

## Euclidean Space

*Euclidean space* is a mathematical construct that encompasses the line, the plane, and three-dimensional space as special cases. Its elements are called *vectors*. Vectors can be understood in various ways: as arrows, as quantities with magnitude and direction, as displacements, or as points. However, along with a sense of what vectors are, we also need to emphasize how they interact. The axioms in Section 2.1 capture the idea that vectors can be added together and can be multiplied by scalars, with both of these operations obeying familiar laws of algebra. Section 2.2 expresses the geometric ideas of length and angle in Euclidean space in terms of vector algebra. Section 2.3 discusses continuity for functions (also called *mappings*) whose inputs and outputs are vectors rather than scalars. Section 2.4 introduces a special class of sets in Euclidean space, the *compact* sets, and shows that compact sets are preserved under continuous mappings.

### 2.1 Algebra: Vectors

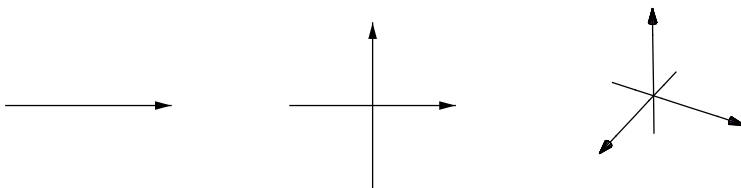
Let  $n$  be a positive integer. The set of all ordered  $n$ -tuples of real numbers,

$$\mathbb{R}^n = \{(x_1, \dots, x_n) : x_i \in \mathbb{R} \text{ for } i = 1, \dots, n\},$$

constitutes  **$n$ -dimensional Euclidean space**. When  $n = 1$ , the parentheses and subscript in the notation  $(x_1)$  are superfluous, so we simply view the elements of  $\mathbb{R}^1$  as real numbers  $x$  and write  $\mathbb{R}$  for  $\mathbb{R}^1$ . Elements of  $\mathbb{R}^2$  and of  $\mathbb{R}^3$  are written  $(x, y)$  and  $(x, y, z)$  to avoid needless subscripts. These first few Euclidean spaces,  $\mathbb{R}$ ,  $\mathbb{R}^2$ , and  $\mathbb{R}^3$ , are conveniently visualized as the line, the plane, and space itself. (See Figure 2.1.)

Elements of  $\mathbb{R}$  are called **scalars**, of  $\mathbb{R}^n$ , **vectors**. The **origin** of  $\mathbb{R}^n$ , denoted  $\mathbf{0}$ , is defined to be

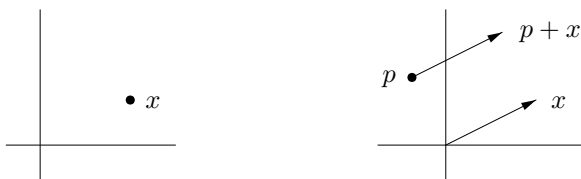
$$\mathbf{0} = (0, \dots, 0).$$



**Figure 2.1.** The first few Euclidean spaces

Sometimes the origin of  $\mathbb{R}^n$  will be denoted  $\mathbf{0}_n$  to distinguish it from other *origins* that we will encounter later.

In the first few Euclidean spaces  $\mathbb{R}$ ,  $\mathbb{R}^2$ ,  $\mathbb{R}^3$ , one can visualize a vector as a point  $x$  or as an arrow. The arrow can have its tail at the origin and its head at the point  $x$ , or its tail at any point  $p$  and its head correspondingly translated to  $p + x$ . (See Figure 2.2. Most illustrations will depict  $\mathbb{R}$  or  $\mathbb{R}^2$ .)



**Figure 2.2.** Various ways to envision a vector

To a mathematician, the word *space* doesn't connote volume but instead refers to a set endowed with some structure. Indeed, Euclidean space  $\mathbb{R}^n$  comes with two algebraic operations. The first is **vector addition**,

$$+ : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n,$$

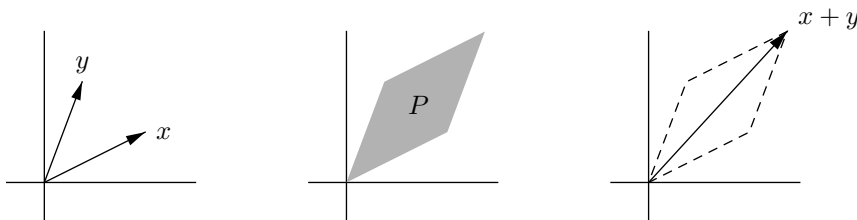
defined by adding the scalars at each component of the vectors,

