BACKGROUND NOTES FOR APPROXIMATION THEORY
MATH 441, FALL 2009

THOMAS SHORES

Contents

Last Rev.: 10/22/09

1. Vector Spaces

Here is the general definition of a vector space:

Definition 1.1. An *(abstract) vector space* is a nonempty set V of elements called vectors, together with operations of vector addition (+) and scalar multiplication ($\cdot$), such that the following laws hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and scalars $a, b \in \mathbb{F}$:

(1): (Closure of vector addition) $\mathbf{u} + \mathbf{v} \in V$.
(2): (Commutativity of addition) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
(3): (Associativity of addition) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
(4): (Additive identity) There exists an element $\mathbf{0} \in V$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u} = \mathbf{0} + \mathbf{u}$.
(5): (Additive inverse) There exists an element $-\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0} = (-\mathbf{u}) + \mathbf{u}$.
(6): (Closure of scalar multiplication) $a \cdot \mathbf{u} \in V$.
(7): (Distributive law) $a \cdot (\mathbf{u} + \mathbf{v}) = a \cdot \mathbf{u} + a \cdot \mathbf{v}$.

(8): (Distributive law) $(a+b)\cdot\mathbf{u} = a\cdot\mathbf{u} + b\cdot\mathbf{u}$.
(9): (Associative law) $(ab)\cdot\mathbf{u} = a\cdot(b\cdot\mathbf{u})$.
(10): (Monoidal law) $1\cdot\mathbf{u} = \mathbf{u}$.

About notation: just as in matrix arithmetic, for vectors $\mathbf{u}, \mathbf{v} \in V$, we understand that $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$. We also suppress the dot $(\cdot)$ of scalar multiplication and usually write $a\mathbf{u}$ instead of $a\cdot\mathbf{u}$.

About scalars: the only scalars that we will use in 441 are the real numbers $\mathbb{R}$.

Definition 1.2. Given a positive integer n, we define the *standard vector space of dimension n over the reals* to be the set of vectors

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \in \mathbb{R}\}$$

together with the standard vector addition and scalar multiplication. (Recall that $(x_1, x_2, \ldots, x_n)$ is shorthand for the column vector $[x_1, x_2, \ldots, x_n]^T$.)

We see immediately from the definition that the required closure properties of vector addition and scalar multiplication hold, so these really are vector spaces in the sense defined above. The standard real vector spaces are often called the real Euclidean vector spaces once the notion of a norm (a notion of length covered in the next chapter) is attached to them.

Example 1.3. Let $C[a,b]$ denote the set of all real-valued functions that are continuous on the interval $[a,b]$ and use the standard function addition and scalar multiplication for these functions. That is, for $f(x), g(x) \in C[a,b]$ and real number c, we define the functions $f+g$ and cf by

$$\begin{aligned}(f+g)(x) &= f(x) + g(x)\\ (cf)(x) &= c(f(x)).\end{aligned}$$

Show that $C[a,b]$ with the given operations is a vector space.

We set $V = C[a,b]$ and check the vector space axioms for this V. For the rest of this example, we let f, g, h be arbitrary elements of V. We know from calculus that the sum of any two continuous functions is continuous and that any constant times a continuous function is also continuous. Therefore the closure of addition and that of scalar multiplication hold. Now for all x such that $a \le x \le b$, we have from the definition and the commutative law of real number addition that

$$(f+g)(x) = f(x) + g(x) = g(x) + f(x) = (g+f)(x).$$

Since this holds for all x, we conclude that $f+g = g+f$, which is the commutative law of vector addition. Similarly,

$$\begin{aligned}((f+g)+h)(x) &= (f+g)(x) + h(x) = (f(x)+g(x)) + h(x)\\ &= f(x) + (g(x)+h(x)) = (f+(g+h))(x).\end{aligned}$$

Since this holds for all x, we conclude that $(f+g)+h = f+(g+h)$, which is the associative law for addition of vectors.

Next, if 0 denotes the constant function with value 0, then for any $f \in V$ we have that for all $a \le x \le b$,

$$(f+0)(x) = f(x) + 0 = f(x).$$

(We don't write the zero element of this vector space in boldface because it's customary not to write functions in bold.) Since this is true for all x we have that $f+0 = f$, which establishes the additive identity law. Also, we define $(-f)(x) = -(f(x))$ so that for all $a \leq x \leq b$,

$$(f + (-f))(x) = f(x) - f(x) = 0,$$

from which we see that $f + (-f) = 0$. The additive inverse law follows. For the distributive laws note that for real numbers c, d and continuous functions $f, g \in V$, we have that for all $a \leq x \leq b$,

$$c(f + g)(x) = c(f(x) + g(x)) = cf(x) + cg(x) = (cf + cg)(x),$$

which proves the first distributive law. For the second distributive law, note that for all $a \leq x \leq b$,

$$((a + b)g)(x) = (a + b)g(x) = ag(x) + bg(x) = (ag + bg)(x),$$

and the second distributive law follows. For the scalar associative law, observe that for all $a \leq x \leq b$,

$$((cd)f)(x) = (cd)f(x) = c(df(x)) = (c(df))(x),$$

so that $(cd)f = c(df)$, as required. Finally, we see that

$$(1f)(x) = 1f(x) = f(x),$$

from which we have the monoidal law $1f = f$. Thus, $C[a, b]$ with the prescribed operations is a vector space. □

A similar argument shows that, for example, the set of polynomial functions on $[a, b]$ is a vector space, but is it much easier to simply appeal to the Subspace Test, now that we know that $C[a, b]$ is a vector space. Recall from Math 314:

Definition 1.4. A *subspace* of the vector space V is a subset W of V such that W, together with the binary operations it inherits from V, forms a vector space (over the same field of scalars as V) in its own right.

Given a subset W of the vector space V, we can apply the definition of vector space directly to the subset W to obtain the following very useful test.

Theorem 1.5. *Let W be a subset of the vector space V. Then W is a subspace of V if and only if*

- **(1):** *W contains the zero element of V.*
- **(2):** *(Closure of addition) For all $\mathbf{u}, \mathbf{v} \in W$, $\mathbf{u} + \mathbf{v} \in W$.*
- **(3):** *(Closure of scalar multiplication) For all $\mathbf{u} \in W$ and scalars c, $c\mathbf{u} \in W$.*

It is easy to see that the

Example 1.6. Let $C^{(n)}[a, b]$ denote the set of all real-valued functions whose n-th derivative $f^{(n)}(x)$ is continuous on the interval $[a, b]$. Show that $C^{(n)}[a, b]$ is a subspace of the vector space $C[a, b]$.

Notice that with the notation of this example, $C[a, b] = C^{(0)}[a, b]$. Furthermore, one can check easily the the set $\mathcal{P}$ of all polynomial functions and the sets $\mathcal{P}_m$ of polynomial functions of degree at most m, are subspaces of any $C^{(n)}[a, b]$. Finally, we recall some key ideas about linear combinations of vectors.

Definition 1.7. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be vectors in the vector space V. The *span* of these vectors, denoted by span $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, is the subset of V consisting of all possible linear combinations of these vectors, i.e.,

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \mid c_1, c_2, \ldots, c_n \text{ are scalars}\}$$

Recall that one can show that spans are always subspaces of the containing vector space. A few more fundamental ideas:

Definition 1.8. The vectors $v_1, v_2, \ldots, v_n$ are said to be *linearly dependent* if there exist scalars $c_1, c_2, \ldots, c_n$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = 0. \quad (1.1)$$

Otherwise, the vectors are called *linearly independent*.

Definition 1.9. A *basis* for the vector space V is a spanning set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ that is a linearly independent set.

In particular, we have the following key facts about bases:

Theorem 1.10. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis of the vector space V. Then every $\mathbf{v} \in V$ can be expressed uniquely as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, up to order of terms.*

Definition 1.11. The vector space V is called *finite-dimensional* if V has a finite spanning set.

It's easy to see that every finite-dimensional space has a basis: simply start with a spanning set and throw away elements until you have reduced the set to a minimal spanning set. Such a set will be linearly independent.

Theorem 1.12. *Let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ be a linearly independent set in the space V and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis of V. Then $r \leq n$ and we may substitute all of the $\mathbf{w}_i$'s for r of the $\mathbf{v}_j$'s in such a way that the resulting set of vectors is still a basis of V.*

As a consequence of the Steinitz substitution principle above, we obtain this central result:

Theorem 1.13. *Let V be a finite-dimensional vector space. Then any two bases of V have the same number of elements, which is called the dimension of the vector space and denoted by* $\dim V$

Exercises

Exercise 1.1. Show that the subset $\mathcal{P}$ of $C[a,b]$ consisting of all polynomial functions is a subspace of $C[a,b]$ and that the subset $\mathcal{P}_n$ consisting of all polynomials of degree at most n on the interval $[a,b]$ is a subspace of $\mathcal{P}$.

Exercise 1.2. Show that $\mathcal{P}_n$ is a finite dimensional vector space of dimension n, but that $\mathcal{P}$ is not a finite dimensional space, that is, does not have a finite vector basis (linearly independent spanning set).

2. Norms

The basic definition:

Definition 2.1. A *norm* on the vector space V is a function $\|\cdot\|$ that assigns to each vector $\mathbf{v} \in V$ a real number $\|\mathbf{v}\|$ such that for c a scalar and $\mathbf{u}, \mathbf{v} \in V$ the following hold:

- **(1):** $\|\mathbf{u}\| \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$.
- **(2):** $\|c\mathbf{u}\| = |c| \, \|\mathbf{u}\|$.

(3): (Triangle Inequality) $\|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$.

A vector space V, together with a norm $\|\cdot\|$ on the space V, is called a *normed vector (or linear) space.* If $\mathbf{u}, \mathbf{v} \in V$, the distance between $\mathbf{u}$ and $\mathbf{v}$ is defined to be $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Notice that if V is a normed space and W is any subspace of V, then W automatically becomes a normed space if we simply use the norm of V on elements of W. Obviously all the norm laws still hold, since they hold for elements of the bigger space V.

Of course, we have already studied some very important examples of normed spaces, namely the standard vector space $\mathbb{R}^n$, or any subspace thereof, together with the standard norms given by

$$\|(z_1, z_2, \ldots, z_n)\| = \left(|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2\right)^{1/2}.$$

Since our vectors are real then we can drop the conjugate bars. This norm is actually one of a family of norms that are commonly used.

