would suffice.

If we always use orthonormal basis sets, we can always find the component of vector $\boldsymbol{x}$ along the j^{th} basis vector by performing a simple inner product operation with that j^{th} basis vector. That is, if $\{\boldsymbol{e}_i\}$, $i = 1, \ldots, n$ is an orthonormal basis for space $\mathbf{X}$ and if $\boldsymbol{x} \in \mathbf{X}$, then $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i$ and

$$
\begin{aligned}
x_j &= \langle \boldsymbol{e}_j, \boldsymbol{x} \rangle = \sum_{i=1}^{n} x_i \langle \boldsymbol{e}_j, \boldsymbol{e}_i \rangle \\
&= x_1 \cancelto{0}{\langle \boldsymbol{e}_j, \boldsymbol{e}_1 \rangle} + x_2 \cancel{\langle \boldsymbol{e}_j, \boldsymbol{e}_2 \rangle} + \cdots + x_j \cancelto{1}{\langle \boldsymbol{e}_j, \boldsymbol{e}_j \rangle} + \cdots + x_n \cancel{\langle \boldsymbol{e}_j, \boldsymbol{e}_n \rangle}
\end{aligned}
\tag{2.66}
$$

Furthermore, if $\boldsymbol{x}$ is expressed in an orthonormal basis $\{\boldsymbol{e}_i\},\ i = 1, \ldots, n$, then the norm of $\boldsymbol{x}$ can be found using a form of *Parseval's theorem*.

THEOREM (Parseval's theorem): If a vector in an n-dimensional linear space is given in *any* orthonormal basis by the representation $\{c_i\}$, then the euclidean norm of that vector can be expressed as

$$
\|\boldsymbol{x}\| = \left[\sum_{i=1}^{n} c_i^{\,2} \right]^{1/2} \tag{2.67}
$$

PROOF: If a vector is given in an orthonormal basis by the expression $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i$, then

$$
\begin{aligned}
\|\boldsymbol{x}\| &= \langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2} \\
&= \left\langle \sum_{j=1}^{n} c_j \boldsymbol{e}_j, \sum_{i=1}^{n} c_i \boldsymbol{e}_i \right\rangle^{1/2} \\
&= \left[\sum_{i=1}^{n} c_i \left(\left\langle \sum_{j=1}^{n} c_j \boldsymbol{e}_j, \boldsymbol{e}_i \right\rangle \right) \right]^{1/2} \\
&= \left[\sum_{i=1}^{n} c_i (c_i) \right]^{1/2} = \left[\sum_{i=1}^{n} (c_i)^2 \right]^{1/2}
\end{aligned}
$$

Therefore for the vector $\boldsymbol{x}$ above, we can write:

$$\|\boldsymbol{x}\| = \left[\sum_{i=1}^{n} x_i^2\right]^{1/2} \tag{2.68}$$

Note that this is the formula often given for the norm of any vector, but it may not be true if the vector is not expressed in orthonormal bases. See Example 2.7 (Change of Basis) on page 64. Parseval's theorem is particularly useful when finding the norm of a *function* that has been expanded into an orthonormal set of basis functions (see Example 2.11 on page 79); the theorem applies to infinite-dimensional spaces as well.

2.4 Subspaces and the Projection Theorem

Often, we have a linear vector space that contains a particular portion in which we are interested. For example, although we are sometimes interested in the mechanics of a particle moving in a plane, we cannot forget that that plane is a *subspace* of the larger three-dimensional space of physical motion. In words, a *subspace* is a *subset* of a linear vector space that itself qualifies as a vector space. To make this notion mathematically general, we must give a definition for a subspace:

2.4.1 Subspaces

> ***Subspace:*** A *subspace* **M** of a vector space **X** is a subset of **X** such that if $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{M}$, α and β are scalars, and $\boldsymbol{z} = \alpha\boldsymbol{x} + \beta\boldsymbol{y}$, then $\boldsymbol{z} \in \mathbf{M}$. (2.69)

Strictly, this definition specifies only that the subspace is closed under scalar multiplication and vector addition. Some of the other properties that must be satisfied for the subspace to be considered a space unto itself are automatically satisfied given the fact that vectors in subspace **M** are gathered from the space **X**. Other properties can be inferred from closure. For example, that the subspace by necessity contains a zero vector is implied by the above definition if one chooses $\alpha = \beta = 0$.

