

**LINEAR ALGEBRA NOTES FOR APPROXIMATION THEORY  
MATH 441, FALL 2009**

THOMAS SHORES

Last Rev.: 9/1/09

1. VECTOR SPACES

Here is the general definition of a vector space:

**Definition 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 2.** Given a positive integer  $n$ , we define the *standard vector space of dimension  $n$  over the reals* to be the set of vectors

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

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

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 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  $a \leq x \leq b$ , 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  $a \leq x \leq b$ ,

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

(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.  $\square$

We could have abbreviated the work above by using the Subspace Test. Recall from Math 314:

**Definition 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 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  $c$ ,  $c\mathbf{u} \in W$ .

Finally, we recall some key ideas about linear combinations of vectors.

**Definition 6.** Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  be vectors in the vector space  $V$ . The *span* of these vectors, denoted by  $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ , is the subset of  $V$  consisting of all possible linear combinations of these vectors, i.e.,

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

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

**Definition 7.** The vectors  $v_1, v_2, \dots, v_n$  are said to be *linearly dependent* if there exist scalars  $c_1, c_2, \dots, c_n$ , not all zero, such that

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

Otherwise, the vectors are called *linearly independent*.

**Definition 8.** A *basis* for the vector space  $V$  is a spanning set of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  that is a linearly independent set.

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

**Theorem 9.** *Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  be a basis of the vector space  $V$ . Then every  $\mathbf{v} \in V$  can be expressed uniquely as a linear combination of  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , up to order of terms.*

**Definition 10.** 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 11.** *Let  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r$  be a linearly independent set in the space  $V$  and let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  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 12.** *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.** 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 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 13.** 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)  $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ .

A vector space  $V$ , together with a norm  $\|\cdot\|$  on the space  $V$ , is called a *normed (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

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

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

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

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  $p \rightarrow \infty$ . To keep matters simple, we'll supply a separate definition for this norm.

**Definition 15.** Let  $V$  be one of the standard spaces  $\mathbb{R}^n$  or  $\mathbb{C}^n$ . The  $\infty$ -norm of a vector in  $V$  is defined by the formula

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

**Example 16.** Calculate  $\|\mathbf{v}\|_p$ , where  $p = 1, 2$ , or  $\infty$  and  $\mathbf{v} = (1, -3, 2, -1) \in \mathbb{R}^4$ .

We calculate:

$$\|(1, -3, 2, -1)\|_1 = |1| + |-3| + |2| + |-1| = 7$$

$$\|(1, -3, 2, -1)\|_2 = \sqrt{|1|^2 + |-3|^2 + |2|^2 + |-1|^2} = \sqrt{15}$$

$$\square \quad \|(1, -3, 2, -1)\|_\infty = \max\{|1|, |-3|, |2|, |-1|\} = 3.$$

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

Let  $c$  be a scalar, and let  $\mathbf{u} = (z_1, z_2, \dots, z_n)$ , and  $\mathbf{v} = (w_1, w_2, \dots, w_n)$  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

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

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

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

**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

$$(2.1) \quad \mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

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

**Example 18.** Construct a unit vector in the direction of  $\mathbf{v} = (1, -3, 2, -1)$ , where the 1-norm, 2-norm, and  $\infty$ -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 19.** 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 *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 20.** Sketch the unit balls centered at the origin for the 1-norm, 2-norm, and  $\infty$ -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  $\infty$ -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  $\infty$ -norms have corners, unlike the 2-norm. See Figure 2.1 for a picture of these balls.  $\square$

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,  $\lim_{n \rightarrow \infty} \mathbf{v}_n = \mathbf{v}$  was taken to mean that  $\lim_{n \rightarrow \infty} \|\mathbf{v}_n - \mathbf{v}\| = 0$ . In this case we said that the sequence  $\mathbf{v}_1, \mathbf{v}_2, \dots$  *converges* to  $\mathbf{v}$ . Will we have to have a different notion of limits for different norms? For *finite-dimensional* spaces, the somewhat

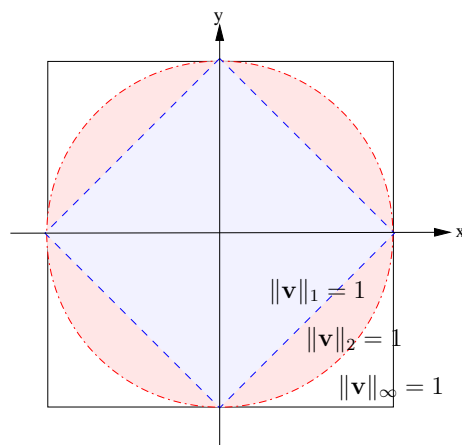


FIGURE 2.1. Boundaries of unit balls in various norms.

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  $\mathbf{v}$ ,

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

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 21.** Verify that  $\lim_{n \rightarrow \infty} \mathbf{v}_n$  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

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

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

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

and

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

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.  $\square$

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

**Definition 22.** 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 23.** 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.

**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 24.** 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

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

are also in  $S$ .