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n).$$

For example,  $(1, 2, 3) + (4, 5, 6) = (5, 7, 9)$ . Note that the meaning of the “+” sign is now overloaded: on the left of the displayed equality, it denotes the new operation of vector addition, whereas on the right side it denotes the old addition of real numbers. The multiple meanings of the plus sign shouldn't cause problems, because the meaning of “+” is clear from context, i.e., the

meaning of “+” is clear from whether it sits between vectors or scalars. (An expression such as “ $(1, 2, 3) + 4$ ,” with the plus sign between a vector and a scalar, makes no sense according to our grammar.)

The interpretation of vectors as arrows gives a geometric description of vector addition, at least in  $\mathbb{R}^2$ . To add the vectors  $x$  and  $y$ , draw them as arrows starting at  $\mathbf{0}$  and then complete the parallelogram  $P$  that has  $x$  and  $y$  as two of its sides. The diagonal of  $P$  starting at  $\mathbf{0}$  is then the arrow depicting the vector  $x + y$ . (See Figure 2.3.) The proof of this is a small argument with similar triangles, left to the reader as Exercise 2.1.2.



**Figure 2.3.** The parallelogram law of vector addition

The second operation on Euclidean space is **scalar multiplication**,

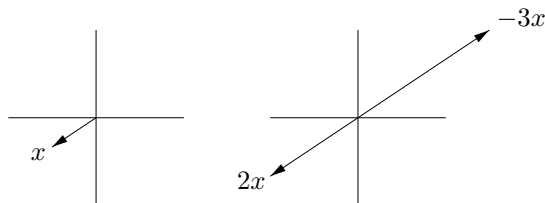
$$\cdot : \mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}^n,$$

defined by

$$a \cdot (x_1, \dots, x_n) = (ax_1, \dots, ax_n).$$

For example,  $2 \cdot (3, 4, 5) = (6, 8, 10)$ . We will almost always omit the symbol “ $\cdot$ ” and write  $ax$  for  $a \cdot x$ . With this convention, juxtaposition is overloaded as “+” was overloaded above, but again this shouldn’t cause problems.

Scalar multiplication of the vector  $x$  (viewed as an arrow) by  $a$  also has a geometric interpretation: it simply stretches (i.e., scales)  $x$  by a factor of  $a$ . When  $a$  is negative,  $ax$  turns  $x$  around and stretches it in the other direction by  $|a|$ . (See Figure 2.4.)



**Figure 2.4.** Scalar multiplication as stretching

With these two operations and distinguished element  $\mathbf{0}$ , Euclidean space satisfies the following algebraic laws.

**Theorem 2.1.1 (Vector space axioms).**

- (A1) *Addition is associative:  $(x + y) + z = x + (y + z)$  for all  $x, y, z \in \mathbb{R}^n$ .*
- (A2)  *$\mathbf{0}$  is an additive identity:  $\mathbf{0} + x = x$  for all  $x \in \mathbb{R}^n$ .*
- (A3) *Existence of additive inverses: for each  $x \in \mathbb{R}^n$  there exists  $y \in \mathbb{R}^n$  such that  $y + x = \mathbf{0}$ .*
- (A4) *Addition is commutative:  $x + y = y + x$  for all  $x, y \in \mathbb{R}^n$ .*
- (M1) *Scalar multiplication is associative:  $a(bx) = (ab)x$  for all  $a, b \in \mathbb{R}$ ,  $x \in \mathbb{R}^n$ .*
- (M2) *1 is a multiplicative identity:  $1x = x$  for all  $x \in \mathbb{R}^n$ .*
- (D1) *Scalar multiplication distributes over scalar addition:  $(a + b)x = ax + bx$  for all  $a, b \in \mathbb{R}$ ,  $x \in \mathbb{R}^n$ .*
- (D2) *Scalar multiplication distributes over vector addition:  $a(x + y) = ax + ay$  for all  $a \in \mathbb{R}$ ,  $x, y \in \mathbb{R}^n$ .*

