12. INNER PRODUCT SPACES

12.1. Vector spaces

A real vector space is a set of objects that you can do two things with: you can add two of them together to get another such object, and you can multiply one of them by any real number to get another such object. There’s a set of axioms that a vector space must satisfy; you can find these in other textbooks. Similarly, a complex vector space is a set of objects that you can do two things with: you can add two of them together to get another such object, and you can multiply one of them by any complex number to get another such object. If you are speaking about a real vector space, you call any real number a scalar. If you are speaking about a complex vector space, you call any complex number a scalar. If you have a (finite) list of vectors $v_1, v_2, v_3, \ldots, v_n$, the most general way that you can combine these vectors together to get another vector is to choose a list of scalars $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$ and form the linear combination

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \cdots + \alpha_n v_n$$

Examples of real vector spaces include $\mathbb{R}^N$, where the vectors are $N$-tuples of real numbers (most familiarly, $\mathbb{R}^2$ – the plane – which is the set of ordered pairs of real numbers and $\mathbb{R}^3$ – three-dimensional space – which is the set of ordered triples of real numbers). The basic operations look like this: if $x = (x_0, x_1, \ldots, x_{N-1})$, $y = (y_0, y_1, \ldots, y_{N-1})$, and $\alpha$ is any real number, then

$$x + y = (x_0 + y_0, x_1 + y_1, \ldots, x_{N-1} + y_{N-1})$$

and

$$\alpha x = (\alpha x_0, \alpha x_1, \ldots, \alpha x_{N-1})$$

The zero vector is $(0, 0, \ldots, 0)$. Other examples include many different varieties of vector spaces whose members are real valued functions on a certain domain. When you deal with such function spaces, you have to be careful of a few things. One is that each function must have the same domain – each function must be defined “everywhere” (actually, for spaces where membership is determined by some kind of integrability condition, “almost everywhere” is good enough). As an example, let’s consider a set of real-valued functions suitable for feeding into the Laplace transform: the set of “locally integrable” functions on the interval $[0, \infty)$ that “don’t grow too fast at $\infty$”. Is $f(t) = \ln t$ a member of this set? Yes, although its domain isn’t quite all of $[0, \infty)$. It blows up at 0, but not so badly as to spoil its integrability there. Is $f(t) = \sqrt{1 - t^2}$ a member of this set? No – because we need to know what it does between 1 and $\infty$. As such, it would be silly to ask what the Laplace transform of this function is. You also must make sure that your set of functions really is a vector space. This primarily means that you have to check that the sum of any two such functions is also such a function, and that a constant times any such function is still such a function. Examples of function spaces include the following:

<table>
<thead>
<tr>
<th>Description</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous functions on $[a, b]$</td>
<td>$C[a, b]$</td>
</tr>
<tr>
<td>Periodic continuous functions</td>
<td>$C(\mathbb{T})$</td>
</tr>
<tr>
<td>Continuous functions on the whole real line</td>
<td>$C(\mathbb{R})$</td>
</tr>
<tr>
<td>Functions with one continuous derivative on a set $S$ ($S$ can be $[a, b], \mathbb{T}, \mathbb{R}$, etc.)</td>
<td>$C^1(S)$</td>
</tr>
<tr>
<td>Bounded (actually, “essentially bounded”), functions on a set $S$</td>
<td>$L^\infty(S)$</td>
</tr>
<tr>
<td>Absolutely integrable functions on $S$</td>
<td>$L^1(S)$</td>
</tr>
<tr>
<td>Square integrable functions on $S$</td>
<td>$L^2(S)$</td>
</tr>
</tbody>
</table>

Examples of complex vector spaces include $\mathbb{C}^N$, the set of $N$-tuples of complex numbers, which is just like $\mathbb{R}^N$, except that both the entries (components) of the vectors and the scalars are allowed.
to be complex numbers, and various spaces of complex valued functions. Be sure to understand
that we are talking about having the \textit{values} of the function being complex – we are not assuming
that these are functions of a complex \textit{variable}. As an example, $e^{inx}$ is a complex-valued function
of the \textit{real} variable $x$. Our list of complex function spaces that we are likely to encounter is exactly
the same as the list above, and when we say \textit{exactly}, we mean that we are using the exact same
symbols to name these spaces. When we use a symbol like $C(T_P)$, we are not committing ourselves
as to whether we mean the real-valued functions or the complex-valued functions – we either have
to make that clear in the surrounding context, or else admit that in the particular case we have in
mind it doesn’t much matter which one we mean.

The study of vector spaces in general falls under the label \textit{linear algebra}. In that study, sets
of vectors (that is, sets of elements of some vector space) are probed as to what their span is and
as to whether or not they are \textit{linearly independent} (\textit{independent} for short). For further details, see
any reasonably-written textbook on the subject. Here are two definitions out of all of that study:
if you have a function (or mapping, or map, \textit{etc} – the same idea has many names) $T$ that has as
its domain – the things that you can feed this mapping – all of one vector space, and that has as its
output elements of some other (well, possibly other – no rule says it can’t be the same one) vector
space, then $T$ is a \textit{linear transformation} (or linear mapping or linear \textit{whatzit} or just plain \textit{linear}) if:

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v)$$

(12.1)

A \textit{linear functional} is something whose domain is some vector space – which takes vectors as
input – that gives scalars (real numbers if it is a real vector space and complex numbers if it is a
complex vector space) as output, and that satisfies equation (12.1).

