Part 2

Vector spaces of functions

part:vector-spaces

The main goal for this part of the course is to establish a framework where we can treat functions in the same way that we treated vectors in the multivariable course. To do this, we first introduce the idea of a *vector space*, a generalization of the Euclidean space $\mathbb{R}^n$. While there are many interesting examples of vector spaces, we focus primarily on collections of functions. We show how many features of the Euclidean space $\mathbb{R}^n$ can be generalized to these other vector spaces.

CHAPTER 5

Introduction to vector spaces

5.1. Vector spaces

From our multivariable calculus course, we know many things we can do with vectors. The most basic properties of vectors in Euclidean space is that they can be *scaled* and can be *added*. This motivates the following definition. A collection V of mathematical objects is called a ***real vector space*** if

DefineRealVectorSpace

(1) for any two items v_1 and v_2 in V the sum $v_1 + v_2$ is also in V,
(2) for any item v in V and real number α the rescaled item αv is also in V,

and if all the "usual rules" (associativity, commutativity, distributive property, etc.) hold true; see Exercise 5.1.

Here are some examples.

Example 5.1. *The collection of vectors in $\mathbb{R}^n$ forms a real vector space.*

Example 5.2. *The collection of all polynomials $p(x)$ such that $p(1) = 0$ forms a vector space. To see this, notice that if p_1 and p_2 are such polynomials, then $p(x) = p_1(x)+p_2(x)$ is also a polynomial such that $p(1) = 0$. Similarly, if $p(1) = 0$, then any multiple of p is also a polynomial having that property.*

Example 5.3. *The collection of all continuous functions $\mathbb{R} \to \mathbb{R}$ forms a vector space that is denoted $C^0(\mathbb{R})$.*

Example 5.4. *The collection of all functions $\mathbb{R} \to \mathbb{R}$ with continuous derivative forms a vector space that is denoted $C^1(\mathbb{R})$.*

VS:ode-solutions

Example 5.5. *The collection of solutions $u(t)$ to the differential equation*

$$\frac{d^2u}{dt^2} + 17u = 0$$

form a vector space. (In differential equations course, the fact that the space of solutions is a vector space is as the linearity principle or the principle of superposition.)

define-l2 EXAMPLE 5.6. *The collection of sequences a_k such that the sum $\sum_{k=1}^{\infty}(a_k)^2$ converges is a vector space, and is given the $l^2(\mathbb{N})$.*

To verify that it is a vector space, we need to verify that if a_k and b_k are in $l^2(\mathbb{N})$ and α is a real number, then the sums

$$\sum_{k=1}^{\infty}(\alpha a_k)^2 \quad \text{and} \quad \sum_{k=1}^{\infty}(a_k + b_k)^2$$

converge. To see this, we consider the N^{th} partial sums, noting that

$$\sum_{k=1}^{N}(\alpha a_k)^2 \leq \alpha^2 \sum_{k=1}^{N}(a_k)^2$$
$$\sum_{k=1}^{N}(a_k + b_k)^2 \leq 2\sum_{k=1}^{N}(a_k)^2 + 2\sum_{k=1}^{N}(b_k)^2.$$

The desired convergences now follows from the fact that a_k and b_k are in $l^2(\mathbb{N})$.

We remark that one may also consider sequences a_k where the index k rangers over all integers. The corresponding vector space is denoted $l^2(\mathbb{Z})$.

2-Interval-VectorSpace EXAMPLE 5.7. *The collection of functions $u\colon [-1, 1] \to \mathbb{R}$ such that*

$$\int_{-1}^{1} |u(x)|^2 dx < \infty \tag{5.1}$$ L2IntervalCondition

is a vector space. This vector space is given the symbol $L^2([-1, 1])$; see Exercise 5.3.

Notice that $u(x)$ does not need to be continuous – for example the following function is in $L^2([-1, 1])$:

$$u(x) = \begin{cases} x^{-1/4} & \text{if } > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

These examples show that the concept of a vector space is very general; indeed in linear algebra class one learns many tools for studying general vector spaces.

We make these remarks:

- Example 5.5 is a hint that the concept of a vector space is very useful for studying differential equations.
- We sometimes refer to the objects in a vector space V as "vectors", even when they are other types of mathematical gadgets. The numbers α are sometimes called "scalars" as they serve to rescale the vectors in V.

- We use the convention that vectors (and vector-valued functions) in $\mathbb{R}^n$ are typeset in boldface. However, vectors in other vector spaces are not distinguished in any way. Pay close attention to contextual clues in order to determine which objects are being considered as vectors.
- Sometimes it is convenient to use complex numbers for scalars; in this case we say that V is a ***complex vector space***.

It is also useful to consider some examples of collections of objects that are *not* vector spaces.

EXAMPLE 5.8. *The collection of unit vectors (i.e. vectors of magnitude* 1*) in* $\mathbb{R}^n$ *is not a vector space. This is straightforward to see: If two vectors each have magnitude* 1*, then their sum will (almost always) not have magnitude* 1*. Furthermore, if we rescale a unit vector, then it is no longer unit!*

EXAMPLE 5.9. *The collection of polynomials* $p(x)$ *such that* $p(1) = 2$ *is not a vector space. If polynomials* $p_1(x)$ *and* $p_2(x)$ *have this property, then their sum is a polynomial, but will not have the desired property.*

EXAMPLE 5.10. *The collection of solutions* $u(t)$ *to the differential equation*

$$\frac{d^2u}{dt^2} + 17u = \cos t$$

does not form a vector space. (Recall from differential equations class that inhomogeneous equations satisfy a somewhat more complicated version of the superposition principle.)

5.2. Subspaces

Suppose that $v_1, \ldots, v_k$ are elements of some vector space V. For any scalars $\alpha_1, \ldots, \alpha_k$, the quantity

$$\alpha_1 v_1 + \cdots + \alpha_k v_k$$

is called a ***linear combination*** of the objects $v_1, \ldots, v_k$. The collection of all possible linear combinations is called the ***span*** of the vectors, and is denoted $\operatorname{span}\{v_1, \ldots, v_k\}$.

EXAMPLE 5.11. *The span of the polynomials* $p_0(x) = 1$, $p_1(x) = x$*, and* $p_2(x) = x^2$ *is the collection of all polynomials of degree less than or equal to* 2*.*

An important fact is that is that the span of any collection of vectors is itself a vector space. Since all of the vectors in this vector space are also in the vector space V, we say that $\operatorname{span}\{v_1, \ldots, v_k\}$ is a "subspace" of V.

It is important that the phrase "linear combination" only refers to *finite* sums!

Here are two examples.

EXAMPLE 5.12. *Consider the vectors* $\mathbf{v}_1 = \langle 1, 1, 1\rangle$ *and* $\mathbf{v}_2 = \langle 1, 0, -1\rangle$ *in the vector space* $\mathbb{R}^3$. *The subspace* $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ *is a plane in* $\mathbb{R}^3$*; the equation of the plane is* $x - 2y + z = 0$.

VS:poly-degree-three

EXAMPLE 5.13. *The collection of polynomials of degree less than* 4 *is* $\operatorname{span}\{1, x, x^2, x^3\}$, *a subspace of the vector space of polynomials.*

More generally, we say that a vector space W is a ***subspace*** of vector space V if

- W is a vector space, and
- all objects in W are also in V.

EXAMPLE 5.14. *The collection of all functions* $u(x)$ *such that* $u(0) = 0$ *is a subspace of* $C^0(\mathbb{R})$.