All of these are consequences of how “+” and “.” and  $\mathbf{0}$  are defined for  $\mathbb{R}^n$  in conjunction with the fact that the real numbers, in turn endowed with “+” and “.” and containing 0 and 1, satisfy the field axioms (see Section 1.1). For example, to prove that  $\mathbb{R}^n$  satisfies (M1), take any scalars  $a, b \in \mathbb{R}$  and any vector  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ . Then

$$\begin{aligned}
 a(bx) &= a(b(x_1, \dots, x_n)) && \text{by definition of } x \\
 &= a(bx_1, \dots, bx_n) && \text{by definition of scalar multiplication} \\
 &= (a(bx_1), \dots, a(bx_n)) && \text{by definition of scalar multiplication} \\
 &= ((ab)x_1, \dots, (ab)x_n) && \text{by } n \text{ applications of (m1) in } \mathbb{R} \\
 &= (ab)(x_1, \dots, x_n) && \text{by definition of scalar multiplication} \\
 &= (ab)x && \text{by definition of } x.
 \end{aligned}$$

The other vector space axioms for  $\mathbb{R}^n$  can be shown similarly, by unwinding vectors to their coordinates, quoting field axioms coordinatewise, and then bundling the results back up into vectors (see Exercise 2.1.3). Nonetheless, the vector space axioms do not perfectly parallel the field axioms, and you are encouraged to spend a little time comparing the two axiom sets to get a feel for where they are similar and where they are different (see Exercise 2.1.4). Note in particular that

*For  $n > 1$ ,  $\mathbb{R}^n$  is not endowed with vector-by-vector multiplication.*

Although one can define vector multiplication on  $\mathbb{R}^n$  componentwise, this multiplication does not combine with vector addition to satisfy the field axioms except when  $n = 1$ . The multiplication of complex numbers makes  $\mathbb{R}^2$  a field, and in Section 3.10 we will see an interesting noncommutative multiplication of vectors for  $\mathbb{R}^3$ , but these are special cases.

One benefit of the vector space axioms for  $\mathbb{R}^n$  is that they are phrased **intrinsically**, meaning that they make no reference to the scalar coordinates of the vectors involved. Thus, once you use coordinates to establish the vector space axioms, your vector algebra can be intrinsic thereafter, making it lighter and more conceptual. Also, in addition to being intrinsic, the vector space axioms are general. While  $\mathbb{R}^n$  is the prototypical set satisfying the vector space axioms, it is by no means the only one. In coming sections we will encounter other sets  $V$  (whose elements may be, for example, functions) endowed with their own addition, multiplication by elements of a field  $F$ , and distinguished element  $\mathbf{0}$ . If the vector space axioms are satisfied with  $V$  and  $F$  replacing  $\mathbb{R}^n$  and  $\mathbb{R}$  then we say that  $V$  is a **vector space over  $F$** .

The pedagogical point here is that although the similarity between vector algebra and scalar algebra may initially make vector algebra seem uninspiring, in fact the similarity is exciting. It makes mathematics easier, because familiar algebraic manipulations apply in a wide range of contexts. The same symbol-patterns have more meaning. For example, we use intrinsic vector algebra to prove a result from Euclidean geometry, that the three medians of a triangle intersect. (A median is a segment from a vertex to the midpoint of the opposite edge.) Consider a triangle with vertices  $x$ ,  $y$ , and  $z$ , and form the average of the three vertices,

$$p = \frac{x + y + z}{3}.$$

This algebraic average will be the **geometric center** of the triangle, where the medians meet. (See Figure 2.5.) Indeed, rewrite  $p$  as

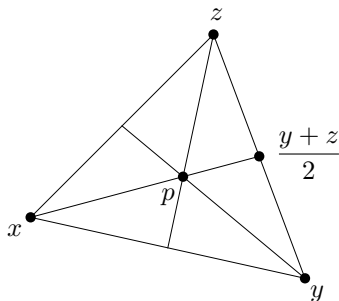
$$p = x + \frac{2}{3} \left( \frac{y + z}{2} - x \right).$$

The displayed expression for  $p$  shows that it is two-thirds of the way from  $x$  along the line segment from  $x$  to the average of  $y$  and  $z$ , i.e., that  $p$  lies on the triangle median from vertex  $x$  to side  $yz$ . (Again see the figure. The idea is that  $(y + z)/2$  is being interpreted as the midpoint of  $y$  and  $z$ , each of these viewed as a point, while on the other hand, the little mnemonic

*head minus tail*

helps us to remember quickly that  $(y + z)/2 - x$  can be viewed as the arrow-vector from  $x$  to  $(y + z)/2$ .) Since  $p$  is defined symmetrically in  $x$ ,  $y$ , and  $z$ , and it lies on one median, it therefore lies on the other two medians as well. In fact, the vector algebra has shown that it lies two-thirds of the way along each median. (As for how a person might find this proof, it is a matter of hoping that the geometric center  $(x + y + z)/3$  lies on the median by taking the form  $x + c((y + z)/2 - x)$  for some  $c$  and then seeing that indeed  $c = 2/3$  works.)