By now you should be familiar with the principle that "∧" – the thing that takes in functions
and gives back Fourier coefficients – is, by any reasonable definition, linear.

\subsection*{12.2. \hspace{1em} Real inner product spaces}

The vector space $\mathbb{R}_N$ comes equipped with a very special geometrically-inspired algebraic op-
eration called the \textit{dot product}. The dot product takes in two vectors and gives you back a real
number – a scalar. Synonyms for the words dot product include inner product and scalar product.
We will write the dot product of the two vectors $u$ and $v$ in at least two different notations: as
$u \cdot v$ or as $(u, v)$. The dot product on $\mathbb{R}^N$ is defined as follows: if $x = (x_0, x_1, \ldots, x_{N-1})$ and
$y = (y_0, y_1, \ldots, y_{N-1})$, then

$$x \cdot y = (x, y) = x_0y_0 + x_1y_1 + \cdots + x_{N-1}y_{N-1} = \sum_{k=0}^{N-1} x_k y_k$$

(12.2)

The geometric description of the dot product is the following: if $\theta$ is the angle between $x$ and
$y$, then

$$\theta = \cos^{-1} \left( \frac{x \cdot y}{\|x\| \|y\|} \right)$$

(12.3)

where $\|x\| = \|x\|_2 = \sqrt{x_0^2 + x_1^2 + \cdots + x_{N-1}^2} = \sqrt{x \cdot x}$

(12.4)

is the \textit{length} or \textit{norm} of the vector $x$. If the dot product of two vectors is zero, then the angle
between them is a right angle and we call the two vectors \textit{orthogonal}. We also notice in (12.4) that
computations of lengths can use the dot product, as there seems to be a close relationship. (There
is also the familiar idea that in any problem involving lengths and the Pythagorean Theorem, it is frequently easier to work with the square of a distance than it is to work with the distance itself.)

What if we have a real vector space that is not \( \mathbb{R}^N \) – such as one of the function spaces? We may, under some circumstances, be able to define something called an \textit{inner product} on that vector space. An inner product on a real vector space is something that has all of the vital properties of the dot product on \( \mathbb{R}^N \) – if we can figure out what those properties are. The accepted line of jargon for a real inner product is that it is a \textit{positive-definite symmetric bilinear form}. What does this mean? It means that it is an object that takes as input two elements of the vector space and gives back a real number - for vectors \( u \) and \( v \), let us write the inner product as \( \langle u, v \rangle \). The other words in this description have the following meanings:

\textit{Symmetric:}

\[ \langle u, v \rangle = \langle v, u \rangle \text{ for all vectors } u \text{ and } v \quad (12.5) \]

\textit{Bilinear:}

\[ \langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle \text{ and } \]

\[ \langle u, \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle \text{ for all } u, v, w, \alpha \text{ and } \beta \quad (12.6) \]

\textit{Positive definite:}

\[ \langle u, u \rangle \geq 0 \text{ always and } \langle u, u \rangle > 0 \text{ for } u \neq 0 \quad (12.7) \]

The best example that we will have of something that satisfies all of these properties is to take a space whose elements are functions and to let \( \langle f, g \rangle = \int fg \), or something very much like that. For instance, for functions periodic of period \( P \), let us define the standard real inner product of \( f \) and \( g \) as:

\[ \langle f, g \rangle = \frac{1}{P} \int_0^P f(x)g(x) \, dx \quad (12.8) \]

It is not hard to show that this satisfies properties (12.5), (12.6), and (12.7) - except possibly for a little fudging on the second part of (12.7), but let’s not worry too much about that now.

Any real vector space on which a real inner product has been defined is a \textit{real inner product space}.

\subsection*{12.3. Complex inner product spaces}

We’d like to do this for complex vector spaces - but we realize that there are going to have to be some subtle modifications of the detail. In particular, if we want something positive definite – something like (12.7) – we are going to have to use something with a lot of complex conjugates and absolute values in it. Conventional wisdom has settled on just what we need, and it is the following: a complex inner product is something that takes as input two vectors from a complex vector space and gives as output a complex number, and it is a \textit{positive-definite Hermitian sesquilinear form}. Actually, “sesquilinear” is a part of “Hermitian”, but I threw it in to make the phrase sound more impressive. “Sesquilinear” means linear in one factor and “conjugate-linear” in the other factor - that’s something like linear, but with some stray complex conjugates hanging around. It makes no earthly difference which factor is which, but we have to come to some choice and stick to it. As accident would have it, mathematicians have fallen into the rut of always putting the complex conjugate on the second factor, while physicists have fallen into the rut of putting the complex conjugate on the first factor. In what follows, we will follow the mathematician’s convention - we suppose that a physicist will just have to read it in a mirror. Here’s what these words mean:
Hermitian: 
\[ \langle u, v \rangle = \overline{\langle u, v \rangle} \] for all vectors \( u \) and \( v \) \hspace{1cm} (12.9)

