Chapter 5

Vector spaces and signal space

In the previous chapter, we showed that any $\mathcal{L}_2$ function $u(t)$ can be expanded in various orthogonal expansions, using such sets of orthogonal functions as the T-spaced truncated sinusoids or the sinc-weighted sinusoids. Thus $u(t)$ may be specified (up to $\mathcal{L}_2$ equivalence) by a countably infinite sequence such as $\{u_{k,m}; -\infty < k, m < \infty\}$ of coefficients in such an expansion.

In engineering, n-tuples of numbers are often referred to as *vectors*, and the use of vector notation is very helpful in manipulating these n-tuples. The collection of n-tuples of real numbers is called $\mathbb{R}^n$ and that of complex numbers $\mathbb{C}^n$. It turns out that the most important properties of these n-tuples also apply to countably infinite sequences of real or complex numbers. It should not be surprising, after the results of the previous sections, that these properties also apply to $\mathcal{L}_2$ waveforms.

A vector space is essentially a collection of objects (such as the collection of real n-tuples) along with a set of rules for manipulating those objects. There is a set of axioms describing precisely how these objects and rules work. Any properties that follow from those axioms must then apply to any vector space, *i.e.*, any set of objects satisfying those axioms. $\mathbb{R}^n$ and $\mathbb{C}^n$ satisfy these axioms, and we will see that countable sequences and $\mathcal{L}_2$ waveforms also satisfy them.

Fortunately, it is just as easy to develop the general properties of vector spaces from these axioms as it is to develop specific properties for the special case of $\mathbb{R}^n$ or $\mathbb{C}^n$ (although we will constantly use $\mathbb{R}^n$ and $\mathbb{C}^n$ as examples). Fortunately also, we can use the example of $\mathbb{R}^n$ (and particularly $\mathbb{R}^2$) to develop geometric insight about general vector spaces.

The collection of $\mathcal{L}_2$ functions, viewed as a vector space, will be called *signal space*. The signal-space viewpoint has been one of the foundations of modern digital communication theory since its popularization in the classic text of Wozencraft and Jacobs[35].

The signal-space viewpoint has the following merits:

- Many insights about waveforms (signals) and signal sets do not depend on time and frequency (as does the development up until now), but depend only on vector relationships.
- Orthogonal expansions are best viewed in vector space terms.
- Questions of limits and approximation are often easily treated in vector space terms. It is for this reason that many of the results in Chapter 4 are proved here.

5.1 The axioms and basic properties of vector spaces

A *vector space* $\mathcal{V}$ is a set of elements, $\boldsymbol{v} \in \mathcal{V}$, called *vectors*, along with a set of rules for operating on both these vectors and a set of ancillary elements called *scalars*. For the treatment here, the set $\mathbb{F}$ of scalars[1] will either be the set of real numbers $\mathbb{R}$ or the set of complex numbers $\mathbb{C}$. A vector space with real scalars is called a *real vector space*, and one with complex scalars is called a *complex vector space*.

The most familiar example of a real vector space is $\mathbb{R}^n$. Here the vectors are represented as n-tuples of real numbers.[2] $\mathbb{R}^2$ is represented geometrically by a plane, and the vectors in $\mathbb{R}^2$ by points in the plane. Similarly, $\mathbb{R}^3$ is represented geometrically by three-dimensional Euclidean space.

The most familiar example of a complex vector space is $\mathbb{C}^n$, the set of n-tuples of complex numbers.

The axioms of a vector space $\mathcal{V}$ are listed below; they apply to arbitrary vector spaces, and in particular to the real and complex vector spaces of interest here.

1. Addition: For each $\boldsymbol{v} \in \mathcal{V}$ and $\boldsymbol{u} \in \mathcal{V}$, there is a unique vector $\boldsymbol{v} + \boldsymbol{u} \in \mathcal{V}$, called the sum of $\boldsymbol{v}$ and $\boldsymbol{u}$, satisfying
   (a) Commutativity: $\boldsymbol{v} + \boldsymbol{u} = \boldsymbol{u} + \boldsymbol{v}$,
   (b) Associativity: $\boldsymbol{v} + (\boldsymbol{u} + \boldsymbol{w}) = (\boldsymbol{v} + \boldsymbol{u}) + \boldsymbol{w}$ for each $\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w} \in \mathcal{V}$.
   (c) Zero: There is a unique element $\mathbf{0} \in \mathcal{V}$ satisfying $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathcal{V}$,
   (d) Negation: For each $\boldsymbol{v} \in \mathcal{V}$, there is a unique $-\boldsymbol{v} \in \mathcal{V}$ such that $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}$.
2. Scalar multiplication: For each scalar[3] α and each $\boldsymbol{v} \in \mathcal{V}$ there is a unique vector $\alpha\boldsymbol{v} \in \mathcal{V}$ called the scalar product of α and $\boldsymbol{v}$ satisfying
   (a) Scalar associativity: $\alpha(\beta\boldsymbol{v}) = (\alpha\beta)\boldsymbol{v}$ for all scalars α, β, and all $\boldsymbol{v} \in \mathcal{V}$,
   (b) Unit multiplication: for the unit scalar 1, $1\boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathcal{V}$.
3. Distributive laws:
   (a) For all scalars α and all $\boldsymbol{v}, \boldsymbol{u} \in \mathcal{V}$, $\alpha(\boldsymbol{v} + \boldsymbol{u}) = \alpha\boldsymbol{v} + \alpha\boldsymbol{u}$;
   (b) For all scalars α, β and all $\boldsymbol{v} \in \mathcal{V}$, $(\alpha + \beta)\boldsymbol{v} = \alpha\boldsymbol{v} + \beta\boldsymbol{v}$.

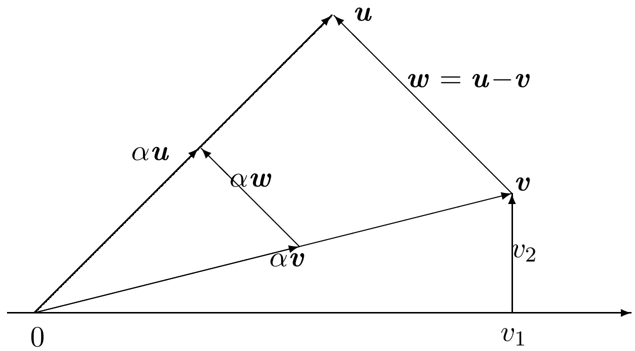
Example 5.1.1. For $\mathbb{R}^n$, a vector $\boldsymbol{v}$ is an n-tuple $(v_1, \dots, v_n)$ of real numbers. Addition is defined by $\boldsymbol{v} + \boldsymbol{u} = (v_1{+}u_1, \dots, v_n{+}u_n)$. The zero vector is defined by $\mathbf{0} = (0, \dots, 0)$. The scalars α are the real numbers, and $\alpha\boldsymbol{v}$ is defined to be $(\alpha v_1, \dots, \alpha v_n)$. This is illustrated geometrically in Figure 5.1.1 for $\mathbb{R}^2$.

Example 5.1.2. The vector space $\mathbb{C}^n$ is the same as $\mathbb{R}^n$ except that $\boldsymbol{v}$ is an n-tuple of complex numbers and the scalars are complex. Note that $\mathbb{C}^2$ can not be easily illustrated geometrically, since a vector in $\mathbb{C}^2$ is specified by 4 real numbers. The reader should verify the axioms for both $\mathbb{R}^n$ and $\mathbb{C}^n$.

[1] More generally, vector spaces can be defined in which the scalars are elements of an arbitrary *field*. It is not necessary here to understand the general notion of a field.

[2] Many people prefer to define $\mathbb{R}^n$ as the class of real vector spaces of dimension n, but almost everyone visualizes $\mathbb{R}^n$ as the space of n-tuples. More importantly, the space of n-tuples will be constantly used as an example and $\mathbb{R}^n$ is a convenient name for it.

[3] Addition, subtraction, multiplication, and division between scalars is done according to the familiar rules of $\mathbb{R}$ or $\mathbb{C}$ and will not be restated here. Neither $\mathbb{R}$ nor $\mathbb{C}$ includes ∞.



Vectors are represented by points or directed lines.

The scalar multiple $\alpha\boldsymbol{u}$ and $\boldsymbol{u}$ lie on the same line from 0.

The distributive law says that triangles scale correctly.

Figure 5.1: Geometric interpretation of $\mathbb{R}^2$. The vector $\boldsymbol{v} = (v_1, v_2)$ is represented as a point in the Euclidean plane with abscissa v_1 and ordinate v_2. It can also be viewed as the directed line from $\mathbf{0}$ to the point $\boldsymbol{v}$. Sometimes, as in the case of $\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{v}$, a vector is viewed as a directed line from some nonzero point ($\boldsymbol{v}$ in this case) to another point $\boldsymbol{u}$. This geometric interpretation also suggests the concepts of length and angle, which are not included in the axioms. This is discussed more fully later.

Example 5.1.3. There is a trivial vector space for which the only element is the zero vector $\mathbf{0}$. Both for real and complex scalars, $\alpha\mathbf{0} = \mathbf{0}$. The vector spaces of interest here are non-trivial spaces, *i.e.*, spaces with more than one element, and this will usually be assumed without further mention.

Because of the commutative and associative axioms, we see that a finite sum $\sum_j \alpha_j \boldsymbol{v}_j$, where each α_j is a scalar and $\boldsymbol{v}_j$ a vector, is unambiguously defined without the need for parentheses. This sum is called a *linear combination* of the vectors $\{\boldsymbol{v}_i\}$.

We next show that the set of finite-energy complex waveforms can be viewed as a complex vector space.[4] When we view a waveform $v(t)$ as a vector, we denote it by $\boldsymbol{v}$. There are two reasons for this: first, it reminds us that we are viewing the waveform is a vector; second, $v(t)$ sometimes denotes a function and sometimes denotes the value of that function at a particular argument t. Denoting the function as $\boldsymbol{v}$ avoids this ambiguity.

The vector sum $\boldsymbol{v} + \boldsymbol{u}$ is defined in the obvious way as the waveform for which each t is mapped into $v(t) + u(t)$; the scalar product $\alpha\boldsymbol{v}$ is defined as the waveform for which each t is mapped into $\alpha v(t)$. The vector $\mathbf{0}$ is defined as the waveform that maps each t into 0.

The vector space axioms are not difficult to verify for this space of waveforms. To show that the sum $\boldsymbol{v} + \boldsymbol{u}$ of two finite energy waveforms $\boldsymbol{v}$ and $\boldsymbol{u}$ also has finite energy, recall first that the sum of two measurable waveforms is also measurable. Next, recall that if v and u are complex numbers, then $|v + u|^2 \leq 2|v|^2 + 2|u|^2$. Thus,

$$\int_{-\infty}^{\infty} |v(t) + u(t)|^2 \, dt \leq \int_{-\infty}^{\infty} 2|v(t)|^2 \, dt + \int_{-\infty}^{\infty} 2|u(t)|^2 \, dt < \infty. \tag{5.1}$$

Similarly, if $\boldsymbol{v}$ has finite energy, then $\alpha\boldsymbol{v}$ has $|\alpha|^2$ times the energy of $\boldsymbol{v}$ which is also finite. The other axioms can be verified by inspection.

The above argument has shown that the set of finite-energy waveforms, along with the definitions of addiition and complex scalar multiplication, form a complex vector space. The set of real

[4]There is a small but important technical difference between the vector space being defined here and what we will later define to be the vector space $\mathcal{L}_2$. This difference centers on the notion of $\mathcal{L}_2$ equivalence, and will be discussed later.

finite-energy waveforms along with the analogous addition and real scalar multiplication form a real vector space.

5.1.1 Finite-dimensional vector spaces

A set of vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathcal{V}$ *spans* $\mathcal{V}$ (and is called a *spanning set* of $\mathcal{V}$) if every vector $\boldsymbol{v} \in \mathcal{V}$ is a linear combination of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$. For the R^n example, let $\boldsymbol{e}_1 = (1, 0, 0, \dots, 0)$, $\boldsymbol{e}_2 = (0, 1, 0, \dots, 0), \dots, \boldsymbol{e}_n = (0, \dots 0, 1)$ be the n *unit vectors* of $\mathbb{R}^n$. The unit vectors span R^n since every vector $\boldsymbol{v} \in \mathbb{R}^n$ can be expressed as a linear combination of the unit vectors, *i.e.*,

$$\boldsymbol{v} = (\alpha_1, \dots, \alpha_n) = \sum_{j=1}^{n} \alpha_j \boldsymbol{e}_j.$$

A vector space $\mathcal{V}$ is *finite-dimensional* if a finite set of vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ exist that span $\mathcal{V}$. Thus $\mathbb{R}^n$ is finite-dimensional since it is spanned by $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$. Similarly, $\mathbb{C}^n$ is finite-dimensional. If $\mathcal{V}$ is not finite-dimensional, then it is *infinite-dimensional*; we will soon see that $\mathcal{L}_2$ is infinite-dimensional.

A set of vectors, $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathcal{V}$ is *linearly dependent* if $\sum_{j=1}^{n} \alpha_j \boldsymbol{v}_j = 0$ for some set of scalars not all equal to 0. This implies that each vector $\boldsymbol{v}_k$ for which $\alpha_k \neq 0$ is a linear combination of the others, *i.e.*,

$$\boldsymbol{v}_k = \sum_{j \neq k} \frac{-\alpha_j}{\alpha_k} \boldsymbol{v}_j.$$

A set of vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathcal{V}$ is linearly independent if it is not linearly dependent, *i.e.*, if $\sum_{j=1}^{n} \alpha_j \boldsymbol{v}_j = 0$ implies that each α_j is 0. For brevity we often omit the word "linear" when we refer to independence or dependence.