The **standard basis** of  $\mathbb{R}^n$  is the set of vectors



**Figure 2.5.** Three medians of a triangle

$$\{e_1, e_2, \dots, e_n\}$$

where

$$e_1 = (1, 0, \dots, 0), \quad e_2 = (0, 1, \dots, 0), \quad \dots, \quad e_n = (0, 0, \dots, 1).$$

(Thus each  $e_i$  is itself a vector, not the  $i$ th scalar entry of a vector.) Every vector  $x = (x_1, x_2, \dots, x_n)$  (where the  $x_i$  are scalar entries) decomposes as

$$\begin{aligned} x &= (x_1, x_2, \dots, x_n) \\ &= (x_1, 0, \dots, 0) + (0, x_2, \dots, 0) + \cdots + (0, 0, \dots, x_n) \\ &= x_1(1, 0, \dots, 0) + x_2(0, 1, \dots, 0) + \cdots + x_n(0, 0, \dots, 1) \\ &= x_1 e_1 + x_2 e_2 + \cdots + x_n e_n, \end{aligned}$$

or more succinctly,

$$x = \sum_{i=1}^n x_i e_i. \tag{2.1}$$

Note that in equation (2.1),  $x$  and the  $e_i$  are vectors, while the  $x_i$  are scalars. The equation shows that every  $x \in \mathbb{R}^n$  is expressible as a **linear combination** (sum of scalar multiples) of the standard basis vectors. The expression is unique, for if also  $x = \sum_{i=1}^n x'_i e_i$  for some scalars  $x'_1, \dots, x'_n$  then the equality says that  $x = (x'_1, x'_2, \dots, x'_n)$ , so that  $x'_i = x_i$  for  $i = 1, \dots, n$ .

(The reason that the geometric-sounding word *linear* is used here and elsewhere in this chapter to describe properties having to do with the algebraic operations of addition and scalar multiplication will be explained in Chapter 3.)

The standard basis is handy in that it is a finite set of vectors from which each of the infinitely many vectors of  $\mathbb{R}^n$  can be obtained in exactly one way as a linear combination. But it is not the only such set, nor is it always the optimal one.

**Definition 2.1.2 (Basis).** A set of vectors  $\{f_i\}$  is a **basis** of  $\mathbb{R}^n$  if every  $x \in \mathbb{R}^n$  is uniquely expressible as a linear combination of the  $f_i$ .

For example, the set  $\{f_1, f_2\} = \{(1, 1), (1, -1)\}$  is a basis of  $\mathbb{R}^2$ . To see this, consider an arbitrary vector  $(x, y) \in \mathbb{R}^2$ . This vector is expressible as a linear combination of  $f_1$  and  $f_2$  if and only if there are scalars  $a$  and  $b$  such that

$$(x, y) = af_1 + bf_2.$$

Since  $f_1 = (1, 1)$  and  $f_2 = (1, -1)$ , this vector equation is equivalent to a pair of scalar equations,

$$\begin{aligned}x &= a + b, \\y &= a - b.\end{aligned}$$

Add these equations and divide by 2 to get  $a = (x + y)/2$ , and similarly  $b = (x - y)/2$ . In other words, we have found that

$$(x, y) = \frac{x + y}{2}(1, 1) + \frac{x - y}{2}(1, -1),$$

and the coefficients  $a = (x + y)/2$  and  $b = (x - y)/2$  on the right side of the equation are the only possible coefficients  $a$  and  $b$  for the equation to hold. That is, scalars  $a$  and  $b$  exist to express the vector  $(x, y)$  as a linear combination of  $\{f_1, f_2\}$ , and the scalars are uniquely determined by the vector. Thus  $\{f_1, f_2\}$  is a basis of  $\mathbb{R}^2$ , as claimed.

The set  $\{g_1, g_2\} = \{(1, 3), (2, 6)\}$  is not a basis of  $\mathbb{R}^2$ , because every linear combination  $ag_1 + bg_2$  takes the form  $(a + 2b, 3a + 6b)$ , with the second entry equal to three times the first. The vector  $(1, 0)$  is therefore not a linear combination of  $g_1$  and  $g_2$ .