Sesquilinear: 
\[ \langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle \] and
\[ \langle u, \alpha v + \beta w \rangle = \overline{\alpha} \langle u, v \rangle + \overline{\beta} \langle u, w \rangle \] for all \( u, v, w, \alpha \) and \( \beta \) \hspace{1cm} (12.10)

Positive definite: 
\[ \langle u, u \rangle \geq 0 \] always and \( \langle u, u \rangle > 0 \) for \( u \neq 0 \) \hspace{1cm} (12.11)

We will give two examples: the standard inner product on \( \mathbb{C}^N \) is defined as follows: if \( z = (z_0, z_1, \ldots, z_{N-1}) \) and \( w = (w_0, w_1, \ldots, w_{N-1}) \), then
\[ \langle z, w \rangle = z_0 \overline{w_0} + z_1 \overline{w_1} + \cdots + z_{N-1} \overline{w_{N-1}} = \sum_{k=0}^{N-1} z_k \overline{w_k} \] \hspace{1cm} (12.12)

On a function space the inner product of \( f \) and \( g \) will be \( \langle f, g \rangle = \int f \overline{g} \), or to make it specific to functions periodic of period \( P \), the standard complex inner product of \( f \) and \( g \) is:
\[ \langle f, g \rangle = \frac{1}{P} \int_0^P f(x) \overline{g(x)} \, dx \] \hspace{1cm} (12.13)

Any complex vector space on which a complex inner product as been defined is called a complex inner product space.

12.4. The real Cauchy-Buniakovski-Schwarz inequality

Theorem – The Cauchy-Schwarz inequality: in any real inner product space, for any two vectors \( u \) and \( v \),
\[ \langle u, v \rangle^2 \leq \langle u, u \rangle \langle v, v \rangle \] \hspace{1cm} (12.14)
with equality holding if and only if one of these vectors is a scalar multiple of the other.

The proof uses nothing but the properties of an inner product – that is, (12.5), (12.6), and (12.7) – and the technique of completing the square. If \( v = 0 \), there is nothing to prove, both sides of the inequality being zero, so we assume that \( v \neq 0 \). Consider the vector \( u + \lambda v \) for any real number \( \lambda \). By property (12.7), the inner product of this vector with itself is always greater or equal to zero (with equality only if it is the zero vector). Thus:
\[ 0 \leq \langle u - \lambda v, u - \lambda v \rangle \]
\[ = \langle u, u \rangle - \langle u, \lambda v \rangle - \langle \lambda v, u \rangle + \langle \lambda v, \lambda v \rangle \] \hspace{1cm} by (12.6)
\[ = \langle u, u \rangle - 2\lambda \langle u, v \rangle + \lambda^2 \langle v, v \rangle \] \hspace{1cm} by (12.5) and (12.6)

Since we know that \( \langle v, v \rangle \) is positive (by (12.7) again), we can divide both sides of this inequality by it, getting:
\[ \lambda^2 - \frac{2\langle u, v \rangle}{\langle v, v \rangle} \lambda + \frac{\langle u, u \rangle}{\langle v, v \rangle} \geq 0 \]
\[ \lambda^2 - 2\frac{\langle u, v \rangle}{\langle v, v \rangle} \lambda \geq -\frac{\langle u, u \rangle}{\langle v, v \rangle} \]
Now complete the square – that is, use formula (11.1) from the previous chapter:

\[
\lambda^2 - \frac{2\langle u, v \rangle}{\langle v, v \rangle} \lambda + \left( \frac{\langle u, v \rangle}{\langle v, v \rangle} \right)^2 \geq \left( \frac{\langle u, v \rangle}{\langle v, v \rangle} \right)^2 - \frac{\langle u, u \rangle}{\langle v, v \rangle}
\]

This is true for all real \( \lambda \). If we let \( \lambda \) be that value which minimizes the left hand side – that is, if we let \( \lambda = \frac{\langle u, v \rangle}{\langle v, v \rangle} \), then the left hand side is zero and the right hand side must be less than or equal to zero. Hence,

\[
0 \geq \frac{\langle u, v \rangle^2 - \langle u, u \rangle \langle v, v \rangle}{\langle v, v \rangle^2}
\]

or \( \langle u, v \rangle^2 \leq \langle u, u \rangle \langle v, v \rangle \)

As it turns out, there is another notation in which this fact is usually expressed. From property (12.7), the inner product of a vector with itself is positive and hence has a real, positive square root. We call this square root the norm of the vector - by analogy to (12.4). That is, in any real inner product space we define the norm of a vector \( u \) to be

\[
\|u\| = \sqrt{\langle u, u \rangle}
\]  

(12.15)

Using this terminology, we take the square root of both sides of the inequality (12.14) to get:

**The Cauchy-Schwarz inequality (norm version):**

\[
|\langle u, v \rangle| \leq \|u\| \|v\|
\]

(12.16)

One side effect of this is that we can form the fraction \( \frac{\langle u, v \rangle}{\|u\| \|v\|} \) and be assured that it is between \(-1\) and \(1\). Hence it has an arccosine, and we call that arccosine the angle between the two vectors.