It can be seen that the unit vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, as elements of $\mathbb{R}^n$, are linearly independent. Similarly, they are linearly independent as elements of $\mathbb{C}^n$,

A set of vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathcal{V}$ is defined to be a *basis* for $\mathcal{V}$ if the set both spans $\mathcal{V}$ and is linearly independent. Thus the unit vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ form a basis for $\mathbb{R}^n$. Similarly, the unit vectors, as elements of $\mathbb{C}^n$, form a basis for $\mathbb{C}^n$.

The following theorem is both important and simple; see Exercise 5.1 or any linear algebra text for a proof.

Theorem 5.1.1 (Basis for finite-dimensional vector space). *Let $\mathcal{V}$ be a non-trivial finite-dimensional vector space.*[5] *Then*

- *If $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ span $\mathcal{V}$ but are linearly dependent, then a subset of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ forms a basis for $\mathcal{V}$ with $n < m$ vectors.*
- *If $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ are linearly independent but do not span $\mathcal{V}$, then there exists a basis for $\mathcal{V}$ with $n > m$ vectors that includes $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$.*
- *Every basis of $\mathcal{V}$ contains the same number of vectors.*

[5]The trivial vector space whose only element is $\mathbf{0}$ is conventionally called a zero-dimensional space and could be viewed as having an empty basis.

The *dimension* of a finite-dimensional vector space may now be defined as the number of vectors in a basis. The theorem implicitly provides two conceptual algorithms for finding a basis. First, start with any linearly independent set (such as a single nonzero vector) and successively add independent vectors until reaching a spanning set. Second, start with any spanning set and successively eliminate dependent vectors until reaching a linearly independent set.

Given any basis, $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$, for a finite-dimensional vector space $\mathcal{V}$, any vector $\boldsymbol{v} \in \mathcal{V}$ can be represented as

$$\boldsymbol{v} = \sum_{j=1}^{n} \alpha_j \boldsymbol{v}_j\,, \qquad \text{where } \alpha_1, \dots, \alpha_n \text{ are scalars.} \tag{5.2}$$

In terms of the given basis, each $\boldsymbol{v} \in \mathcal{V}$ can be uniquely represented by the n-tuple of coefficients $(\alpha_1, \dots, \alpha_n)$ in (5.2). Thus any n-dimensional vector space $\mathcal{V}$ over $\mathbb{R}$ or $\mathbb{C}$ may be viewed (relative to a given basis) as a version[6] of $\mathbb{R}^n$ or $\mathbb{C}^n$. This leads to the elementary vector/matrix approach to linear algebra. What is gained by the axiomatic ("coordinate-free") approach is the ability to think about vectors without first specifying a basis. We see the value of this shortly when we define subspaces and look at finite-dimensional subspaces of infinite-dimensional vector spaces such as $\mathcal{L}_2$.

5.2 Inner product spaces

The vector space axioms above contain no inherent notion of length or angle, although such geometric properties are clearly present in Figure 5.1.1 and in our intuitive view of $\mathbb{R}^n$ or $\mathbb{C}^n$. The missing ingredient is that of an *inner product*.

An *inner product* on a complex vector space $\mathcal{V}$ is a complex-valued function of two vectors, $\boldsymbol{v}, \boldsymbol{u} \in \mathcal{V}$, denoted by $\langle \boldsymbol{v}, \boldsymbol{u} \rangle$, that satisfies the following axioms:

(a) Hermitian symmetry: $\langle \boldsymbol{v}, \boldsymbol{u} \rangle = \langle \boldsymbol{u}, \boldsymbol{v} \rangle^*$;

(b) Hermitian bilinearity: $\langle \alpha \boldsymbol{v} + \beta \boldsymbol{u}, \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{v}, \boldsymbol{w} \rangle + \beta \langle \boldsymbol{u}, \boldsymbol{w} \rangle$
(and consequently $\langle \boldsymbol{v}, \alpha \boldsymbol{u} + \beta \boldsymbol{w} \rangle = \alpha^* \langle \boldsymbol{v}, \boldsymbol{u} \rangle + \beta^* \langle \boldsymbol{v}, \boldsymbol{w} \rangle$);

(c) Strict positivity: $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$, with equality if and only if $\boldsymbol{v} = \boldsymbol{0}$.

A vector space with an inner product satisfying these axioms is called an *inner product space*.

The same definition applies to a real vector space, but the inner product is always real and the complex conjugates can be omitted.

The *norm* or *length* $\|\boldsymbol{v}\|$ of a vector $\boldsymbol{v}$ in an inner product space is defined as

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}.$$

Two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ are defined to be *orthogonal* if $\langle \mathbf{v}, \boldsymbol{u} \rangle = 0$. Thus we see that the important geometric notions of length and orthogonality are both defined in terms of the inner product.

[6] More precisely $\mathcal{V}$ and $\mathbb{R}^n$ ($\mathbb{C}^n$) are isomorphic in the sense that that there is a one-to one correspondence between vectors in $\mathcal{V}$ and n-tuples in $\mathbb{R}^n$ (C^n) that preserves the vector space operations. In plain English, solvable problems concerning vectors in $\mathcal{V}$ can always be solved by first translating to n-tuples in a basis and then working in $\mathbb{R}^n$ or $\mathbb{C}^n$.

5.2.1 The inner product spaces $\mathbb{R}^n$ and $\mathbb{C}^n$

For the vector space $\mathbb{R}^n$ of real n-tuples, the inner product of vectors $\boldsymbol{v} = (v_1, \dots v_n)$ and $\boldsymbol{u} = (u_1, \dots, u_n)$ is usually defined (and is defined here) as

$$\langle \boldsymbol{v}, \boldsymbol{u} \rangle = \sum_{j=1}^{n} v_j u_j.$$

You should verify that this definition satisfies the inner product axioms above.

The length $\|\boldsymbol{v}\|$ of a vector $\boldsymbol{v}$ is then $\sqrt{\sum_j v_j^2}$, which agrees with Euclidean geometry. Recall that the formula for the cosine between two arbitrary nonzero vectors in $\mathbb{R}^2$ is given by

$$\cos(\angle(\boldsymbol{v}, \boldsymbol{u})) = \frac{v_1 u_1 + v_2 u_2}{\sqrt{v_1^2 + v_2^2}\,\sqrt{u_1^2 + u_1^2}} = \frac{\langle \boldsymbol{v}, \boldsymbol{u} \rangle}{\|\boldsymbol{v}\|\,\|\boldsymbol{u}\|}, \tag{5.3}$$

where the final equality also expresses this as an inner product. Thus the inner product determines the angle between vectors in $\mathbb{R}^2$. This same inner product formula will soon be seen to be valid in any real vector space, and the derivation is much simpler in the coordinate free environment of general vector spaces than in the unit vector context of $\mathbb{R}^2$.

For the vector space $\mathbb{C}^n$ of complex n-tuples, the inner product is defined as

$$\langle \boldsymbol{v}, \boldsymbol{u} \rangle = \sum_{j=1}^{n} v_j u_j^* \tag{5.4}$$

The norm, or length, of $\boldsymbol{v}$ is then $\sqrt{\sum_j |v_j|^2} = \sqrt{\sum_j [\Re(v_j)^2 + \Im(v_j)^2]}$. Thus, as far as length is concerned, a complex n-tuple $\boldsymbol{u}$ can be regarded as the real $2n$-vector formed from the real and imaginary parts of $\boldsymbol{u}$. Warning: although a complex n-tuple can be viewed as a real $2n-$tuple for some purposes, such as length, many other operations on complex n-tuples are very different from those operations on the corresponding real $2n$-tuple. For example, scalar multiplication and inner products in $\mathbb{C}^n$ are very different from those operations in $\mathbb{R}^{2n}$.

5.2.2 One-dimensional projections

An important problem in constructing orthogonal expansions is that of breaking a vector $\boldsymbol{v}$ into two components relative to another vector $\boldsymbol{u} \neq \boldsymbol{0}$ in the same inner-product space. One component, $\boldsymbol{v}_{\perp\boldsymbol{u}}$, is to be orthogonal (*i.e.*, perpendicular) to $\boldsymbol{u}$ and the other, $\boldsymbol{v}_{|\boldsymbol{u}}$, is to be collinear with $\boldsymbol{u}$ (two vectors $\boldsymbol{v}_{|\boldsymbol{u}}$ and $\boldsymbol{u}$ are collinear if $\boldsymbol{v}_{|\boldsymbol{u}} = \alpha\boldsymbol{u}$ for some scalar α). Figure 5.2 illustrates this decomposition for vectors in $\mathbb{R}^2$. We can view this geometrically as dropping a perpendicular from $\boldsymbol{v}$ to $\boldsymbol{u}$. From the geometry of Figure 5.2, $\|\boldsymbol{v}_{|\boldsymbol{u}}\| = \|\boldsymbol{v}\| \cos(\angle(\boldsymbol{v}, \boldsymbol{u}))$. Using (5.3), $\|\boldsymbol{v}_{|\boldsymbol{u}}\| = \langle \boldsymbol{v}, \boldsymbol{u} \rangle / \|\boldsymbol{u}\|$. Since $\boldsymbol{v}_{|\boldsymbol{u}}$ is also collinear with $\boldsymbol{u}$, it can be seen that

$$\boldsymbol{v}_{|\boldsymbol{u}} = \frac{\langle \boldsymbol{v}, \boldsymbol{u} \rangle}{\|\boldsymbol{u}\|^2}\,\boldsymbol{u}. \tag{5.5}$$

The vector $\boldsymbol{v}_{|\boldsymbol{u}}$ is called the *projection* of $\boldsymbol{v}$ onto $\boldsymbol{u}$.

Rather surprisingly, (5.5) is valid for any inner product space. The general proof that follows is also simpler than the derivation of (5.3) and (5.5) using plane geometry.

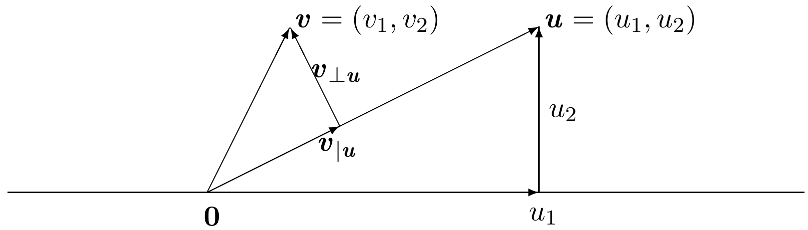


Figure 5.2: Two vectors, $\boldsymbol{v} = (v_1, v_2)$ and $\boldsymbol{u} = (u_1, u_2)$ in $\mathbb{R}^2$. Note that $\|\boldsymbol{u}\|^2 = \langle \boldsymbol{u}, \boldsymbol{u} \rangle = u_1^2 + u_2^2$ is the squared length of $\boldsymbol{u}$. The vector $\boldsymbol{v}$ is also expressed as $\boldsymbol{v} = \boldsymbol{v}_{|\boldsymbol{u}} + \boldsymbol{v}_{\perp \boldsymbol{u}}$ where $\boldsymbol{v}_{|\boldsymbol{u}}$ is collinear with $\boldsymbol{u}$ and $\boldsymbol{v}_{\perp \boldsymbol{u}}$ is perpendicular to $\boldsymbol{u}$.

Theorem 5.2.1 (One-dimensional projection theorem). *Let $\boldsymbol{v}$ and $\boldsymbol{u}$ be arbitrary vectors with $\boldsymbol{u} \neq \mathbf{0}$ in a real or complex inner product space. Then there is a unique scalar α for which $\langle \boldsymbol{v} - \alpha\boldsymbol{u}, \boldsymbol{u} \rangle = 0$. That α is given by $\alpha = \langle \boldsymbol{v}, \boldsymbol{u} \rangle / \|\boldsymbol{u}\|^2$.*

Remark: The theorem states that $\boldsymbol{v} - \alpha\boldsymbol{u}$ is perpendicular to $\boldsymbol{u}$ if and only if $\alpha = \langle \boldsymbol{v}, \boldsymbol{u} \rangle / \|\boldsymbol{u}\|^2$. Using that value of α, $\boldsymbol{v} - \alpha\boldsymbol{u}$ is called the perpendicular to $\boldsymbol{u}$ and is denoted as $\boldsymbol{v}_{\perp \boldsymbol{u}}$; similarly $\alpha\boldsymbol{u}$ is called the projection of $\boldsymbol{v}$ on $\boldsymbol{u}$ and is denoted as $\boldsymbol{u}_{|\boldsymbol{u}}$. Finally, $\boldsymbol{v} = \boldsymbol{v}_{\perp \boldsymbol{u}} + \boldsymbol{v}_{|\boldsymbol{u}}$, so $\boldsymbol{v}$ has been split into a perpendicular part and a collinear part.

Proof: Calculating $\langle \boldsymbol{v} - \alpha\boldsymbol{u}, \boldsymbol{u} \rangle$ for an arbitrary scalar α, the conditions can be found under which this inner product is zero:

$$\langle \boldsymbol{v} - \alpha\boldsymbol{u}, \boldsymbol{u} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle - \alpha \langle \boldsymbol{u}, \boldsymbol{u} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle - \alpha \|\boldsymbol{u}\|^2,$$

which is equal to zero if and only if $\alpha = \langle \boldsymbol{v}, \boldsymbol{u} \rangle / \|\boldsymbol{u}\|^2$. □

The reason why $\|\boldsymbol{u}\|^2$ is in the denominator of the projection formula can be understood by rewriting (5.5) as

$$\boldsymbol{v}_{|\boldsymbol{u}} = \langle \boldsymbol{v}, \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|} \rangle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}.$$

In words, the projection of $\boldsymbol{v}$ on $\boldsymbol{u}$ is the same as the projection of $\boldsymbol{v}$ on the normalized version of $\boldsymbol{u}$. More generally, the value of $\boldsymbol{v}_{|\boldsymbol{u}}$ is invariant to scale changes in $\boldsymbol{u}$, *i.e.*,

$$\boldsymbol{v}_{|\beta\boldsymbol{u}} = \frac{\langle \boldsymbol{v}, \beta\boldsymbol{u} \rangle}{\|\beta\boldsymbol{u}\|^2} \beta\boldsymbol{u} = \frac{\langle \boldsymbol{v}, \boldsymbol{u} \rangle}{\|\boldsymbol{u}\|^2} \boldsymbol{u} = \boldsymbol{v}_{|\boldsymbol{u}}. \tag{5.6}$$

This is clearly consistent with the geometric picture in Figure 5.2 for $\mathbb{R}^2$, but it is also valid for complex vector spaces where such figures cannot be drawn.