Nor is the set  $\{h_1, h_2, h_3\} = \{(1, 0), (1, 1), (1, -1)\}$  a basis of  $\mathbb{R}^2$ , because  $h_3 = 2h_1 - h_2$ , so that  $h_3$  is a nonunique linear combination of the  $h_j$ .

See Exercises 2.1.9 and 2.1.10 for practice with bases.

## Exercises

**2.1.1.** Write down any three specific nonzero vectors  $u, v, w$  from  $\mathbb{R}^3$  and any two specific nonzero scalars  $a, b$  from  $\mathbb{R}$ . Compute  $u + v$ ,  $aw$ ,  $b(v + w)$ ,  $(a + b)u$ ,  $u + v + w$ ,  $abw$ , and the additive inverse to  $u$ .

**2.1.2.** Working in  $\mathbb{R}^2$ , give a geometric proof that if we view the vectors  $x$  and  $y$  as arrows from  $\mathbf{0}$  and form the parallelogram  $P$  with these arrows as two of its sides, then the diagonal  $z$  starting at  $\mathbf{0}$  is the vector sum  $x + y$  viewed as an arrow.

**2.1.3.** Verify that  $\mathbb{R}^n$  satisfies vector space axioms (A2), (A3), (D1).

**2.1.4.** Are all the field axioms used in verifying that Euclidean space satisfies the vector space axioms?

**2.1.5.** Show that  $\mathbf{0}$  is the unique additive identity in  $\mathbb{R}^n$ . Show that each vector  $x \in \mathbb{R}^n$  has a unique additive inverse, which can therefore be denoted  $-x$ . (And it follows that vector subtraction can now be defined,

$$- : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n, \quad x - y = x + (-y) \quad \text{for all } x, y \in \mathbb{R}^n.)$$

Show that  $0x = \mathbf{0}$  for all  $x \in \mathbb{R}^n$ .

**2.1.6.** Repeat the previous exercise, but with  $\mathbb{R}^n$  replaced by an arbitrary vector space  $V$  over a field  $F$ . (Work with the axioms.)

**2.1.7.** Show the uniqueness of the additive identity and the additive inverse using only (A1), (A2), (A3). (This is tricky; the opening pages of some books on group theory will help.)

**2.1.8.** Let  $x$  and  $y$  be noncollinear vectors in  $\mathbb{R}^3$ . Give a geometric description of the set of all linear combinations of  $x$  and  $y$ .

**2.1.9.** Which of the following sets are bases of  $\mathbb{R}^3$ ?

$$S_1 = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\},$$

$$S_2 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\},$$

$$S_3 = \{(1, 1, 0), (0, 1, 1)\},$$

$$S_4 = \{(1, 1, 0), (0, 1, 1), (1, 0, -1)\}.$$

How many elements do you think a basis for  $\mathbb{R}^n$  must have? Give (without proof) geometric descriptions of all bases of  $\mathbb{R}^2$ , of  $\mathbb{R}^3$ .

**2.1.10.** Recall the field  $\mathbb{C}$  of complex numbers. Define **complex  $n$ -space**  $\mathbb{C}^n$  analogously to  $\mathbb{R}^n$ :

$$\mathbb{C}^n = \{(z_1, \dots, z_n) : z_i \in \mathbb{C} \text{ for } i = 1, \dots, n\},$$

and endow it with addition and scalar multiplication defined by the same formulas as for  $\mathbb{R}^n$ . You may take for granted that under these definitions,  $\mathbb{C}^n$  satisfies the vector space axioms with scalar multiplication by scalars from  $\mathbb{R}$ , and also  $\mathbb{C}^n$  satisfies the vector space axioms with scalar multiplication by scalars from  $\mathbb{C}$ . That is, using language that was introduced briefly in this section,  $\mathbb{C}^n$  can be viewed as a vector space over  $\mathbb{R}$  and also, separately, as a vector space over  $\mathbb{C}$ . Give a basis for each of these vector spaces.

## Brief Pedagogical Interlude

Before continuing, a few comments about how to work with these notes may be helpful.

The subject-matter of Chapters 2 through 5 is largely cumulative, with the main theorem of Chapter 5 being proved with main results of Chapters 2, 3, and 4. Each chapter is largely cumulative internally as well. To acquire detailed command of so much material and also a large-scale view of how it fits together, the trick is to focus on each section's techniques while studying that section and working its exercises, but thereafter to use the section's ideas freely by reference. Specifically, after the scrutiny of vector algebra in the previous section, one's vector manipulations should be fluent from now on, freeing one to concentrate on vector geometry in the next section, after which the geometry should also be light while one is concentrating on the analytical ideas of the following section, and so forth.