Note that space **X** can be considered a subspace of itself. In fact, if the dimension of **M**, $\dim(\mathbf{M})$, is the same as the dimension of **X**, then $\mathbf{X} = \mathbf{M}$. If $\dim(\mathbf{M}) < \dim(\mathbf{X})$, then **M** is called a *proper subspace* of **X**. Because of the closure requirement and the fact that each subspace must contain the zero vector, we can make the following observations on euclidean spaces:

1. All proper subspaces of $\Re^2$ are straight lines that pass through the origin. (In addition, one could consider the trivial set consisting of only the zero vector to be a subspace.)

2. All proper subspaces of $\Re^3$ are either straight lines that pass through the origin, or planes that pass through the origin.
3. All proper subspaces of $\Re^n$ are surfaces of dimension $n-1$ or less, and all must pass through the origin. Such subspaces that we refer to as "surfaces" but that are not two-dimensional, as we usually think of surfaces, are called *hypersurfaces*.

2.4.2 The Projection Theorem

Often, we will encounter a vector in a relatively large-dimensional space, and a subspace of which the vector is not a member. The most easily envisioned example is a vector in a three-dimensional euclidean space and a two-dimensional subspace consisting of a plane. If the vector of interest is not in the plane, we can ask the question: Which vector that *does* lie in the plane is closest to the given vector, which does *not* lie in the plane? We will use the projection theorem to answer this question, which arises in the solution of simultaneous algebraic equations, as encountered in Chapter 3.

> THEOREM: Suppose $\mathbf{U}$ is a proper subspace of $\mathbf{X}$, so that $\dim(\mathbf{U}) < \dim(\mathbf{X})$. Then for every $\boldsymbol{x} \in \mathbf{X}$, there exists a vector $\boldsymbol{u} \in \mathbf{U}$ such that $\langle \boldsymbol{x} - \boldsymbol{u}, \boldsymbol{y} \rangle = 0$ for every $\boldsymbol{y} \in \mathbf{U}$. The vector $\boldsymbol{u}$ is *the orthogonal projection* of $\boldsymbol{x}$ into $\mathbf{U}$. (2.70)

This situation is depicted in Figure 2.4. In the figure, vector $\boldsymbol{x}$ is in the three-dimensional space defined by the coordinate axes shown.

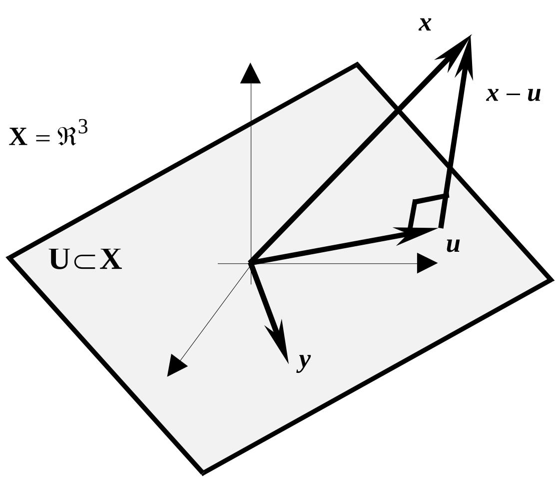

Figure 2.4 Illustration of the orthogonal projection theorem. Vectors $\boldsymbol{u}$ and $\boldsymbol{y}$ lie in the shaded plane. Vector $\boldsymbol{x}$ extends off the plane, and $\boldsymbol{x} - \boldsymbol{u}$ is orthogonal to $\boldsymbol{u}$ and every other vector in the plane.

The shaded plane that passes through the origin is the subspace **U**. Vector $\boldsymbol{x}$ does not lie in the plane of **U**. The vector $\boldsymbol{x}-\boldsymbol{u}$ referred to in the theorem is orthogonal to the plane itself; i.e., it is orthogonal to every vector in the plane, such as the vector $\boldsymbol{y}$. It is this orthogonality condition that enables us to find the projection.

In the following example, we will practice using the projection theorem by considering subspaces that are the *orthogonal complements* of one another. For this we give the definition:

> ***Orthogonal Complement:*** If **W** is a subspace of **X**, the *orthogonal complement* of **W**, denoted $\mathbf{W}^{\perp}$, is also a subspace of **X**, all of whose vectors are orthogonal to the set **W**. (2.71)

Example 2.10: Projection of a Vector

Consider vectors in $\Re^4$: $\boldsymbol{f}=[4\quad 0\quad 2\quad -1]^{\mathrm{T}}$, $\boldsymbol{x}_1=[1\quad 2\quad 2\quad 1]^{\mathrm{T}}$, and $\boldsymbol{x}_2=[0\quad 0\quad 1\quad 1]^{\mathrm{T}}$. Decompose the vector $\boldsymbol{f}$ into a sum $\boldsymbol{f}=\boldsymbol{g}+\boldsymbol{h}$, where $\boldsymbol{g}\in\mathbf{W}=span\{\boldsymbol{x}_1,\boldsymbol{x}_2\}$ and $\boldsymbol{h}$ is in the *orthogonal complement* of **W**.