In $\mathbb{R}^2$, the cosine formula can be rewritten as

$$\cos(\angle(\boldsymbol{u}, \boldsymbol{v})) = \langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}, \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \rangle. \tag{5.7}$$

That is, the cosine of $\angle(\boldsymbol{u}, \boldsymbol{v})$ is the inner product of the normalized versions of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Another well known result in $\mathbb{R}^2$ that carries over to any inner product space is the Pythagorean theorem: If $\boldsymbol{v}$ and $\boldsymbol{u}$ are orthogonal, then

$$\|\boldsymbol{v} + \boldsymbol{u}\|^2 = \|\boldsymbol{v}\|^2 + \|\boldsymbol{u}\|^2. \tag{5.8}$$

To see this, note that

$$\langle \boldsymbol{v}+\boldsymbol{u}, \boldsymbol{v}+\boldsymbol{u}\rangle = \langle \boldsymbol{v}, \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{u}\rangle + \langle \boldsymbol{u}, \boldsymbol{v}\rangle + \langle \boldsymbol{u}, \boldsymbol{u}\rangle.$$

The cross terms disappear by orthogonality, yielding (5.8).

Theorem 5.2.1 has an important corollary, called the *Schwarz inequality*:

Corollary 5.2.1 (Schwarz inequality). *Let $\boldsymbol{v}$ and $\boldsymbol{u}$ be vectors in a real or complex inner product space. Then*

$$|\langle \boldsymbol{v}, \boldsymbol{u}\rangle| \le \|\boldsymbol{v}\|\ \|\boldsymbol{u}\|. \tag{5.9}$$

Proof: Assume $\boldsymbol{u} \neq \boldsymbol{0}$ since (5.9) is obvious otherwise. Since $\boldsymbol{v}_{|\boldsymbol{u}}$ and $\boldsymbol{v}_{\perp\boldsymbol{u}}$ are orthogonal, (5.8) shows that

$$\|\boldsymbol{v}\|^2 = \|\boldsymbol{v}_{|\boldsymbol{u}}\|^2 + \|\boldsymbol{v}_{\perp\boldsymbol{u}}\|^2.$$

Since $\|\boldsymbol{v}_{\perp\boldsymbol{u}}\|^2$ is nonnegative, we have

$$\|\boldsymbol{v}\|^2 \quad \ge \quad \|\boldsymbol{v}_{|\boldsymbol{u}}\|^2 = \left|\frac{\langle \boldsymbol{v}, \boldsymbol{u}\rangle}{\|\boldsymbol{u}\|^2}\right|^2 \|\boldsymbol{u}\|^2 = \frac{|\langle \boldsymbol{v}, \boldsymbol{u}\rangle|^2}{\|\boldsymbol{u}\|^2},$$

which is equivalent to (5.9). □

For $\boldsymbol{v}$ and $\boldsymbol{u}$ both nonzero, the Schwarz inequality may be rewritten in the form

$$\left|\langle \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}, \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}\rangle\right| \le 1.$$

In $\mathbb{R}^2$, the Schwarz inequality is thus equivalent to the familiar fact that the cosine function is upperbounded by 1.

As shown in Exercise 5.6, the triangle inequality below is a simple consequence of the Schwarz inequality.

$$\|\boldsymbol{v}+\boldsymbol{u}\| \le \|\boldsymbol{v}\| + \|\boldsymbol{u}\|. \tag{5.10}$$

5.2.3 The inner product space of $\mathcal{L}_2$ functions

Consider the set of complex finite energy waveforms again. We attempt to define the inner product of two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ in this set as

$$\langle \boldsymbol{v}, \boldsymbol{u}\rangle = \int_{-\infty}^{\infty} v(t)u^*(t)dt. \tag{5.11}$$

It is shown in Exercise 5.8 that $\langle \boldsymbol{v}, \boldsymbol{u}\rangle$ is always finite. The Schwarz inequality cannot be used to prove this, since we have not yet shown that this satisfies the axioms of an inner product space. However, the first two inner product axioms follow immediately from the existence and finiteness of the inner product, *i.e.*, the integral in (5.11). This existence and finiteness is a vital and useful property of $\mathcal{L}_2$.

The final inner product axiom is that $\langle \boldsymbol{v}, \boldsymbol{v}\rangle \ge 0$, with equality if and only if $\boldsymbol{v} = \boldsymbol{0}$. This axiom does not hold for finite-energy waveforms, because as we have already seen, if a function $v(t)$ is zero almost everywhere, then its energy is 0, even though the function is not the zero function. This is a nit-picking issue at some level, but axioms cannot be ignored simply because they are inconvenient.

The resolution of this problem is to define *equality* in an $\mathcal{L}_2$ inner product space as $\mathcal{L}_2$-equivalence between $\mathcal{L}_2$ functions. What this means is that each *element* of an $\mathcal{L}_2$ inner product space is the *equivalence class* of $\mathcal{L}_2$ functions that are equal almost everywhere. For example, the zero equivalence class is the class of zero-energy functions, since each is $\mathcal{L}_2$ equivalent to the all-zero function. With this modification, the inner product axioms all hold.

Viewing a vector as an equivalence class of $\mathcal{L}_2$ functions seems very abstract and strange at first. From an engineering perspective, however, the notion that all zero-energy functions are the same is more natural than the notion that two functions that differ in only a few isolated points should be regarded as different.

From a more practical viewpoint, it will be seen later that $\mathcal{L}_2$ functions (in this equivlence class sense) can be represented by the coefficients in any orthogonal expansion whose elements span the $\mathcal{L}_2$ space. Two ordinary functions have the same coefficients in such an orthogonal expansion if and only if they are $\mathcal{L}_2$ equivalent. Thus each element $\boldsymbol{u}$ of the $\mathcal{L}_2$ inner product space is in one-to-one correspondence to a finite-energy sequence $\{u_k;\ k \in \mathbb{Z}\}$ of coefficients in an orthogonal expansion. Thus we can now avoid the awkwardness of having many $\mathcal{L}_2$ equivalent ordinary functions map into a single sequence of coefficients and having no very good way of going back from sequence to function. Once again engineering common sense and sophisticated mathematics come together.

From now on we will simply view $\mathcal{L}_2$ as an inner product space, referring to the notion of $\mathcal{L}_2$ equivalence only when necessary. With this understanding, we can use all the machinery of inner product spaces, including projections and the Schwarz inequality.

5.2.4 Subspaces of inner product spaces

A *subspace* $\mathcal{S}$ of a vector space $\mathcal{V}$ is a subset of the vectors in $\mathcal{V}$ which forms a vector space in its own right (over the same set of scalars as used by $\mathcal{V}$). An equivalent definition is that for all $\boldsymbol{v}$ and $\boldsymbol{u} \in \mathcal{S}$, the linear combination $\alpha\boldsymbol{v} + \beta\boldsymbol{u}$ is in $\mathcal{S}$ for all scalars α and β. If $\mathcal{V}$ is an inner product space, then it can be seen that $\mathcal{S}$ is also an inner product space using the same inner product definition as $\mathcal{V}$.

Example 5.2.1 (Subspaces of $\mathbb{R}^3$). Consider the real inner product space $\mathbb{R}^3$, namely the inner product space of real 3-tuples $\boldsymbol{v} = (v_1, v_2, v_3)$. Geometrically, we regard this as a space in which there are three orthogonal coordinate directions, defined by the three unit vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. The 3-tuple v_1, v_2, v_3 then specifies the length of $\boldsymbol{v}$ in each of those directions, so that $\boldsymbol{v} = v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 + v_3\boldsymbol{e}_3$.

Let $\boldsymbol{u} = (1, 0, 1)$ and $\boldsymbol{w} = (0, 1, 1)$ be two fixed vectors, and consider the subspace of $\mathbb{R}^3$ composed of all linear combinations, $\boldsymbol{v} = \alpha\boldsymbol{u} + \beta\boldsymbol{w}$, of $\boldsymbol{u}$ and $\boldsymbol{w}$. Geometrically, this subspace is a plane going through the points $\boldsymbol{0}, \boldsymbol{u}$, and $\boldsymbol{w}$. In this plane, as in the original vector space, $\boldsymbol{u}$ and $\boldsymbol{w}$ each have length $\sqrt{2}$ and $\langle \boldsymbol{u}, \boldsymbol{w} \rangle = 1$.

Since neither $\boldsymbol{u}$ nor $\boldsymbol{w}$ is a scalar multiple of the other, they are linearly independent. They span $\mathcal{S}$ by definition, so $\mathcal{S}$ is a two-dimensional subspace with a basis $\{\boldsymbol{u}, \boldsymbol{w}\}$.

The projection of $\boldsymbol{u}$ on $\boldsymbol{w}$ is $\boldsymbol{u}_{|\boldsymbol{w}} = (0, 1/2, 1/2)$, and the perpendicular is $\boldsymbol{u}_{\perp\boldsymbol{w}} = (1, -1/2, 1/2)$.

These vectors form an orthogonal basis for $\mathcal{S}$. Using these vectors as an orthogonal basis, we can view $\mathcal{S}$, pictorially and geometrically, in just the same way as we view vectors in R^2.

Example 5.2.2 (General 2D subspace). Let $\mathcal{V}$ be an arbitrary non-trivial real or complex inner product space, and let $\boldsymbol{u}$ and $\boldsymbol{w}$ be arbitrary noncollinear vectors. Then the set $\mathcal{S}$ of linear combinations of $\boldsymbol{u}$ and $\boldsymbol{w}$ is a two-dimensional subspace of $\mathcal{V}$ with basis $\{\boldsymbol{u}, \boldsymbol{w}\}$. Again, $\boldsymbol{u}_{|\boldsymbol{w}}$ and $\boldsymbol{u}_{\perp \boldsymbol{w}}$ forms an orthogonal basis of $\mathcal{S}$. We will soon see that this procedure for generating subspaces and orthogonal bases from two vectors in an arbitrary inner product space can be generalized to orthogonal bases for subspaces of arbitrary dimension.

Example 5.2.3 ($\mathbb{R}^2$ is a subset but not a subspace of $\mathbb{C}^2$). Consider the complex vector space $\mathbb{C}^2$. The set of real 2-tuples is a subset of $\mathbb{C}^2$, but this subset is not closed under multiplication by scalars in $\mathbb{C}$. For example, the real 2-tuple $\boldsymbol{u} = (1, 2)$ is an element of $\mathbb{C}^2$ but the scalar product $i\boldsymbol{u}$ is the vector $(i, 2i)$, which is not a real 2-tuple. More generally, the notion of linear combination (which is at the heart of both the use and theory of vector spaces) depends on what the scalars are.

We cannot avoid dealing with both complex and real $\mathcal{L}_2$ waveforms without enormously complicating the subject (as a simple example, consider using the sine and cosine forms of the Fourier transform and series). We also cannot avoid inner product spaces without great complication. Finally we cannot avoid going back and forth between complex and real $\mathcal{L}_2$ waveforms. The price of this is frequent confusion between real and complex scalars. The reader is advised to use considerable caution with linear combinations and to be very clear about whether real or complex scalars are involved.

5.3 Orthonormal bases and the projection theorem

In an inner product space, a set of vectors $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots$ is *orthonormal* if

$$\langle \boldsymbol{\phi}_j, \boldsymbol{\phi}_k \rangle = \begin{cases} 0 & \text{for} \quad j \neq k \\ 1 & \text{for} \quad j = k. \end{cases} \tag{5.12}$$

In other words, an orthonormal set is a set of nonzero orthogonal vectors where each vector is *normalized* to unit length. It can be seen that if a set of vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots$ is orthogonal, then the set

$$\boldsymbol{\phi}_j = \frac{1}{\|\boldsymbol{u}_j\|} \boldsymbol{u}_j$$

is orthonormal. Note that if two nonzero vectors are orthogonal, then any scaling (including normalization) of each vector maintains orthogonality.

If a vector $\boldsymbol{v}$ is projected onto a normalized vector $\boldsymbol{\phi}$, then the one-dimensional projection theorem states that the projection is given by the simple formula

$$\boldsymbol{v}_{|\boldsymbol{\phi}} = \langle \boldsymbol{v}, \boldsymbol{\phi} \rangle \boldsymbol{\phi}. \tag{5.13}$$

Furthermore, the theorem asserts that $\boldsymbol{v}_{\perp \boldsymbol{\phi}} = \boldsymbol{v} - \boldsymbol{v}_{|\boldsymbol{\phi}}$ is orthogonal to $\boldsymbol{\phi}$. We now generalize the Projection Theorem to the projection of a vector $\boldsymbol{v} \in \mathcal{V}$ onto any finite dimensional subspace $\mathcal{S}$ of $\mathcal{V}$.