Definition 2.2. Let V be one of the standard spaces $\mathbb{R}^n$ and $p \ge 1$ a real number. The p-norm of a vector in V is defined by the formula

$$\|(z_1, z_2, \ldots, z_n)\|_p = \left(|z_1|^p + |z_2|^p + \cdots + |z_n|^p\right)^{1/p}.$$

Notice that when $p = 2$ we have the familiar example of the standard norm. Another important case is that in which $p = 1$. The last important instance of a p-norm is one that isn't so obvious: $p = \infty$. It turns out that the value of this norm is the limit of p-norms as $p \to \infty$. To keep matters simple, we'll supply a separate definition for this norm.

Definition 2.3. Let V be one of the standard spaces $\mathbb{R}^n$ or $\mathbb{C}^n$**.** The ∞-norm of a vector in V is defined by the formula

$$\|(z_1, z_2, \ldots, z_n)\|_\infty = \max\{|z_1|, |z_2|, \ldots, |z_n|\}.$$

Example 2.4. Calculate $\|\mathbf{v}\|_p$, where $p = 1$, 2, or ∞ and $\mathbf{v} = (1, -3, 2, -1) \in \mathbb{R}^4$.

We calculate:

$$\begin{aligned}
\|(1,-3,2,-1)\|_1 &= |1| + |-3| + |2| + |-1| = 7 \\
\|(1,-3,2,-1)\|_2 &= \sqrt{|1|^2 + |-3|^2 + |2|^2 + |-1|^2} = \sqrt{15} \\
\|(1,-3,2,-1))\|_\infty &= \max\{|1|, |-3|, |2|, |-1|\} = 3.
\end{aligned}$$

□

It may seem a bit odd at first to speak of the same vector as having different lengths. You should take the point of view that choosing a norm is a bit like choosing a measuring stick. If you choose a yard stick, you won't measure the same number as you would by using a meter stick on an object.

Example 2.5. Verify that the norm properties are satisfied for the p-norm in the case that $p = \infty$.

Let c be a scalar, and let $\mathbf{u} = (z_1, z_2, \ldots, z_n)$, and $\mathbf{v} = (w_1, w_2, \ldots, w_n)$ be two vectors. Any absolute value is nonnegative, and any vector whose largest component in absolute value is zero must have all components equal to zero. Property (1) follows. Next, we have that

$$\begin{aligned}\|c\mathbf{u}\|_\infty &= \|(cz_1, cz_2, \ldots, cz_n)\|_\infty \\ &= \max\{|cz_1|, |cz_2|, \ldots, |cz_n|\} \\ &= |c| \max\{|z_1|, |z_2|, \ldots, |z_n|\} = |c| \, \|\mathbf{u}\|_\infty ,\end{aligned}$$

which proves (2). For (3) we observe that

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|_\infty &= \max\{|z_1| + |w_1|, |z_2| + |w_2|, \ldots, |z_n| + |w_n|\} \\ &\leq \max\{|z_1|, |z_2|, \ldots, |z_n|\} + \max\{|w_1|, |w_2|, \ldots, |w_n|\} \\ &\leq \|\mathbf{u}\|_\infty + \|\mathbf{v}\|_\infty .\end{aligned}$$

□

2.1. **Unit Vectors.** Sometimes it is convenient to deal with vectors whose length is one. Such a vector is called a *unit vector*. We saw in Chapter 3 that it is easy to concoct a unit vector $\mathbf{u}$ in the same direction as a given nonzero vector $\mathbf{v}$ when using the standard norm, namely take

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}. \tag{2.1}$$

The same formula holds for any norm whatsoever because of norm property (2).

Example 2.6. Construct a unit vector in the direction of $\mathbf{v} = (1, -3, 2, -1)$, where the 1-norm, 2-norm, and ∞-norms are used to measure length.

We already calculated each of the norms of $\mathbf{v}$. Use these numbers in equation (2.1) to obtain unit-length vectors

$$\begin{aligned}\mathbf{u}_1 &= \frac{1}{7}(1, -3, 2, -1) \\ \mathbf{u}_2 &= \frac{1}{\sqrt{15}}(1, -3, 2, -1) \\ \mathbf{u}_\infty &= \frac{1}{3}(1, -3, 2, -1).\end{aligned}$$

From a geometric point of view there are certain sets of vectors in the vector space V that tell us a lot about distances. These are the so-called *balls about a vector (or point)* $\mathbf{v}_0$ *of radius* r, whose definition is as follows:

Definition 2.7. The (closed) ball of radius r centered at the vector $\mathbf{v}_0$ *is the set of vectors*

$$B_r(\mathbf{v}_0) = \{\mathbf{v} \in V \mid \|\mathbf{v} - \mathbf{v}_0\| \leq r\}.$$

The set of vectors

$$B_r^o(\mathbf{v}_0) = \{\mathbf{v} \in V \mid \|\mathbf{v} - \mathbf{v}_0\| < r\}.$$

is the *open* ball of radius r centered at the vector $\mathbf{v}_0$.

Here is a situation very important to approximation theory in which these balls are helpful: imagine trying to find the distance from a given vector $\mathbf{v}_0$ to a closed (this means it contains all points on its boundary) set S of vectors that need not be a subspace. One way to accomplish this is to start with a ball centered at $\mathbf{v}_0$ so small that the ball avoids S. Then expand this ball by increasing its radius until you have found a least radius r such that the ball $B_r(\mathbf{v}_0)$ intersects S nontrivially. Then the distance from $\mathbf{v}_0$ to this set is this number r. Actually, this is a reasonable

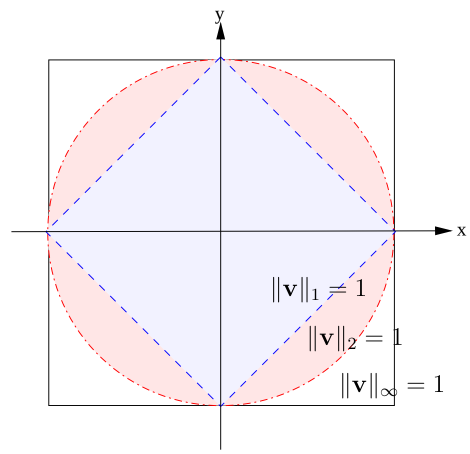

FIGURE 2.1. Boundaries of unit balls in various norms.

definition of the distance from $\mathbf{v}_0$ to the set S. One expects these balls, for a given norm, to have the same shape, so it is sufficient to look at the unit balls, that is, the case $r = 1$.

Example 2.8. Sketch the unit balls centered at the origin for the 1-norm, 2-norm, and ∞-norms in the space $V = \mathbb{R}^2$.

In each case it's easiest to determine the boundary of the ball $B_1(0)$, i.e., the set of vectors $\mathbf{v} = (x, y)$ such that $\|\mathbf{v}\| = 1$. These boundaries are sketched in Figure 2.1, and the ball consists of the boundaries plus the interior of each boundary. Let's start with the familiar 2-norm. Here the boundary consists of points (x, y) such that

$$1 = \|(x, y)\|_2 = \sqrt{x^2 + y^2},$$

which is the familiar circle of radius 1 centered at the origin. Next, consider the 1-norm, in which case

$$1 = \|(x, y)\|_1 = |x| + |y|.$$

It's easier to examine this formula in each quadrant, where it becomes one of the four possibilities

$$\pm x \pm y = 1.$$

For example, in the first quadrant we get $x + y = 1$. These equations give lines that connect to form a square whose sides are diagonal lines. Finally, for the ∞-norm we have

$$1 = |(x, y)|_\infty = \max\{|x|, |y|\},$$

which gives four horizontal and vertical lines $x = \pm 1$ and $y = \pm 1$. These intersect to form another square. Thus we see that the unit "balls" for the 1- and ∞-norms have corners, unlike the 2-norm. See Figure 2.1 for a picture of these balls. □

One of the important applications of the norm concept is that it enables us to make sense out of the idea of limits and convergence of vectors. In a nutshell, $\lim_{n\to\infty} \mathbf{v}_n = \mathbf{v}$ was taken to mean that $\lim_{n\to\infty} \|\mathbf{v}_n - \mathbf{v}\| = 0$. In this case we said that the sequence $\mathbf{v}_1, \mathbf{v}_2, \ldots$ *converges* to $\mathbf{v}$. Will we have to have a different notion of limits for different norms? For *finite-dimensional* spaces, the somewhat surprising answer is no. The reason is that given any two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on a finite-dimensional vector space, it is always possible to find positive real constants c and d such that for any vector $\mathbf{v}$,

$$\|\mathbf{v}\|_a \leq c \cdot \|\mathbf{v}\|_b \text{ and } \|\mathbf{v}\|_b \leq d\,\|\mathbf{v}\|_a .$$

Hence, if $\|\mathbf{v}_n - \mathbf{v}\|$ tends to 0 in one norm, it will tend to 0 in the other norm. For this reason, any two norms satisfying these inequalities are called *equivalent*. It can be shown that all norms on a finite-dimensional vector space are equivalent. Indeed, it can be shown that the condition that $\|\mathbf{v}_n - \mathbf{v}\|$ tends to 0 in any one norm is equivalent to the condition that each coordinate of $\mathbf{v}_n$ converges to the corresponding coordinate of $\mathbf{v}$. We will verify the limit fact in the following example.

Example 2.9. Verify that $\lim_{n\to\infty} \mathbf{v}_n$ exists and is the same with respect to both the 1-norm and 2-norm, where

$$\mathbf{v}_n = \begin{bmatrix} (1-n)/n \\ e^{-n}+1 \end{bmatrix}.$$

Which norm is easier to work with?

First we have to know what the limit will be. Let's examine the limit in each coordinate. We have

$$\lim_{n\to\infty} \frac{1-n}{n} = \lim_{n\to\infty} \frac{1}{n} - 1 = 0 - 1 = -1 \text{ and } \lim_{n\to\infty} e^{-n} + 1 = 0 + 1 = 1.$$

So we try to use $\mathbf{v} = (-1, 1)$ as the limiting vector. Now calculate

$$\mathbf{v} - \mathbf{v}_n = \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1-n}{n} \\ e^{-n}+1 \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \\ e^{-n} \end{bmatrix},$$

so that

$$\|\mathbf{v} - \mathbf{v}_n\|_1 = \left|\frac{1}{n}\right| + \left|e^{-n}\right| \underset{n\to\infty}{\longrightarrow} 0$$

and

$$\|\mathbf{v} - \mathbf{v}_n\| = \sqrt{\left(\frac{1}{n}\right)^2 + \left(e^{-n}\right)^2} \underset{n\to\infty}{\longrightarrow} 0,$$

which shows that the limits are the same in either norm. In this case the 1-norm appears to be easier to work with, since no squaring and square roots are involved. □

Here are two examples of norms defined on nonstandard vector spaces:

Definition 2.10. The *p-norm* on $C[a, b]$ is defined by $\|f\|_p = \left\{\int_a^b |f(x)|^p dx\right\}^{1/p}$.