**12.5. The complex Cauchy-Buniakowski-Schwarz inequality**

**Theorem – the Cauchy-Schwarz inequality:** in any real inner product space, for any two vectors \( u \) and \( v \),

\[
|\langle u, v \rangle|^2 \leq \langle u, u \rangle \langle v, v \rangle
\]

(12.17)

with equality holding if and only if one of these vectors is a scalar multiple of the other.

What's the difference between this and equation (12.14)? Two things: the fact that the vectors involved belong to a complex rather than a real vector space, and the way that we had to make it the square of the absolute value of something on the left hand side – we can't just put the square of a complex number into an inequality and have it be meaningful.

The proof uses nothing but the properties of an inner product – that is, (12.9), (12.10), and (12.11) – and the technique of completing the square for a complex variable. If \( v = 0 \), there is nothing to prove, both sides of the inequality being zero, so we assume that \( v \neq 0 \). Consider the vector \( u + \lambda v \) for any complex number \( \lambda \). By property (12.1), the inner product of this vector with itself is always greater or equal to zero (with equality only if it is the zero vector). Thus:

\[
0 \leq \langle u - \lambda v, u - \lambda v \rangle
\]
\[ \langle u, u \rangle - \langle u, \lambda v \rangle - \langle \lambda v, u \rangle + \langle \lambda v, \lambda v \rangle \quad \text{by (12.6)} \]

\[ = \langle u, u \rangle - \lambda \langle u, v \rangle - \lambda \langle v, u \rangle + \lambda \overline{\langle v, v \rangle} \quad \text{by (12.9) and (12.10)} \]

Since we know that \( \langle v, v \rangle \) is positive (by (12.11) again), we can divide both sides of this inequality by it, getting:

\[ |\lambda|^2 - \overline{\langle u, v \rangle \langle v, v \rangle} - \lambda \overline{\langle u, v \rangle \langle v, v \rangle} + \langle u, u \rangle \overline{\langle v, v \rangle} \geq 0 \]

Next, employ equation (11.7) of the “Completing the Square” chapter:

\[ |\lambda|^2 - \overline{\langle u, v \rangle \langle v, v \rangle} - \lambda \overline{\langle u, v \rangle \langle v, v \rangle} + \langle u, u \rangle \overline{\langle v, v \rangle} \geq 0 \]

\[ \left( \lambda - \frac{\langle u, v \rangle}{\langle v, v \rangle} \right) \left( \overline{\lambda} - \frac{\overline{\langle u, v \rangle}}{\langle v, v \rangle} \right) - \left| \frac{\langle u, v \rangle}{\langle v, v \rangle} \right|^2 \geq 0 \]

This is true for all possible values of \( \lambda \), especially including that value which minimizes the left hand side – namely, \( \lambda = \frac{\langle u, v \rangle}{\langle v, v \rangle} \). If we choose that value of \( \lambda \), then the numerator of the second fraction must be greater or equal to zero - that is, (12.17) must be true.

We now repeat the way we finished off the previous section: we let the norm of the vector be the square root of its inner product with itself – that is, in a complex inner product space,

\[ \|u\| = \sqrt{\langle u, u \rangle} \quad \text{(12.18)} \]

Given this convention, we take the square root of both sides of (12.17) to get

**The Cauchy-Schwarz inequality (norm version, complex vector space):**

\[ |\langle u, v \rangle| \leq \|u\| \|v\| \quad \text{(12.19)} \]

### 12.6. Orthonormal sets in a real inner product space

The driving principle of the vast majority of mathematical work in inner product spaces is orthogonality. The definition of orthogonal is the same in both real and complex inner product spaces: two vectors are orthogonal if and only if their inner product is zero. With that in mind, we define the concepts of an orthogonal set and an orthonormal set of vectors:

**Definition:** Let \( \{e_j\} \) be a set of vectors in a (real or complex) inner product space. The variable \( j \) – the index to this list of vectors – runs through some set of possible values. We are at the moment being deliberately vague as to whether that index set is finite or infinite. Then \( \{e_j\} \) is an orthogonal set if and only if:

\[ \langle e_j, e_k \rangle = \begin{cases} 
0 & \text{if } j \neq k \\
> 0 & \text{if } j = k
\end{cases} \quad \text{(12.20)} \]
The same set is an orthonormal set if and only if:

\[
\langle e_j, e_k \rangle = \begin{cases} 
0 & \text{if } j \neq k \\
1 & \text{if } j = k 
\end{cases}
\]  
(12.21)

An orthonormal set is clearly a special case of an orthogonal set. On the other hand, any orthogonal set may be turned into an orthonormal set by merely dividing each element by its own norm. Note also that, although the zero vector is always orthogonal to everything, we don’t want it as a member of anything that we are willing to call an orthogonal set. Our central problem is this: suppose we have a finite orthonormal set \(\{e_j\}\) (now we are specifying that the set of indices – the set of possible values for \(j\) – be a finite set, although we are willing to let it be a very large finite set.) Let \(v\) be any vector in the inner product space. How closely can we approximate \(v\) by a linear combination of the elements of the orthonormal set? More specifically, how can we choose the coefficients \(\{\alpha_j\}\) so that

\[
\sum_j \alpha_j e_j \quad \text{is as close as possible to } v.
\]