5.3.1 Finite-dimensional projections

If $\mathcal{S}$ is a subspace of an inner product space $\mathcal{V}$, and $\boldsymbol{v} \in \mathcal{V}$, then a *projection of* $\boldsymbol{v}$ *on* $\mathcal{S}$ is defined to be a vector $\boldsymbol{v}_{|\mathcal{S}} \in \mathcal{S}$ such that $\boldsymbol{v} - \boldsymbol{v}_{|\mathcal{S}}$ is orthogonal to all vectors in $\mathcal{S}$. The theorem to follow shows that $\boldsymbol{v}_{|\mathcal{S}}$ always exists and has a unique value given in the theorem.

The earlier definition of projection is a special case of that here in which $\mathcal{S}$ is taken to be the one dimensional subspace spanned by a vector $\boldsymbol{u}$ (the orthonormal basis is then $\boldsymbol{\phi} = \boldsymbol{u}/\|\boldsymbol{u}\|$).

Theorem 5.3.1 (Projection theorem). *Let $\mathcal{S}$ be an n-dimensional subspace of an inner product space $\mathcal{V}$ and assume that $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_n\}$ is an orthonormal basis for $\mathcal{S}$. Then for any $\boldsymbol{v} \in \mathcal{V}$, there is a unique vector $\boldsymbol{v}_{|\mathcal{S}} \in \mathcal{S}$ such that $\langle \boldsymbol{v} - \boldsymbol{v}_{|\mathcal{S}}, \boldsymbol{s}\rangle = 0$ for all $\boldsymbol{s} \in \mathcal{S}$. Furthermore, $\boldsymbol{v}_{|\mathcal{S}}$ is given by*

$$\boldsymbol{v}_{|\mathcal{S}} = \sum_{j=1}^{n} \langle \boldsymbol{v}, \boldsymbol{\phi}_j\rangle \boldsymbol{\phi}_j. \tag{5.14}$$

Remark: The theorem assumes that $\mathcal{S}$ has a set of orthonormal vectors as a basis. It will be shown later that any non-trivial finite-dimensional inner product space has such an orthonormal basis, so that the assumption does not restrict the generality of the theorem.

Proof: Let $\boldsymbol{w} = \sum_{j=1}^{n} \alpha_j \boldsymbol{\phi}_j$ be an arbitrary vector in $\mathcal{S}$. First consider the conditions on $\boldsymbol{w}$ under which $\boldsymbol{v} - \boldsymbol{w}$ is orthogonal to all vectors $\boldsymbol{s} \in \mathcal{S}$. It can be seen that $\boldsymbol{v} - \boldsymbol{w}$ is orthogonal to all $\boldsymbol{s} \in \mathcal{S}$ if and only if

$$\langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{\phi}_j\rangle = 0, \quad \text{for all } j,\ 1 \le j \le n,$$

or equivalently if and only if

$$\langle \boldsymbol{v}, \boldsymbol{\phi}_j\rangle = \langle \boldsymbol{w}, \boldsymbol{\phi}_j\rangle, \quad \text{for all } j,\ 1 \le j \le n. \tag{5.15}$$

Since $\boldsymbol{w} = \sum_{\ell=1}^{n} \alpha_\ell \boldsymbol{\phi}_\ell$,

$$\langle \boldsymbol{w}, \boldsymbol{\phi}_j\rangle = \sum_{\ell=1}^{n} \alpha_\ell \langle \boldsymbol{\phi}_\ell, \boldsymbol{\phi}_j\rangle = \alpha_j, \quad \text{for all } \ j,\ 1 \le j \le n. \tag{5.16}$$

Combining this with (5.15), $\boldsymbol{v} - \boldsymbol{w}$ is orthogonal to all $\boldsymbol{s} \in \mathcal{S}$ if and only if $\alpha_j = \langle \boldsymbol{v}, \boldsymbol{\phi}_j\rangle$ for each j, *i.e.*, if and only if $\boldsymbol{w} = \sum_j \langle \boldsymbol{v}, \boldsymbol{\phi}_j\rangle \boldsymbol{\phi}_j$. Thus $\boldsymbol{v}_{|\mathcal{S}}$ as given in (5.14) is the unique vector $\boldsymbol{w} \in \mathcal{S}$ for which $\boldsymbol{v} - \boldsymbol{v}_{|\mathcal{S}}$ is orthogonal to all $\boldsymbol{s} \in \mathcal{S}$. □

The vector $\boldsymbol{v} - \boldsymbol{v}_{|\mathcal{S}}$ is denoted as $\boldsymbol{v}_{\perp\mathcal{S}}$, the *perpendicular from* $\boldsymbol{v}$ *to* $\mathcal{S}$. Since $\boldsymbol{v}_{|\mathcal{S}} \in \mathcal{S}$, we see that $\boldsymbol{v}_{|\mathcal{S}}$ and $\boldsymbol{v}_{\perp\mathcal{S}}$ are orthogonal. The theorem then asserts that $\boldsymbol{v}$ can be uniquely split into two orthogonal components, $\boldsymbol{v} = \boldsymbol{v}_{|\mathcal{S}} + \boldsymbol{v}_{\perp\mathcal{S}}$ where the projection $\boldsymbol{v}_{|\mathcal{S}}$ is in $\mathcal{S}$ and the perpendicular $\boldsymbol{v}_{\perp\mathcal{S}}$ is orthogonal to all vectors $\boldsymbol{s} \in \mathcal{S}$.

5.3.2 Corollaries of the projection theorem

There are three important corollaries of the projection theorem that involve the norm of the projection. First, for any scalars $\alpha_1, \dots, \alpha_n$, the squared norm of $\boldsymbol{w} = \sum_j \alpha_j \boldsymbol{\phi}_j$ is given by

$$\|\boldsymbol{w}\|^2 = \langle \boldsymbol{w}, \sum_{j=1}^{n} \alpha_j \boldsymbol{\phi}_j\rangle = \sum_{j=1}^{n} \alpha_j^* \langle \boldsymbol{w}, \boldsymbol{\phi}_j\rangle = \sum_{j=1}^{n} |\alpha_j|^2,$$

where (5.16) has been used in the last step. For the projection $\boldsymbol{v}_{|\mathcal{S}}$, $\alpha_j = \langle \boldsymbol{v}, \boldsymbol{\phi}_j \rangle$, so

$$\|\boldsymbol{v}_{|\mathcal{S}}\|^2 = \sum_{j=1}^{n} |\langle \boldsymbol{v}, \boldsymbol{\phi}_j \rangle|^2. \tag{5.17}$$

Also, since $\boldsymbol{v} = \boldsymbol{v}_{|\mathcal{S}} + \boldsymbol{v}_{\perp\mathcal{S}}$ and $\boldsymbol{v}_{|\mathcal{S}}$ is orthogonal to $\boldsymbol{v}_{\perp\mathcal{S}}$, It follows from the Pythagorean theorem (5.8) that

$$\|\boldsymbol{v}\|^2 = \|\boldsymbol{v}_{|\mathcal{S}}\|^2 + \|\boldsymbol{v}_{\perp\mathcal{S}}\|^2. \tag{5.18}$$

Since $\|\boldsymbol{v}_{\perp\mathcal{S}}\|^2 \geq 0$, the following corollary has been proven:

Corollary 5.3.1 (norm bound).

$$0 \leq \|\boldsymbol{v}_{|\mathcal{S}}\|^2 \leq \|\boldsymbol{v}\|^2, \tag{5.19}$$

with equality on the right if and only if $\boldsymbol{v} \in \mathcal{S}$ *and equality on the left if and only if* $\boldsymbol{v}$ *is orthogonal to all vectors in* $\mathcal{S}$.

Substituting (5.17) into (5.19), we get *Bessel's inequality*, which is the key to understanding the convergence of orthonormal expansions.

Corollary 5.3.2 (Bessel's inequality). *Let* $\mathcal{S} \subseteq \mathcal{V}$ *be the subspace spanned by the set of orthonormal vectors* $\{\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_n\}$. *For any* $\boldsymbol{v} \in \mathcal{V}$

$$0 \leq \sum_{j=1}^{n} |\langle \boldsymbol{v}, \boldsymbol{\phi}_j \rangle|^2 \leq \|\boldsymbol{v}\|^2,$$

with equality on the right if and only if $\boldsymbol{v} \in \mathcal{S}$ *and equality on the left if and only if* $\boldsymbol{v}$ *is orthogonal to all vectors in* $\mathcal{S}$.

Another useful characterization of the projection $\boldsymbol{v}_{|\mathcal{S}}$ is that it is the vector in $\mathcal{S}$ that is closest to $\boldsymbol{v}$. In other words, using some $\boldsymbol{s} \in \mathcal{S}$ as an approximation to $\boldsymbol{v}$, the squared error is $\|\boldsymbol{v} - \boldsymbol{s}\|^2$. The following corollary says that $\boldsymbol{v}_{|\mathcal{S}}$ is the choice for $\boldsymbol{s}$ that yields the least squared error (LS).

Corollary 5.3.3 (LS error property). *The projection* $\boldsymbol{v}_{|\mathcal{S}}$ *is the unique closest vector in* $\mathcal{S}$ *to* $\boldsymbol{v}$*; i.e., for all* $\boldsymbol{s} \in \mathcal{S}$,

$$\|\boldsymbol{v} - \boldsymbol{v}_{|\mathcal{S}}\|^2 \leq \|\boldsymbol{v} - \boldsymbol{s}\|^2,$$

with equality if and only if $\boldsymbol{s} = \boldsymbol{v}_{|\mathcal{S}}$.

Proof: Decomposing $\boldsymbol{v}$ into $\boldsymbol{v}_{|\mathcal{S}} + \boldsymbol{v}_{\perp\mathcal{S}}$, we have $\boldsymbol{v} - \boldsymbol{s} = [\boldsymbol{v}_{|\mathcal{S}} - \boldsymbol{s}] + \boldsymbol{v}_{\perp\mathcal{S}}$. Since $\boldsymbol{v}_{|\mathcal{S}}$ and $\boldsymbol{s}$ are in $\mathcal{S}$, $\boldsymbol{v}_{|\mathcal{S}} - \boldsymbol{s}$ is also in $\mathcal{S}$, so by Pythagoras,

$$\|\boldsymbol{v} - \boldsymbol{s}\|^2 = \|\boldsymbol{v}_{|\mathcal{S}} - \boldsymbol{s}\|^2 + \|\boldsymbol{v}_{\perp\mathcal{S}}\|^2 \geq \|\boldsymbol{v}_{\perp\mathcal{S}}\|^2,$$

with equality if and only if $\|\boldsymbol{v}_{|\mathcal{S}} - \boldsymbol{s}\|^2 = 0$, *i.e.*, if and only if $\boldsymbol{s} = \boldsymbol{v}_{|\mathcal{S}}$. Since $\boldsymbol{v}_{\perp\mathcal{S}} = \boldsymbol{v} - \boldsymbol{v}_{|\mathcal{S}}$, this completes the proof. □

5.3.3 Gram-Schmidt orthonormalization

Theorem 5.3.1, the projection theorem, assumed an orthonormal basis $\{\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_n\}$ for any given n-dimensional subspace $\mathcal{S}$ of $\mathcal{V}$. The use of orthonormal bases simplifies almost everything concerning inner product spaces, and for infinite-dimensional expansions, orthonormal bases are even more useful.

This section presents the Gram-Schmidt procedure, which, starting from an arbitrary basis $\{\boldsymbol{s}_1, \dots, \boldsymbol{s}_n\}$ for an n-dimensional inner product subspace $\mathcal{S}$, generates an orthonormal basis for $\mathcal{S}$. The procedure is useful in finding orthonormal bases, but is even more useful theoretically, since it shows that such bases always exist. In particular, since every n-dimensional subspace contains an orthonormal basis, the projection theorem holds for each such subspace.

The procedure is almost obvious in view of the previous subsections. First an orthonormal basis, $\boldsymbol{\phi}_1 = \boldsymbol{s}_1/\|\boldsymbol{s}_1\|$, is found for the one-dimensional subspace $\mathcal{S}_1$ spanned by $\boldsymbol{s}_1$. Projecting $\boldsymbol{s}_2$ onto this one-dimensional subspace, a second orthonormal vector can be found. Iterating, a complete orthonormal basis can be constructed.

In more detail, let $(\boldsymbol{s}_2)_{|S_1}$ be the projection of $\boldsymbol{s}_2$ onto $\mathcal{S}_1$. Since $\boldsymbol{s}_2$ and $\boldsymbol{s}_1$ are linearly independent, $(\boldsymbol{s}_2)_{\perp\mathcal{S}_1} = \boldsymbol{s}_2 - (\boldsymbol{s}_2)_{|S_1}$ is nonzero. It is orthogonal to $\boldsymbol{\phi}_1$ since $\boldsymbol{\phi}_1 \in \mathcal{S}_1$. It is normalized as $\boldsymbol{\phi}_2 = (\boldsymbol{s}_2)_{\perp\mathcal{S}_1}/\|(\boldsymbol{s}_2)_{\perp\mathcal{S}_1}\|$. Then $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ span the space $\mathcal{S}_2$ spanned by $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Now, using induction, suppose that an orthonormal basis $\{\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_k\}$ has been constructed for the subspace $\mathcal{S}_k$ spanned by $\{\boldsymbol{s}_1, \dots, \boldsymbol{s}_k\}$. The result of projecting $\boldsymbol{s}_{k+1}$ onto $\mathcal{S}_k$ is $(\boldsymbol{s}_{k+1})_{|\mathcal{S}_k} = \sum_{j=1}^{k} \langle \boldsymbol{s}_{k+1}, \boldsymbol{\phi}_j\rangle \boldsymbol{\phi}_j$. The perpendicular, $(\boldsymbol{s}_{k+1})_{\perp\mathcal{S}_k} = \boldsymbol{s}_{k+1} - (\boldsymbol{s}_{k+1})_{|\mathcal{S}_k}$ is given by

$$(\boldsymbol{s}_{k+1})_{\perp\mathcal{S}_k} = \boldsymbol{s}_{k+1} - \sum_{j=1}^{k} \langle \boldsymbol{s}_{k+1}, \boldsymbol{\phi}_j\rangle \boldsymbol{\phi}_j. \tag{5.20}$$

This is nonzero since $\boldsymbol{s}_{k+1}$ is not in $\mathcal{S}_k$ and thus not a linear combination of $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_k$. Normalizing,

$$\boldsymbol{\phi}_{k+1} = \frac{(\boldsymbol{s}_{k+1})_{\perp\mathcal{S}_k}}{\|(\boldsymbol{s}_{k+1})_{\perp\mathcal{S}_k}\|} \tag{5.21}$$

From (5.20) and (5.21), $\boldsymbol{s}_{k+1}$ is a linear combination of $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_{k+1}$ and $\boldsymbol{s}_1, \dots, \boldsymbol{s}_k$ are linear combinations of $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_k$, so $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_{k+1}$ is an orthonormal basis for the space $\mathcal{S}_{k+1}$ spanned by $\boldsymbol{s}_1, \dots, \boldsymbol{s}_{k+1}$.