The following is a very important example.

EXAMPLE 5.15. *A function* $u\colon [-1, 1] \to \mathbb{R}$ *is said to be* ***piecewise continuous*** *if* u *is continuous, except at a finite number of points in* $[-1, 1]$. *Furthermore, the function is required to be bounded.*

The collection of piecewise continuous functions $u\colon [-1, 1] \to \mathbb{R}$ *such that the integral* (5.1) *is finite forms a subspace of* $L^2([-1, 1])$. *We denote this subspace by* $L^2_{\mathrm{PC}}([-1, 1])$.

An example of a piecewise continuous function is the following:

$$u(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{if } x < 0. \end{cases}$$

The function $u(x)$ *defined in Example 5.7 is not an example of a piecewise continuous function, because it is not bounded.*

A type of subspace that is important in this class is a subspace defined by imposing a *boundary condition*. The following example appears repeatedly in this class.

ex:PeriodicTrig EXAMPLE 5.16 (PERIODIC BOUNDARY CONDITIONS). *A function* $u\colon [-1,1] \to \mathbb{R}$ *is said to satisfy* ***periodic boundary conditions*** *if*

$$u(-1) = u(1) \quad \text{and} \quad u'(-1) = u'(1).$$

The reason that these conditions are called "periodic" is that any function satisfying these conditions can be periodically extended to a larger domain in such a way that both the function, and its first derivative, are continuous.

Functions that satisfy periodic boundary conditions and also satisfy the condition (5.1) *form a subspace of* $L^2_{\mathrm{PC}}([-1,1])$. *We explore this subspace via the following activities.*

(1) *Show that the function* $f(x) = x^3 - x$ *satisfies periodic boundary conditions. Draw a sketch of the graph of* f *on the interval* $[-1,1]$. *Then extend the graph to a larger domain by periodic repeating. Plot the periodic extension of* $f'(x)$ *on this extended domain.*
(2) *For what values of* ω *does the function* $\sin(\omega x)$ *satisfy the periodic boundary condition?*
(3) *For what values of* ω *does the function* $\cos(\omega x)$ *satisfy the periodic boundary condition?*

Two other boundary conditions – the Dirichlet and Neumann conditions – are explored in the exercises.

5.3. Finite and infinite dimensional vector spaces

Consider the vector space of all polynomials, and let $W = \operatorname{span}\{1, x, x^2, x^3\}$ be the subspace from Example 5.13. Notice that W is also the span of the polynomials

$$1, \quad 1+x, \quad 1+x+x^2, \quad 1+x+x^2+x^3, \tag{5.2}$$ Alt-poly-basis

meaning that any polynomial of degree less than 4 can be realized as a linear combination of these four polynomials, just as any such polynomial can be realized as a linear combination of

$$1, x, x^2, x^3. \tag{5.3}$$ usual-poly-basis

EXAMPLE 5.17. *Let* $p(x) = 3 + 2x - 4x^3$. *We have*

$$\begin{aligned} p(x) &= 3\cdot 1 + 2\cdot x + 0\cdot x^2 + (-4)\cdot x^3 \\ &= 1\cdot(1) + 2\cdot(1+x) + 4\cdot(1+x+x^2) + (-4)\cdot(1+x+x^2+x^3) \end{aligned}$$

After a bit of thought, you ought to be able to convince yourself that not only can any polynomial of degree less than 4 be realized as a linear combination of the polynomials in (5.2), but that for any polynomial $p(x)$ there is *exactly one* such combination. (The same is true of the collection in (5.3), of course.)

In general, when there is a *finite* list of objects such that any element in a vector space can be realized as a linear combination of those objects (meaning that the vector space is the span of that list of objects), then we say that the vector space is ***finite dimensional***. If a vector space is not finite dimensional, then we say that it is ***infinite dimensional***. The notion of dimension is discussed in more detail below in Chapter 7.

Example 5.18. *The vector space of all polynomials is infinite dimensional (can you prove it?), while the vector space of polynomials of degree less than* 42 *is finite dimensional.*

Example 5.19. *The vector space of solutions to the differential equation*

$$\frac{d^2u}{dt^2} + 4u = 0$$

is finite dimensional. We know this from differential equations class, where we learned that all solutions take the form

$$\alpha \cos(2t) + \beta \sin(2t).$$

Example 5.20. *The vector space of functions $u(t)$ such that $u(0) = 0$ is infinite dimensional.*

Finite dimensional vector spaces are studied in detail in the linear algebra course. In this course, however, many (if not most) of the vector spaces we consider are infinite dimensional.

Exercises

`Ex:UsualRules` **Exercise 5.1.** Look up the definition of a vector space in a linear algebra textbook (or on the internet). What are the "usual rules" referred to in the definition of a vector space on page 33?

Exercise 5.2. Let U be a subset of $\mathbb{R}^n$. Show that the following are vector spaces:

- $C^k(U)$: the collection of all functions $U \to \mathbb{R}$ such that derivatives up to order k exist and are continuous,

- $C^\infty(U)$: the collection of all functions $U \to \mathbb{R}$ such that derivatives of all order exist and are continuous,
- $C_0^\infty(U)$: the collection of functions $u \in C^\infty(U)$ such that $u = 0$ in a neighborhood of the boundary.

Functions in $C^\infty(U)$ are called *smooth functions*; functions in $C_0^\infty(U)$ are called *smooth test functions*. (The reason for the phrase "test function" is explained in Chapter 6.)

One can also consider complex-valued functions; in this case we write $C^k(U, \mathbb{C})$, etc.

ex:DefineL1andL2 **Exercise 5.3.** Let U be a subset of $\mathbb{R}^n$. Here dV represents the appropriate length / area / volume element (depending on the dimension n).

(1) Show that the following are vector spaces.
  - $L^1(U)$: the collection of all functions $u : U \to \mathbb{R}$ such that
  $$\int_U |u|\, dV$$
  is finite. Functions in $L^1(U)$ are sometimes called *integrable*.
  - $L^2(U)$: the collection of all functions $u : U \to \mathbb{R}$ such that
  $$\int_U |u|^2\, dV$$
  is finite. Functions in $L^2(U)$ are sometimes called *square integrable*.

(2) For what values of p is the function $u(x) = x^p$ in $L^1([1, \infty))$? For what values of p is it in $L^2([1, \infty))$?

ex:DirichletTrig **Exercise 5.4.** Let U be a subset of $\mathbb{R}^n$. A function $u\colon U \to \mathbb{R}$ is said to satisfy the ***Dirichlet boundary condition*** if $u = 0$ at all points along the boundary of U.

(1) Show that the collection of all functions in $L^2_{\mathrm{PC}}(U)$ satisfying the Dirichlet boundary condition form a vector space.
(2) Suppose that $U = [-1, 1]$. Explain how the Dirichlet boundary condition is simply that
$$u(-1) = 0 \quad \text{and} \quad u(1) = 0. \tag{5.4}$$
BC:UnitDirichlet

Show that the collection of all functions satisfying (5.4) form a vector space.
(3) For what values of ω does the function $\sin(\omega x)$ satisfy the Dirichlet boundary condition (5.4)?
(4) For what values of ω does the function $\cos(\omega x)$ satisfy the Dirichlet boundary condition (5.4)?

ex:NeumannTrig **Exercise 5.5.** Let U be a subset of $\mathbb{R}^n$ and let $\hat{n}$ be the unit inward-pointing normal vector. A function $u\colon U \to \mathbb{R}$ is said to satisy the ***Neumann boundary condition*** if $\operatorname{grad} u \cdot \hat{n} = 0$ at all points along the boundary of U.