Although this is a common form of the definition, a better form that is often used is

$$\|f\|_p = \left\{\frac{1}{b-a}\int_a^b |f(x)|^p dx\right\}^{1/p}.$$

This form is better in the sense that it scales the size of the interval.

Definition 2.11. The *uniform* (*or* infinity) *norm* on $C[a, b]$ is defined by $\|f\|_\infty = \max_{a\leq x\leq b} |f(x)|$.

This norm is well defined by the extreme value theorem, which guarantees that the maximum value of a continuous function on a closed interval exists. We leave verification of the norm laws as an exercise.

2.2. **Convexity.** The basic idea is that if a set in a linear space is convex, then the line connecting any two points in the set should lie entirely inside the set. Here's how to say this in symbols:

Definition 2.12. A set S in the vector space V is convex if, for any vectors $\mathbf{u}, \mathbf{v} \in S$, all vectors of the form

$$\lambda \mathbf{u} + (1 - \lambda)\,\mathbf{v},\ 0 \le \lambda \le 1,$$

are also in S.

Definition 2.13. A set S in the normed linear space V is strictly convex if, for any vectors $\mathbf{u}, \mathbf{v} \in S$, all vectors of the form

$$\mathbf{w} = \lambda \mathbf{u} + (1 - \lambda)\,\mathbf{v},\ 0 < \lambda < 1$$

are in the interior of S, that is, for each $\mathbf{w}$ there exists a positive r such that the ball $B_r(\mathbf{w})$ is entirely contained in S.

Exercises

Exercise 2.1. Show that the uniform norm on $C[a, b]$ satisfies the norm properties.

Exercise 2.2. Show that for positive r and $\mathbf{v}_0 \in V$, a normed linear space, the ball $B_r(\mathbf{v}_0)$ is a convex set. Show by example that it need not be strictly convex.

3. Inner Product Spaces

This dot product of calculus and Math 314 amounted to the "standard" inner product of the two standard vectors. We now extend this idea to a setting that allows for abstract vector spaces.

Definition 3.1. An (abstract) *inner product* on the vector space V is a function $\langle \cdot, \cdot \rangle$ that assigns to each pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a scalar $\langle \mathbf{u}, \mathbf{v} \rangle$ such that for c a scalar and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ the following hold:

(1): $\langle \mathbf{u}, \mathbf{u} \rangle \ge 0$ with $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.
(2): $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$
(3): $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
(4): $\langle \mathbf{u}, c\mathbf{v} \rangle = c\,\langle \mathbf{u}, \mathbf{v} \rangle$

A vector space V, together with an inner product $\langle \cdot, \cdot \rangle$ on the space V, is called an *inner product space*. Notice that in the case of the more common vector spaces over *real scalars*, property (2) becomes the commutative law: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$. Also observe that if V is an inner product space and W is any subspace of V, then W automatically becomes an inner product space if we simply use the inner product of V on elements of W. For all the inner product laws still hold, since they hold for elements of the larger space V.

Of course, we have the standard examples of inner products, namely the dot products on $\mathbb{R}^n$ and $\mathbb{C}^n$.

Example 3.2. For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, with $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, define

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \mathbf{u}^T \mathbf{v}.$$

This is just the standard dot product, and one can verify that all the inner product laws are satisfied by application of the laws of matrix arithmetic.

Here is an example of a nonstandard inner product on a standard space that is useful in certain engineering problems.

Example 3.3. For vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in $V = \mathbb{R}^2$, define an inner product by the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2.$$

Show that this formula satisfies the inner product laws.

First we see that

$$\langle \mathbf{u}, \mathbf{u} \rangle = 2u_1^2 + 3u_2^2,$$

so the only way for this sum to be 0 is for $u_1 = u_2 = 0$. Hence (1) holds. For (2) calculate

$$\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2 = 2v_1u_1 + 3v_2u_2 = \langle \mathbf{v}, \mathbf{u} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle},$$

since all scalars in question are real. For (3) let $\mathbf{w} = (w_1, w_2)$ and calculate

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle &= 2u_1 (v_1 + w_1) + 3u_2 (v_2 + w_2) \\ &= 2u_1v_1 + 3u_2v_2 + 2u_1w_1 + 3u_2 = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle . \end{aligned}$$

For the last property, check that for a scalar c,

□ $$\langle \mathbf{u}, c\mathbf{v} \rangle = 2u_1cv_1 + 3u_2cv_2 = c (2u_1v_1 + 3u_2v_2) = c \langle \mathbf{u}, \mathbf{v} \rangle .$$

It follows that this "weighted" inner product is indeed an inner product according to our definition. In fact, we can do a whole lot more with even less effort. Consider this example, of which the preceding is a special case.

Example 3.4. Let A be an $n \times n$ Hermitian matrix ($A = A^*$) and define the product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^* A \mathbf{v}$ for all $\mathbf{u}, \mathbf{v} \in V$, where V is $\mathbb{R}^n$ or $\mathbb{C}^n$. Show that this product satisfies inner product laws (2), (3), and (4) and that if, in addition, A is positive definite, then the product satisfies (1) and is an inner product.

As usual, let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and let c be a scalar. For (2), remember that for a 1×1 scalar quantity q, $q^* = \overline{q}$, so we calculate

$$\langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^* A \mathbf{u} = (\mathbf{u}^* A \mathbf{v})^* = \langle \mathbf{u}, \mathbf{v} \rangle^* = \overline{\langle \mathbf{u}, \mathbf{v} \rangle}.$$

For (3), we calculate

$$\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \mathbf{u}^* A(\mathbf{v} + \mathbf{w}) = \mathbf{u}^* A \mathbf{v} + \mathbf{u}^* A \mathbf{w} = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle .$$

For (4), we have that

$$\langle \mathbf{u}, c\mathbf{v} \rangle = \mathbf{u}^* A c \mathbf{v} = c \mathbf{u}^* A \mathbf{v} = c \langle \mathbf{u}, \mathbf{v} \rangle .$$

Finally, if we suppose that A is also positive definite, then by definition,

$$\langle \mathbf{u}, \mathbf{u} \rangle = \mathbf{u}^* A \mathbf{u} > 0, \text{ for } \mathbf{u} \neq \mathbf{0},$$

which shows that inner product property (1) holds. Hence, this product defines an inner product. □

We leave it to the reader to check that if we take

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix},$$

we obtain the inner product of the first example above.

Here is an example of an inner product space that is useful in approximation theory:

Example 3.5. Let $V = C[a,b]$, the space of continuous functions on the interval $[a,b]$ with the usual function addition and scalar multiplication. Show that the formula

$$\langle f,g\rangle = \int_a^b f(x)g(x)\,dx$$

defines an inner product on the space V.

Certainly $\langle f,g\rangle$ is a real number. Now if $f(x)$ is a continuous function then $f(x)^2$ is nonnegative on $[a,b]$ and therefore $\int_0^1 f(x)^2dx = \langle f,f\rangle \geq 0$. Furthermore, if $f(x)$ is nonzero, then the area under the curve $y = f(x)^2$ must also be positive since $f(x)$ will be positive and bounded away from 0 on some subinterval of $[a,b]$. This establishes property (1) of inner products.

Now let $f(x), g(x), h(x) \in V$. For property (2), notice that

$$\langle f,g\rangle = \int_a^b f(x)g(x)dx = \int_a^b g(x)f(x)dx = \langle g,f\rangle .$$

Also,

$$\begin{aligned}\langle f,g+h\rangle &= \int_a^b f(x)(g(x)+h(x))dx\\ &= \int_a^b f(x)g(x)dx + \int_a^b f(x)h(x)dx = \langle f,g\rangle + \langle f,h\rangle ,\end{aligned}$$

which establishes property (3). Finally, we see that for a scalar c,

$$\langle f,cg\rangle = \int_a^b f(x)cg(x)\,dx = c\int_a^b f(x)g(x)\,dx = c\langle f,g\rangle ,$$

which shows that property (4) holds. □

We shall refer to this inner product on a function space as the *standard inner product* on the function space $C[a,b]$. (Most of our examples and exercises involving function spaces will deal with polynomials, so we remind the reader of the integration formula $\int_a^b x^m\,dx = \frac{1}{m+1}\left(b^{m+1} - a^{m+1}\right)$ and special case $\int_0^1 x^m\,dx = \frac{1}{m+1}$ for $m \geq 0$.) There is a slight variant on the standard inner product that is frequently useful in approximation theory, namely the *weighted* inner product given by

$$\langle f,g\rangle_w = \int_a^b w(x)\,f(x)\,g(x)\,dx,$$

where the weight function $w(x)$ is continuous and positive on (a,b) and integrable on $[a,b]$. The proof that this gives an inner product is essential the same as the case $w(x) = 1$ above.

Following are a few simple facts about inner products that we will use frequently. The proofs are left to the exercises.

Theorem 3.6. *Let V be an inner product space with inner product $\langle\cdot,\cdot\rangle$. Then we have that for all $\mathbf{u},\mathbf{v},\mathbf{w} \in V$ and scalars a,*

- **(1):** $\langle \mathbf{u},\mathbf{0}\rangle = 0 = \langle \mathbf{0},\mathbf{u}\rangle$,
- **(2):** $\langle \mathbf{u}+\mathbf{v},\mathbf{w}\rangle = \langle \mathbf{u},\mathbf{w}\rangle + \langle \mathbf{v},\mathbf{w}\rangle$,
- **(3):** $\langle a\mathbf{u},\mathbf{v}\rangle = \overline{a}\langle \mathbf{u},\mathbf{v}\rangle$.

3.1. **Induced Norms and the CBS Inequality.** It is a striking fact that we can accomplish all the goals we set for the standard inner product using general inner products: we can introduce the ideas of angles, orthogonality, projections, and so forth. We have already seen much of the work that has to be done, though it was stated in the context of the standard inner products. As a first step, we want to point out that every inner product has a "natural" norm associated with it.

Definition 3.7. Let V be an inner product space. For vectors $\mathbf{u} \in V$, the norm defined by the equation

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$$

is called the *norm induced by the inner product* $\langle \cdot, \cdot \rangle$ *on* V.