In summary, given any n-dimensional subspace $\mathcal{S}$ with a basis $\{\boldsymbol{s}_1, \dots, \boldsymbol{s}_n\}$, the Gram-Schmidt orthonormalization procedure produces an orthonormal basis $\{\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_n\}$ for $\mathcal{S}$.

Note that if a set of vectors is not necessarily independent, then the procedure will automatically find any vector $\boldsymbol{s}_j$ that is a linear combination of previous vectors via the projection theorem. It can then simply discard such a vector and proceed. Consequently it will still find an orthonormal basis, possibly of reduced size, for the space spanned by the original vector set.

5.3.4 Orthonormal expansions in $\mathcal{L}_2$

The background has now been developed to understand countable orthonormal expansions in $\mathcal{L}_2$. We have already looked at a number of *orthogonal* expansions, such as those used in the sampling theorem, the Fourier series, and the T-spaced truncated or sinc-weighted sinusoids. Turning these into *orthonormal* expansions involves only minor scaling changes.

The Fourier series will be used both to illustrate these changes and as an example of a general orthonormal expansion. The vector space view will then allow us to understand the Fourier series at a deeper level. Define $\theta_k(t) = e^{2\pi ikt/T}\text{rect}(\frac{t}{T})$ for $k \in \mathbb{Z}$. The set $\{\theta_k(t); k \in \mathbb{Z}\}$ of functions is orthogonal with $\|\boldsymbol{\theta}_k\|^2 = T$. The corresponding orthonormal expansion is obtained by scaling each $\boldsymbol{\theta}_k$ by $\sqrt{1/T}$; *i.e.*,

$$\phi_k(t) = \sqrt{\frac{1}{T}}\, e^{2\pi ikt/T}\text{rect}(\frac{t}{T}). \tag{5.22}$$

The Fourier series of an $\mathcal{L}_2$ function $\{v(t) : [-T/2, T/2] \to \mathbb{C}\}$ then becomes $\sum_k \alpha_k\phi_k(t)$ where $\alpha_k = \int v(t)\phi_k^*(t)\,dt = \langle \boldsymbol{v}, \boldsymbol{\phi}_k\rangle$. For any integer $n > 0$, let $\mathcal{S}_n$ be the $(2n+1)$-dimensional subspace spanned by the vectors $\{\boldsymbol{\phi}_k, -n \le k \le n\}$. From the projection theorem, the projection $\boldsymbol{v}_{|\mathcal{S}_n}$ of $\boldsymbol{v}$ on $\mathcal{S}_n$ is

$$\boldsymbol{v}_{|\mathcal{S}_n} = \sum_{k=-n}^{n} \langle \boldsymbol{v}, \boldsymbol{\phi}_k\rangle \boldsymbol{\phi}_k.$$

That is, the projection $\boldsymbol{v}_{|\mathcal{S}_n}$ is simply the approximation to $\boldsymbol{v}$ resulting from truncating the expansion to $-n \le k \le n$. The error in the approximation, $\boldsymbol{v}_{\perp\mathcal{S}_n} = \boldsymbol{v} - \boldsymbol{v}_{|\mathcal{S}_n}$, is orthogonal to all vectors in $\mathcal{S}_n$, and from the LS error property, $\boldsymbol{v}_{|\mathcal{S}_n}$ is the closest point in $\mathcal{S}_n$ to $\boldsymbol{v}$. As n increases, the subspace $\mathcal{S}_n$ becomes larger and $\boldsymbol{v}_{|\mathcal{S}_n}$ gets closer to $\boldsymbol{v}$ (*i.e.*, $\|\boldsymbol{v} - \boldsymbol{v}_{|\mathcal{S}_n}\|$ is nonincreasing).

As the analysis above applies equally well to any orthonormal sequence of functions, the general case can now be considered. The main result of interest is the following infinite-dimensional generalization of the projection theorem.

Theorem 5.3.2 (Infinite-dimensional projection). *Let $\{\boldsymbol{\phi}_m, 1{\le}m{<}\infty\}$ be a sequence of orthonormal vectors in $\mathcal{L}_2$, and let $\boldsymbol{v}$ be an arbitrary $\mathcal{L}_2$ vector. Then there exists a unique[7] $\mathcal{L}_2$ vector $\boldsymbol{u}$ such that $\boldsymbol{v} - \boldsymbol{u}$ is orthogonal to each $\boldsymbol{\phi}_m$ and*

$$\lim_{n\to\infty} \|\boldsymbol{u} - \sum_{m=1}^{n} \alpha_m\boldsymbol{\phi}_m\| \;=\; 0 \qquad \textit{where} \quad \alpha_m = \langle \boldsymbol{v}, \boldsymbol{\phi}_m\rangle \tag{5.23}$$

$$\|\boldsymbol{u}\|^2 \;=\; \sum |\alpha_m|^2. \tag{5.24}$$

Conversely, for any complex sequence $\{\alpha_m;\, 1{\le}m{\le}\infty\}$ such that $\sum_k |\alpha_k|^2 < \infty$, an $\mathcal{L}_2$ function $\boldsymbol{u}$ exists satisfying (5.23) and (5.24)

Remark: This theorem says that the orthonormal expansion $\sum_m \alpha_m\boldsymbol{\phi}_m$ converges in the $\mathcal{L}_2$ sense to an $\mathcal{L}_2$ function $\boldsymbol{u}$, which we later interpret as the projection of $\boldsymbol{v}$ onto the infinite-dimensional subspace $\mathcal{S}$ spanned by $\{\boldsymbol{\phi}_m, 1{\le}m{<}\infty\}$. For example, in the Fourier series case, the orthonormal functions span the subspace of $\mathcal{L}_2$ functions time-limited to $[-T/2, T/2]$, and $\mathbf{u}$ is then $v(t)\,\text{rect}(\frac{t}{T})$. The difference $v(t) - v(t)\,\text{rect}(\frac{t}{T})$ is then $\mathcal{L}_2$ equivalent to $\mathbf{0}$ over $[-T/2,\, T/2]$ and thus orthogonal to each $\boldsymbol{\phi}_m$.

[7]Recall that the vectors in the $\mathcal{L}_2$ class of functions are equivalence classes, so this uniqueness specifies only the equivalence class and not an individual function within that class.

Proof: Let $\mathcal{S}_n$ be the subspace spanned by $\{\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_n\}$. From the finite-dimensional projection theorem, the projection of $\boldsymbol{v}$ on $\mathcal{S}_n$ is then $\boldsymbol{v}_{|\mathcal{S}_n} = \sum_{k=1}^{n} \alpha_k \boldsymbol{\phi}_k$. From (5.17),

$$\|\boldsymbol{v}_{|\mathcal{S}_n}\|^2 = \sum_{k=1}^{n} |\alpha_k|^2 \qquad \text{where} \quad \alpha_k = \langle \boldsymbol{v}, \boldsymbol{\phi}_k \rangle. \tag{5.25}$$

This quantity is nondecreasing with n, and from Bessel's inequality, it is upperbounded by $\|\boldsymbol{v}\|^2$, which is finite since $\boldsymbol{v}$ is $\mathcal{L}_2$. It follows that for any n and any $m > n$,

$$\|\boldsymbol{v}_{|\mathcal{S}_m} - \boldsymbol{v}_{|\mathcal{S}_n}\|^2 = \sum_{n<|k|\le m} |\alpha_k|^2 \le \sum_{|k|>n} |\alpha_k|^2 \quad \stackrel{n\to\infty}{\longrightarrow} 0. \tag{5.26}$$

This says that the projections $\{\boldsymbol{v}_{|\mathcal{S}_n}; n \in \mathbb{Z}^+\}$ approach each other as $n \to \infty$ in terms of their energy difference.

A sequence whose terms approach each other is called a *Cauchy sequence*. The Riesz-Fischer theorem[8] is a central theorem of analysis stating that any Cauchy sequence of $\mathcal{L}_2$ waveforms has an $\mathcal{L}_2$ limit. Taking $\boldsymbol{u}$ to be this $\mathcal{L}_2$ limit, *i.e.*, $\boldsymbol{u} = \underset{n\to\infty}{\text{l.i.m.}}\ \boldsymbol{v}_{|\mathcal{S}_n}$, we have (5.23) and (5.24).[9]

Essentially the same use of the Riesz-Fischer theorem establishes (5.23) and (5.24) starting with the sequence $\alpha_1, \alpha_2, \dots$. □

Let $\mathcal{S}$ be the space of functions (or, more precisely, of equivalence classes) that can be represented as $\text{l.i.m.} \sum_k \alpha_k \phi_k(t)$ over all sequences $\alpha_1, \alpha_2, \dots$ such that $\sum_k |\alpha_k|^2 < \infty$. It can be seen that this is an inner product space. It is the *space spanned* by the orthonormal sequence $\{\boldsymbol{\phi}_k; k \in \mathbb{Z}\}$.

The following proof of the Fourier series theorem illustrates the use of the infinite dimensional projection theorem and infinite dimensional spanning sets.

Proof of Theorem 4.4.1: Let $\{v(t) : [-T/2, T/2]] \to \mathbb{C}\}$ be an arbitrary $\mathcal{L}_2$ function over $[-T/2, T/2]$. We have already seen that $v(t)$ is $\mathcal{L}_1$, that $\hat{v}_k = \frac{1}{T}\int v(t)e^{-2\pi ikt/T}\,dt$ exists and that $|\hat{v}_k| \le \int |v(t)|\,dt$ for all $k \in \mathbb{Z}$. From Theorem 5.3.2, there is an $\mathcal{L}_2$ function $u(t) = \text{l.i.m.} \sum_k \hat{v}_k e^{2\pi ikt/T}\text{rect}(t/T)$ such that $v(t) - u(t)$ is orthogonal to $\theta_k(t) = e^{2\pi ikt/T}\text{rect}(t/T)$ for each $k \in \mathbb{Z}$.

We now need an additional basic fact:[10] the above set of orthogonal functions $\{\theta_k(t) = e^{2\pi ikt/T}\text{rect}(t/T); k \in \mathbb{Z}\}$ span the space of $\mathcal{L}_2$ functions over $[-T/2, T/2]$, *i.e.*, there is no function of positive energy over $[-T/2, T/2]$ that is orthogonal to each $\theta_k(t)$. Using this fact, $v(t) - u(t)$ has zero energy and is equal to 0 a.e. Thus $v(t) = \text{l.i.m.} \sum_k \hat{v}_k e^{2\pi ikt/T}\text{rect}(t/T)$. The energy equation then follows from (5.24). The final part of the theorem follows from the final part of Theorem 5.3.2. □

As seen by the above proof, the infinite dimensional projection theorem can provide simple and intuitive proofs and interpretations of limiting arguments and the approximations suggested by those limits. The appendix uses this theorem to prove both parts of the Plancherel theorem, the sampling theorem, and the aliasing theorem.

Another, more pragmatic, use of the theorem lies in providing a uniform way to treat all orthonormal expansions. As in the above Fourier series proof, though, the theorem doesn't nec-

[8] See any text on real and complex analysis, such as Rudin[26].

[9] An inner product space in which all Cauchy sequences have limits is said to be *complete*, and is called a *Hilbert space*. Thus the Riesz-Fischer theorem states that $\mathcal{L}_2$ is a Hilbert space.

[10] Again, see any basic text on real and complex analysis.

essarily provide a simple characterization of the space spanned by the orthonormal set. Fortunately, however, knowing that the truncated sinusoids span $[-T/2, T/2]$ shows us, by duality, that the T-spaced sinc functions span the space of baseband-limited $\mathcal{L}_2$ functions. Similarly, both the T-spaced truncated and the sinc-weighted sinusoids span all of $\mathcal{L}_2$.

5.4 Summary

The theory of $\mathcal{L}_2$ waveforms, viewed as vectors in the inner product space known as signal space, has been developed. The most important consequence of this viewpoint is that *all* orthonormal expansions in $\mathcal{L}_2$ may be viewed in a common framework. The Fourier series is simply one example.

Another important consequence is that, as additional terms are added to a partial orthonormal expansion of an $\mathcal{L}_2$ waveform, the partial expansion changes by increasingly small amounts, approaching a limit in $\mathcal{L}_2$. A major reason for restricting attention to finite-energy waveforms (in addition to physical reality) is that as their energy gets used up in different degrees of freedom (*i.e.*, expansion coefficients), there is less energy available for other degrees of freedom, so that some sort of convergence must result. The $\mathcal{L}_2$ limit above simply make this intuition precise.