(1) Explain why the collection of functions in $L^2_{\mathrm{PC}}(U)$ satisfying the Neumann boundary condition forms a vector space.

(2) Suppose $U = [-1, 1]$. Explain how the Neumann boundary condition reduces to

$$u'(-1) = 0 \quad \text{and} \quad u'(1) = 0. \tag{5.5}$$ BC:UnitNeumann

(3) For what values of ω does the function $\sin(\omega x)$ satisfy the Neumann boundary condition (5.5)?

(4) For what values of ω does the function $\cos(\omega x)$ satisfy the Neumann boundary condition (5.5)?

Exercise 5.6.

(1) Show that the collection of odd functions forms a subspace of $L^2_{\mathrm{PC}}([-L, L])$.

(2) Show that the collection of even functions forms a subspace of $L^2_{\mathrm{PC}}([-L, L])$.

Exercise 5.7.

(1) Show that the collection of polynomials of degree less than or equal to n forms a subspace of $L^2_{\mathrm{PC}}([-L, L])$. What is the dimension of this subspace?

(2) Show that the collection of polynomials forms a subspace of $L^2_{\mathrm{PC}}([-L, L])$. What is the dimension of this subspace?

Exercise 5.8. A function of the form

$$p(x) = a_0 + a_1 \cos x + a_2 \cos 2x + \cdots + a_n \cos nx \\ + b_1 \sin x + b_2 \sin 2x + \cdots + b_n \sin nx$$

is called a ***(real) trigonometric polynomial*** of degree n.

(1) Show that the collection of trigonometric polynomials of degree (less than or equal to) n forms a subspace of $L^2_{\mathrm{PC}}([-L, L])$. What is the dimension of this subspace?

(2) Show that the collection of all trigonometric polynomials forms a subspace of $L^2_{\mathrm{PC}}([-L, L])$. What is the dimension of this subspace?

Exercise 5.9. A function of the form

$$p(x) = a_0 + a_1 \cos x + a_2 \cos 2x + \cdots + a_n \cos nx \\ + i\,(b_1 \sin x + b_2 \sin 2x + \cdots + b_n \sin nx)$$

is called a ***complex trigonometric polynomial*** of degree (less than or equal to) n.

(1) Show that the collection of complex trigonometric polynomials of degree less than (or equal to) n forms a subspace of $L^2_{\mathrm{PC}}([-L, L]; \mathbb{C})$. What is the dimension of this subspace?

(2) Show that any trigonometric polynomial of degree less than or equal to n can be written in the form

$$p(x) = c_{-n}e^{-nix} + \cdots + c_{-2}e^{-2ix} + c_{-1}e^{-ix} + c_0 + c_1e^{ix} + \cdots + c_ne^{nix}$$

for some constants $c_{-n}, \ldots, c_n$.

CHAPTER 6

Inner products

ch:inner-products

6.1. Motivation and definition

The Euclidean spaces $\mathbb{R}^n$ have an important structure that is used extensively in the vector calculus course – the dot product. Our first task of this section is to think about how to define dot-product-like structures for other vector spaces. To do this, let's extract the essential features of the dot product.

We can view the dot product as a function, taking as an inputs pairs of vectors in $\mathbb{R}^n$ and returning numbers as outputs. This function has the following properties:

- *Linearity property.* We have $(\alpha_1\mathbf{v} + \alpha_2\mathbf{v}_2) \cdot \mathbf{w} = \alpha_1(\mathbf{v}_1 \cdot \mathbf{w}) + \alpha_2(\mathbf{v}_2 \cdot \mathbf{w})$ for all scalars α_1, α_2 and for all vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w}$.
- *Symmetry property.* We have $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$ for all vectors $\mathbf{v}, \mathbf{w}$.
- *Positivity property.* We have $\mathbf{v} \cdot \mathbf{v} \geq 0$ for all vectors $\mathbf{v}$.
- *Definite property.* We have $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

These last two properties are often combined in to a single *positive definite property*.

Recall that we can geometrically interpret the dot product as measuring the extent to which two vectors are colinear. This is captured in the *Law of Cosines formula*

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \, \|\mathbf{w}\| \, \cos\theta, \tag{6.1}$$

DotProduct-LawOfCosine

where θ is the unique number in the interval $[0, \pi)$ that describes the angle between the lines spanned by the two vectors. The formula (6.1) makes use of an important application of the dot product: the *Euclidean norm* of a vector, which is defined by

$$\|\mathbf{v}\| = (\mathbf{v} \cdot \mathbf{v})^{1/2}. \tag{6.2}$$

DotProduct-Norm

Notice that the formulas (6.1) and (6.2) implicitly make use of the properties listed above. For example, without the positivity property, the norm would not be properly defined for all vectors; without the definite property, it would be possible for a vector to have norm zero without being the zero vector, an undesirable circumstance. Likewise, the symmetry property ensures that the angle between two vectors is well defined. Finally, the linearity property implies that the function $\mathbf{v} \mapsto \mathbf{v} \cdot \mathbf{w}$ is linear for any fixed $\mathbf{w}$.

When working in $\mathbb{R}^n$, the dot product allows us to measure the size of vectors, as well as determine the angle between vectors. We would like to be able to accomplish these same tasks in other vector spaces. Thus we need a mathematical gadget that fulfills the role of the dot product; such gadgets are called *inner products*. Here is the technical definition:

An ***inner product*** on a real vector space V is a function that takes in two elements v and w of V and returns a number, which given the symbol $\langle u, v\rangle$; this function must satisfy the following properties:

- *Linearity property.* We have $\langle \alpha_1 v_1 + \alpha_2 v_2, w\rangle = \alpha_1\langle v_1, w\rangle + \alpha_2\langle v_2, w\rangle$ for all scalars α_1, α_2 and for all vectors v_1, v_2, w in V.
- *Symmetry property.* We have $\langle v, w\rangle = \langle w, v\rangle$ for all v, w in V.
- *Positivity property.* We have $\langle v, v\rangle \geq 0$ for all v in V.
- *Definite property.* We have $\langle v, v\rangle = 0$ if and only if $v = t0$.

If we have an inner product on vector space V, then we can use it to define the ***norm*** of an element in V by
$$\|v\| = \sqrt{\langle v, v\rangle}.$$
The positive definite property of the inner product implies that $\|v\| = 0$ only when $v = 0$.

In this course, the most important example of an inner product is the following.

Example-DefineL2

EXAMPLE 6.1. *Suppose Ω is a region in $\mathbb{R}^n$. Recall that $L^2(\Omega)$ is the vector space of functions $u\colon \Omega \to \mathbb{R}$ such that*
$$\int_\Omega |u|^2\, dV < \infty.$$
We define an inner product on $L^2(\Omega)$ by
$$\langle u, v\rangle = \int_\Omega u\, v\, dV. \tag{6.3}$$
define-L2-inner-produc

It is easy to check that this satisfies the necessary properties and is thus an inner product.

Note that the condition defining $L^2(\Omega)$ is precisely the condition that the norm of u is finite. Thus we could have first defined the inner product, and then defined the vector space $L^2(\Omega)$ to simply be all functions having finite norm.