Solution:

We will use two approaches to solve this problem. In the first, we use the projection theorem without first orthonormalizing the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. To directly apply the projection theorem, we propose that the projection of $\boldsymbol{f}$ onto **W** be written in the following form, with as yet undetermined coefficients α_i:

$$P\boldsymbol{f}=\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2$$

Then by the projection theorem, we must have $(\boldsymbol{f}-P\boldsymbol{f})\perp\mathbf{W}$, so that $\langle\boldsymbol{f}-P\boldsymbol{f},\boldsymbol{x}_1\rangle=0.$ But $\langle\boldsymbol{f}-P\boldsymbol{f},\boldsymbol{x}_1\rangle=\langle\boldsymbol{f},\boldsymbol{x}_1\rangle-\langle P\boldsymbol{f},\boldsymbol{x}_1\rangle=0$, so

$$\langle\boldsymbol{f},\boldsymbol{x}_1\rangle=\langle P\boldsymbol{f},\boldsymbol{x}_1\rangle=\alpha_1\langle\boldsymbol{x}_1,\boldsymbol{x}_1\rangle+\alpha_2\langle\boldsymbol{x}_2,\boldsymbol{x}_1\rangle$$

Similarly,

$$\langle\boldsymbol{f},\boldsymbol{x}_2\rangle=\langle P\boldsymbol{f},\boldsymbol{x}_2\rangle=\alpha_1\langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle+\alpha_2\langle\boldsymbol{x}_2,\boldsymbol{x}_2\rangle$$

Together, these two equations give the pair of simultaneous equations:

$$\begin{aligned} 7 &= \alpha_1 10 + \alpha_2 3 \\ 1 &= \alpha_1 3 + \alpha_2 2 \end{aligned}$$

that can be solved (see Chapter 3) to give $\alpha_1 = 1$ and $\alpha_2 = -1$. Therefore,

$$\boldsymbol{g} = P\boldsymbol{f} = \boldsymbol{x}_1 + (-1) \cdot \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

This being the desired vector in **W**, it is a simple matter to compute:

$$\boldsymbol{h} = \boldsymbol{f} - \boldsymbol{g} = \begin{bmatrix} 3 \\ -2 \\ 1 \\ -1 \end{bmatrix} \in \mathbf{W}^{\perp}$$

As an alternative path to the solution, which avoids the simultaneous equation solving at the expense of the Gram-Schmidt operations, we could first orthonormalize the $\boldsymbol{x}_i$ vectors. Denote the orthonormalized version as $\{\boldsymbol{v}_i\}$ and compute them as described:

$$\boldsymbol{v}_1 = \boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}$$

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - a\boldsymbol{v}_1$$

where

$$a = \frac{\langle \boldsymbol{v}_1, \boldsymbol{x}_2 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} = 0.3$$

giving

$$\boldsymbol{v}_2 = \begin{bmatrix} -0.3 \\ -0.6 \\ 0.4 \\ 0.7 \end{bmatrix}$$

Normalizing the two new basis vectors:

$$\hat{\boldsymbol{v}}_1 = \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} = \begin{bmatrix} 0.3162 \\ 0.6325 \\ 0.6325 \\ 0.3162 \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{v}}_2 = \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} = \begin{bmatrix} -0.2860 \\ -0.5721 \\ 0.3814 \\ 0.6674 \end{bmatrix}$$

Then to find the components of $\boldsymbol{f}$ along these directions,

$$f_1 = \langle \boldsymbol{f}, \hat{\boldsymbol{v}}_1 \rangle = 2.2136$$
$$f_2 = \langle \boldsymbol{f}, \hat{\boldsymbol{v}}_2 \rangle = -1.0488$$

giving

$$\boldsymbol{g} = P\boldsymbol{f} = f_1\hat{\boldsymbol{v}}_1 + f_2\hat{\boldsymbol{v}}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

exactly as before.