Another consequence is the realization that if $\mathcal{L}_2$ functions are represented by orthonormal expansions, or approximated by partial orthonormal expansions, then there is no further need to deal with sophisticated mathematical issues such as $\mathcal{L}_2$ equivalence. Of course, how the truncated expansions converge may be tricky mathematically, but the truncated expansions themselves are very simple and friendly.

5A Appendix: Supplementary material and proofs

The first part of the appendix uses the inner-product results of this chapter to prove the theorems about Fourier transforms in Chapter 4. The second part uses inner-products to prove the theorems in Chapter 4 about sampling and aliasing. The final part discusses prolate spheroidal waveforms; these provide additional insight about the degrees of freedom in a time/bandwidth region.

5A.1 The Plancherel theorem

Proof of Theorem 4.5.1 (Plancherel 1): The idea of the proof is to expand the time-waveform $\boldsymbol{u}$ into an orthonormal expansion for which the partial sums have known Fourier transforms; the $\mathcal{L}_2$ limit of these transforms is then identified as the $\mathcal{L}_2$ transform $\hat{\mathbf{u}}$ of $\boldsymbol{u}$.

First expand an arbitrary $\mathcal{L}_2$ function $u(t)$ in the T-spaced truncated sinusoid expansion, using $T = 1$. This expansion spans $\mathcal{L}_2$ and the orthogonal functions $e^{2\pi ikt}\mathrm{rect}(t-m)$ are orthonormal since $T = 1$. Thus the infinite dimensional projection, as specified by Theorem 5.3.2, is[11]

$$\begin{aligned} u(t) &= \operatorname*{l.i.m.}_{n\to\infty} u^{(n)}(t) \qquad \text{where} \quad u^{(n)}(t) = \sum_{m=-n}^{n}\sum_{k=-n}^{n} \hat{u}_{k,m}\theta_{k,m}(t), \\ \theta_{k,m}(t) &= e^{2\pi ikt}\mathrm{rect}(t-m) \qquad \text{and} \quad \hat{u}_{k,m} = \int u(t)\theta^*_{k,m}(t)\,dt. \end{aligned}$$

Since $u^{(n)}(t)$ is time-limited, it is $\mathcal{L}_1$, and thus has a continuous Fourier transform which is defined pointwise by

$$\hat{u}^{(n)}(f) = \sum_{m=-n}^{n}\sum_{k=-n}^{n} \hat{u}_{k,m}\psi_{k,m}(f), \tag{5.27}$$

where $\psi_{k,m}(f) = e^{2\pi ifm}\mathrm{sinc}(f-k)$ is the k,m term of the T-spaced sinc-weighted orthonormal set with $T = 1$. By the final part of Theorem 5.3.2, the sequence of vectors $\hat{\boldsymbol{u}}^{(n)}$ converges to an $\mathcal{L}_2$ vector $\hat{\boldsymbol{u}}$ (equivalence class of functions) denoted as the Fourier transform of $u(t)$ and satisfying

$$\lim_{n\to\infty} \|\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}^{(n)}\| = 0. \tag{5.28}$$

This must now be related to the functions $u_A(t)$ and $\hat{u}_A(f)$ in the theorem. First, for each integer $\ell > n$ define

$$\hat{u}^{(n,\ell)}(f) = \sum_{m=-n}^{n}\sum_{k=-\ell}^{\ell} \hat{u}_{k,m}\psi_{k,m}(f), \tag{5.29}$$

Since this is a more complete partial expansion than $\hat{u}^{(n)}(f)$,

$$\|\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}^{(n)}\| \geq \|\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}^{(n,\ell)}\|$$

In the limit $\ell \to \infty$, $\hat{\boldsymbol{u}}^{(n,\ell)}$ is the Fourier transform $\hat{u}_A(f)$ of $u_A(t)$ for $A = n + \frac{1}{2}$. Combining this with (5.28),

$$\lim_{n\to\infty} \|\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}_{n+\frac{1}{2}}\| = 0. \tag{5.30}$$

Finally, taking the limit of the finite dimensional energy equation,

$$\|\boldsymbol{u}^{(n)}\|^2 = \sum_{k=-n}^{n}\sum_{m=-n}^{n} |\hat{u}_{k,m}|^2 = \|\hat{\boldsymbol{u}}^{(n)}\|^2,$$

we get the $\mathcal{L}_2$ energy equation, $\|\boldsymbol{u}\|^2 = \|\hat{\boldsymbol{u}}\|^2$. This also shows that $\|\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}_A\|$ is monotonic in A so that (5.30) can be replaced by

$$\lim_{A\to\infty} \|\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}_{n+\frac{1}{2}}\| = 0.$$

□

[11] Note that $\{\theta_{k,m}; k,m \in \mathbb{Z}\}$ is a countable set of orthonormal vectors, and they have been arranged in an order so that, for all $n \in \mathbb{Z}^+$, all terms with $|k| \leq n$ and $|m| \leq n$ come before all other terms.

Proof of Theorem 4.5.2 (Plancherel 2): By time/frequency duality with Theorem 4.5.1, we see that $\text{l.i.m.}_{B\to\infty} u_B(t)$ exists and we denote this by $\mathcal{F}^{-1}(\hat{u}(f))$. The only remaining thing to prove is that this inverse transform is $\mathcal{L}_2$ equivalent to the original $u(t)$. Note first that the Fourier transform of $\theta_{0,0}(t) = \text{rect}(t)$ is $\text{sinc}(f)$ and that the inverse transform, defined as above, is $\mathcal{L}_2$ equivalent to $\text{rect}(t)$. By time and frequency shifts, we see that $u^{(n)}(t)$ is the inverse transform, defined as above, of $\hat{u}^{(n)}(f)$. It follows that $\lim_{n\to\infty} \|\mathcal{F}^{-1}(\hat{\mathbf{u}}) - \boldsymbol{u}^{(n)}\| = 0$, so we see that $\|\mathcal{F}^{-1}(\hat{\mathbf{u}}) - \boldsymbol{u}\| = 0$. $\square$

As an example of the Plancherel theorem, let $h(t)$ be 1 on the rationals in (0, 1) and be zero elsewhere. Then $\boldsymbol{h}$ is both $\mathcal{L}_1$ and $\mathcal{L}_2$ and has a Fourier transform $\hat{h}(f) = 0$ which is continuous, $\mathcal{L}_1$, and $\mathcal{L}_2$. The inverse transform is also 0 and equal to $h(t)$ a.e.

The function $h(t)$ above is in some sense trivial since it is $\mathcal{L}_2$ equivalent to the zero function. The next example to be discussed is $\mathcal{L}_2$, nonzero only on and $\mathcal{L}_1$, but all members of its equivalence class are discontinuous everywhere and unbounded in every interval.

We now discuss an example of a real $\mathcal{L}_2$ function that is nonzero only on the interval $(0, 1)$. This function is $\mathcal{L}_1$, has a continuous Fourier transform, but all functions in its equivalence class are discontinuous everywhere and unbounded over every open interval within $(0, 1)$. This example will illustrate how truly Bizarre functions can have nice Fourier transforms and *vice versa*. It will also be used later to illustrate some properties of $\mathcal{L}_2$ functions.

Example 5A.1 (A Bizarre $\mathcal{L}_2$ and $\mathcal{L}_1$ function)). List the rationals in (0,1) by increasing denominator, *i.e.*, as a_1=1/2, a_2=1/3, a_3=2/3, a_4=1/4, a_5=3/4, a_6=1/5, $\cdots$. Define

$$g_n(t) = \begin{cases} 1 & \text{for} \quad a_n \le t < a_n + 2^{-n-1} \\ 0 & \text{elsewhere,} \end{cases}$$

$$g(t) = \sum_{n=1}^{\infty} g_n(t).$$

Thus $g(t)$ is a sum of rectangular functions, one for each rational number, with the width of the function going to zero rapidly with the index of the rational number (see Figure 5.3). The integral of $g(t)$ can be calculated as

$$\int_0^1 g(t)\,dt = \sum_{n=1}^{\infty} \int g_n(t)\,dt = \sum_{n=1}^{\infty} 2^{-n-1} = \frac{1}{2}.$$

Thus $g(t)$ is an $\mathcal{L}_1$ function as illustrated in Figure 5.3.

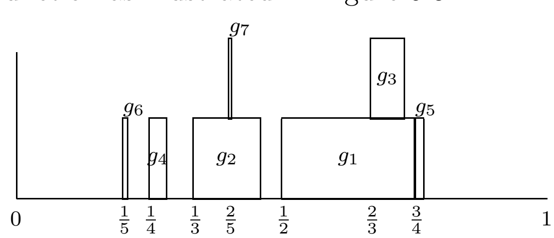


Figure 5.3: First 7 terms of $\sum_i g_i(t)$

Consider the interval $[\frac{2}{3}, \frac{2}{3} + \frac{1}{8})$ corresponding to the rectangle g_3 in the figure. Since the rationals are dense over the real line, there is a rational, say a_j, in the interior of this interval, and thus a new interval starting at a_j over which g_1, g_3, and g_j all have value 1; thus $g(t) \ge 3$ within this new interval. Moreover, this same argument can be repeated within this new interval, which again contains a rational, say $a_{j'}$. Thus there is an interval starting at $a_{j'}$ where g_1, g_3, g_j, and $g_{j'}$ are 1 and thus $g(t) \geq 4$.

Iterating this argument, we see that $[\frac{2}{3}, \frac{2}{3} + \frac{1}{8})$ contains subintervals within which $g(t)$ takes on arbitrarily large values. In fact, by taking the limit $a_1, a_3, a_j, a_{j'}, \dots$, we find a limit point a for which $g(a) = \infty$. Moreover, we can apply the same argument to any open interval within $(0, 1)$ to show that $g(t)$ takes on infinite values within that interval.[12] More explicitly, for every $\varepsilon > 0$ and every $t \in (0, 1)$, there is a t' such that $|t - t'| < \varepsilon$ and $g(t') = \infty$. This means that $g(t)$ is discontinuous and unbounded in each region of $(0, 1)$.

The function $g(t)$ is also in $\mathcal{L}_2$ as seen below:

$$\int_0^1 g^2(t)\,dt = \sum_{n,m} \int g_n(t) g_m(t)\,dt \tag{5.31}$$

$$= \sum_n \int g_n^2(t)\,dt + 2 \sum_n \sum_{m=n+1}^{\infty} \int g_n(t)\, g_m(t)\,dt \tag{5.32}$$

$$\leq \frac{1}{2} + 2 \sum_n \sum_{m=n+1}^{\infty} \int g_m(t)\,dt = \frac{3}{2}, \tag{5.33}$$

where in (5.33) we have used the fact that $g_n^2(t) = g_n(t)$ in the first term and $g_n(t) \leq 1$ in the second term.

In conclusion, $g(t)$ is both $\mathcal{L}_1$ and $\mathcal{L}_2$ but is discontinuous everywhere and takes on infinite values at points in every interval. The transform $\hat{g}(f)$ is continuous and $\mathcal{L}_2$ but not $\mathcal{L}_1$. The inverse transform, $g_B(t)$ of $\hat{g}(f)\text{rect}(\frac{f}{2B})$ is continuous, and converges in $\mathcal{L}_2$ to $g(t)$ as $B \to \infty$. For $B = 2^k$, the function $g_B(t)$ is roughly approximated by $g_1(t) + \cdots + g_k(t)$, all somewhat rounded at the edges.

This is a nice example of a continuous function $\hat{g}(f)$ which has a bizarre inverse Fourier transform. Note that $g(t)$ and the function $h(t)$ that is 1 on the rationals in (0,1)and 0 elsewhere are both discontinuous everywhere in (0,1). However, the function $h(t)$ is 0 a.e., and thus is weird only in an artificial sense. For most purposes, it is the same as the zero function. The function $g(t)$ is weird in a more fundamental sense. It cannot be made respectable by changing it on a countable set of points.

One should not conclude from this example that intuition cannot be trusted, or that it is necessary to take a few graduate math courses before feeling comfortable with functions. One can conclude, however, that the simplicity of the results about Fourier transforms and orthonormal expansions for $\mathcal{L}_2$ functions is truly extraordinary in view of the bizarre functions included in the $\mathcal{L}_2$ class.

In summary, Plancherel's theorem has taught us two things. First, Fourier transforms and inverse transforms exist for *all* $\mathcal{L}_2$ functions. Second, finite-interval and finite-bandwidth approximations become arbitrarily good (in the sense of $\mathcal{L}_2$ convergence) as the interval or the bandwidth becomes large.

[12] The careful reader will observe that $g(t)$ is not really a function $\mathbb{R} \to \mathbb{R}$, but rather a function from $\mathbb{R}$ to the extended set of real values including ∞ and $-\infty$. The set of t on which $g(t) = \infty$ has zero measure and this can be ignored in Lebesgue integration. Do not confuse a function that takes on an infinite value at some isolated point with a unit impulse at that point. The first integrates to 0 around the singularity, whereas the second is a generalized function that by definition integrates to 1.

5A.2 The sampling and aliasing theorems

This section contains proofs of the sampling and aliasing theorems. The proofs are important and not available elsewhere in this form. However, they involve some careful mathematical analysis that might be beyond the interest and/or background of many students.