As a matter of fact, this idea is not really new. Recall that we introduced the standard inner product on $V = \mathbb{R}^n$ or $\mathbb{C}^n$ with an eye toward the standard norm. At the time it seemed like a nice convenience that the norm could be expressed in terms of the inner product. It is, and so much so that we have turned this cozy relationship into a definition. Just calling the induced norm a norm doesn't make it so. Is the induced norm really a norm? We have some work to do. The first norm property is easy to verify for the induced norm: from property (1) of inner products we see that $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, with equality if and only if $\mathbf{u} = 0$. This confirms norm property (1). Norm property (2) isn't too hard either: let c be a scalar and check that

$$\|c\mathbf{u}\| = \sqrt{\langle c\mathbf{u}, c\mathbf{u} \rangle} = \sqrt{c\bar{c} \langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{|c|^2} \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = |c| \, \|\mathbf{u}\| \, .$$

Norm property (3), the triangle inequality, remains. This one isn't easy to verify from first principles. We need a tool called the the Cauchy–Bunyakovsky–Schwarz (CBS) inequality.

Theorem 3.8. *(CBS Inequality) Let* V *be an inner product space. For* $\mathbf{u}, \mathbf{v} \in V$, *if we use the inner product of* V *and its induced norm, then*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| \, .$$

Henceforth, when the norm sign $\|\cdot\|$ is used in connection with a given inner product, it is understood that this norm is the induced norm of this inner product, unless otherwise stated.

Just as with the standard dot products, we can formulate the following definition thanks to the CBS inequality.

Definition 3.9. For vectors $\mathbf{u}, \mathbf{v} \in V$, a real inner product space, we define the *angle* between $\mathbf{u}$ and $\mathbf{v}$ to be any angle θ satisfying

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|}.$$

We know that $|\langle \mathbf{u}, \mathbf{v} \rangle| / (\|\mathbf{u}\| \, \|\mathbf{v}\|) \leq 1$, so that this formula for $\cos\theta$ makes sense.

Example 3.10. Let $\mathbf{u} = (1, -1)$ and $\mathbf{v} = (1, 1)$ be vectors in $\mathbb{R}^2$. Compute an angle between these two vectors using the inner product of Example 3.3. Compare this to the angle found when one uses the standard inner product in $\mathbb{R}^2$.

Solution. According to 3.3 and the definition of angle, we have

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|} = \frac{2 \cdot 1 \cdot 1 + 3 \cdot (-1) \cdot 1}{\sqrt{2 \cdot 1^2 + 3 \cdot (-1)^2} \sqrt{2 \cdot 1^2 + 3 \cdot 1^2}} = \frac{-1}{5}.$$

Hence the angle in radians is

$$\theta = \arccos\left(\frac{-1}{5}\right) \approx 1.7722.$$

On the other hand, if we use the standard norm, then

$$\langle \mathbf{u}, \mathbf{v}\rangle = 1 \cdot 1 + (-1) \cdot 1 = 0,$$

from which it follows that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal and $\theta = \pi/2 \approx 1.5708$. □

In the previous example, it shouldn't be too surprising that we can arrive at two different values for the "angle" between two vectors. Using different inner products to measure angle is somewhat like measuring length with different norms. Next, we extend the perpendicularity idea to arbitrary inner product spaces.

Definition 3.11. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in the same inner product space are *orthogonal* if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.

Note that if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$, then $\langle \mathbf{v}, \mathbf{u}\rangle = \overline{\langle \mathbf{u}, \mathbf{v}\rangle} = 0$. Also, this definition makes the zero vector orthogonal to every other vector. It also allows us to speak of things like "orthogonal functions." One has to be careful with new ideas like this. Orthogonality in a function space is not something that can be as easily visualized as orthogonality of geometrical vectors. Inspecting the graphs of two functions may not be quite enough. If, however, graphical data is tempered with a little understanding of the particular inner product in use, orthogonality can be detected.

Example 3.12. Show that $f(x) = x$ and $g(x) = x - \frac{2}{3}$ are orthogonal elements of $C[0,1]$ with the inner product of Example 3.5 and provide graphical evidence of this fact.

Solution. According to the definition of inner product in this space,

$$\langle f, g\rangle = \int_0^1 f(x)g(x)dx = \int_0^1 x\left(x - \frac{2}{3}\right)dx = \left(\frac{x^3}{3} - \frac{x^2}{3}\right)\Big|_0^1 = 0.$$

It follows that f and g are orthogonal to each other. For graphical evidence, sketch $f(x)$, $g(x)$, and $f(x)g(x)$ on the interval $[0,1]$ as in Figure 3.1. The graphs of f and g are not especially enlightening; but we can see in the graph that the area below $f \cdot g$ and above the x-axis to the right of $(2/3, 0)$ seems to be about equal to the area to the left of $(2/3, 0)$ above $f \cdot g$ and below the x-axis. Therefore the integral of the product on the interval $[0,1]$ might be expected to be zero, which is indeed the case.

Some of the basic ideas from geometry that fuel our visual intuition extend very elegantly to the inner product space setting. One such example is the famous Pythagorean theorem, which takes the following form in an inner product space.

Theorem 3.13. *Let $\mathbf{u}, \mathbf{v}$ be orthogonal vectors in an inner product space V. Then* $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$.

Proof. Compute

$$\begin{aligned}\|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.\end{aligned}$$

Here is an example of another standard geometrical fact that fits well in the abstract setting. This is equivalent to the law of parallelograms, which says that

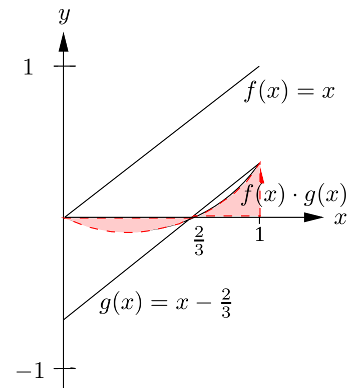

FIGURE 3.1. Graphs of f, g, and $f \cdot g$ on the interval $[0, 1]$.

the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of all four sides.

Example 3.14. Use properties of inner products to show that if we use the induced norm, then

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\right).$$

The key to proving this fact is to relate induced norm to inner product. Specifically,

$$\|\mathbf{u}+\mathbf{v}\|^2 = \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle,$$

while

$$\|\mathbf{u}-\mathbf{v}\|^2 = \langle \mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle - \langle \mathbf{u}, \mathbf{v}\rangle - \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle.$$

Now add these two equations and obtain by using the definition of induced norm again that

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\langle \mathbf{u}, \mathbf{u}\rangle + 2\langle \mathbf{v}, \mathbf{v}\rangle = 2\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\right),$$

which is what was to be shown. □

It would be nice to think that every norm on a vector space is induced from some inner product. Unfortunately, this is not true, as the following example shows.

Example 3.15. Use the result of Example 3.14 to show that the infinity norm on $V = \mathbb{R}^2$ is not induced by any inner product on V.

Solution. Suppose the infinity norm were induced by some inner product on V. Let $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (0, 1/2)$. Then we have

$$\|\mathbf{u}+\mathbf{v}\|_\infty^2 + \|\mathbf{u}-\mathbf{v}\|_\infty^2 = \|(1, 1/2)\|_\infty^2 + \|1, -1/2\|_\infty^2 = 2,$$

while

$$2\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\right) = 2\,(1 + 1/4) = 5/2.$$

This contradicts Example 3.14, so that the infinity norm cannot be induced from an inner product. □

The proof of the following key facts and their corollaries are the same as those of for standard dot products. All we have to do is replace dot products by inner products. The observations that followed the proof of this theorem are valid for general inner products as well. We omit the proofs.

Theorem 3.16. *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be vectors in an inner product space with* $\mathbf{v} \neq \mathbf{0}$*. Define the projection of* $\mathbf{u}$ *along* $\mathbf{v}$ *as*

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v}$$

and let $\mathbf{p} = \operatorname{proj}_{\mathbf{v}} \mathbf{u}$*,* $\mathbf{q} = \mathbf{u} - \mathbf{p}$*. Then* $\mathbf{p}$ *is parallel to* $\mathbf{v}$*,* $\mathbf{q}$ *is orthogonal to* $\mathbf{v}$*, and* $\mathbf{u} = \mathbf{p} + \mathbf{q}$*.*

As with the standard inner product, it is customary to call the vector $\operatorname{proj}_{\mathbf{v}} \mathbf{u}$ of this theorem the *(parallel) projection of* $\mathbf{u}$ *along* $\mathbf{v}$. Likewise, components and orthogonal projections are defined as in the standard case. In summary, we have the two vector and one scalar quantities

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v},$$

$$\operatorname{orth}_{\mathbf{v}} \mathbf{u} = \mathbf{u} - \operatorname{proj}_{\mathbf{v}} \mathbf{u},$$

$$\operatorname{comp}_{\mathbf{v}} \mathbf{u} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\|}.$$

3.2. Orthogonal Sets of Vectors.

Definition 3.17. The set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in an inner product space is said to be an *orthogonal set* if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$. If, in addition, each vector has unit length, i.e., $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1$ for all i, then the set of vectors is said to be an *orthonormal set* of vectors.

In general, the problem of finding the coordinates of a vector relative to a given basis is a nontrivial problem. If the basis is an orthogonal set, however, the problem is much simpler, as the following theorem shows.

Theorem 3.18. *Let* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *be an orthogonal set of nonzero vectors and suppose that* $\mathbf{v} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$*. Then* $\mathbf{v}$ *can be expressed uniquely (up to order) as a linear combination of* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$*, namely*

$$\mathbf{v} = \frac{\langle \mathbf{v}_1, \mathbf{v} \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 + \frac{\langle \mathbf{v}_2, \mathbf{v} \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 + \cdots + \frac{\langle \mathbf{v}_n, \mathbf{v} \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle} \mathbf{v}_n.$$

Corollary 3.19. *Every orthogonal set of nonzero vectors is linearly independent.*

Another useful corollary is the following fact about the length of a vector, whose proof is left as an exercise. Think of this as a generalized Pythagorean theorem.

Corollary 3.20. *If* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *is an orthogonal set of vectors and* $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$*, then*

$$\|\mathbf{v}\|^2 = c_1^2 \|\mathbf{v}_1\|^2 + c_2^2 \|\mathbf{v}_2\|^2 + \cdots + c_n^2 \|\mathbf{v}_n\|^2.$$

We can extend the idea of projection of one vector along another in the following way. Notice that in the case of $n = 1$ this next definition amounts to the projection of $\mathbf{u}$ along the vector $\mathbf{v}_1$.

Definition 3.21. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be an orthogonal basis for the subspace V of the inner product space W. For any $\mathbf{u} \in W$, the *(parallel) projection of* $\mathbf{u}$ *along the subspace* V is the vector

$$\operatorname{proj}_V \mathbf{u} = \frac{\langle \mathbf{v}_1, \mathbf{u} \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 + \frac{\langle \mathbf{v}_2, \mathbf{u} \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 + \cdots + \frac{\langle \mathbf{v}_n, \mathbf{u} \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle} \mathbf{v}_n.$$

Clearly $\operatorname{proj}_V \mathbf{u} \in V$, and from Theorem 3.18, we see that if $\mathbf{u} \in V$, then $\operatorname{proj}_V \mathbf{u} = \mathbf{u}$. It appears that for vectors $\mathbf{u}$ not in V the definition of proj_V depends on the basis vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, but we shall see in the next section that that this is not the case.