But what do we mean by “as close as possible”? Surely we mean that the size of the difference is as small as possible. But what do we mean by “size”? Well, let’s see – every inner product space, real or complex, has a built-in notion of size: the norm, as defined in (12.15) and (12.18). That is, our problem is to find \(\alpha_j\) so that

\[
\|v - \sum_j \alpha_j e_j\| \quad \text{is minimized.}
\]

To minimize this, it is sufficient to minimize its square. This isn’t a new idea, of course – it is used in nearly every calculus problem that asks that a distance be maximized or minimized. The Pythagorean theorem just makes working with the squares of distances easier than working with the distances themselves. So here we go:

\[
\|v - \sum_j \alpha_j e_j\|^2 = \left\langle v, v - \sum_j \alpha_j e_j - \sum_k \alpha_k e_j \right\rangle
\]  
(12.22)

We used two different names – \(j\) and \(k\) – for the index variable because we had to. If you multiply a sum of, for instance, seven terms by itself, the result will be a sum with 49 terms – you’ve got to take all possible products of any of the terms with any of the other terms. So far, we are assuming that we are working in a real inner product space. Use (12.5) and (12.6) to simplify (12.22):

\[
\|v - \sum_j \alpha_j e_j\|^2 = \langle v, v \rangle - \sum_j \langle v, \alpha_j e_j \rangle - \sum_j \langle \alpha_j e_j, v \rangle + \sum_j \langle \alpha_j e_j, \alpha_k e_k \rangle
\]

\[
= \langle v, v \rangle - 2\sum_j \alpha_j \langle v, e_j \rangle + \sum_{j,k} \alpha_j \alpha_k \langle e_j, e_k \rangle
\]

\[
= \langle v, v \rangle - 2\sum_j \alpha_j \langle v, e_j \rangle + \sum_j \alpha_j^2 \quad \text{by } 12.21
\]

88
\[= \sum_j \left( \alpha_j^2 - 2\alpha_j \langle \mathbf{v}, \mathbf{e}_j \rangle \right) + \| \mathbf{v} \|^2\]

We complete the square, using formula (11.1):

\[= \sum_j \left( \alpha_j^2 - 2\alpha_j \langle \mathbf{v}, \mathbf{e}_j \rangle + \langle \mathbf{v}, \mathbf{e}_j \rangle^2 \right) + \| \mathbf{v} \|^2\]

\[= \sum_j (\alpha_j - \langle \mathbf{v}, \mathbf{e}_j \rangle)^2 - \sum_j \langle \mathbf{v}, \mathbf{e}_j \rangle^2 + \| \mathbf{v} \|^2 \]  \hspace{1cm} (12.23)

We’re trying to minimize this quantity, and we get to choose the \( \alpha_j \) any way we want to. It is clear that the first summation on the right hand side of (12.23), being the sum of squares, is always greater than or equal to zero – but we can make it zero if we just choose the \( \alpha_j \)'s right. The right choice is to let \( \alpha_j \) be equal to \( \langle \mathbf{v}, \mathbf{e}_j \rangle \). These values for the coefficients are called (for reasons which will eventually become clear) the generalized Fourier coefficients for \( \mathbf{v} \) with respect to this orthonormal set. Since the left hand side of (12.22) is the square of a norm, it must always be greater than or equal to zero. If we let \( \alpha_j \) be equal to \( \langle \mathbf{v}, \mathbf{e}_j \rangle \) for each \( j \), the remaining portion of the right hand side of (12.23) must be nonnegative. That is, we have the following inequality, generally known as Bessel’s inequality:

\[\sum_j \langle \mathbf{v}, \mathbf{e}_j \rangle^2 \leq \| \mathbf{v} \|^2 \]  \hspace{1cm} (12.24)

If we are able to write \( \mathbf{v} \) as a linear combination of the \( \mathbf{e}_j \)'s, then the minimum value of the left hand side of (12.22) would be zero and the inequality in (12.24) would actually be an equality.

What if the orthonormal set is actually an infinite set rather than a finite set? Then such sums as appear in (12.22) and (12.24) would have to be interpreted as infinite series. Fortunately, the very workings of this problem – notably Bessel’s inequality – help to assure us that these series converge in some appropriate sense. (Actually, to get everything that we might ask for in terms of these series being meaningful, we’ll have to have our inner product space be a complete metric space, which makes it a Hilbert space – unfortunately, this is not the same meaning of the word complete as we are about to use below.) Bessel’s inequality will always be true, and the norm of the difference between \( \mathbf{v} \) and our linear combination of the \( \mathbf{e}_j \)’s will always be minimized if we choose our coefficients to be the generalized Fourier coefficients. A new issue now arises: does our orthonormal set have enough elements in it that we can write any vector \( \mathbf{v} \) as the limit of linear combinations of – that is, as an infinite series based on – that set? If so, we call the orthonormal set complete. (This is also known as having an orthonormal basis.) It happens – and the calculations above provide the framework for this argument, too – that an orthonormal set is complete if and only if the only vector orthogonal to every element in it is the zero vector.