Example 2.11: Finite Fourier Series

Let **V** be a linear vector space consisting of all piecewise continuous, real-valued functions of time defined over the interval $t \in [-\pi, \pi]$. A valid inner product for this space is

$$\langle \boldsymbol{f}(t), \boldsymbol{g}(t) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \boldsymbol{f}(t)\boldsymbol{g}(t)\, dt \tag{2.72}$$

Now consider a subspace of this space, $\mathbf{W} \subset \mathbf{V}$ consisting of all functions that can be formed by linear combinations of the functions 1, $\cos t$, $\sin t$, $\cos 2t$, and $\sin 2t$, i.e.,

$$\mathbf{W} = span\{1,\ \cos t,\ \sin t,\ \cos 2t,\ \sin 2t\}.$$

1. From the inner product given, induce a norm on this space.
2. If it is known that the vectors that span **W** are linearly independent and therefore form a basis for this 5-dimensional subspace, show that the set of vectors $\{1,\ \cos t,\ \sin t,\ \cos 2t,\ \sin 2t\}$ is an orthogonal basis for **W**. Normalize this set to make it an orthonormal basis.
3. Find the orthogonal projection of the following function, $\boldsymbol{f}(t)$, defined in Figure 2.5, onto the space **W.**

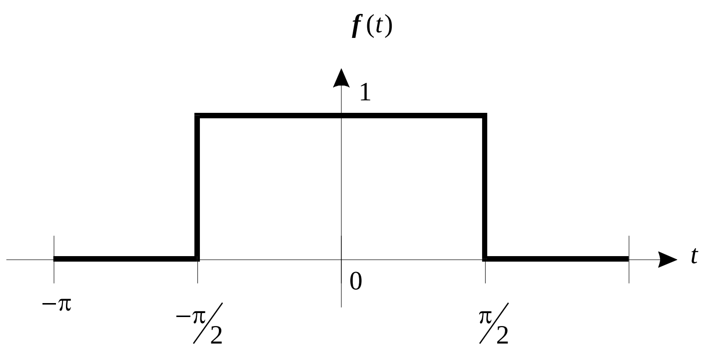

Figure 2.5 Square wave signal for Example 2.11.

Solution:

To induce a norm, we use

$$\|\boldsymbol{f}(t)\| = \langle \boldsymbol{f}(t), \boldsymbol{f}(t)\rangle^{1/2} = \left[\frac{1}{\pi}\int_{-\pi}^{\pi} \boldsymbol{f}^2(t)\, dt\right]^{1/2} \tag{2.73}$$

Use the notation:

$$\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4, \boldsymbol{e}_5\} = \{1,\ \cos t,\ \sin t,\ \cos 2t,\ \sin 2t\}$$

If this set is linearly independent (can you show this?), then it automatically qualifies as a basis. To show that the set is orthogonal, we must demonstrate that condition (2.60) holds. Finding all these inner products,

$$
\begin{aligned}
\langle \boldsymbol{e}_1, \boldsymbol{e}_1 \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} 1\, dt = 2 \\
\langle \boldsymbol{e}_2, \boldsymbol{e}_2 \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2 t \;dt = 1 \\
\langle \boldsymbol{e}_3, \boldsymbol{e}_3 \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2 t \;dt = 1 \\
\langle \boldsymbol{e}_4, \boldsymbol{e}_4 \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2 2t \;dt = 1 \\
\langle \boldsymbol{e}_5, \boldsymbol{e}_5 \rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2 2t \;dt = 1 \\
\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle &= \langle \boldsymbol{e}_2, \boldsymbol{e}_1 \rangle = 0 \\
&\vdots \\
\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle &= 0 \quad \text{for all } i \neq j
\end{aligned}
$$

which indeed satisfies (2.60).

To normalize the set $\{\boldsymbol{e}_i\}$, we simply have to divide each basis vector by its own norm, as defined by the induced norm above. These computations were mostly done while showing orthogonality, and it was found that the only vector with a nonunity norm was $\boldsymbol{e}_1$, with $\|\boldsymbol{e}_1\| = \sqrt{\langle \boldsymbol{e}_1, \boldsymbol{e}_1 \rangle} = \sqrt{2}$. Therefore an orthonormal set of basis vectors can be written as:

$$
\{\hat{\boldsymbol{e}}_i\} = \left\{ \frac{1}{\sqrt{2}},\ \cos t,\ \sin t,\ \cos 2t,\ \sin 2t \right\}
$$

To find the orthogonal projection of $\boldsymbol{f}(t)$ onto the space spanned by any orthonormal basis, we have only to find the components of the given vector along those basis vectors, which, according to (2.66), can simply be found via an inner product. Suppose we denote this projected vector (in **W**) as:

$$
g(t) = C_1\hat{\boldsymbol{e}}_1 + C_2\hat{\boldsymbol{e}}_2 + C_3\hat{\boldsymbol{e}}_3 + C_4\hat{\boldsymbol{e}}_4 + C_5\hat{\boldsymbol{e}}_5
$$