L2-ConcreteExample

EXAMPLE 6.2. *Consider the unit disk D^2 in $\mathbb{R}^2$. The functions $u(x, y) = x$, $v(x, y) = x^2 + y^2$, $w(x, y) = e^{x^2+y^2}$ are elements in $L^2(D^2)$. We compute*

$$\|u\| = \left(\int_{D^2} |u|^2 \, dA\right)^{1/2} = \left(\int_{-\pi}^{\pi} \int_0^1 r^2 \cos^2\theta \, r \, dr \, d\theta\right)^{1/2} = \frac{\sqrt{\pi}}{2},$$

$$\|v\| = \left(\int_{D^2} |v|^2 \, dA\right)^{1/2} = \left(\int_{-\pi}^{\pi} \int_0^1 r^4 \, r \, dr \, d\theta\right)^{1/2} = \sqrt{\frac{2\pi}{5}}.$$

$$\|w\| = \left(\int_{D^2} |w|^2 \, dA\right)^{1/2} = \left(\int_{-\pi}^{\pi} \int_0^1 e^{2r^2} \, r \, dr \, d\theta\right)^{1/2} = \frac{1}{2}\sqrt{e^2 - 1}.$$

We also compute

$$\langle u, v\rangle = \int_{D^2} u \, v \, dA = \int_{-\pi}^{\pi} \int_0^1 r\cos\theta \, r^2 \, r \, dr \, d\theta = 0,$$

$$\langle v, w\rangle = \int_{D^2} v \, w \, dA = \int_{-\pi}^{\pi} \int_0^1 e^{r^2} r^2 \, r \, dr \, d\theta = \frac{1}{2}.$$

At this stage, I need to make some remarks about the inner product on $L^2(\Omega)$ defined in Example 6.1. Suppose one had a function that is everywhere zero except at one point, where the function is nonzero. Since integrals do not "see" the value of a function at a single point, the norm of this function is zero, which violates the definite property that all inner products must satisfy. While it might be tempting to simply ignore such "pathological" functions by restricting to "nice functions" (for example, perhaps continuous functions), there are good reasons for wanting to include a large variety (including discontinuous) functions in our vector space.

The fancy way to deal with this situation is to define something called a *distribution* or *generalized function*. Such "functions" are typically defined in terms of what they do upon integration, and thus two functions with a "negligible difference" (such as the two functions discussed above) are considered to be the same. (Those who have taken the Discrete Math course should think about this in terms of equivalence classes.) In (the optional) Exercise 6.12 you can examine a sequence of functions in $L^2([0, 1])$ in order to explore these issues a little bit further.

Unfortunately, a systematic treatment of distributions is beyond the scope of this course, in part because the theory of distributions requires a more sophisticated theory of integration than what we have from our first-year calculus course. (The crux of the matter is how to give a proper definition of "negligible difference.") Thus for the purposes of this class we focus primarily on a the subspace $L^2_{\mathrm{PC}}(\Omega)$ of piecewise continuous functions. This subspace contains "enough" functions to handle most physically interesting situations. (We do, however, in a later section encounter one special distribution — the infamous Dirac distribution.) We furthermore say that two functions are ***the same in the L^2 sense*** if the L^2 norm of their difference is the zero.

EXAMPLE 6.3. *The functions*

$$u(x) = \begin{cases} 7 & \text{if } x \neq 2 \\ 3 & \text{if } x = 2 \end{cases} \qquad \text{and} \qquad v(x) = 7$$

are both elements of $L^2_{\mathrm{PC}}([0,10])$. Since

$$\|u - v\| = \left[\int_0^{10} (u(x) - v(x))^2 dx\right]^{1/2} = 0$$

the two functions are the same in the L^2 sense.

6.2. Weighted inner products

Sometimes it is convenient to have an inner product that "weights" certain parts of the domain differently from others. For example, if we have functions $u\colon \mathbb{R} \to \mathbb{R}$, we might want to "discount" the parts of the functions near $\pm\infty$. An easy way to accomplish this is to pick a positive function $w\colon \mathbb{R} \to \mathbb{R}$ and define a "weighted inner product" by

$$\langle u, v\rangle_w = \int_{-\infty}^{\infty} u(x)\, v(x)\, w(x)\, dx.$$

In this integral, regions where $w(x)$ is large are emphasized, while regions where $w(x)$ is small are discounted.

EXAMPLE 6.4. *Here are four examples of (weighted) inner product spaces that are important later in this course. The names come from the associated "special functions," which are discussed later.*

"Legendre": *The vector space $L^2_{PC}([-1,1])$ with inner product*

$$\langle u, v\rangle_P = \int_{-1}^{1} u(x)\, v(x)\, dx. \tag{6.4}$$ Legendre-IP

The corresponding norm we denote by $\|u\|_P$.

"Hermite": *The vector space $L^2_{PC}(\mathbb{R}, e^{-x^2})$ with inner product*

$$\langle u, v\rangle_H = \int_{-\infty}^{\infty} u(x)\, v(x)\, e^{-x^2}\, dx. \tag{6.5}$$

Hermite-IP

The corresponding norm we denote by $\|u\|_H$.

"Laguerre": *The vector space $L^2_{PC}([0,\infty), e^{-x})$ with inner product*

$$\langle u, v\rangle_L = \int_0^{\infty} u(x)\, v(x)\, e^{-x}\, dx. \tag{6.6}$$

Laguerre-IP

The corresponding norm we denote by $\|u\|_L$.

"Chebyshev": *The vector space $L^2_{PC}([-1,1], \frac{1}{\sqrt{1-x^2}})$ with inner product*

$$\langle u, v\rangle_T = \int_{-1}^{1} u(x)\, v(x)\, \frac{1}{\sqrt{1-x^2}}\, dx. \tag{6.7}$$

Chebyshev-IP

The corresponding norm we denote by $\|u\|_T$.

6.3. Orthogonality

When working in $\mathbb{R}^n$, we called two vectors *orthgonal* or perpendicular when their dot product is zero. We extend this definition to inner products: If $\langle u, v\rangle = 0$ then we say that u and v are ***orthogonal***. This concept of orthogonality is extremely important in the rest of the course.

Example 6.5. *In Example 6.2, the functions u and v are orthogonal, while the functions v and w are not.*

Perhaps the most well-known collection mutually orthogonal vectors are the standard coordinate vectors in $\mathbb{R}^3$:

$$\mathbf{e}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}. \tag{6.8}$$

R3-coordinate-vectors

In an infinite dimensional inner product space it is possible to have an infinite list of vectors that are all orthogonal to one another. If $\{v_k\} = \{v_1, v_2, v_3, \dots\}$ is an infinite list of vectors such that $\langle v_i, v_j\rangle = 0$ whenever $i \neq j$, then $\{v_k\}$ is called an ***orthogonal sequence***.

Example 6.6. *In Example 5.16 The list of sine functions $w_1 = \sin(\pi x)$, $w_2 = \sin(2\pi x)$, $w_3 = \sin(3\pi x)$,... that were constructed in Example 5.16 in fact form an orthogonal sequence.*

In the next two sections we discuss two important applications of orthogonality.

6.4. When is a vector zero?

Suppose we have a vector $\mathbf{u}$ in $\mathbb{R}^3$. Under what conditions can we deduce that $\mathbf{u}$ is actually the zero vector?

Certainly if $\mathbf{u}$ is orthogonal to every vector in $\mathbb{R}^3$, then we must have $\mathbf{u} = =0$. The reason is that if $\mathbf{u}$ is orthogonal to all vectors, then we have $\|\mathbf{u}\|^2 = \mathbf{u} \cdot \mathbf{u} = 0$ and thus we conclude that $\mathbf{u} = \mathbf{0}$ from the definite property of the dot product.