About notation: we take the point of view that a projection along a subspace given above is "parallel" to the subspace, but be warned that it is common to call this the *orthogonal* projection of a vector *into* the subspace, thus reversing the usage of the terms "parallel" and "orthogonal" in this context.

We have seen that orthogonal bases have some very pleasant properties, such as easy coordinate calculations. Our next goal is the following: given a subspace V of some inner product space and a basis $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ of V, to turn this basis into an orthogonal basis. The tool we need is the Gram–Schmidt algorithm.

Theorem 3.22. *Let* $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ *be a basis of the inner product space* V*. Define vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *recursively by the formula*

$$\mathbf{v}_k = \mathbf{w}_k - \frac{\langle \mathbf{v}_1, \mathbf{w}_k \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_k \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 - \cdots - \frac{\langle \mathbf{v}_{k-1}, \mathbf{w}_k \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle} \mathbf{v}_{k-1}, \quad k = 1, \ldots, n.$$

Then

(1): *The vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ *form an orthogonal set.*

(2): *For each index* $k = 1, \ldots, n$*,*

$$\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}.$$

Proof. In the case $k = 1$, we have that the single vector $\mathbf{v}_1 = \mathbf{w}_1$ is an orthogonal set and certainly $\operatorname{span}\{\mathbf{w}_1\} = \operatorname{span}\{\mathbf{v}_1\}$. Now suppose that for some index $k > 1$ we have shown that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}$ is an orthogonal set such that $\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{k-1}\} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}\}$. Then it is true that $\langle \mathbf{v}_r, \mathbf{v}_s \rangle = 0$ for any indices r, s both less than k. Take the inner product of $\mathbf{v}_k$, as given by the formula above, with the vector $\mathbf{v}_j$, where $j < k$, and we obtain

$$\begin{aligned}
\langle \mathbf{v}_j, \mathbf{v}_k \rangle &= \left\langle \mathbf{v}_j, \mathbf{w}_k - \frac{\langle \mathbf{v}_1, \mathbf{w}_k \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_k \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 - \cdots - \frac{\langle \mathbf{v}_{k-1}, \mathbf{w}_k \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle} \mathbf{v}_{k-1} \right\rangle \\
&= \langle \mathbf{v}_j, \mathbf{w}_k \rangle - \langle \mathbf{v}_1, \mathbf{w}_k \rangle \frac{\langle \mathbf{v}_j, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} - \cdots - \langle \mathbf{v}_{k-1}, \mathbf{w}_k \rangle \frac{\langle \mathbf{v}_j, \mathbf{v}_{k-1} \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle} \\
&= \langle \mathbf{v}_j, \mathbf{w}_k \rangle - \langle \mathbf{v}_j, \mathbf{w}_k \rangle \frac{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} = 0.
\end{aligned}$$

It follows that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is an orthogonal set. The Gram–Schmidt formula show us that one of $\mathbf{v}_k$ or $\mathbf{w}_k$ can be expressed as a linear combination of the other and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}$. Therefore

$$\begin{aligned}
\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{k-1}, \mathbf{w}_k\} &= \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}, \mathbf{w}_k\} \\
&= \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_k\},
\end{aligned}$$

which is the second part of the theorem. Repeat this argument for each index $k = 2, \ldots, n$ to complete the proof of the theorem. □

The Gram–Schmidt formula is easy to remember: subtract from the vector $\mathbf{w}_k$ all of the projections of $\mathbf{w}_k$ along the directions $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}$ to obtain the vector $\mathbf{v}_k$.

Example 3.23. Let $C[0,1]$ be the space of continuous functions on the interval $[0,1]$ with the usual function addition and scalar multiplication, and (standard) inner product given by

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

Let $V = \mathcal{P}_2 = \operatorname{span}\{1, x, x^2\}$ and apply the Gram–Schmidt algorithm to the basis $1, x, x^2$ to obtain an orthogonal basis for the space of quadratic polynomials.

Solution. Set $\mathbf{w}_1 = 1$, $\mathbf{w}_2 = x$, $\mathbf{w}_3 = x^2$ and calculate the Gram–Schmidt formulas:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{w}_1 = 1, \\
\mathbf{v}_2 &= \mathbf{w}_2 - \frac{\langle \mathbf{v}_1, \mathbf{w}_2 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 = x - \frac{1/2}{1}1 = x - \frac{1}{2}, \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\langle \mathbf{v}_1, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 \\
&= x^2 - \frac{1/3}{1}1 - \frac{1/12}{1/12}(x - \frac{1}{2}) = x^2 - x + \frac{1}{6}.
\end{aligned}$$

□

Had we used $C[-1,1]$ and required that each polynomial have value 1 at $x = 1$, the same calculations would have given us the first three well-known functions called *Legendre polynomials*. These polynomials are used extensively in approximation theory and applied mathematics.

If we prefer to have an orthonormal basis rather than an orthogonal basis, then, as a final step in the orthogonalizing process, simply replace each vector $\mathbf{v}_k$ by the normalized vector $\mathbf{u}_k = \mathbf{v}_k / \|\mathbf{v}_k\|$.

3.3. **Best Approximations and Least Squares Problems.** The problem we consider here is the following:

Approximation Problem: Given a subspace V of the inner product space W, and an element $\mathbf{f} \in W$, find a vector $\mathbf{v}^* \in V$ that minimizes $\|\mathbf{f} - \mathbf{v}\|$, that is, a vector $\mathbf{v}^* \in V$ such that

$$\|\mathbf{f} - \mathbf{v}^*\| = \min_{\mathbf{v} \in V} \|\mathbf{f} - \mathbf{v}\|.$$

Of course, there may be no such vector. Here is a characterization of what the solution to an approximation problem must satisfy:

Theorem 3.24. *The vector* $\mathbf{v}$ *in the subspace* V *of the inner product space* W *minimizes the distance from a vector* $\mathbf{f} \in W$ *if and only if* $\mathbf{f} - \mathbf{v}$ *is orthogonal to every* $\mathbf{f} \in V$.

Proof. First observe that minimizing $\|\mathbf{f} - \mathbf{v}\|$ over $\mathbf{v} \in V$, is equivalent to minimizing $\|\mathbf{f} - \mathbf{v}\|^2$. Let $\mathbf{v} \in V$. Suppose that $\mathbf{p}$ is the projection of $\mathbf{f} - \mathbf{v}$ to any vector in V. Use the Pythagorean theorem to obtain that

$$\|\mathbf{f} - \mathbf{v}\|^2 = \|\mathbf{f} - \mathbf{v} - \mathbf{p}\|^2 + \|\mathbf{p}\|^2 = \|\mathbf{f} - (\mathbf{v} + \mathbf{p})\|^2 + \|\mathbf{p}\|^2.$$

However, $\mathbf{v} + \mathbf{p} \in V$, so that $\|\mathbf{f} - \mathbf{v}\|$ is the minimum distance from $\mathbf{f}$ to a vector in V if and only if $\|\mathbf{p}\| = 0$ for all possible $\mathbf{p}$, which is equivalent to the condition that $\mathbf{f} - \mathbf{v}$ is orthogonal to every vector in V. □

We can now solve the approximation problem completely in the case of a finite dimensional subspace V.

Theorem 3.25. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be an orthogonal basis for the subspace V of the inner product space W. For any $\mathbf{f} \in \mathbf{W}$, the vector $\mathbf{v}^* = \operatorname{proj}_V \mathbf{f}$ is the unique vector in V that minimizes $\|\mathbf{f} - \mathbf{v}\|$.*

Proof. Suppose that $\mathbf{v} \in V$ minimizes $\|\mathbf{f} - \mathbf{v}\|^2$. It follows from Theorem 3.24 that $\mathbf{f} - \mathbf{v}$ is orthogonal to any vector in V. Now let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be an *orthogonal* basis of V and express the vector $\mathbf{v}$ in the form

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Then for each $\mathbf{v}_k$ we must have

$$\begin{aligned} 0 &= \langle \mathbf{v}_k, \mathbf{f} - \mathbf{v} \rangle = \langle \mathbf{v}_k, \mathbf{f} - c_1\mathbf{v}_1 - c_2\mathbf{v}_2 - \cdots - c_n\mathbf{v}_n \rangle \\ &= \langle \mathbf{v}_k, \mathbf{f} \rangle - c_1 \langle \mathbf{v}_k, \mathbf{v}_1 \rangle - c_2 \langle \mathbf{v}_k, \mathbf{v}_2 \rangle - \cdots c_n \langle \mathbf{v}_k, \mathbf{v}_n \rangle \\ &= \langle \mathbf{v}_k, \mathbf{f} \rangle - c_k \langle \mathbf{v}_k, \mathbf{v}_k \rangle , \end{aligned}$$

from which we deduce that $c_k = \langle \mathbf{v}_k, \mathbf{f} \rangle / \langle \mathbf{v}_k, \mathbf{v}_k \rangle$. It follows that

$$\mathbf{v} = \frac{\langle \mathbf{v}_1, \mathbf{f} \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{v}_2, \mathbf{f} \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{v}_n, \mathbf{f} \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n = \operatorname{proj}_V \mathbf{f}.$$

This proves that there can be only one solution to the projection problem, namely the vector $\mathbf{v}$ given by the projection formula above. Finally, note that for any j,

$$\langle \mathbf{v}_j, \mathbf{f} - \mathbf{v} \rangle = \langle \mathbf{v}_j, \mathbf{f} \rangle - \sum_{k=1}^{n} \left\langle \mathbf{v}_j, \frac{\langle \mathbf{v}_k, \mathbf{f} \rangle}{\langle \mathbf{v}_k, \mathbf{v}_k \rangle}\mathbf{v}_k \right\rangle = \langle \mathbf{v}_j, \mathbf{f} \rangle - \frac{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \langle \mathbf{v}_j, \mathbf{f} \rangle = 0.$$

It follows that $\mathbf{f} - \mathbf{v}$ is orthogonal to each basis vector $\mathbf{v}_j$ of V and therefore to any linear combination of these vectors, that is, any vector in V. Hence by Theorem 3.24 the projection of $\mathbf{f}$ into V is unique vector in V that minimizes $\|\mathbf{f} - \mathbf{v}\|$. □

Since the definition of best approximation from a subspace is independent of any basis of the subspace, we can now confirm that our definition of $\operatorname{proj}_V \mathbf{f}$ does not depend on any particular basis of V.