Then the undetermined coefficients C_i can be found by (2.66) as

$$C_1 = \langle \boldsymbol{f}(t), \hat{\boldsymbol{e}}_1 \rangle = \left\langle \boldsymbol{f}(t), \frac{1}{\sqrt{2}} \right\rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{\sqrt{2}} \boldsymbol{f}(t)\, dt = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{\sqrt{2}}\, dt = \frac{1}{\sqrt{2}}$$

$$C_2 = \langle \boldsymbol{f}(t), \hat{\boldsymbol{e}}_2 \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos t \;\; \boldsymbol{f}(t)\, dt = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos t \;\; dt = \frac{2}{\pi}$$

$$C_3 = \langle \boldsymbol{f}(t), \hat{\boldsymbol{e}}_3 \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin t \;\; \boldsymbol{f}(t)\, dt = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \sin t \;\; dt = 0$$

$$C_4 = \langle \boldsymbol{f}(t), \hat{\boldsymbol{e}}_4 \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos 2t \;\; \boldsymbol{f}(t)\, dt = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos 2t \;\; dt = 0$$

$$C_5 = \langle \boldsymbol{f}(t), \hat{\boldsymbol{e}}_5 \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin 2t \;\; \boldsymbol{f}(t)\, dt = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \sin 2t \;\; dt = 0$$

so that the projection of $\boldsymbol{f}(t)$ onto **W** is

$$\boldsymbol{g}(t) = \frac{1}{2} + \frac{2}{\pi} \cos t \tag{2.74}$$

Note that this is simply the five-term Fourier series approximation for the given signal $\boldsymbol{f}(t)$. The original function and its projection are shown in Figure 2.6 below.

It is useful at this point to also remember Parseval's theorem. Because the projection $\boldsymbol{g}(t)$ is written in an orthonormal basis, its norm can easily be found from the coefficients in this basis as:

$$\|\boldsymbol{g}(t)\| = \left[\sum_{i=1}^{5} C_i^2 \right]^{1/2} = \left[\frac{1}{2} + \frac{4}{\pi^2} \right]^{1/2} \approx 0.95 \tag{2.75}$$

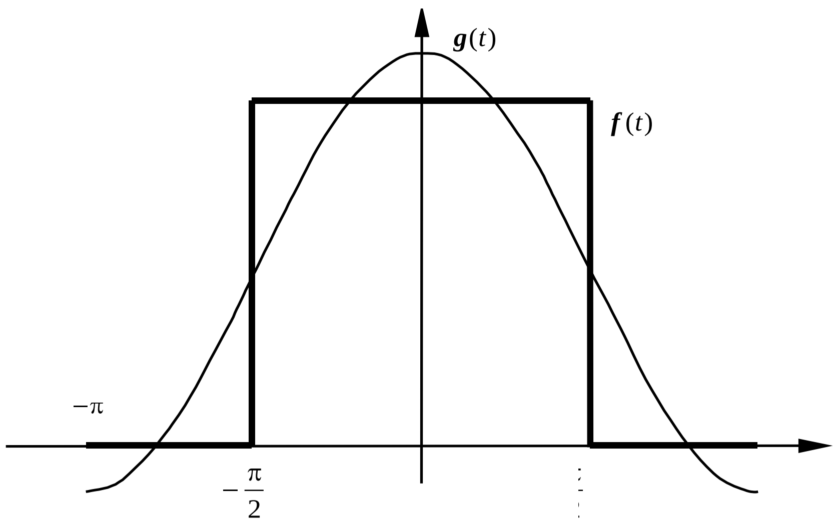

Figure 2.6 Original function $\boldsymbol{f}(t)$ and its projection $\boldsymbol{g}(t)$ onto the subspace **W**.

2.5 Linear Algebras

We will close this chapter by defining a term that is often used to refer to the general mathematical process of manipulating vectors, matrices, and simultaneous equations:

> ***Linear Algebra:*** A *linear algebra* is a linear vector space that has, in addition to all its other requirements for being a linear space, an operation called *multiplication*. This operation satisfies the following rules for all vectors $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$, and scalars α from its field:
>
> 1. Associativity of scalar multiplication, i.e.,
>
> $$\alpha(\boldsymbol{xy}) = (\alpha\boldsymbol{x})\boldsymbol{y} = \boldsymbol{x}(\alpha\boldsymbol{y})$$
>
> 2. Associativity of vector multiplication, i.e.,
>
> $$(\boldsymbol{xy})\boldsymbol{z} = \boldsymbol{x}(\boldsymbol{yz})$$
>
> 3. Distributivity of vector multiplication,
>
> $$(\boldsymbol{x}+\boldsymbol{y})\boldsymbol{z} = \boldsymbol{xz} + \boldsymbol{yz} \ .$$
>
> Note that the vector multiplication need not be commutative in a linear algebra. (2.76)