Let’s summarize our findings:

**Fourier coefficients:**

\[\left\| \mathbf{v} - \sum_j \alpha_j \mathbf{e}_j \right\| \text{ is minimized if each } \alpha_j = \langle \mathbf{v}, \mathbf{e}_j \rangle \]  \hspace{1cm} (12.25)

**Bessel’s inequality:**

\[\sum_j |\langle \mathbf{v}, \mathbf{e}_j \rangle|^2 \leq \| \mathbf{v} \|^2 \]  \hspace{1cm} (12.26)
If the orthonormal set is also complete, then:

**Generalized Fourier series:**
\[ v = \sum_k \langle v, e_j \rangle e_j \]  
(12.27)

**Generalized Parseval’s identity:**
\[ \sum_j |\langle v, e_j \rangle|^2 = \|v\|^2 \]  
(12.28)

### 12.7. Orthonormal sets in a complex inner product space

Now we suppose that \( \{e_j\} \) is a (finite, for the time being) orthonormal set in a complex inner product space. Without further ado, let’s repeat what was in the last section:

\[
\left\| v - \sum_j \alpha_j e_j \right\|^2 = \left\langle v - \sum_j \alpha_j e_j, v - \sum_k \alpha_k e_k \right\rangle
\]

\[
= \langle v, v \rangle - \sum_j \langle v, \alpha_j e_j \rangle - \sum_j \langle \alpha_j e_j, v \rangle - \sum_{j,k} \langle \alpha_j e_j, \alpha_k e_k \rangle
\]

\[
= \langle v, v \rangle - \sum_j \overline{\alpha_j} \langle v, e_j \rangle - \sum_j \alpha_j \langle v, e_j \rangle + \sum_{j,k} \alpha_j \overline{\alpha_k} \langle e_j, e_k \rangle
\]

\[
= \langle v, v \rangle - \sum_j \overline{\alpha_j} \langle v, e_j \rangle - \sum_j \alpha_j \langle v, e_j \rangle + \sum_j |\alpha_j|^2
\]

We complete the square, using formula (11.7):

\[
= \sum_j \left( |\alpha_j|^2 - \overline{\alpha_j} \langle v, e_j \rangle - \alpha_j \langle v, e_j \rangle + |\langle v, e_j \rangle|^2 - |\langle v, e_j \rangle|^2 \right) + \|v\|^2
\]

\[
= \sum_j |\alpha_j - \langle v, e_j \rangle|^2 - \sum_j |\langle v, e_j \rangle|^2 + \|v\|^2
\]  
(12.29)

Now we are ready to draw conclusions from this, exactly as before.

*Formulas (12.25), (12.26), (12.27), and (12.28) are also valid without change in a complex inner product space.*

Of course, we were clever enough to include some well-chosen absolute values in these statements!

### 12.8. Square integrable functions on \( \mathbb{T}_P \)

Consider a function \( f \) (real or complex valued) that is defined on the line so as to be periodic of period \( P \). We call such a function *square integrable*, and say that it belongs to \( L^2(\mathbb{T}_P) \), provided the following integral converges:

\[
\frac{1}{P} \int_0^P |f(x)|^2 \, dx < \infty
\]
Just to give you a taste of this condition – it doesn’t require that the function be bounded, but it is harder to satisfy than mere integrability, or even absolute integrability. As an example, take the function \( f(x) = \frac{1}{\sqrt{|x|}} \) for \( 0 < |x| \leq \frac{P}{2} \). This function is integrable on \( \left[ -\frac{P}{2}, \frac{P}{2} \right] \) (and hence absolutely integrable since it is positive), but if we square it, we get \( \frac{1}{|x|} \), which is not integrable.

We say that this function belongs to \( L^1 \) but not to \( L^2 \).

We now define the inner product. Please recognize that our decision to divide out front by \( P \) represents one of several possible notational choices, and may not necessarily be reflected in other works.

The inner product for real-valued functions:

If \( f, g \in L^2(T_P) \), then

\[
\langle f, g \rangle = \frac{1}{P} \int_{-P}^{P} f(x)g(x) \, dx
\]  

(12.30)

The inner product for complex-valued functions:

If \( f, g \in L^2(T_P) \), then

\[
\langle f, g \rangle = \frac{1}{P} \int_{-P}^{P} f(x)\overline{g(x)} \, dx
\]  

(12.31)

In either case, the norm is as follows:

\[
\|f\|_2 = \langle f, f \rangle = \left( \frac{1}{P} \int_{-P}^{P} |f(x)|^2 \, dx \right)^{\frac{1}{2}}
\]  

(12.32)

The Cauchy-Schwarz inequality in this case reads as follows:

\[
\left| \frac{1}{P} \int_{-P}^{P} f(x)\overline{g(x)} \, dx \right| \leq \left( \frac{1}{P} \int_{-P}^{P} |f(x)|^2 \, dx \right)^{\frac{1}{2}} \left( \frac{1}{P} \int_{-P}^{P} |g(x)|^2 \, dx \right)^{\frac{1}{2}} = \|f\|_2\|g\|_2
\]  