Admittedly, the model that one has internalized all the prior material before moving on is idealized. For that matter, so is the model that a body of interplaying ideas is linearly cumulative. In practice, focusing entirely on the details of whichever topics are currently active while using previous ideas by reference isn't always optimal. One might engage with the details of previous ideas because one is coming to understand them better, or because the current ideas showcase the older ones in a new way. Still, the paradigm of technical emphasis on the current ideas and fluent use of the earlier material does help a person who is navigating a large body of mathematics to conserve energy and synthesize a larger picture.

## 2.2 Geometry: Length and Angle

The geometric notions of length and angle in  $\mathbb{R}^n$  are readily described in terms of the algebraic notion of inner product.

**Definition 2.2.1 (Inner product).** *The inner product is a function from pairs of vectors to scalars,*

$$\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R},$$

*defined by the formula*

$$\langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = \sum_{i=1}^n x_i y_i.$$

For example,

$$\begin{aligned} \langle (1, 1, \dots, 1), (1, 2, \dots, n) \rangle &= \frac{n(n+1)}{2}, \\ \langle x, e_j \rangle &= x_j \quad \text{where } x = (x_1, \dots, x_n) \text{ and } j \in \{1, \dots, n\}, \\ \langle e_i, e_j \rangle &= \delta_{ij} \text{ (this means 1 if } i = j, 0 \text{ otherwise).} \end{aligned}$$

**Proposition 2.2.2 (Inner product properties).**

- (IP1) *The inner product is positive definite:  $\langle x, x \rangle \geq 0$  for all  $x \in \mathbb{R}^n$ , with equality if and only if  $x = \mathbf{0}$ .*
- (IP2) *The inner product is symmetric:  $\langle x, y \rangle = \langle y, x \rangle$  for all  $x, y \in \mathbb{R}^n$ .*
- (IP3) *The inner product is bilinear:*

$$\begin{aligned}\langle x + x', y \rangle &= \langle x, y \rangle + \langle x', y \rangle, & \langle ax, y \rangle &= a\langle x, y \rangle, \\ \langle x, y + y' \rangle &= \langle x, y \rangle + \langle x, y' \rangle, & \langle x, by \rangle &= b\langle x, y \rangle\end{aligned}$$

for all  $a, b \in \mathbb{R}$ ,  $x, x', y, y' \in \mathbb{R}^n$ .

*Proof.* Exercise 2.2.4. □

The reader should be aware that:

*In general,  $\langle x + x', y + y' \rangle$  does not equal  $\langle x, y \rangle + \langle x', y' \rangle$ .*

Indeed, expanding  $\langle x + x', y + y' \rangle$  carefully with the inner product properties shows that the cross-terms  $\langle x, y' \rangle$  and  $\langle x', y \rangle$  are present in addition to  $\langle x, y \rangle$  and  $\langle x', y' \rangle$ .

Like the vector space axioms, the inner product properties are phrased intrinsically, although they need to be proved using coordinates. As mentioned in the previous section, intrinsic methods are neater and more conceptual than using coordinates. More importantly:

*The rest of the results of this section are proved by reference to the inner product properties, with no further reference to the inner product formula.*

The notion of an inner product generalizes beyond Euclidean space—this will be demonstrated in Exercise 2.3.4, for example—and thanks to the displayed sentence, once the properties (IP1) through (IP3) are established for any inner product, all of the pending results in the section will follow automatically with no further work. (But here a slight disclaimer is necessary. In the displayed sentence, the word *results* does not refer to the pending graphic figures. The fact that the length and angle to be defined in this section will agree with prior notions of length and angle in the plane, or in three-dimensional space, does depend on the specific inner product formula. In Euclidean space, the inner product properties do not determine the inner product formula uniquely. This point will be addressed in Exercise 3.5.1.)