The euclidean vectors that we have been using as examples throughout this chapter are in fact *not* linear algebras because we have defined no multiplication for them. Nevertheless, some common vector spaces that we will discuss will indeed be linear algebras. An example is the space of $n \times n$ matrices, which we will introduce as representations of linear operators in the next chapter. The familiar procedure for matrix multiplication will be the multiplication operator that makes the *space* of matrices into an *algebra* of matrices. The interpretation of these matrices as operators on vector spaces is the topic of the next chapter.

2.6 Summary

In this chapter we have reviewed the basics of vector spaces and linear algebra. These fundamental concepts are rightfully the domain of mathematics texts, but it is useful to present them here in order that we can stress engineering relevance and the terminology we will find important later. The most important concepts include the following:

- Vectors need not be thought of as "magnitude and direction" pairs, but this is a good way to imagine them in low dimensions.
- All linear finite dimensional vector spaces, such as time functions or rational polynomials, may be treated as vectors of n-tuples because they are isomorphic to n-tuples. We then treat their representations as we would n-tuples, unifying the vector-matrix operations we must perform into one computational set.
- The basis of a space is an important generalization of the "coordinate system" that we are familiar with from basic mechanics and physics. Basis expansions and change-of-basis operations can be performed with vector-matrix arithmetic.
- The rank of a matrix represents the dimension of the space formed by its columns, i.e., it is the number of linearly independent columns the matrix possesses.
- Inner products are used to relate the differences in direction of two vectors, and norms are used to gauge their "size."
- Orthonormal bases are the most convenient kind of basis for computational purposes and can be constructed with the Gram-Schmidt procedure.
- There exists a simple theorem that enables us to find the vector closest to a subspace but not actually in it. This process will be important in engineering approximations wherein we must find the best solution to a problem when an exact solution is unavailable or impractical.

In the next chapter, we will consider linear operators, which will help us solve simultaneous equations, map vectors from one space into another or into itself, and compute other functions performed on vectors.

2.7 Problems

2.1 The set **W** of $n \times n$ matrices with real entries is known to be a linear vector space. Determine which of the following sets are subspaces of **W**:

a) The set of $n \times n$ skew-symmetric matrices?

b) The set of $n \times n$ diagonal matrices?

c) The set of $n \times n$ upper-diagonal matrices?

d) The set of $n \times n$ singular matrices?

2.2 Consider the set of all polynomials in s of degree less than or equal to four, with real coefficients, with the conditions:

$$\boldsymbol{p}'(0) + \boldsymbol{p}(0) = 0$$
$$\boldsymbol{p}(1) = 0$$

Show that this set is a linear vector space, and find its dimension. Construct a basis for this space.

2.3 For Example 2.4 on page 54 where we tested the linear independence of vectors of rational polynomials, show that indeed, if a_1 and a_2 are real, then they must both be zero.

2.4 Are the following vectors in $\Re^4$ linearly independent over the reals?

$$\begin{bmatrix} 2 \\ 0 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 4 \\ 0 \end{bmatrix}, \text{ and } \begin{bmatrix} 1 \\ 1 \\ -2 \\ 2 \end{bmatrix}$$

2.5 Which of the following sets of vectors are linearly independent?

a) $\left\{ \begin{bmatrix} -1 \\ -9 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} \right\}$ in the space of real 3-tuples over the field of reals.

b) $\left\{ \begin{bmatrix} 2-i \\ -i \end{bmatrix}, \begin{bmatrix} 1+2i \\ -i \end{bmatrix}, \begin{bmatrix} -i \\ 3+4i \end{bmatrix} \right\}$ in the space of complex pairs over the field of reals.

c) $\left\{ 2s^2+2s-1,\ -2s^2+2s+1,\ s^2-s-5 \right\}$ in the space of polynomials over the field of reals.

2.6 Prove or disprove the following claims: if $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ are linearly independent in vector space **V**, then so are

a) $\boldsymbol{u}$, $\boldsymbol{u}+\boldsymbol{v}$, and $\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}$.

b) $\boldsymbol{u}+2\boldsymbol{v}-\boldsymbol{w}$, $\boldsymbol{u}-2\boldsymbol{v}-\boldsymbol{w}$, and $4\boldsymbol{v}$.

c) $\boldsymbol{u}-\boldsymbol{v}$, $\boldsymbol{v}-\boldsymbol{w}$, and $\boldsymbol{w}-\boldsymbol{u}$.

d) $-\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}$, $\boldsymbol{u}-\boldsymbol{v}+\boldsymbol{w}$, and $-\boldsymbol{u}+\boldsymbol{v}-\boldsymbol{w}$.