However, it turns out that we can deduce that a vector $\mathbf{u}$ is zero from much less information. If we know that $\mathbf{u}$ is orthogonal to the three vectors $\mathbf{e}_i$ in (6.8), then it must be that $\mathbf{u} = \mathbf{0}$. The reason is that if $\mathbf{u} = \langle u_1, u_2, u_3 \rangle$ then for each $i = 1, 2, 3$ our knowledge that $\mathbf{u} \cdot \mathbf{e}_i = 0$ implies that $u_i = 0$.

We now develop an analogous concept for the vector spaces $L^2_{\mathrm{PC}}(\Omega)$. In our discussion, we focus primarily on the case where $\Omega \subset \mathbb{R}$. However, it is rather straightforward to extend the discussion to general domains Ω in $\mathbb{R}^n$.

Our first order of business is to define a collection of functions that play a role similar to that played by the vectors (6.8). We say that a function $\phi \colon \Omega \to \mathbb{R}$ is a ***test function*** if

- $\phi \geq 0$,
- ϕ can be differentiated as many times as we like,
- $\phi = 0$ in a neighborhood of the boundary of Ω.

Figure 6.1 shows an example of a test function on the domain $\Omega = [-2, 2]$, as well as two nonexamples.

Test functions can be used to locally determine whether a function is zero.

theorem:test-functions

Theorem 6.1. *Suppose $f \in L^2_{\mathrm{PC}}(\Omega)$ is such that $\langle f, \phi \rangle = 0$ for all test functions ϕ on Ω. Then $f = 0$.*

The proof of Theorem 6.1, which we present here in the case that $\Omega \subset \mathbb{R}$, is the following: Suppose $f(x)$ is a continuous function that is not zero at some point. (If f is only piecewise continuous, we can restrict attention to a region where f is continuous.) Since the function f is continuous, there exists a small region where the $f(x)$ is nonzero in that region. Choose a test function $\phi(x)$ that is

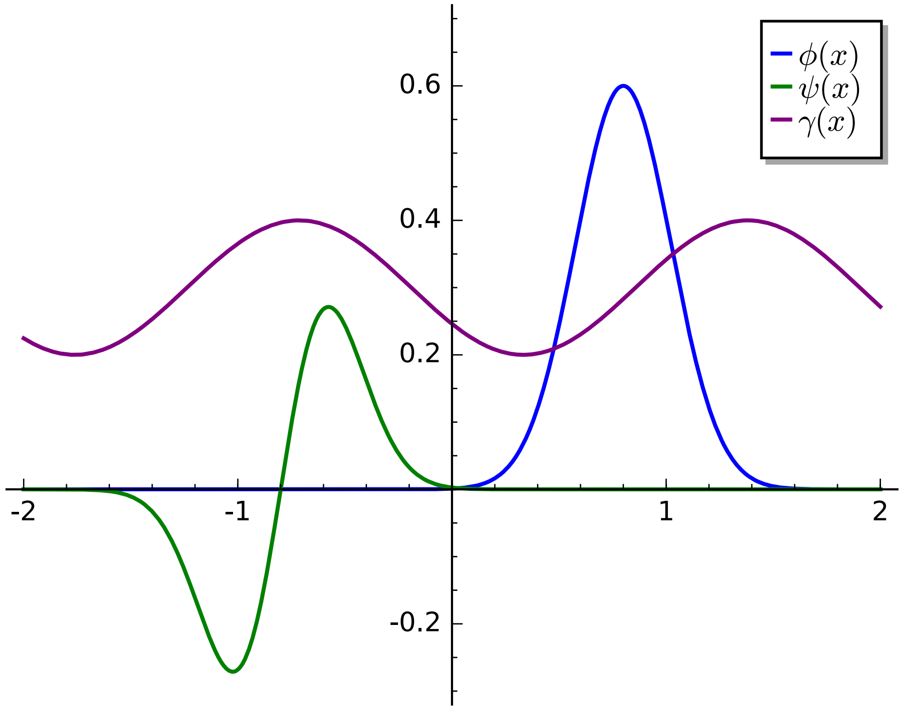

FIGURE 6.1. The function $\phi(x)$ is a test function on the domain $\Omega = [-2, 2]$. The function $\psi(x)$ is not a test function because it is not positive, while the function $\gamma(x)$ is not a test function because it does not vanish in a neighborhood of the boundary of Ω. `fig:bump-no-bump`

zero outside this small region, and that is nonzero in the middle of the region. Then the product $f(x)\phi(x)$ is not everywhere zero, and is either nonnegative or nonpositive (depending on the sign of f in the region under consideration); see Figure 6.2. Hence $\langle f, \phi \rangle$, which is the integral of the product $f(x)\phi(x)$ is either strictly positive or strictly negative. This contradicts our assumption that f is not zero at some point and hence we conclude that f is zero.

Theorem 6.1 can be used to justify a key step we made when deriving the wave equation in Chapter 3. There we argued that

$$\int_{-L}^{L} \frac{\partial u}{\partial t} \left\{ \frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} \right\} dx = 0 \tag{6.9}$$ `recall1D-conserve-ener`

implies that

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0. \tag{6.10}$$ `recall1D-PreWave`

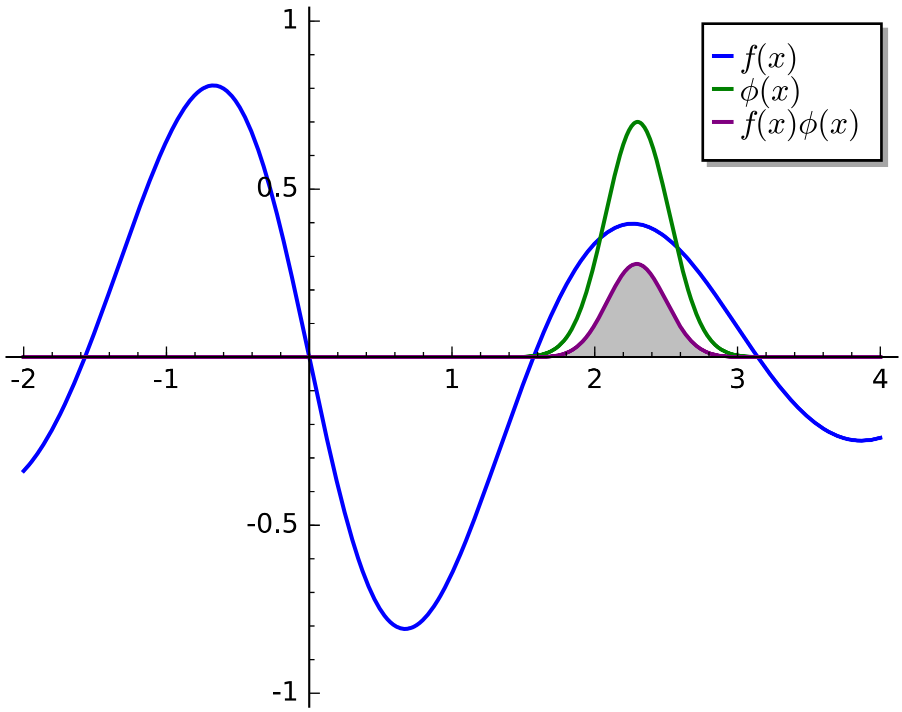

FIGURE 6.2. For each region where $f(x)$ is not zero we can find a test function $\phi(x)$ such that the integral of $f(x)\phi(x)$ is either strictly positive (as shown above) or strictly negative.

fig:bump-function-appl

We can now be a bit more precise, interpreting (6.9) as the condition that

$$\left\langle \frac{\partial u}{\partial t}, \frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} \right\rangle = 0$$

in the inner product space $L^2_{\mathrm{PC}}([-L, L])$. We want this to hold for any possible velocity function $\frac{\partial u}{\partial t}$; in particular, we want this to hold when the velocity function is a test function. Thus Theorem 6.1 implies that we must have (6.10). This same logic arises again in the next part of the course.