**Definition 2.2.3 (Modulus).** *The modulus (or absolute value) of a vector  $x \in \mathbb{R}^n$  is defined as*

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

Thus the modulus is defined in terms of the inner product, rather than by its own formula. The inner product formula shows that the modulus formula is

$$|(x_1, \dots, x_n)| = \sqrt{x_1^2 + \dots + x_n^2},$$

so that some particular examples are

$$|(1, 2, \dots, n)| = \sqrt{\frac{n(n+1)(2n+1)}{6}},$$

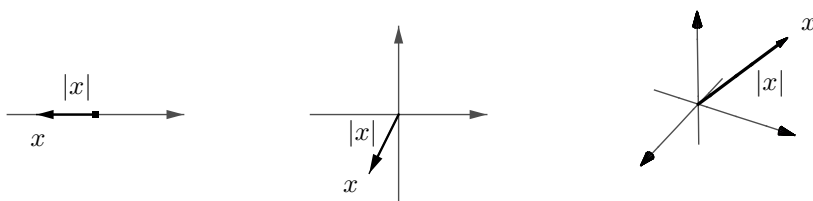
$$|e_i| = 1.$$

However, the definition of the modulus in terms of inner product combines with the inner product properties to show, with no reference to the inner product formula or the modulus formula, that the modulus satisfies the following properties (Exercise 2.2.5).

**Proposition 2.2.4 (Modulus properties).**

- (Mod1) *The modulus is positive:  $|x| \geq 0$  for all  $x \in \mathbb{R}^n$ , with equality if and only if  $x = \mathbf{0}$ .*
- (Mod2) *The modulus is absolute-homogeneous:  $|ax| = |a||x|$  for all  $a \in \mathbb{R}$  and  $x \in \mathbb{R}^n$ .*

Like other symbols, the absolute value signs are now overloaded, but their meaning can be inferred from context, as in property (Mod2). When  $n$  is 1, 2, or 3, the modulus  $|x|$  gives the distance from  $\mathbf{0}$  to the point  $x$ , or the length of  $x$  viewed as an arrow. (See Figure 2.6.)



**Figure 2.6.** Modulus as length

The following relation between inner product and modulus will help to show that distance in  $\mathbb{R}^n$  behaves as it should, and that angle in  $\mathbb{R}^n$  makes sense. Since the relation is not obvious, its proof is a little subtle.

**Theorem 2.2.5 (Cauchy–Schwarz inequality).** *For all  $x, y \in \mathbb{R}^n$ ,*

$$|\langle x, y \rangle| \leq |x| |y|,$$

*with equality if and only if one of  $x, y$  is a scalar multiple of the other.*

Note that the absolute value signs mean different things on each side of the Cauchy–Schwarz inequality. On the left side, the quantities  $x$  and  $y$  are vectors, their inner product  $\langle x, y \rangle$  is a scalar, and  $|\langle x, y \rangle|$  is its scalar absolute value, while on the right side,  $|x|$  and  $|y|$  are the scalar absolute values of vectors, and  $|x| |y|$  is their product. That is, the Cauchy–Schwarz inequality says:

*The size of the product is at most the product of the sizes.*

The Cauchy–Schwarz inequality can be written out in coordinates if we temporarily abandon the principle that we should avoid reference to formulas,

$$(x_1 y_1 + \cdots + x_n y_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2).$$

And this inequality can be proved unconceptually as follows (the reader is encouraged only to skim the following computation). Rewrite the desired inequality as

$$\left( \sum_i x_i y_i \right)^2 \leq \sum_i x_i^2 \cdot \sum_j y_j^2,$$

where the indices of summation run from 1 to  $n$ . Expand the square to get

$$\sum_i x_i^2 y_i^2 + \sum_{\substack{i,j \\ i \neq j}} x_i y_i x_j y_j \leq \sum_{i,j} x_i^2 y_j^2,$$

and canceling the terms common to both sides reduces it to

$$\sum_{i \neq j} x_i y_i x_j y_j \leq \sum_{i \neq j} x_i^2 y_j^2,$$

or

$$\sum_{i \neq j} (x_i^2 y_j^2 - x_i y_i x_j y_j) \geq 0.$$

Rather than sum over all pairs  $(i, j)$  with  $i \neq j$ , sum over the pairs with  $i < j$ , collecting the  $(i, j)$ -term and the  $(j, i)$ -term for each such pair, and the previous inequality becomes

$$\sum_{i < j} (x_i^2 y_j^2 + x_j^2 y_i^2 - 2x_i y_i x_j y_j) \geq 0.$$

Thus the desired inequality has reduced to a true inequality,

$$\sum_{i < j} (x_i y_j - x_j y_i)^2 \geq 0.$$

So the main proof is done, although there is still the question of when equality holds.