Proof of Theorem 4.6.2 (Sampling Theorem): Let $\hat{u}(f)$ be an $\mathcal{L}_2$ function that is zero outside of $[-\mathsf{W}, \mathsf{W}]$. From Theorem 4.3.2, $\hat{u}(f)$ is $\mathcal{L}_1$, so by Lemma 4.5.1,

$$u(t) = \int_{-\mathsf{W}}^{\mathsf{W}} \hat{u}(f) e^{2\pi i f t}\, df \tag{5.34}$$

holds at each $t \in \mathbb{R}$. We want to show that the sampling theorem expansion also holds at each t. By the DTFT theorem,

$$\hat{u}(f) = \operatorname*{l.i.m.}_{\ell\to\infty}\ \hat{u}^{(\ell)}(f), \quad \text{where} \quad \hat{u}^{(\ell)}(f) = \sum_{k=-\ell}^{\ell} u_k \hat{\phi}_k(f) \tag{5.35}$$

and where $\hat{\phi}_k(f) = e^{-2\pi i k f/(2\mathsf{W})}\mathrm{rect}\left(\frac{f}{2\mathsf{W}}\right)$ and

$$u_k = \frac{1}{2\mathsf{W}} \int_{-\mathsf{W}}^{\mathsf{W}} \hat{u}(f) e^{2\pi i k f/(2\mathsf{W})}\, df. \tag{5.36}$$

Comparing (5.34) and (5.36), we see as before that $2\mathsf{W}u_k = u(\frac{k}{2\mathsf{W}})$. The functions $\hat{\phi}_k(f)$ are in $\mathcal{L}_1$, so the finite sum $\hat{u}^{(\ell)}(f)$ is also in $\mathcal{L}_1$. Thus the inverse Fourier transform

$$u^{(\ell)}(t) = \int \hat{u}^{(\ell)}(f)\, df = \sum_{k=-\ell}^{\ell} u(\frac{k}{2\mathsf{W}})\, \mathrm{sinc}(2\mathsf{W}t - k)$$

is defined pointwise at each t. For each $t \in \mathbb{R}$, the difference $u(t) - u^{(\ell)}(t)$ is then

$$u(t) - u^{(\ell)}(t) = \int_{-\mathsf{W}}^{\mathsf{W}} [\hat{u}(f) - \hat{u}^{(\ell)}(f)] e^{2\pi i f t}\, df.$$

This integral can be viewed as the inner product of $\hat{u}(f) - \hat{u}^{(\ell)}(f)$ and $e^{-2\pi i f t}\mathrm{rect}[\frac{f}{2\mathsf{W}}]$, so, by the Schwarz inequality, we have

$$|u(t) - u^{(\ell)}(t)| \le \sqrt{2\mathsf{W}}\|\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}^{(\ell)}\|.$$

From the $\mathcal{L}_2$ convergence of the DTFT, the right side approaches 0 as $\ell \to \infty$, so the left side also approaches 0 for each t, establishing pointwise convergence. □

Proof of Theorem 4.6.3 (Sampling theorem for transmission): For a given W, assume that the sequence $\{u(\frac{k}{2\mathsf{W}}); k \in \mathbb{Z}\}$ satisfies $\sum_k |u(\frac{k}{2\mathsf{W}})|^2 < \infty$. Define $u_k = \frac{1}{2\mathsf{W}}u(\frac{k}{2\mathsf{W}})$ for each $k \in \mathbb{Z}$. By the DTFT theorem, there is a frequency function $\hat{u}(f)$, nonzero only over $[-\mathsf{W}, \mathsf{W}]$, that satisfies (4.60) and (4.61). By the sampling theorem, the inverse transform $u(t)$ of $\hat{u}(f)$ has the desired properties. □

Proof of Theorem 4.7.1 (Aliasing theorem): We start by separating $\hat{u}(f)$ into frequency slices $\{\hat{v}_m(f); m \in \mathbb{Z}\}$,

$$\hat{u}(f) = \sum_m \hat{v}_m(f), \quad \text{where} \quad \hat{v}_m(f) = \hat{u}(f)\mathrm{rect}^{\dagger}(fT - m). \tag{5.37}$$

The function $\mathrm{rect}^\dagger(f)$ is defined to equal 1 for $-\frac{1}{2} < f \le \frac{1}{2}$ and 0 elsewhere. It is $\mathcal{L}_2$ equivalent to $\mathrm{rect}(f)$, but gives us pointwise equality in (5.37). For each positive integer n, define $\hat{v}^{(n)}(f)$ as

$$\hat{v}^{(n)}(f) = \sum_{m=-n}^{n} \hat{v}_m(f) = \begin{cases} \hat{u}(f) & \text{for } \frac{2n-1}{2T} < f \le \frac{2n+1}{2T} \\ 0 & \text{elsewhere.} \end{cases} \tag{5.38}$$

It is shown in Exercise 5.16 that the given conditions on $\hat{u}(f)$ imply that $\hat{u}(f)$ is in $\mathcal{L}_1$. In conjunction with (5.38), this implies that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |\hat{u}(f) - \hat{v}^{(n)}(f)|\, df = 0.$$

Since $\hat{u}(f) - \hat{v}^{(n)}(f)$ is in $\mathcal{L}_1$, the inverse transform at each t satisfies

$$\begin{aligned} \left|u(t) - v^{(n)}(t)\right| &= \left|\int_{-\infty}^{\infty} [\hat{u}(f) - \hat{v}^{(n)}(f)] e^{2\pi i f t}\, df\right| \\ &\le \int_{-\infty}^{\infty} \left|\hat{u}(f) - \hat{v}^{(n)}(f)\right|\, df = \int_{|f|\ge(2n+1)/(2T)} |\hat{u}(f)|\, df. \end{aligned}$$

Since $\hat{u}(f)$ is in $\mathcal{L}_1$, the final integral above approaches 0 with increasing n. Thus, for each t, we have

$$u(t) = \lim_{n\to\infty} v^{(n)}(t). \tag{5.39}$$

Next define $\hat{s}_m(f)$ as the frequency slice $\hat{v}_m(f)$ shifted down to baseband, *i.e.*,

$$\hat{s}_m(f) = \hat{v}_m(f - \frac{m}{T}) = \hat{u}(f - \frac{m}{T})\mathrm{rect}^\dagger(fT). \tag{5.40}$$

Applying the sampling theorem to $v_m(t)$, we get

$$v_m(t) = \sum_k v_m(kT)\, \mathrm{sinc}(\frac{t}{T} - k) e^{2\pi i m t/T}. \tag{5.41}$$

Applying the frequency shift relation to (5.40), we see that $s_m(t) = v_m(t)e^{-2\pi i f t}$, and thus

$$s_m(t) = \sum_k v_m(kT)\, \mathrm{sinc}(\frac{t}{T} - k). \tag{5.42}$$

Now define $\hat{s}^{(n)}(f) = \sum_{m=-n}^{n} \hat{s}_m(f)$. From (5.40), we see that $\hat{s}^{(n)}(f)$ is the aliased version of $\hat{v}^{(n)}(f)$, as illustrated in Figure 4.10. The inverse transform is then

$$s^{(n)}(t) = \sum_{k=-\infty}^{\infty} \sum_{m=-n}^{n} v_m(kT)\, \mathrm{sinc}(\frac{t}{T} - k). \tag{5.43}$$

We have interchanged the order of summation, which is valid since the sum over m is finite. Finally, define $\hat{s}(f)$ to be the "folded" version of $\hat{u}(f)$ summing over all m, *i.e.*,

$$\hat{s}(f) = \underset{n\to\infty}{\mathrm{l.i.m.}}\ \hat{s}^{(n)}(f). \tag{5.44}$$

Exercise 5.16 shows that this limit converges in the $\mathcal{L}_2$ sense to an $\mathcal{L}_2$ function $\hat{s}(f)$. Exercise 4.38 provides an example where $\hat{s}(f)$ is not in $\mathcal{L}_2$ if the condition $\lim_{|f|\to\infty} \hat{u}(f)|f|^{1+\varepsilon} = 0$ is not satisfied.

Since $\hat{s}(f)$ is in $\mathcal{L}_2$ and is 0 outside $[-\frac{1}{2T}, \frac{1}{2T}]$, the sampling theorem shows that the inverse transform $s(t)$ satisfies

$$s(t) = \sum_k s(kT)\mathrm{sinc}(\frac{t}{T} - k). \tag{5.45}$$

Combining this with (5.43),

$$s(t) - s^{(n)}(t) = \sum_k \left[s(kT) - \sum_{m=-n}^{n} v_m(kT) \right] \mathrm{sinc}(\frac{t}{T} - k). \tag{5.46}$$

From (5.44), we see that $\lim_{n\to\infty} \|\boldsymbol{s} - \boldsymbol{s}^{(n)}\| = 0$, and thus

$$\lim_{n\to\infty} \sum_k |s(kT) - v^{(n)}(kT)|^2 = 0.$$

This implies that $s(kT) = \lim_{n\to\infty} v^{(n)}(kT)$ for each integer k. From (5.39), we also have $u(kT) = \lim_{n\to\infty} v^{(n)}(kT)$, and thus $s(kT) = u(kT)$ for each $k \in \mathbb{Z}$.

$$s(t) = \sum_k u(kT)\mathrm{sinc}(\frac{t}{T} - k). \tag{5.47}$$

This shows that (5.44) implies (5.47). Since $s(t)$ is in $\mathcal{L}_2$, it follows that $\sum_k |u(kT)|^2 < \infty$. Conversely, (5.47) defines a unique $\mathcal{L}_2$ function, and thus its Fourier transform must be $\mathcal{L}_2$ equivalent to $\hat{s}(f)$ as defined in (5.44). □

5A.3 Prolate spheroidal waveforms

The prolate spheroidal waveforms (see [29]) are a set of orthonormal functions that provide a more precise way to view the degree-of-freedom arguments of Section 4.7.2. For each choice of baseband bandwidth W and time interval $[-T/2, T/2]$, these functions form an orthonormal set $\{\boldsymbol{\phi}_0(t), \boldsymbol{\phi}_1(t), \dots,\}$ of real $\mathcal{L}_2$ functions time-limited to $[-T/2, T/2]$. In a sense to be described, these functions have the maximum possible energy in the frequency band $(-\mathsf{W}, \mathsf{W})$ subject to their constraint to $[-T/2, T/2]$.

To be more precise, for each $n \geq 0$ let $\hat{\phi}_n(f)$ be the Fourier transform of $\phi_n(t)$, and define

$$\hat{\theta}_n(f) = \begin{cases} \hat{\phi}_n(f) & \text{for } -\mathsf{W}<t<\mathsf{W} \\ 0 & \text{elsewhere.} \end{cases} \tag{5.48}$$

That is, $\theta_n(t)$ is $\phi_n(t)$ truncated in frequency to $(-\mathsf{W}, \mathsf{W})$; equivalently, $\theta_n(t)$ may be viewed as the result of passing $\phi_n(t)$ through an ideal low-pass filter.

The function $\phi_0(t)$ is chosen to be the normalized function $\phi_0(t) : (-T/2, T/2) \to \mathbb{R}$ that maximizes the energy in $\theta_0(t)$. We will not show how to solve this optimization problem. However, $\phi_0(t)$ turns out to resemble $\sqrt{1/T}\,\mathrm{rect}(\frac{t}{T})$, except that it is rounded at the edges to reduce the out-of-band energy.

Similarly, for each $n > 0$, the function $\phi_n(t)$ is chosen to be the normalized function $\{\phi_n(t) : (-T/2, T/2) \to \mathbb{R}\}$ that is orthonormal to $\phi_m(t)$ for each $m < n$ and, subject to this constraint, maximizes the energy in $\theta_n(t)$.

Finally, define $\lambda_n = \|\boldsymbol{\theta}_n\|^2$. It can be shown that $1 > \lambda_0 > \lambda_1 > \cdots$. We interpret λ_n as the fraction of energy in $\boldsymbol{\phi}_n$ that is baseband-limited to $(-\mathsf{W}, \mathsf{W})$. The number of degrees of freedom in $(-T/2, T/2), (-\mathsf{W}, \mathsf{W})$ is then reasonably defined as the largest n for which λ_n is close to 1.

The values λ_n depend on the product $T\mathsf{W}$, so they can be denoted by $\lambda_n(T\mathsf{W})$. The main result about prolate spheroidal wave functions, which we do not prove, is that for any $\varepsilon > 0$,

$$\lim_{T\mathsf{W}\to\infty} \lambda_n(T\mathsf{W}) = \begin{cases} 1 & \text{for } n < 2T\mathsf{W}(1-\varepsilon) \\ 0 & \text{for } n > 2T\mathsf{W}(1+\varepsilon). \end{cases}$$

This says that when $T\mathsf{W}$ is large, there are close to $2T\mathsf{W}$ orthonormal functions for which most of the energy in the time-limited function is also frequency-limited, but there are not significantly more orthonormal functions with this property.

The prolate spheroidal wave functions $\phi_n(t)$ have many other remarkable properties, of which we list a few:

- For each n, $\phi_n(t)$ is continuous and has n zero crossings.
- $\phi_n(t)$ is even for n even and odd for n odd.
- $\theta_n(t)$ is an orthogonal set of functions.
- In the interval $(-T/2, T/2)$, $\theta_n(t) = \lambda_n \phi_n(t)$.

5.E Exercises

5.1. (basis) Prove Theorem 5.1.1 by first suggesting an algorithm that establishes the first item and then an algorithm to establish the second item.

5.2. Show that the $\mathbf{0}$ vector can be part of a spanning set but cannot be part of a linearly indepenendent set.

5.3. (basis) Prove that if a set of n vectors *uniquely* spans a vector space $\mathcal{V}$, in the sense that each $\boldsymbol{v} \in \mathcal{V}$ has a unique representation as a linear combination of the n vectors, then those n vectors are linearly independent and $\mathcal{V}$ is an n-dimensional space.

5.4. ($\mathbb{R}^2$) (a) Show that the vector space $\mathbb{R}^2$ with vectors $\{\boldsymbol{v} = (v_1, v_2)\}$ and inner product $\langle \boldsymbol{v}, \boldsymbol{u}\rangle = v_1u_1 + v_2u_2$ satisfies the axioms of an inner product space.