2.7 Determine a basis for the space $\mathbf{W} \subset \Re^{\mathbf{4}}$ spanned by the four vectors

$$\{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4\} = \left\{ \begin{bmatrix} -1 \\ 1 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 7 \\ -8 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 10 \\ -11 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \\ 15 \\ -15 \end{bmatrix} \right\}$$

Is $\boldsymbol{z} = \begin{bmatrix} 3 \\ -1 \\ 13 \\ -17 \end{bmatrix} \in \mathbf{W}$? Is $\boldsymbol{u} = \begin{bmatrix} 4 \\ 9 \\ 12 \\ -8 \end{bmatrix} \in \mathbf{W}$? Is $\boldsymbol{v} = \begin{bmatrix} -1 \\ -1 \\ -3 \\ 3 \end{bmatrix} \in \mathbf{W}$?

2.8 Use the matrix form of the change of basis operation, i.e., Equation (2.28), to perform the change of basis on the vector $x = \begin{bmatrix} 10 & -2 & 2 & 3 \end{bmatrix}^{\mathrm{T}}$ given in Example 2.5 on page 60, arriving at the new representation shown in that example.

2.9 Prove that the number of linearly independent rows in a matrix is equal to the number of linearly independent columns.

2.10 For the following matrices, find the ranks and degeneracies.

$$\text{a)}\ \begin{bmatrix} 1 & 2 & -3 \\ -2 & 2 & 2 \\ 4 & 1 & -4 \end{bmatrix} \qquad \text{b)}\ \begin{bmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \qquad \text{c)}\ \begin{bmatrix} 1 & 6 & 3 & 2 & -2 \\ 2 & 12 & 4 & 6 & -10 \\ 3 & 18 & 0 & 15 & -15 \end{bmatrix}$$

2.11 Prove or disprove the following statement: If $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ is a basis for the vector space $\Re^4$ and $\mathbf{W}$ is a subspace of $\Re^4$ then some subset of $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ forms a basis for $\mathbf{W}$.

2.12 If it is known that a valid basis for the space of polynomials of degree less than or equal to three is $\{1, t, t^2, t^3\}$, show that every such polynomial has a unique representation as

$$\boldsymbol{p}(t) = a_3 t^3 + a_2 t^2 (1-t) + a_1 t(1-t)^2 + a_0 (1-t)^3$$

2.13 Construct an orthogonal basis of the space spanned by the set of vectors

$$\left\{ \begin{bmatrix} -2 \\ -2 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -3 \\ 4 \\ 8 \end{bmatrix} \right\}$$

Repeat for

$$\left\{ \begin{bmatrix} 2 \\ -1 \\ 3 \\ -3 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -14 \\ -3 \\ -19 \\ 11 \end{bmatrix} \right\}$$

2.14 Using the Gram-Schmidt technique, construct an orthonormal basis for the subspace of $\Re^4$ defined by the vectors

$$x_1 = \begin{bmatrix} -1 \\ 4 \\ 1 \\ 1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 4 \\ -3 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad x_3 = \begin{bmatrix} 3 \\ -1 \\ -2 \\ 2 \end{bmatrix}$$

2.15 Let $\mathbf{C}^4$ denote the space of all complex 4-tuples, over the complex numbers. Use the "standard" inner product and the Gram-Schmidt method to orthogonalize the vectors

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ -i \\ 1+i \end{bmatrix}, \quad x_2 = \begin{bmatrix} 2-i \\ i \\ -1 \\ 1+i \end{bmatrix}, \quad x_3 = \begin{bmatrix} 1+i \\ 1-i \\ 2 \\ 1-i \end{bmatrix}$$

2.16 Consider the vector $y = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$. Find the orthogonal projection of y onto the plane $x_1 - 2x_2 + 4x_3 = 0$.

2.17 Let

$$f_1(x) = \begin{cases} 4x & \text{if } 0 \le x \le 0.25 \\ 2-4x & \text{if } 0.25 < x \le 0.5 \\ 0 & \text{otherwise} \end{cases}$$

Also let $f_2(x) = f_1(x-0.25)$ and $f_3(x) = f_1(x-0.5)$. Let **V** be the space of real-valued continuous functions on the interval $x \in [0, 1]$ using the inner product

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

and let **W** be the subspace spanned by the three functions f_1, f_2, and f_3. Find the dimension of the subspace **W**.