We conclude this section by remarking that, unlike the vectors in (6.8), the collection of test functions in $L^2_{\mathrm{PC}}(\Omega)$ are not mutually orthogonal. It is possible to choose a "more efficient" collection of functions that has better orthogonality properties. . . and behaves more like a collection of "basis vectors." One way to do this is through the theory of *wavelets*, a fascinating subject that is extremely useful in the field of image and signal processing.

6.5. Best approximations

We now consider another important application of orthogonality — the problem of best approximations.

We begin with the following problem: Suppose we have two vectors $\mathbf{v}$ and $\mathbf{u}$ in $\mathbb{R}^n$. Which multiple of $\mathbf{v}$ is closest to $\mathbf{u}$? That is, which vector pointing in the direction of $\mathbf{v}$ does the best job of approximating $\mathbf{u}$?

To answer this question, we define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(\alpha) = ||\mathbf{u} - \alpha\mathbf{v}||^2.$$

This function measures the extent to which $\alpha\mathbf{v}$ approximates $\mathbf{u}$; the best approximation corresponds to the value of α that minimizes f. In order to find the minimizer we compute (using the product rule)

$$f'(\alpha) = 2(\mathbf{u} - \alpha\mathbf{v}) \cdot (-\mathbf{v}).$$

Setting $f'(\alpha) = 0$, we find that the optimizer corresponds to

$$\alpha = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{v}||^2}.$$

This value of α represents the *extent to which* $\mathbf{u}$ *coincides with* $\mathbf{v}$ The resulting vector

$$\frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{v}||^2}\mathbf{v}$$

is called the ***projection*** of $\mathbf{u}$ on to the line determined by $\mathbf{v}$.

Let's now make the problem more interesting by supposing that we in fact have three vectors — $\mathbf{u}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ — and ask the question: Which combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ best approximates $\mathbf{u}$? Geometrically, we can interpret this question by noticing that the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ determine a plane. We are asking for the vector in that plane that is closest to $\mathbf{u}$.

We address the question by adapting the approach above. Defining $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(\alpha_1, \alpha_2) = ||\mathbf{u} - \alpha_1\mathbf{v}_1 - \alpha_2\mathbf{v}_2||^2,$$

we seek those α_1 and α_2 that minimize f. This time we need to use the optimization we learned in multivariable calculus — the minimizer occurs when the gradient of

f vanishes. We compute

$$\begin{aligned}\operatorname{grad} f(\alpha_1,\alpha_2) &= \left\langle \frac{\partial f}{\partial \alpha_1}, \frac{\partial f}{\partial \alpha_2} \right\rangle \\ &= \langle -2(\mathbf{u}-\alpha_1\mathbf{v}_1-\alpha_2\mathbf{v}_2)\cdot\mathbf{v}_1, -2(\mathbf{u}-\alpha_1\mathbf{v}_1-\alpha_2\mathbf{v}_2)\cdot\mathbf{v}_2\rangle.\end{aligned}$$

Thus we need to solve the equations

$$\begin{aligned}(\mathbf{u}-\alpha_1\mathbf{v}_1-\alpha_2\mathbf{v}_2)\cdot\mathbf{v}_1 &= 0\\ (\mathbf{u}-\alpha_1\mathbf{v}_1-\alpha_2\mathbf{v}_2)\cdot\mathbf{v}_2 &= 0.\end{aligned}$$

This is two equations with two unknowns, so there is hope that it can be solved... but it does not look to pretty. (Furthermore, if we were to generalize this to a list of vectors $\mathbf{v}_1,\ldots,\mathbf{v}_k$, then things would get very complicated quick quickly!)

To simplify our problem a bit, let us assume that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. In this case our two equations are easily solved; the result is

$$\alpha_1 = \frac{\mathbf{u}\cdot\mathbf{v}_1}{\|\mathbf{v}_1\|^2} \quad\text{and}\quad \alpha_2 = \frac{\mathbf{u}\cdot\mathbf{v}_2}{\|\mathbf{v}_2\|^2}.$$

In fact, we can now see how to easily generalize this to the case where we have lots of vectors $\mathbf{v}_i$. The result is the following.

stApproximationTheorem

Theorem 6.2 (Approximation by orthogonal vectors, version 1). *Suppose $\mathbf{v}_1,\ldots,\mathbf{v}_k$ is a collection of orthogonal vectors in $\mathbb{R}^n$. The combination of these vectors that best approximates vector $\mathbf{u}$ is*

$$\alpha_1\mathbf{v}_1+\cdots+\alpha_k\mathbf{v}_k,$$

where the scalars α_i are given by the formula

$$\alpha_i = \frac{\mathbf{u}\cdot\mathbf{v}_i}{\|\mathbf{v}_i\|^2}.$$

In Exercise 6.4 you write down a proof this theorem.

We now make an important observation: The entire discussion above is valid in any inner product space. Thus we obtain the following.

ndApproximationTheorem

Theorem 6.3 (Approximation by orthogonal vectors, version 2). *Suppose $v_1,\ldots,v_k$ is a collection of orthogonal vectors in inner product space V. The combination of these vectors that best approximates vector u is*

$$\alpha_1 v_1+\cdots+\alpha_k v_k,$$

where the scalars α_i are given by the formula

$$\alpha_i = \frac{\langle u, v_i \rangle}{\|\mathbf{v}_i\|^2}.$$

We conclude this section with a remark for those students who have taken the linear algebra course: The formula for the best approximation should look familiar... it is simply the formula for the projection on to the subspace spanned by $v_1, \ldots, v_k$.

imate-by-periodic-trig

EXAMPLE 6.7 (APPROXIMATION BY PERIODIC COSINES AND SINES). *In Example 5.16 you found a list of cosine and sine functions in $L^2([-1,1])$ that satisfy the periodic boundary conditions. They are*

$$\begin{aligned} v_0 &= 1 & & \\ v_1 &= \cos(\pi x) & w_1 &= \sin(\pi x) \\ v_2 &= \cos(2\pi x) & w_2 &= \sin(2\pi x) \\ &\vdots & &\vdots \\ v_k &= \cos(k\pi x) & w_k &= \sin(k\pi x) \\ &\textit{etc.} & & \end{aligned}$$

It is easy to check that these functions form an orthogonal sequence. Furthermore, we can compute that $\|v_0\|^2 = 2$, and that for $k \geq 1$ we have $\|v_k\|^2 = 1$ and $\|w_k\|^2 = 1$.

Suppose we want to best approximate the function $u(x) = x$ by a linear combination of $v_0, \ldots, v_5$ and $w_1, \ldots, w_5$. We compute

$$\langle v_k, u \rangle = 0 \quad \textit{and} \quad \langle w_k, u \rangle = \frac{(-1)^{k+1}}{k\pi}.$$

Thus the best approximation is

$$\frac{2\sin(\pi x)}{\pi} - \frac{2\sin(2\pi x)}{2\pi} + \frac{2\sin(3\pi x)}{3\pi} - \frac{2\sin(4\pi x)}{4\pi} + \frac{2\sin(5\pi x)}{5\pi};$$

see Figure 6.3.