(b) Show that, in the Euclidean plane, the length of $\boldsymbol{v}$ (*i.e.*, the distance from $\mathbf{0}$ to $\boldsymbol{v}$ is $\|\boldsymbol{v}\|$.

(c) Show that the distance from $\boldsymbol{v}$ to $\boldsymbol{u}$ is $\|\boldsymbol{v} - \boldsymbol{u}\|$.

(d) Show that $\cos(\angle(\boldsymbol{v}, \boldsymbol{u})) = \frac{\langle \boldsymbol{v}, \boldsymbol{u}\rangle}{\|\boldsymbol{v}\|\,\|\boldsymbol{u}\|}$; assume that $\|\boldsymbol{u}\| > 0$ and $\|\boldsymbol{v}\| > 0$.

(e) Suppose that the definition of the inner product is now changed to $\langle \boldsymbol{v}, \boldsymbol{u}\rangle = v_1u_1 + 2v_2u_2$. Does this still satisfy the axioms of an inner product space? Does the length formula and the angle formula still correspond to the usual Euclidean length and angle?

5.5. Consider $\mathbb{C}^n$ and define $\langle \boldsymbol{v}, \boldsymbol{u}\rangle$ as $\sum_{j=1}^n c_j v_j u_j^*$ where $c_1, \dots, c_n$ are complex numbers. For each of the following cases, determine whether $\mathbb{C}^n$ must be an inner product space and explain why or why not.

(a) The c_j are all equal to the same positive real number.

(b) The c_j are all positive real numbers.

(c) The c_j are all non-negative real numbers.

(d) The c_j are all equal to the same nonzero complex number.

(e) The c_j are all nonzero complex numbers.

5.6. (Triangle inequality) Prove the triangle inequality, (5.10). Hint: Expand $\|\boldsymbol{v}+\boldsymbol{u}\|^2$ into four terms and use the Schwarz inequality on each of the two cross terms.

5.7. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be orthonormal vectors in $\mathbb{C}^n$ and let $\boldsymbol{w} = w_u\boldsymbol{u} + w_v\boldsymbol{v}$ and $\boldsymbol{x} = x_u\boldsymbol{u} + x_v\boldsymbol{v}$ be two vectors in the subspace spanned by $\boldsymbol{u}$ and $\boldsymbol{v}$.

(a) Viewing $\boldsymbol{w}$ and $\boldsymbol{x}$ as vectors in the subspace $\mathbb{C}^2$, find $\langle \boldsymbol{w}, \boldsymbol{x}\rangle$.

(b) Now view $\boldsymbol{w}$ and $\boldsymbol{x}$ as vectors in $\mathbb{C}^n$, *e.g.*, $\boldsymbol{w} = (w_1, \dots, w_n)$ where $w_j = w_u u_j + w_v v_j$ for $1 \le j \le n$. Calculate $\langle \boldsymbol{w}, \boldsymbol{x}\rangle$ this way and show that the answer agrees with that in part (a).

5.8. ($\mathcal{L}_2$ inner product) Consider the vector space of $\mathcal{L}_2$ functions $\{u(t) : \mathbb{R} \to \mathbb{C}\}$. Let $\boldsymbol{v}$ and $\boldsymbol{u}$ be two vectors in this space represented as $v(t)$ and $u(t)$. Let the inner product be defined by

$$\langle \boldsymbol{v}, \boldsymbol{u}\rangle = \int_{-\infty}^{\infty} v(t)u^*(t)\,dt.$$

(a) Assume that $u(t) = \sum_{k,m} \hat{u}_{k,m}\theta_{k,m}(t)$ where $\{\theta_{k,m}(t)\}$ is an orthogonal set of functions each of energy T. Assume that $v(t)$ can be expanded similarly. Show that

$$\langle \boldsymbol{u}, \boldsymbol{v}\rangle = T\sum_{k,m} \hat{u}_{k,m}\hat{v}^*_{k,m}.$$

(b) Show that $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ is finite. Do not use the Schwarz inequality, because the purpose of this exercise is to show that $\mathcal{L}_2$ is an inner product space, and the Schwarz inequality is based on the assumption of an inner product space. Use the result in (a) along with the properties of complex numbers (you can use the Schwarz inequality for the one dimensional vector space $\mathbb{C}^1$ if you choose).

(c) Why is this result necessary in showing that $\mathcal{L}_2$ is an inner product space?

5.9. ($\mathcal{L}_2$ inner product) Given two waveforms $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{L}_2$, let $\mathcal{V}$ be the set of all waveforms $\boldsymbol{v}$ that are equi-distant from $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Thus

$$\mathcal{V} = \Big\{ \boldsymbol{v} : \|\boldsymbol{v} - \boldsymbol{u}_1\| = \|\boldsymbol{v} - \boldsymbol{u}_2\| \Big\}.$$

(a) Is $\mathcal{V}$ a vector sub-space of $\mathcal{L}_2$?

(b) Show that

$$\mathcal{V} = \Big\{ \boldsymbol{v} : \Re\left(\langle \boldsymbol{v}, \boldsymbol{u}_2 - \boldsymbol{u}_1 \rangle\right) = \frac{\|\boldsymbol{u}_2\|^2 - \|\boldsymbol{u}_1\|^2}{2} \Big\}.$$

(c) Show that $(\boldsymbol{u}_1 + \boldsymbol{u}_2)/2 \in \mathcal{V}$

(d) Give a geometric interpretation for $\mathcal{V}$.

5.10. (sampling) For any $\mathcal{L}_2$ function $\{u(t) : [-\mathsf{W}, \mathsf{W}] \to \mathbb{C}\}$ and any t, let $a_k = u(\frac{k}{2\mathsf{W}})$ and let $b_k = \operatorname{sinc}(2\mathsf{W}t - k)$. Show that $\sum_k |a_k|^2 < \infty$ and $\sum_k |b_k|^2 < \infty$. Use this to show that $\sum_k |a_k b_k| < \infty$. Use this to show that the sum in the sampling equation (4.65) converges for each t.

5.11. (projection) Consider the following set of functions $\{u_m(t)\}$ for integer $m \geq 0$:

$$u_0(t) = \begin{cases} 1, & 0 \leq t < 1; \\ 0 & \text{otherwise.} \end{cases}$$

$$\vdots$$

$$u_m(t) = \begin{cases} 1, & 0 \leq t < 2^{-m}; \\ 0 & \text{otherwise.} \end{cases}$$

$$\vdots$$

Consider these functions as vectors $\boldsymbol{u}_0, \boldsymbol{u}_1 \dots$, over real $\mathcal{L}_2$ vector space. Note that $\boldsymbol{u}_0$ is normalized; we denote it as $\boldsymbol{\phi}_0 = \boldsymbol{u}_0$.

(a) Find the projection $(\boldsymbol{u}_1)_{|\boldsymbol{\phi}_0}$ of $\boldsymbol{u}_1$ on $\boldsymbol{\phi}_0$, find the perpendicular $(\boldsymbol{u}_1)_{\perp \boldsymbol{\phi}_0}$, and find the normalized form $\boldsymbol{\phi}_1$ of $(\boldsymbol{u}_1)_{\perp \boldsymbol{\phi}_0}$. Sketch each of these as functions of t.

(b) Express $u_1(t - 1/2)$ as a linear combination of $\boldsymbol{\phi}_0$ and $\boldsymbol{\phi}_1$. Express (in words) the subspace of real $\mathcal{L}_2$ spanned by $u_1(t)$ and $u_1(t - 1/2)$. What is the subspace $\mathcal{S}_1$ of real $\mathcal{L}_2$ spanned by $\boldsymbol{\phi}_0$ and $\boldsymbol{\phi}_1$?

(c) Find the projection $(\boldsymbol{u}_2)_{|\mathcal{S}_1}$ of $\boldsymbol{u}_2$ on $\mathcal{S}_1$, find the perpendicular $(\boldsymbol{u}_2)_{\perp \mathcal{S}_1}$, and find the normalized form of $(\boldsymbol{u}_2)_{\perp \mathcal{S}_1}$. Denote this normalized form as $\boldsymbol{\phi}_{2,0}$; it will be clear shortly why a double subscript is used here. Sketch $\boldsymbol{\phi}_{2,0}$ as a function of t.

(d) Find the projection of $u_2(t - 1/2)$ on $\mathcal{S}_1$ and find the perpendicular $u_2(t - 1/2)_{\perp \mathcal{S}_1}$. Denote the normalized form of this perpendicular by $\boldsymbol{\phi}_{2,1}$. Sketch $\boldsymbol{\phi}_{2,1}$ as a function of t and explain why $\langle \boldsymbol{\phi}_{2,0}, \boldsymbol{\phi}_{2,1} \rangle = 0$.

(e) Express $u_2(t-1/4)$ and $u_2(t-3/4)$ as linear combinations of $\{\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \boldsymbol{\phi}_{2,0}, \boldsymbol{\phi}_{2,1}\}$. Let $\mathcal{S}_2$ be the subspace of real $\mathcal{L}_2$ spanned by $\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \boldsymbol{\phi}_{2,0}$, $\boldsymbol{\phi}_{2,1}$ and describe this subspace in words.

(f) Find the projection $(\boldsymbol{u}_3)_{|\mathcal{S}_2}$ of $\boldsymbol{u}_3$ on $\mathcal{S}_2$, find the perpendicular $(\boldsymbol{u}_2)_{\perp\mathcal{S}_1}$, and find its normalized form, $\boldsymbol{\phi}_{3,0}$. Sketch $\boldsymbol{\phi}_{3,0}$ as a function of t.

(g) For $j = 1, 2, 3$, find $u_3(t-j/4)_{\perp\mathcal{S}_2}$ and find its normalized form $\boldsymbol{\phi}_{3,j}$. Describe the subspace $\mathcal{S}_3$ spanned by $\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \boldsymbol{\phi}_{2,0}, \boldsymbol{\phi}_{2,1}, \boldsymbol{\phi}_{3,0}, \dots, \boldsymbol{\phi}_{3,3}$.

(h) Consider iterating this process to form $\mathcal{S}_4, \mathcal{S}_5, \dots$. What is the dimension of $\mathcal{S}_m$? Describe this subspace. Describe the projection of an arbitrary real $\mathcal{L}_2$ function constrained to the interval [0,1) on $\mathcal{S}_m$.

5.12. (Orthogonal subspaces) For any subspace $\mathcal{S}$ of an inner product space $\mathcal{V}$, define $\mathcal{S}^\perp$ as the set of vectors $\boldsymbol{v} \in \mathcal{V}$ that are orthogonal to all $\boldsymbol{w} \in \mathcal{S}$.

(a) Show that $\mathcal{S}^\perp$ is a subspace of $\mathcal{V}$.

(b) Assuming that $\mathcal{S}$ is finite dimensional, show that any $\boldsymbol{u} \in \mathcal{V}$ can be uniquely decomposed into $\boldsymbol{u} = \boldsymbol{u}_{|\mathcal{S}} + \boldsymbol{u}_{\perp S}$ where $\boldsymbol{u}_{|\mathcal{S}} \in \mathcal{S}$ and $\boldsymbol{u}_{\perp S} \in \mathcal{S}^\perp$.

(c) Assuming that $\mathcal{V}$ is finite dimensional, show that $\mathcal{V}$ has an orthonormal basis where some of the basis vectors form a basis for $\boldsymbol{S}$ and the remaining basis vectors form a basis for $\mathcal{S}^\perp$.

5.13. (Orthonormal expansion) Expand the function $\operatorname{sinc}(3t/2)$ as an orthonormal expansion in the set of functions $\{\operatorname{sinc}(t-n)\,;\ -\infty < n < \infty\}$.

5.14. (bizarre function) (a) Show that the pulses $g_n(t)$ in Example 5A.1 of Section 5A.1 overlap each other either completely or not at all.

(b) Modify each pulse $g_n(t)$ to $h_n(t)$ as follows: Let $h_n(t) = g_n(t)$ if $\sum_{i=1}^{n-1} g_i(t)$ is even and let $h_n(t) = -g_n(t)$ if $\sum_{i=1}^{n-1} g_i(t)$ is odd. Show that $\sum_{i=1}^{n} h_i(t)$ is bounded between 0 and 1 for each $t \in (0,1)$ and each $n \geq 1$.

(c) Show that there are a countably infinite number of points t at which $\sum_n h_n(t)$ does not converge.

5.15. (Parseval) Prove Parseval's relation, (4.44) for $\mathcal{L}_2$ functions. Use the same argument as used to establish the energy equation in the proof of Plancherel's theorem.

5.16. (Aliasing theorem) Assume that $\hat{u}(f)$ is $\mathcal{L}_2$ and $\lim_{|f|\to\infty} \hat{u}(f)|f|^{1+\varepsilon} = 0$ for some $\varepsilon > 0$.

(a) Show that for large enough $A > 0$, $|\hat{u}(f)| \leq |f|^{-1-\varepsilon}$ for $|f| > A$.

(b) Show that $\hat{u}(f)$ is $\mathcal{L}_1$. Hint: for the A above, split the integral $\int |\hat{u}(f)|\,df$ into one integral for $|f| > A$ and another for $|f| \leq A$.

(c) Show that, for $T = 1$, $\hat{s}(f)$ as defined in (5.44), satisfies

$$|\hat{s}(f)| \leq \sqrt{(2A+1)\sum\nolimits_{|m|\leq A} |\hat{u}(f+m)|^2} + \sum\nolimits_{m\geq A} m^{-1-\varepsilon}.$$

(d) Show that $\hat{s}(f)$ is $\mathcal{L}_2$ for $T = 1$. Use scaling to show that $\hat{s}(f)$ is $\mathcal{L}_2$ for any $T > 0$.