Now let

$$\boldsymbol{g}(x) = \begin{cases} 1 & \text{if } 0.5 \le x \le 1 \\ 0 & \text{if } 0 \le x < 0.5 \end{cases}$$

Find the projection of $\boldsymbol{g}$ in **W**.

2.18 Let V be the vector space of all polynomials, with inner product

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_0^\infty \boldsymbol{f}(x)\boldsymbol{g}(x)e^{-x}\,dx$$

Let **W** be the subspace of **V** spanned by $\{1, x, x^2\}$.

a) Use Gram-Schmidt to orthogonalize $\{1, x, x^2\}$.

b) Find the orthogonal projection of $\boldsymbol{f}(x) = x^3$ onto **W**.

2.19 The LeGendre polynomials may be defined as the orthogonalization of the basis $\{1, t, t^2, \ldots\}$ for the space of polynomials in *t*, defined on $t \in [-1, 1]$, using the inner product

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-1}^{1} \boldsymbol{f}(t)\boldsymbol{g}(t)\,dt$$

Find the first four such LeGendre polynomials.

2.20 Let **V** be the vector space of continuous real-valued functions of *t* defined on the interval $t \in [0, 3]$. Let **W** be the subspace of **V** consisting of functions that are linear for $0 < t < 1$, linear for $1 < t < 2$, and linear for $2 < t < 3$.

a) Construct a basis for **W**.

b) If an inner product for **W** is given by

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_0^3 \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx$$

compute the orthogonal projection of $\boldsymbol{f}(x) = x^2$ onto **W**.

2.21 Let **V** be an inner product space. If $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{V}$, and $\langle \boldsymbol{x}, \boldsymbol{z} \rangle = \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ for every $\boldsymbol{z} \in \mathbf{V}$, prove that $\boldsymbol{x} = \boldsymbol{y}$.

2.22 Use the complex vector $\boldsymbol{x} = \begin{bmatrix} 1+2i & 2-i \end{bmatrix}^{\mathrm{T}}$ to demonstrate that, in a complex vector space, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}$ is not a valid inner product.

2.23 Find the polynomial $\boldsymbol{p}(t) = t^3 + a_2 t^2 + a_1 t + a_0$ for which the integral

$$J = \int_{-1}^{1} \boldsymbol{p}^2(t)\, dt$$

has the smallest value.

2.24 Let **V** be the space of all polynomials with real coefficients, with the inner product

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-\infty}^{\infty} \boldsymbol{f}(x) \boldsymbol{g}(x) e^{-x^2}\, dx$$

Orthogonalize the set of vectors $\{1, x, x^2, x^3\}$.

2.25 Let **V** be the linear space of real-valued continuous functions defined on the interval $t \in [-1, 1]$, with inner product

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-1}^{1} \boldsymbol{f}(x) \boldsymbol{g}(x)\, dx$$

Let **W** be a subspace consisting of polynomials $\boldsymbol{p}$ of degree less than or equal to four that satisfy $\boldsymbol{p}(1) = \boldsymbol{p}(-1) = 0$. Find the projection of the polynomial $\boldsymbol{p}(t) = 1$ in the subspace **W**.

2.26 Prove that a vector of functions $[\boldsymbol{x}_1(t) \ \cdots \ \boldsymbol{x}_n(t)]^{\mathrm{T}}$ is linearly independent if and only if the Gram determinant is nonzero.

2.8 References and Further Reading

The engineering perspective on vectors and vector spaces presented here is comparable to those in [1], [6], and [7]. Further details and a more mathematical treatment of vector spaces and vector-matrix computations can be found in [2], [3], [5], and [8].

[1] Brogan, William L., *Modern Control Theory*, 3rd edition, Prentice-Hall, 1991.

[2] Golub, Gene H., and Charles F. Van Loan, *Matrix Computations*, Johns Hopkins University Press.

[3] Halmos, Paul R., *Finite Dimensional Vector Spaces*, 2nd edition, D. Van Nostrand, 1958.

[4] Kolmogorov, A. N., and S. V. Fomin, *Introductory Real Analysis*, Dover, 1970.

[5] Naylor, A. W., and G. R. Sell, *Linear Operator Theory in Engineering and Science*, Springer-Verlag, 1982.

[6] Shilov, Georgi E., *Linear Algebra*, Dover, 1977.

[7] Strang, Gilbert, *Linear Algebra and Its Applications*, Academic Press, 1980.

[8] Zadeh, Lotfi, A., and Charles A. Desoer, *Linear System Theory: The State Space Approach*, McGraw-Hill, 1963.