Notice that none of the cosine functions were needed for the approximation. This is because $u(x) = x$ is an odd function and the cosine functions are even.

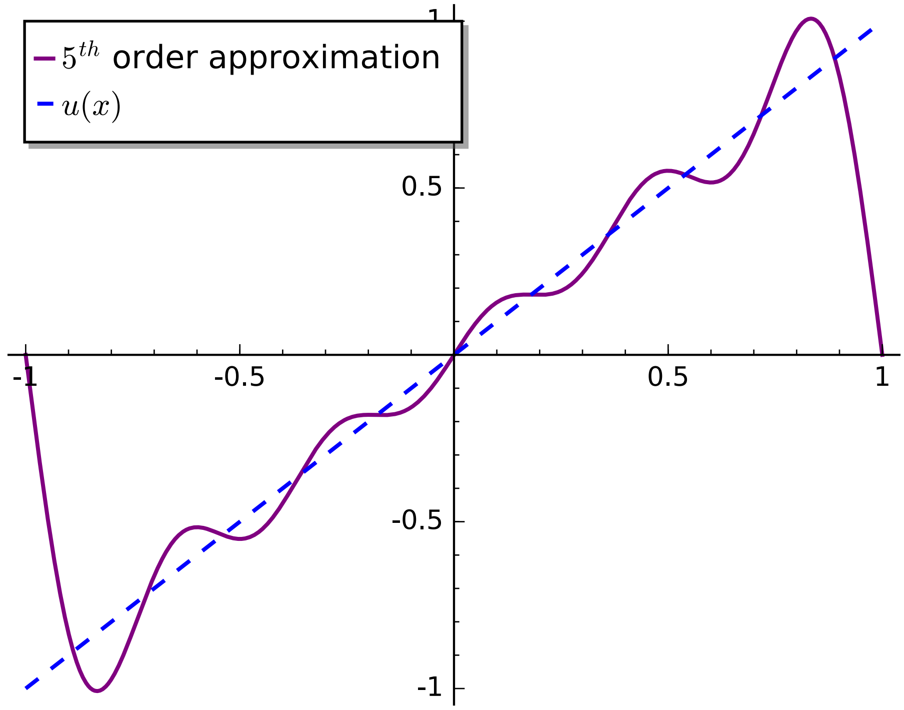

FIGURE 6.3. The 5^{th} order approximation of the function $u(x) = x$ by periodic trigonometric functions.

fig:approximate-x

6.6. Inner products for complex vector spaces

The goal of this section is to introduce inner products for complex vector spaces; recall that this means that scalars come from $\mathbb{C}$. Before we can do this, we need to introduce two concepts regarding complex numbers: modulus and conjugate.

Real numbers have the following property: If we square a real number, the resulting number is greater than or equal to zero, and equal only if the number itself is zero. However, complex numbers do not have this property. For example, $(1-i)^2 = -2i$; we can't even make sense of whether this is "greater than zero" or not.

When working with real vector spaces, the square of a real number is used to determine "how big" the number is (by which we mean "to what extent is the number not zero"). Suppose we have a complex number $z = a + bi$ (where $a, b \in \mathbb{R}$). We can measure the extent to which z is not zero by defining the ***modulus*** of z, which we denote $|z|$, by

$$|z|^2 = a^2 + b^2.$$

For example $|1-i| = \sqrt{(1)^2+(-1)^2} = \sqrt{2}$. It is clearly true that the modulus of a complex number is always greater than or equal to zero, and that it is zero only when the number is zero.

Example 6.8. *We can compute the modulus of $e^{i\pi/7}$ by writing $e^{i\pi/7} = \cos\frac{\pi}{7} + i\sin\frac{\pi}{7}$. We then compute*

$$|e^{i\pi/7}|^2 = \left(\cos\frac{\pi}{7}\right)^2 + \left(\sin\frac{\pi}{7}\right)^2 = 1.$$

Thus $|e^{i\pi/7}| = 1$.

We now introduce a clever way to compute the modulus of a complex number $z = a + ib$. Notice that we can factor

$$|z|^2 = a^2 + b^2 = (a+ib)(a-ib).$$

The first term on the right is the number z itself, while the second term is the number z with the sign changed on the imaginary part. Thus in order to compute the modulus of a complex number, all we need to do is compute

$$(\text{the complex number})(\text{the complex number with } i \text{ replaced by } -i).$$

This motivates the following definition: Suppose we have a complex number $z = a + ib$. The ***complex conjugate*** of z, which is given the symbol $\bar{z}$, is the number obtained by replacing i with $-i$. Thus

$$z = a + ib \quad \Rightarrow \quad \bar{z} = a - ib \quad \Rightarrow \quad |z|^2 = z\bar{z}.$$

Example 6.9. *We compute the modulus of $e^{i\pi/7}$ by*

$$\left|e^{i\pi/7}\right|^2 = e^{i\pi/7}e^{-i\pi/7} = e^0 = 1.$$

Thus $\left|e^{i\pi/7}\right| = 1$. Notice that this was more efficient than the previous method.

We are now ready to discuss inner products for complex vector spaces. Our approach is to modify the definition used for real vector spaces, using complex conjugates. Suppose V is a complex vector space. A function $\langle\cdot,\cdot\rangle$ that takes in two vectors v, w and gives out a complex number is a ***complex inner product*** if

- *Linearity property*. We have $\langle \alpha_1 v_1 + \alpha_2 v_2, w\rangle = \alpha_1\langle v_1, w\rangle + \alpha_2\langle v_2, w\rangle$ for all complex scalars α_1, α_2 and all vectors v_1, v_2, w.
- *Conjugate linearity property*. We have $\langle v, \beta_1 w_1 + \beta_2 w_2\rangle = \bar{\beta}_1\langle v, w_1\rangle + \bar{\beta}_2\langle v, w_2\rangle$ for all complex scalars β_1, β_2 and all vectors v, w_1, w_2.

- *Conjugate symmetry property*. We have $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all vectors v, w.
- *Positive definite property*. For all vectors v the quantity $\langle v, v\rangle$ is real and $\langle v, v\rangle \geq 0$ with equality only when $v = 0$.

The positive definite property means that we can define the norm of a vector by

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

The following is an important example of a complex inner product space.

ComplexL2Interval

EXAMPLE 6.10. *Consider the collection of functions $u\colon [-1, 1] \to \mathbb{C}$ with inner product*

$$\langle u, v\rangle = \int_{-1}^{1} u(x)\,\overline{v(x)}\,dx.$$

It is easy to check that this satisfies all the properties listed above.

We define the complex inner product space $L^2([-1, 1])$ to those functions for which

$$\langle u, u\rangle = \int_{-1}^{1} |u(x)|^2 dx < \infty.$$

(Notice that we give this space the same symbol as the real one – one must know from context whether we are working with real-valued functions or complex-valued functions! See, however, the remark following this example.)