Corollary 3.26. *The definition of projection vector into a finite dimensional subspace V of the inner product space W as given in Definition 3.21 does not depend on the choice of orthogonal basis of V.*

In analogy with the standard inner products, we define the *orthogonal projection of* $\mathbf{f}$ to V by the formula

$$\operatorname{orth}_V \mathbf{f} = \mathbf{f} - \operatorname{proj}_V \mathbf{f}.$$

Just as in the case of a single vector, we have the following fact:

Theorem 3.27. *Let $\mathbf{f}$ be a vector in the nonzero subspace V of the inner product space W. Let $\mathbf{p} = \operatorname{proj}_V \mathbf{f}$ and $\mathbf{q} = \operatorname{orth}_V \mathbf{f}$. Then $\mathbf{p} \in V$, $\mathbf{q}$ is orthogonal to every vector in V, and $\mathbf{f} = \mathbf{p} + \mathbf{q}$.*

Suppose now that we are given a finite basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for a subspace V of the inner product space W and a vector $\mathbf{f} \in W$. Can we solve the approximation problem of minimizing $\|\mathbf{f} - \mathbf{v}\|$, equivalently, minimizing $\|\mathbf{f} - \mathbf{v}\|^2$in terms of this basis without resorting to replacing B by an orthogonal basis? The answer is "yes" and this can be accomplished as follows: We already know from Theorem 3.25 that there actually is a solution vector $\mathbf{v} \in V$. So write

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Then we have

$$\mathbf{f} - \mathbf{v} = \mathbf{f} - c_1\mathbf{v}_1 - c_2\mathbf{v}_2 - \cdots - c_n\mathbf{v}_n.$$

Now require that $\mathbf{f}-\mathbf{v}$ be orthogonal to every basis vector of B, that is, $\langle \mathbf{f} - \mathbf{v}, \mathbf{v}_j \rangle = 0$, for $j = 1, \ldots, n$, so that it will be orthogonal to any linear combination of these basis vectors, in accordance with Theorem 3.24. This leads to the system of equations

$$\langle \mathbf{v}_i, \mathbf{v}_1 \rangle c_1 + \langle \mathbf{v}_i, \mathbf{v}_2 \rangle c_2 + \cdots + \langle \mathbf{v}_i, \mathbf{v}_2 \rangle c_n = \langle \mathbf{v}_i, \mathbf{f} \rangle, \ \ i = 1, 2, \ldots, n.$$

We can write this in matrix form as $G\mathbf{c} = \mathbf{b}$, where $\mathbf{c} = (c_1, \ldots, c_n)$, $\mathbf{b} = (\langle \mathbf{v}_1, \mathbf{f} \rangle, \ldots, \langle \mathbf{v}_1, \mathbf{f} \rangle)$ and $G = [g_{i,j}]$ is given by

$$G = \begin{bmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \ldots & \langle \mathbf{v}_1, \mathbf{v}_j \rangle & \ldots & \langle \mathbf{v}_1, \mathbf{v}_n \rangle \\ \vdots & & \vdots & & \vdots \\ \langle \mathbf{v}_i, \mathbf{v}_1 \rangle & \ldots & \langle \mathbf{v}_i, \mathbf{v}_j \rangle & \ldots & \langle \mathbf{v}_i, \mathbf{v}_n \rangle \\ \vdots & & \vdots & & \vdots \\ \langle \mathbf{v}_n, \mathbf{v}_1 \rangle & \ldots & \langle \mathbf{v}_n, \mathbf{v}_j \rangle & \ldots & \langle \mathbf{v}_n, \mathbf{v}_n \rangle \end{bmatrix} = [\langle \mathbf{v}_i, \mathbf{v}_j \rangle]$$

is the so-called *Gramian matrix* for the basis B. Now we simply solve the system and we have an explicit formula for the best approximation to $\mathbf{f}$ from the subspace spanned by B.

Here is a famous example of a Gramian matrix.

Example 3.28. Let W be the inner product space C$[0, 1]$ with the standard function space inner product. Let $V = \mathcal{P}_n$ so that $B = \left\{1, x, x^2, \ldots, x^n\right\}$ is a basis of V. Compute the Gramian of this basis.

Solution. According to the definition of inner product in this space,

$$\left\langle x^i, x^j \right\rangle = \int_0^1 x^i x^j dx = \int_0^1 x^{i+j} dx = \left.\frac{x^{i+j+1}}{i+j+1}\right|_0^1 = \frac{1}{i+j+1}.$$

It follows that $g_{i,j} = 1/(i+j+1)$ so that the Gramian G looks like

$$G = \left[\frac{1}{i+j+1}\right]_{n+1,n+1} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \frac{1}{n+1} & \frac{1}{n+2} & \frac{1}{n+3} & \cdots & \frac{1}{2n+1} \end{bmatrix} = H_{n+1}.$$

The matrix H_n is known as the n-th order *Hilbert matrix*, and has some interesting numerical properties. □

Exercises

Exercise 3.1. Confirm that $p_1(x) = x$ and $p_2(x) = 3x^2 - 1$ are orthogonal elements of $C[-1,1]$ with the standard inner product and determine whether the following polynomials belong to span $\{p_1(x), p_2(x)\}$ using Theorem **??**.

(a) x^2 (b) $1 + x - 3x^2$ (c) $1 + 3x - 3x^2$

Exercise 3.2. Show that if V an inner product space, $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$, and $\mathbf{v}$ is orthogonal to each vector $\mathbf{v}_i$, $i = 1, \ldots, n$, then $\mathbf{v}$ is orthogonal to any vector $\mathbf{w} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$.

4. Linear Operators

Before giving the definition of linear operator, let us recall some notation that is associated with functions in general. We identify a function f with the notation $f : D \to T$, where D and T are the *domain* and *target* of the function, respectively. This means that for each x in the domain D, the value $f(x)$ is a uniquely determined element in the target T. We want to emphasize at the outset that there is a difference here between the *target* of a function and its *range*. The *range* of the function f is defined as the subset of the target

$$\operatorname{range}(f) = \{y \mid y = f(x) \text{ for some } x \in D\},$$

which is just the set of all possible values of $f(x)$. A function is said to be one-to-one if, whenever $f(x) = f(y)$, then $x = y$. Also, a function is said to be *onto* if the range of f equals its target. For example, we can define a function $f : \mathbb{R} \to \mathbb{R}$ by the formula $f(x) = x^2$. It follows from our specification of f that the target of f is understood to be $\mathbb{R}$, while the range of f is the set of nonnegative real numbers. Therefore, f is not onto. Moreover, $f(-1) = f(1)$ and $-1 \neq 1$, so f is not one-to-one either.

A function that maps elements of one vector space into another, say $f : V \to W$, is sometimes called an *operator* or *transformation*. One of the simplest mappings of a vector space V is the so-called *identity function* $\operatorname{id}_V : V \to V$ given by $\operatorname{id}_V(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in V$. Here domain, range, and target all agree. Of course, matters can become more complicated. For example, the operator $f : \mathbb{R}^2 \to \mathbb{R}^3$ might be given by the formula

$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x^2 \\ xy \\ y^2 \end{bmatrix}.$$

Notice in this example that the target of f is $\mathbb{R}^3$, which is not the same as the range of f, since elements in the range have nonnegative first and third coordinates. From the point of view of linear algebra, this function lacks the essential feature that makes it really interesting, namely linearity.

Definition 4.1. A function $T : V \to W$ from the vector space V into the space W over the same field of scalars is called a *linear operator (mapping, transformation)* if for all vectors $\mathbf{u}, \mathbf{v} \in V$ and scalars c, d, we have

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}).$$

By taking $c = d = 1$ in the definition, we see that a linear function T is *additive,* that is, $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$. Also, by taking $d = 0$ in the definition, we see that a linear function is *outative,* that is, $T(c\mathbf{u}) = cT(\mathbf{u})$. As a matter of fact, these two conditions imply the linearity property, and so are equivalent to it. We leave this fact as an exercise.

If $T : V \to V$ is a linear operator, we simply say that T is a linear operator on V. A linear operator $T : V \to \mathbb{R}$ is called a *linear functional* on V.

By repeated application of the linearity definition, we can extend the linearity property to any linear combination of vectors, not just two terms. This means that for any scalars $c_1, c_2, \ldots, c_n$ and vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, we have

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n).$$

Example 4.2. Determine whether $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear operator, where T is given by the formula

(a) $T((x, y)) = (x^2, xy, y^2)$ or (b) $T((x, y)) = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$.

If T is defined by (a) then we show by a simple example that T fails to be linear. Let us calculate

$$T((1, 0) + (0, 1)) = T((1, 1)) = (1, 1, 1),$$

while

$$T((1, 0)) + T((0, 1)) = (1, 0, 0) + (0, 0, 1) = (1, 0, 1).$$

These two are not equal, so T fails to satisfy the linearity property.

Next consider the operator T defined as in (b). Write

$A = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$,

and we see that the action of T can be given as $T(\mathbf{v}) = A\mathbf{v}$. Now we have already seen that the operation of multiplication by a fixed matrix is a linear operator. $\square$

Recall that an operator $f : V \to W$ is said to be *invertible* if there is an operator $g : W \to V$ such that the composition of functions satisfies $f \circ g = \mathrm{id}_W$ and $g \circ f = \mathrm{id}_V$. In other words, $f(g(\mathbf{w})) = \mathbf{w}$ and $g(f(\mathbf{v})) = \mathbf{v}$ for all $\mathbf{w} \in W$ and $\mathbf{v} \in V$. We write $g = f^{-1}$ and call f^{-1} the inverse of f. One can show that for any operator f, linear or not, being invertible is equivalent to being both one-to-one and onto.

Example 4.3. Show that if $f : V \to W$ is an invertible linear operator on vector spaces, then f^{-1} is also a linear operator.

Solution. We need to show that for $\mathbf{u}, \mathbf{v} \in W$, the linearity property $f^{-1}(c\mathbf{u} + d\mathbf{v}) = cf^{-1}(\mathbf{u}) + df^{-1}(\mathbf{v})$ is valid. Let $\mathbf{w} = cf^{-1}(\mathbf{u}) + df^{-1}(\mathbf{v})$. Apply the function f to both sides and use the linearity of f to obtain that

$$f(\mathbf{w}) = f\left(cf^{-1}(\mathbf{u}) + df^{-1}(\mathbf{v})\right) = cf\left(f^{-1}(\mathbf{u})\right) + df\left(f^{-1}(\mathbf{v})\right) = c\mathbf{u} + d\mathbf{v}.$$

Apply f^{-1} to obtain that $\mathbf{w} = f^{-1}(f(\mathbf{w})) = f^{-1}(c\mathbf{u} + d\mathbf{v})$, which proves the linearity property. $\square$

Abstraction gives us a nice framework for certain key properties of mathematical objects, some of which we have seen before. For example, in calculus we were taught that differentiation has the "linearity property." Now we can express this assertion in a larger context: let V be the space of differentiable functions and define an operator T on V by the rule $T(f(x)) = f'(x)$. Then T is a linear operator on the space V.