(12.33)

A quick corollary of this is that

\[
\|f\|_1 = \frac{1}{P} \int_{-P}^{P} |f(x)| \, dx \leq \left( \frac{1}{P} \int_{-P}^{P} |f(x)|^2 \, dx \right)^{\frac{1}{2}} \left( \frac{1}{P} \int_{-P}^{P} 1 \, dx \right)^{\frac{1}{2}} \leq \|f\|_2
\]  

(12.34)

12.9. Fourier series and orthogonal expansions

\[ \left\{ 1, \cos \left( \frac{2\pi x}{P} \right), \sin \left( \frac{2\pi x}{P} \right), \cos \left( 2 \cdot \frac{2\pi x}{P} \right), \sin \left( 2 \cdot \frac{2\pi x}{P} \right), \cos \left( 3 \cdot \frac{2\pi x}{P} \right), \sin \left( 3 \cdot \frac{2\pi x}{P} \right), \ldots \right\} \]

is an orthogonal set in the real inner product space \( L^2(T_P) \), and almost – but not quite – an orthonormal set. The specific problem with being orthonormal is the normalization: the function 1 has norm equal to 1, but all of the other functions in this set have norm equal to \( \frac{1}{\sqrt{2}} \). We could chase through all of the consequences of that factor, but rather than give you the details, we’ll just give the results.

If we want to approximate a square integrable function \( f \) by an \( N \)th degree trigonometric polynomial, the the closest we can come in the \( L^2 \) norm – the least squares or least root mean square approximation – is to let the coefficients of this polynomial be exactly the Fourier coefficients. That
is, we let this trigonometric polynomial be the \( N \)th partial sum of the Fourier series for \( f \). One consequence of this is – by (12.26) – Bessel’s inequality:

\[
\frac{a[0]^2}{4} + \frac{1}{2} \sum_{k=1}^{N} (a[k]^2 + b[k]^2) \leq \|f\|_2^2
\]  
(12.35)

which can also be written as

\[
\frac{a[0]^2}{2} + \sum_{k=1}^{N} (a[k]^2 + b[k]^2) \leq \frac{2}{P} \int_{w}^{w+P} |f(x)|^2 \, dx
\]  
(12.36)

But then, it turns out that this orthogonal set is complete. We won’t prove that, but basically, our previous convergence theorems for Fourier series make this inevitable. That being the case, we can say that the Fourier series of any square integrable function always converges to that function in the \( L^2 \) sense. Furthermore, by (12.28), we have Parseval’s identity:

\[
\frac{a[0]^2}{2} + \sum_{k=1}^{\infty} (a[k]^2 + b[k]^2) = \|f\|_2^2
\]  
(12.37)

\[
\frac{a[0]^2}{2} + \sum_{k=1}^{\infty} (a[k]^2 + b[k]^2) = \frac{2}{P} \int_{w}^{w+P} |f(x)|^2 \, dx
\]  
(12.38)

Let’s try to repeat this for the complex case. This time, consider the set \( \{e^{2\pi ikx/P}\}_{k=-\infty}^{\infty} \) of complex valued functions on \( \mathbb{T}_P \). This turns out to be an orthonormal set – in fact, demonstrating that fact is far easier for this case than for the real case. This is also a complete orthonormal set – we couldn’t possibly have a different result for the complex case than for the real case – so we have both a Bessel’s inequality and a Parseval’s identity for this case, too.

**Bessel’s inequality:**

\[
\sum_{k=-N}^{N} |\hat{f}[k]|^2 \leq \|f\|_2^2
\]  
(12.39)

**Parseval’s identity:**

\[
\sum_{k=-\infty}^{\infty} |\hat{f}[k]|^2 = \|f\|_2^2
\]  
(12.40)

### 12.10. The trigonometric orthogonal set on \( \mathbb{P}_N \)

On the polygon \( \mathbb{P}_N \), the set of functions \( e_k[n] = e^{2\pi i kn/N} \) form an orthogonal set with respect to the usual complex inner product. That is,

\[
\langle e_j, e_k \rangle = \sum_{n=0}^{N-1} e_j[n]e_k[n] = \begin{cases} 
N & \text{if } j = k \\
0 & \text{otherwise}
\end{cases}
\]  
(12.41)