It is easy to see that the functions $\psi_k = e^{ik\pi x}$ form an orthogonal sequence. We compute

$$\|\psi_k\|^2 = 2.$$

Consider the function $u(x) = x$. We can find the combination of $\psi_{-5}, \ldots, \psi_5$ that best approximates u by computing

$$\alpha_k = \frac{\langle u, \psi_k\rangle}{2} = \frac{1}{2}\int_{-1}^{1} xe^{-ik\pi x}\,dx = \frac{(-1)^k}{k\pi i} \qquad \text{if } k \neq 0$$
$$\alpha_0 = 0.$$

Thus the best approximation is

$$\frac{1}{5\pi i}e^{-i5\pi x} - \frac{1}{4\pi i}e^{-i4\pi x} + \cdots + \frac{1}{\pi i}e^{-i\pi x} - \frac{1}{\pi i}e^{i\pi x} + \cdots + \frac{1}{4\pi i}e^{i4\pi x} - \frac{1}{5\pi i}e^{i5\pi x}.$$

Notice that this is equal to

$$\frac{2}{\pi}\frac{e^{i\pi x}-e^{-i\pi x}}{2i}-\cdots-\frac{2}{4\pi}\frac{e^{i4\pi x}-e^{-i4\pi x}}{2i}+\frac{2}{5\pi}\frac{e^{i5\pi x}-e^{-i5\pi x}}{2i}$$
$$=\frac{2}{\pi}\sin(\pi x)-\cdots-\frac{2}{4\pi}\sin(4\pi x)+\frac{2}{5\pi}\sin(5\pi x),$$

which agrees with the computation in Example 6.7.

The last part of Example 6.10 illustrates in important point: All real-valued functions may be viewed a complex-valued functions whose outputs happen to be real. For these functions the real L^2 inner product is the same as the complex inner product. (This is simply because the conjugate of a real number is itself.) Thus the complex L^2 inner product spaces are simply extensions of the real ones.

We conclude this section with a warning.

WARNING 6.1. *In physics, particularly in quantum mechanics, it is common to use different notations and conventions for complex inner products. In particular,*

(1) *You may see the symbol u^* used for the conjugate, rather than the symbol $\bar{u}$ that we use in this course.*
(2) *It is also common to use the so-called "bra-ket" notation.*
(3) *You may also see the complex conjugate applied to the first, rather than the second, function in the inner product.*

The cumulation of these points is that the following are equal to one another:

$$\textbf{\textit{Math}} \quad \langle\psi,\phi\rangle = \int \psi(x)\overline{\phi(x)}\,dx$$
$$\textbf{\textit{Physics}} \quad \langle\phi|\psi\rangle = \int \phi(x)^*\psi(x)\,dx.$$

Exercises

Exercise 6.1. Consider the vectors

$$\mathbf{u}=\begin{pmatrix}1\\2\\3\end{pmatrix} \quad \mathbf{v}_1=\begin{pmatrix}1\\1\\1\end{pmatrix} \quad \mathbf{v}_2=\begin{pmatrix}1\\-2\\1\end{pmatrix}.$$

Find the combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ that best approximates $\mathbf{u}$.

Draw a picture of $\mathbf{u}$, $\mathbf{v}_1$, $\mathbf{v}_2$, and the approximating vector. Geometrically, what is happening?

ex:FirstLegendreList **Exercise 6.2.** Consider $L^2([-1,1])$ with the Legendre inner product (6.4). Show that the following polynomials are orthogonal:

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2 - 1) \quad P_3(x) = \frac{1}{2}(5x^3 - 3x).$$

Show that the norms of these polynomials are given by $\|P_n\|^2 = \frac{2}{2n+1}$.

Exercise 6.3. Let P_0, P_1, P_2 be as in Exercise 6.2. Find the combination of these polynomials that best approximates the function e^x in the vector space $L^2([-1,1))$.

veApproximationTheorem **Exercise 6.4.** Give a proof of the approximation theorems, Theorem 6.2 and Theorem 6.3 by constructing a function $f(\alpha_1, \alpha_2, \ldots, \alpha_k)$ and finding critical points. Be sure to explain each step!

Exercise 6.5. Establish the Cauchy-Schwartz inequality:

$$|\langle u, v\rangle| \leq \|u\|\,\|v\|. \tag{6.11}$$ CauchySchwartz

To do this, set $a = \frac{\langle u,v\rangle}{\|v\|^2}$. Then use the fact that

$$0 \leq \|u - av\|^2 = \langle u - av, u - av\rangle.$$

se:NormOfApproximation **Exercise 6.6.** Let $v_1, \ldots, v_N$ be a list of orthogonal vectors in an inner product space. Suppose $\beta_1, \ldots, \beta_N$ are constants. Show that

$$\left\|\sum_{k=1}^{N} \beta_k v_k\right\|^2 = \sum_{k=1}^{N} \|\beta_k v_k\|^2.$$

approximation-and-perp **Exercise 6.7.** Let $v_1, \ldots, v_N$ be orthogonal vectors in an inner product space, u some vector, and u_N be the best approximation of u by the v_k:

$$u_N = \sum_{k=1}^{N} \alpha_k v_k, \qquad \alpha_k = \frac{\langle u, v_k\rangle}{\|v_k\|^2}.$$

The goal of this exercise is to show that

$$\|u_N\| \leq \|u\|. \tag{6.12}$$ BasicApproximationBour

(1) First, define $u^\perp = u - u_N$. Show by direct computation that $u^\perp$ is orthogonal to u_N.
(2) Use the fact that $u = u_N + u^\perp$, and the orthogonality of $u^\perp$ and u_N to conclude that

$$\|u\|^2 = \|u_N\|^2 + \|u^\perp\|^2.$$

Explain why this implies (6.12).

Exercise 6.8. In Exercise 5.4 you found a list of cosine and sine functions in $L^2([-1, 1])$ that satisfy the Dirichlet boundary conditions.

(1) Show that these functions are all orthogonal to one another.
(2) Find the combination of the first 5 of these sine functions that best approximates the function $f(x) = x$. Make a plot of the approximation.

Exercise 6.9. In Exercise 5.5 you found a list of cosine and sine functions in $L^2([-1, 1])$ that satisfy the Neumann boundary conditions.

(1) Show that these functions are all orthogonal to one another.
(2) Find the combination of the first 5 of these sine functions that best approximates the function $f(x) = x$. Make a plot of the approximation.

Exercise 6.10. Let $\psi_k = e^{ik\pi x}$ be the orthogonal functions appearing in Example 6.10. Find the combination of $\psi_{-5}, \dots, \psi_5$ that best approximates the function

$$u(x) = \begin{cases} 1 & \text{if } |x| < a, \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 6.11. Consider $L^2([-1, 1])$ with the Legendre inner product (6.4). Explain why any even function must be orthogonal to any odd function.

-pathological-sequence

Exercise 6.12. (Optional) In this exercise we explore some of the technical points of $L^2([0, 1])$ by means of the list of functions f_n defined by

$$f_n(x) = \begin{cases} 1 - nx & 0 \le x \le \frac{1}{n}, \\ 0 & \frac{1}{n} \le x \le 1. \end{cases}$$

(1) Compute $\|f_n\|$ and verify that $\|f_n\| \to 0$ as $n \to \infty$.
(2) Make a plot of the graph of f_n. Using the plot, explain why in the limit $n \to \infty$ we expect that f_n should converge to the function

$$f_\infty(x) = \begin{cases} 1 & x = 0, \\ 0 & 0 < x \le 1. \end{cases}$$

(3) Suppose $g(x)$ is some continuous function. Explain why

$$\int_0^1 f_\infty(x) g(x) \, dx = 0.$$

(4) Explain why it is convenient to consider f_∞ and the zero function to be the same thing.