4.1. **Operator Norms.** The following idea lets us measure the "size" of a norm in terms of how much the operator is capable of "stretching" an argument. This is a very useful notion for approximation theory in the situation where our approximation to the element f can be described as applying an operator T to f to obtain the approximation $T(f)$.

Definition 4.4. Let $T : V \to W$ be a linear operator between normed linear spaces. The *operator norm* of T is defined to be

$$\max_{0 \neq v \in V} \frac{\|T(v)\|}{\|v\|} = \max_{v \in V, \|v\|=1} \|T(v)\|,$$

provided the maximum value exists, in which case the operator is called *bounded*. Otherwise the operator is called *unbounded*.

This "norm" really is a norm on the appropriate space.

Theorem 4.5. *Given normed linear spaces V, W, the set of $L(V,W)$ of all bounded linear operators from V to W is a vector space with the usual function addition and multiplication. Moreover, the operator norm on $L(V,W)$ is a vector norm, so that $L(V,W)$ is also a normed linear space.*

This theorem gives us an idea as to why operator norms are relevant to approximation theory. Suppose you want to approximate functions f in some function space by a method which can be described as aplying a linear operator T to f (we'll see lots of examples of this in approximation theory). One rough measure of how good an approximation we have is that $\|T(f)\|$ should be close to $\|f\|$. Put another way, $\|T(f)\| / \|f\|$ should be close to one. So if $\|T\|$ is much larger than one, we expect that for some functions f, $T(f)$ will be a poor approximation to f.

Bounded linear operators for infinite dimensional spaces is a large subject which forms part of the area of mathematics called functional analysis. However, in the case of finite dimension, the situation is much simpler.

Theorem 4.6. *Let $T : V \to W$ be a linear operator, where V is a finite dimensional space. Then T is bounded, so that $L(V,W)$ consists of all linear operators from V to W.*

Some notation: if both V and W are standard spaces with the same standard p-norm, then the operator norm of T is denoted by $\|T\|_p$. In some cases the operator norm is fairly simple to compute. Here is an example.

Example 4.7. Let $T_A : \mathbb{R}^n \to \mathbb{R}^m$ be the linear operator given by matrix-vector multiplication, $T_A(\mathbf{v}) = A\mathbf{v}$, where $A = [a_{ij}]$ is an $m \times n$ matrix with (i,j)-th entry a_{ij}. Then T_A is bounded by the previous theorem and moreover

$$\|T_A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|.$$

Solution. To see this, observer that in order for a vector $\mathbf{v} \in V$ to have infinity norm 1, all coordinates must be at most 1 in absolute value and at least one must be equal to one. Now choose the row that has the largest row sum of absolute values $\sum_{j=1}^{n} |a_{ij}|$ and let $\mathbf{v}$ be the vector whose j-th coordinate is just the signum of a_{ij}, so that $|a_{ij}| = v_j a_{ij}$ for all j. Then it is easily checked that this is the largest possible value for $\|A\mathbf{v}\|$. □

Exercises

Exercise 4.1. Show that differentiation is a linear operator on $V = C^{\infty}(\mathbb{R})$, the space of functions defined on the real line that are infinitely differentiable.

Exercise 4.2. Show that the operator $T : C[0,1] \to \mathbb{R}$ given by $T(f) = \int_0^1 f(x)\,dx$ is a linear operator.

5. Metric Spaces and Analysis

5.1. **Metric Spaces.** We start with the definition of a metric space which, in general, is rather different from a vector space, although the main examples for us are in fact normed linear spaces and their subsets. Be aware that in this context the term "space" is much broader than a "vector space."

Definition 5.1. A metric space is a nonempty set X of objects called points, together with a function $d(x,y)$, called a metric for X, mapping pairs of points $x, y \in X$ to real numbers and satisfying for all points x, y, z:
(1) $d(x,y) \geq 0$ with equality if and only if $x = y$.
(2) $d(x,y) = d(y,x)$
(3) $d(x,z) \leq d(x,y) + d(y,z)$

Examples abound. This is a more general concept than norm, since metric spaces do not have to be vector spaces, so that every subset of a metric space is also a metric space with the inherited metric. There are many interesting examples that are not subsets of normed linear spaces. One such example can be found in Powell's text, Exercise 1.2. BTW, this example is important for problems of shape recognition.

Example 5.2. Let For a very nonstandard example, consider a finite network in which every pair x, y of nodes is connected by a path of edges, each of which has a positive cost associated with traversing it. Assume that there are no self-edges. Define the distance between any two nodes x and y to be the minimum cost of a path connecting x to y if $x \neq y$, otherwise the distance is 0. This definition turns the network into a metric space.

It is customary to call a subset Y of the metric space X a *subspace* of X. Certainly, it is true that Y, together with the inherited metric $d(\cdot,\cdot)$ is a metric space in its own right. However, you have to be aware that the term"subspace" is more general than a "vector subspace."

Next, we consider some topological ideas that are useful for metric spaces. In all of the following we assume that X is a metric space with metric $d(x,y)$.

Definition 5.3. Given $r > 0$ and a point x in X, the (closed) ball of radius r centered at the point x_0 *is the set of points*

$$B_r(x) = \{y \in X \mid d(x_0, y) \leq r\}.$$

The set of points

$$B_r^o(x) = \{y \in X \mid d(x_0, y) < r\}.$$

is the *open* ball of radius r centered at the point x_0.

Definition 5.4. A subset Y of the metric space X is *open* if for every point $y \in Y$ there exists a ball $B_r(y)$ contained entirely in Y. The set Y is *closed* if its complement $X \backslash Y$ in X is an open set.

Definition 5.5. A subset Y of the metric space X is bounded if there exists a ball ball $B_r(x)$ in X containing Y.

Definition 5.6. A point x is in the *boundary* of the subset Y of the metric space X if every
ball $B_r(x)$, $r > 0$, contains points in Y and points in the complement $X \backslash Y$ of Y. The set of all such points is denoted by ∂Y.

One can show that a set is is open if and only if it contains none of its boundary points and closed if and only if it contains all of its boundary points.

Definition 5.7. A subset Y of the metric space X is *compact* if for every collection of open sets $\{O_\alpha\}$ that covers Y, that is, every element of Y is in some O_α, there is a finite subset of these open sets that also covers Y.

Definition 5.8. A sequence $\{x_n\}_{n=1}^{\infty}$ of elements in a metric space X is said to *converge* to a point $x^* \in X$ if for every number $\epsilon > 0$ there exists an index N such that for $n \geq N$, $d(x, x^*) < \epsilon$. We write

$$\lim_{n\to\infty} x_n = x^*.$$

Definition 5.9. A sequence $\{x_n\}_{n=1}^{\infty}$ of elements in a metric space X is said to be a *Cauchy sequence* if for every number $\epsilon > 0$ there exists an index N such that for $m, n \geq N$, $d(x_m, x_n) < \epsilon$.

Cauchy sequences are sequences of points that "ought to converge to a point." If every Cauchy sequence in the metric space X converges to a point in X, we say that X is a *complete metric space*.

The real numbers form a complete metric space. However, there is another form of the "completeness property" that is traditionally assigned to the reals: it is the property that every nonempty set of real numbers that is bounded above has a *least upper bound* (lub), that is, an upper bound smaller than any other upper bound, and that every nonempty set of real numbers that is bounded below has a *greatest lower bound* (glb), that is, an lower bound larger than any other lower bound. One can show that this definition of the completeness property is equivalent to the convergence of all Cauchy sequences in $\mathbb{R}$.

Definition 5.10. A subset Y of the metric space X is *sequentially compact* if for every sequence $\{y_n\}_{n=1}^{\infty}$ of elements in Y, there is a subsequence $\{y_{n_k}\}_{k=1}^{\infty}$ with $n_1 < n_2 < \cdots$ that converges to a point $y^* \in Y$.

Definition 5.11. A function $f : X \to Y$ of metric spaces X, Y is continuous at the point $x^* \in X$ if for every ball $B_\epsilon(f(x^*))$ in Y there is a ball $B_\delta(x^*)$ in X such that $f(B_\delta(x^*)) \subseteq B_\epsilon(f(x^*))$. The function is *continuous on a subset* S of X if it is continuous at every point of S.

One can show that this definition of continuity at x^* is equivalent to the following condition: if $\lim_{n\to\infty} x_n = x^*$ in X, then $\lim_{n\to\infty} f(x_n) = f(x^*)$ in Y. This definition is a good example of how mathematical notation is just a refined presentation of something that is really very intuitive: what is says is that "f is continuous at x^* if you can make $f(x)$ as close to $f(x^*)$ as you please by making x sufficiently close to x^*.

Example 5.12. Every bounded linear operator $T : V \to W$ of normed linear spaces is continuous.