Equation (12.41) is just a restatement of equations (4.1) and (4.6). The factor of \( N \) that appears in it means that the vectors are orthogonal but not quite orthonormal. Let’s put this another way. Suppose we have an \( N \)-dimensional complex inner product space with the standard inner product with respect to that basis. (The set of all complex-valued functions on \( \mathbb{P}_N \) is just such a space.) We
can then write any vector as an $N \times 1$ column matrix. Suppose we have a set of $M$ such vectors. Create an $N \times M$ matrix $A$ whose $M$ columns are these $M$ vectors. If $M = N$ we will have a square matrix. How could we tell if these $M$ vectors were independent? This would be a question about the rank of the matrix $A$. If its rank is $M$, we have an independent set. In the $M = N$ square matrix case, the $N$ vectors are independent if and only if the determinant of $A$ is not zero. Now how can we tell if the set of vectors is orthogonal? To do this, let $A^*$ be the complex conjugate of the transpose of $A$, and compute the matrix product $A^*A$. This will be the $M \times M$ matrix whose $(j,k)$th entry is precisely the inner product of the $j$th and $k$th vectors of our set. The statement that the set is orthogonal is precisely the statement that $A^*A$ is diagonal: that all of its entries are zero except those on the main diagonal which are not zero. The statement that our set is orthonormal is precisely the statement that $A^*A = I$, the identity matrix. If $A$ is a square matrix such that $A^*A = I$, we call $A$ a unitary matrix.

Use this language to express (12.41). Let $F$ be the $N \times N$ matrix whose columns are the vectors $e_j$ named above. In other words, the $(j,k)$th entry of $F$ is $e^{2\pi ijk/N} = \zeta_{jk}$. Then $F$ is almost but not quite a unitary matrix: $F^*F = NI$.

So we have an orthogonal but not quite orthonormal set. What can we say about this in general? Suppose that $\{e_j\}$ is a finite orthogonal (but not necessarily orthonormal) set in a complex inner product space, and that $v$ is an arbitrary vector in that space. We wish to find the coefficients $\alpha_j$ such that the norm $\|v - \sum_j \alpha_j e_j\|$ is minimized. This minimum is achieved if and only if

$$\alpha_j = \frac{\langle v, e_j \rangle}{\langle e_j, e_j \rangle} \quad (12.42)$$

Bessel’ s inequality in this case turns out to be

$$\sum_j |\langle v, e_j \rangle|^2 \langle e_j, e_j \rangle \leq \|v\|^2 \quad (12.43)$$

with equality (Parseval’ s identity) in the case of the orthogonal set being complete.

We may easily apply (12.42) and (12.43) to the case of the trigonometric basis for functions in $\mathbb{P}_N$. The coefficients in (12.42) turn out to be:

$$\frac{\langle f, e_j \rangle}{\langle e_j, e_j \rangle} = \frac{1}{N} \sum_{n=0}^{N-1} f[n] e^{-2\pi i jn/N} = \hat{f}[j] \quad (12.44)$$

The set of all $N$ of these trigonometric functions is a set of $N$ orthogonal, hence independent, vectors. Therefore, they span the $N$-dimensional vector space of all functions on the polygon and are a complete orthogonal set. Equality necessarily holds in (12.43). A careful working out of the consequences of (12.43) leads us to Parseval’ s identity for this instance:

$$N\|\hat{f}\|^2 = N \sum_{k=0}^{N-1} |\hat{f}[k]|^2 = \sum_{n=0}^{N-1} |f[n]|^2 = \|f\|^2_2 \quad (12.45)$$

12.11. Exercises

12.1. By imitating the derivations around equation (12.29), prove (12.42) and (12.43).
12.2. Which of the following Fourier series represent square-integrable functions on $\mathbb{T}_{2\pi}$, also known as $[-\pi, \pi]$? What else can you say about the functions they represent?

(a) $\sum_{k=1}^{\infty} \frac{\sin kx}{k} = \sum_{k=-\infty}^{\infty} \frac{e^{ikx}}{2ik}$

(b) $\sum_{k=1}^{\infty} \frac{\sin kx}{\sqrt{k}} = \sum_{k=-\infty}^{\infty} \frac{\sqrt{|k|} e^{ikx}}{2ik}$

(c) $\sum_{k=1}^{\infty} \frac{\cos kx}{k} = \sum_{k=-\infty}^{\infty} \frac{e^{ikx}}{2|k|}$

(d) $\sum_{k=-\infty}^{\infty} r^{|k|} e^{ikx}$, where $0 \leq r < 1$

Supplemental exercises:

12.3. Use the Fourier series of piecewise-polynomial functions on $\mathbb{T}$ and either Parseval’s identity or the synthesis equation at well-chosen points to calculate $\sum_{k=1}^{\infty} \frac{1}{k^2}$, $\sum_{k=1}^{\infty} \frac{1}{k^4}$, and $\sum_{k=1}^{\infty} \frac{1}{k^6}$. One possibility: Start with $f_1(x) = 1-2x$ on the interval $(0, 1)$, extended to be periodic of period 1. Build a sequence of functions $f_n$ such that $f_{n+1}'(x) = f_n(x)$. (That is, $f_{n+1}$ is an antiderivative of $f_n$.) Choose the constant of integration so that $\hat{f}_{n+1}[0] = \int_{0}^{1} f_{n+1}(x) \, dx = 0$.

12.4. Find a way to automate the calculations in exercise 12.3, using a computer to help. Among other things, you’ll need the ability to work with arbitrary precision rational numbers (hence arbitrarily long integers). DERIVE, MAPLE, and MATHEMATICA have this capability. Calculate the exact value of $\sum_{k=1}^{\infty} \frac{1}{k^{26}}$. (Why 26? Because that was as far as Euler got, working the problem by hand.)