**Definition 25.** 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

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

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

**Exercise 4.** 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 26.** 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):  $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$  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 27.** For vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ , with  $\mathbf{u} = (u_1, u_2, \dots, u_n)$  and  $\mathbf{v} = (v_1, v_2, \dots, v_n)$ , define

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \dots + u_nv_n = \mathbf{u}^T \mathbf{v}.$$

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 28.** 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

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle &= 2u_1(v_1 + w_1) + 3u_2(v_2 + w_2) \\ &= 2u_1v_1 + 3u_2v_2 + 2u_1w_1 + 3u_2w_2 = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle. \end{aligned}$$

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

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

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 29.** 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  $1 \times 1$  scalar quantity  $q$ ,  $q^* = \bar{q}$ , 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.  $\square$

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 30.** 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  $\int_0^1 f(x)^2 dx = \langle f, f \rangle \geq 0$ . 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.  $\square$

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  $\int_a^b x^m dx = \frac{1}{m+1} (b^{m+1} - a^{m+1})$  and special case  $\int_0^1 x^m dx = \frac{1}{m+1}$  for  $m \geq 0$ .)

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

**Theorem 31.** *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):  $\langle a\mathbf{u}, \mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle$ .

**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 32.** 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  $\mathbf{u} = \mathbf{0}$ . 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 33.** (*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 34.** 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  $\theta$  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 35.** 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 28. Compare this to the angle found when one uses the standard inner product in  $\mathbb{R}^2$ .

**Solution.** According to 28 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  $\theta = \pi/2 \approx 1.5708$ .  $\square$

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 36.** 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 37.** 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 30 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.

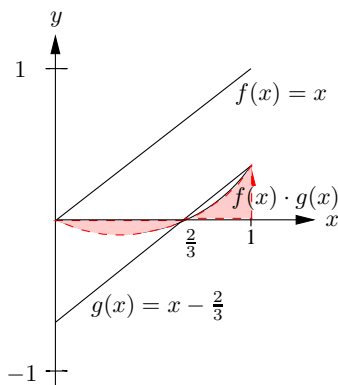


FIGURE 3.1. Graphs of  $f$ ,  $g$ , and  $f \cdot g$  on the interval  $[0, 1]$ .

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

**Example 39.** 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.  $\square$

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 40.** Use the result of Example 39 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 39, so that the infinity norm cannot be induced from an inner product.  $\square$

**Definition 41.** The set of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  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.

The proof of the following key fact and its corollary 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 42.** Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  be an orthogonal set of nonzero vectors and suppose that  $\mathbf{v} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ . Then  $\mathbf{v}$  can be expressed uniquely (up to order) as a linear combination of  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , namely

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

**Corollary 43.** 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.

**Corollary 44.** If  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  is an orthogonal set of vectors and  $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$ , then

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

## Exercises

**Exercise 5.** 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  $\text{span}\{p_1(x), p_2(x)\}$  using Theorem 42.

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

## 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  $f : D \rightarrow T$ , 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

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

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  $f : \mathbb{R} \rightarrow \mathbb{R}$  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  $f : V \rightarrow W$ , is sometimes called an *operator* or *transformation*. One of the simplest mappings of a vector space  $V$  is the so-called *identity function*  $\text{id}_V : V \rightarrow V$  given by  $\text{id}_V(\mathbf{v}) = \mathbf{v}$ , for all  $\mathbf{v} \in V$ . Here domain, range, and target all agree. Of course, matters can become more complicated. For example, the operator  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  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 45.** A function  $T : V \rightarrow W$  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  $T : V \rightarrow V$  is a linear operator, we simply say that  $T$  is a linear operator on  $V$ . A linear operator  $T : V \rightarrow \mathbb{R}$  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  $c_1, c_2, \dots, c_n$  and vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , we have

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

**Example 46.** Determine whether  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  is a linear operator, where  $T$  is given by the formula

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

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

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

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  $f : V \rightarrow W$  is said to be *invertible* if there is an operator  $g : W \rightarrow V$  such that the composition of functions satisfies  $f \circ g = \text{id}_W$  and  $g \circ f = \text{id}_V$ . 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 47.** Show that if  $f : V \rightarrow W$  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

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

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  $V$ .

*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 48.** Let  $T : V \rightarrow W$  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 49.** *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 applying 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 50.** *Let  $T : V \rightarrow W$  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 51.** Let  $T_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  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, observe 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  $\|A\mathbf{v}\|$ .  $\square$

**Exercises**

**Exercise 6.** 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 7.** Show that the operator  $T : C[0, 1] \rightarrow \mathbb{R}$  given by  $T(f) = \int_0^1 f(x) dx$  is a linear operator.

5. METRIC SPACES AND ANALYSIS

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.

**Definition 52.** 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 53.** 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.

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 54.** 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 55.** 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  $X \setminus Y$  in  $X$  is an open set.

(I'll finish this later)

**Definition 56.** bounded set

**Definition 57.** boundary

**Definition 58.** compact, sequentially compact

**Definition 59.** continuous function

**Example 60.** Linear operators

The key theorems we need are

**Theorem 61.** For a metric space  $X$ , compactness and sequential compactness are equivalent.

**Theorem 62.** Heine-Borel theorem

**Theorem 63.** EVT

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

**Theorem 64.** (Weierstrauss) Given any  $f \in C[a, b]$  and number  $\epsilon > 0$ , there exists a polynomial  $p(x)$  such that

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

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

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

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.