The key theorems we need are

Theorem 5.13. *For a metric space X, compactness and sequential compactness are equivalent.*

Theorem 5.14. *(Heine-Borel) A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Technically, the Heine-Borel theorem only applies to the standard normed linear space $\mathbb{R}^n$, but the proof of it is virtually identical to the proof for any finite-dimensional normed linear space. Thus we have:

Theorem 5.15. *A subset of a finite-dimensional normed linear space is compact if and only if it is closed and bounded.*

Finally, we have the following famous theorem of analysis, which tells us that the polynomials are a dense subset of $C[a,b]$ with the infinity norm, that is, every open ball in $C[a,b]$ with respect to this norm contains a polynomial:

Theorem 5.16. *(Weierstrauss) Given any $f \in C[a,b]$ and number $\epsilon > 0$, there exists a polynomial $p(x)$ such that*

$$\|f - p\|_\infty < \epsilon.$$

5.2. **Calculus and Analysis.** Here are two theorems often quoted in calculus courses in the special case of closed bounded intervals on the real line:

Theorem 5.17. *(Intermediate Value Theorem – IVT) Let $f \in C[a,b]$. Then $f(x)$ assumes every value between the numbers $f(a)$ and $f(b)$.*

Theorem 5.18. *(Extreme Value Theorem – EVT) The continuous function $f : X \to \mathbb{R}$ from compact metric space X to the reals assumes its extreme values, that is, there are points x_{min} and x_{max} in X such that for all $x \in X$,*

$$f(x_{min}) \leq f(x) \leq f(x_{max}).$$

Proof. Consider the case of a maximum value. First note that the image of X under f, $f(X)$, must be bounded from above. For if not, one could find a sequence of points $x_n \in X$ such that $f(x_n) \geq n$. By compactness, the sequence $\{x_n\}_{n=1}^{\infty}$ has a convergent subsequence in X, say $\lim_{k\to\infty} x_{n_k} = x^*$. By continuity,

$$\lim_{k\to\infty} f(x_{n_k}) = f(x^*).$$

But the left-hand side is unbounded, hence the limit does not exist. This is impossible, so we conclude that $f(X)$ is bounded from above. Consequently, by the completeness property of the real numbers, this set has a least upper bound, say $lub f(X) = M$. By definition of lub, for any positive integer n, there is a point z_n such that

$$M - \frac{1}{n} \leq f(z_n) \leq M.$$

Clearly $\lim_{n\to\infty} f(z_n) = M$. Now choose a subsequence of points $\{z_{n_k}\}_{k=1}^{\infty}$ that converges to the point $x_{max} \in X$. By continuity, $\lim_{k\to\infty} f(z_{n_k}) = f(x_{max}) = M$, which is what we wanted to show. The minimum value is handled similarly. □

Here is a handy consequence of the EVT, which finds much use in interpolation theory:

Theorem 5.19. *(Rolle's Theorem) Let $f(x) \in C[a,b]$. Suppose that $f'(x)$ is defined for $x \in (a,b)$ and that $f(a) = 0 = f(b)$. Then there exists a number c with $a < c < b$ such that $f'(c) = 0$.*

Proof. Note that there must a point c in the open interval (a,b) at which $f(x)$ achieves its maximum or minimum value, for if they are both at endpoints, then $f(x)$ must be identically zero. Say it is a maximum at $x = c$. We are given that $f'(c)$ exists. Also, in some open interval about c we must have $f(c) \geq f(x)$ for x in the interval. Take the limit as $x \to c$ on the right (so that $x > c$) and obtain that

$$f'(c) = \lim_{x \to c_+} \frac{f(x) - f(c)}{x - c} \leq 0,$$

while if we take the limit on the left (so that $x < c$), we obtain that

$$f'(c) = \lim_{x \to c_-} \frac{f(x) - f(c)}{x - c} \geq 0,$$

from which we conclude that we must actually have $f'(c) = 0$, as desired. □

One can use Rolle's theorem to obtain another useful fact for calculus which, in plain words, simply says that the slope of the line joining two points on the graph of a function on an interval is just the derivative of the function at some point in the interval.

Theorem 5.20. *(Mean Value Theorem – MVT) Let $f(x) \in C[a,b]$. Suppose that $f'(x)$ is defined for $x \in (a,b)$. Then there exists a number c with $a < c < b$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Another very useful fact from calculus that uses the MVT is the following. Recall that $C^{(n)}[a,b]$ is the space of all real-valued functions whose n-th derivative $f^{(n)}(x)$ is continuous on the interval $[a,b]$.

Theorem 5.21. *(Taylor's Formula) Let $f(x) \in C^{(n+1)}[a,b]$ and let $c \in (a,b)$. For any $x \in (a,b)$ there exists a number ξ between c and x such that*

$$f(x) = f(c) + \frac{f'(c)}{1!}(x - c) + \cdots + \frac{f^{(n)}(c)}{n!}(x - c)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - c)^{n+1}.$$

Exercises

Exercise 8. Show that every normed linear space V with norm $\|\cdot\|$ is a metric space with metric given by

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

6. Numerical Analysis

6.1. **Big Oh Notation.** The complexity of an algorithm is a fundamental issue in computer science. "Complexity" is measured in many ways, ranging from simple-minded benchmark timing to sophisticated counting algorithms. Basically, complexity should be a measure of time and/or space – that is, the time required to execute an algorithm and the space in memory that the algorithm requires.

Of course, the question is, "What do you count?". In numerical analysis, perhaps the most common measure of complexity as far as time is concerned is a matter of counting "flops", that is, floating point operations, defined as follows.

Definition 6.1. A flop is a single operation of floating point multiplication, division, addition or subtraction.

Of course, there can be a whole lot more to a numerical algorithm than the number of flops required to complete it. In the past, floating point operations so dominated computations that it was sufficient to count flops. Nowadays, floating point operations approach the speed of integer arithmetic, and in multiprocessing issues such as bus communication times are also significant. So, counting flops is a rougher meaure of time complexity than it used to be, but it is still very important.

Example 6.2. Consider the problem of solving a linear $n \times n$ system $Ax = b$ of n equations in n unknowns and a unique solution. One can compute the expense of Gaussian elimination or, equivalently, LU factorization, and arrive at a formula for the number of flops as roughly $2n^3/3$. More precisely, this means the exact count is a polynomial in $p(n)$ of degree three whose leading coefficient is $2/3$. For large n, the leading term dominates, so $2n^3/3$ is good enough.

There is a standard notation for situations like this that give us a good idea of complexity, namely the "big oh" notation, which is defined as follows.

Definition 6.3. We say that the function $f(x)$ is *big oh* of $g(x)$ as $x \to a$ if there exists a positive constant M such that

$$|f(x)| \leq M |g(x)|$$

for x in some neighborhood (open interval) about a. In this case we write

$$f(x) = \mathcal{O}(g(x))$$

as $x \to a$.

In reference to our linear systems example, we could say that the number of flops required to solve the system is $\mathcal{O}(n^3)$. We don't bother to add "as $n \to \infty$" as that this is understood. Here is another example of this notation.

The big oh notation is also used as a measure of error in numerical analysis. This time, we are typically concerned with some continuous variable, such as step size used in an algorithm, for which accuracy increases as we decrease the size of the variable.

Example 6.4. Suppose that the error in an approximating algorithm depending on a parameter h decreases as $h \to 0$, and in fact, is estimated to be something like $e(h) = C(1 - \cos(h))$ for some positive constant C. Show that

$$e(h) = \mathcal{O}(h^2), \ h \to 0.$$

To see this, recall the Taylor series for $\cos h$ about $h = 0$ is given by

$$\cos h = 1 - \frac{h^2}{2!} + \frac{h^4}{4!} - \cdots$$

so that

$$1 - \cos h = \frac{h^2}{2!} - \frac{h^4}{4!} + \cdots = h^2 \left(\frac{1}{2!} - \frac{h^2}{4!} + \cdots \right) = h^2 p(h),$$

where $p(h)$ is continuous with $p(0) = 1/2$. Therefore, in some interval about $h = 0$, we must have $|p(h)| \leq 1$, say. It follows that

$$|e(h)| = |C(1 - \cos h)| \leq Ch^2$$

in that interval, so that we conclude from definition of big oh that $e(h) = \mathcal{O}(h^2)$, $h \to 0$.

We can get clues about the correct order of magnitude of a power-type big oh term by benchmarking. Roughly, for example, if an algorithm is $\mathcal{O}(n^3)$, where n is the "size" of the problem, then when we solve an equivalent problem of size $2n$, we expect the complexity to increase by a factor of 8 roughly. (Replace n by $2n$ and factor out the 2^3.

We can also get an idea of the correct order of magnitude of an approximation error term depending on parameter h by applying the algorithm to a known function (so that we can calculate the actual error of approximation), and then applying the algorithm to the same function with $h/2$ in place of h. Roughly, for example, if the algorithm error is $\mathcal{O}(h^2)$ as $h \to 0$, then we would expect the error to reduce by a factor of 4. (Replace h by $h/2$ and factor out the $1/2^2$.

6.2. **Numerical Differentiation and Integration.** One can obtain a simple numerical estimate of derivatives simply by going back to the definition of derivative:

Definition. If the function $f(x)$ is defined in an open interval containing the point a, then the derivative of f at $x = a$ is defined to be the following limit, if it exists:

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

It is the second form of the derivative that is of interest for numerical purposes. Simply truncate the limiting process and write

$$f'(a) \approx \frac{f(a+h) - f(a)}{h},$$

for a "sufficiently small" choice of h, whatever that might be. In fact, we can estimate the size of the error using Taylor's formula. Assume that $f(x) \in C^{(2)}[a,b]$ so that there exists a positive number M such that $|f''(\xi)| \leq M$ for all $a \leq \xi \leq b$. According to Taylor's formula, for some ξ between a and x we have

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(\xi)}{2}(x-a)^2,$$

or equivalently

$$\frac{f(x) - f(a)}{x - a} - f'(a) = \frac{f''(\xi)}{2}(x-a).$$

It follows that

$$\left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| \leq \frac{|f''(\xi)|}{2}|x-a| \leq \frac{M}{2}|x-a|.$$

Now set $x = a + h$ and we can express the resulting inequality in the language of big-oh as

$$\frac{f(a+h) - f(a)}{h} - f'(a) = \mathcal{O}(|h|)$$

In regards to integration, there are many techniques for achieving highly accurate approximations to a given definite integral. One of the most elementary is one that comes from calculus

(Simple) Trapezoidal Rule: $\displaystyle\int_a^b f(x)\,dx \approx \frac{f(a) + f(b)}{2(b-a)}$

This is simply the area of the trapezoid obtained by connecting the points $(a, 0)$, $(a, f(a))$, $(b, f(b))$ and $(b, 0)$. It isn't too difficult to verify that the error is $\mathcal{O}\left((b-a)^3\right)$ by assuming the f has a continuous second derivative and using Taylor's formula on $F(x) = \int_x^b f(t)\,dt$.

Next, break the interval up into equal sized subintervals using step size $h = (b-a)/N$ and partition points

$$x_j = a + (j-1)\,h,\ \ j = 0, 1, \ldots, N.$$

Apply the simple trapezoidal rule to each subinterval and add up the results to obtain

$$\text{(Composite) Trapezoidal Rule:}\ \int_a^b f(x)\,dx \approx \frac{b-a}{2N}\left\{f(x_0) + 2\sum_{k=1}^{N-1} f(x_k) + f(x_N)\right\}.$$

One can show that the error of the composite rule is $\mathcal{O}\left(h^2\right)$ by summing the errors of each subinterval.

Exercises

Exercise 6.1. Determine the order of $f(h) = \sin\left(h^2\right)$ as $h \to 0$ in terms of h.

References

[1] M.J.D. Powell, *Approximation Theory and Methods*, Cambridge University Press, Cambridge, 1981.

[2] T. Shores, *Applied Linear Algebra and Matrix Analysis*, Springer, New York, 2007.