But surely the previous paragraph is not the graceful way to argue. The computation draws on the minutiae of the formulas for the inner product and

the modulus, rather than using their properties. It is uninformative, making the Cauchy–Schwarz inequality look like a low-level accident. It suggests that larger-scale mathematics is just a matter of bigger and bigger formulas. To prove the inequality in a way that is enlightening and general, we should work intrinsically, keeping the scalars  $\langle x, y \rangle$  and  $|x|$  and  $|y|$  notated in their concise forms, and we should use properties, not formulas. The idea is that the calculation in coordinates reduces to the fact that squares are nonnegative. That is, the Cauchy–Schwarz inequality is somehow *quadratically hard*, and its verification amounted to completing many squares. The argument to be given here is guided by this insight to prove the inequality by citing facts about quadratic polynomials, facts established by completing one square back in high-school algebra at the moment that doing so was called for. Thus we eliminate redundancy and clutter. So the argument to follow will involve an auxiliary object, a judiciously chosen quadratic polynomial, but in return it will become coherent.

*Proof.* The result is clear when  $x = \mathbf{0}$ , so assume  $x \neq \mathbf{0}$ . For every  $a \in \mathbb{R}$ ,

$$\begin{aligned}
 0 &\leq \langle ax - y, ax - y \rangle && \text{by positive definiteness} \\
 &= a\langle x, ax - y \rangle - \langle y, ax - y \rangle && \text{by linearity in the first variable} \\
 &= a^2\langle x, x \rangle - a\langle x, y \rangle - a\langle y, x \rangle + \langle y, y \rangle && \text{by linearity in the second variable} \\
 &= a^2|x|^2 - 2a\langle x, y \rangle + |y|^2 && \text{by symmetry, definition of modulus.}
 \end{aligned}$$

View the right side as a quadratic polynomial in the scalar variable  $a$ , where the scalar coefficients of the polynomial depend on the generic but fixed vectors  $x$  and  $y$ ,

$$f(a) = |x|^2 a^2 - 2\langle x, y \rangle a + |y|^2.$$

We have shown that  $f(a)$  is always nonnegative, so  $f$  has at most one root. Thus by the quadratic formula its discriminant is nonpositive,

$$4\langle x, y \rangle^2 - 4|x|^2|y|^2 \leq 0,$$

and the Cauchy–Schwarz inequality  $|\langle x, y \rangle| \leq |x||y|$  follows. Equality holds exactly when the quadratic polynomial  $f(a) = |ax - y|^2$  has a root  $a$ , i.e., exactly when  $y = ax$  for some  $a \in \mathbb{R}$ .  $\square$

Geometrically, the condition for equality in Cauchy–Schwarz is that the vectors  $x$  and  $y$ , viewed as arrows at the origin, are parallel, though perhaps pointing in opposite directions. A geometrically conceived proof of Cauchy–Schwarz is given in Exercise 2.2.15 to complement the algebraic argument that has been given here.

The Cauchy–Schwarz inequality shows that the modulus function satisfies the *triangle inequality*.

**Theorem 2.2.6 (Triangle inequality).** *For all  $x, y \in \mathbb{R}^n$ ,*

$$|x + y| \leq |x| + |y|,$$

with equality if and only if one of  $x, y$  is a nonnegative scalar multiple of the other.

*Proof.* To show this, compute

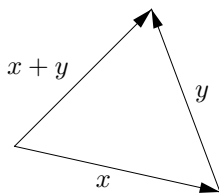
$$\begin{aligned} |x + y|^2 &= \langle x + y, x + y \rangle \\ &= |x|^2 + 2\langle x, y \rangle + |y|^2 \quad \text{by bilinearity} \\ &\leq |x|^2 + 2|x||y| + |y|^2 \quad \text{by Cauchy-Schwarz} \\ &= (|x| + |y|)^2, \end{aligned}$$

proving the inequality. Equality holds exactly when  $\langle x, y \rangle = |x||y|$ , or equivalently when  $|\langle x, y \rangle| = |x||y|$  and  $\langle x, y \rangle \geq 0$ . These hold when one of  $x, y$  is a scalar multiple of the other and the scalar is nonnegative.  $\square$

While the Cauchy-Schwarz inequality says that the size of the product is at most the product of the sizes, the triangle inequality says:

*The size of the sum is at most the sum of the sizes.*

The triangle inequality's name is explained by its geometric interpretation in  $\mathbb{R}^2$ . View  $x$  as an arrow at the origin,  $y$  as an arrow with tail at the head of  $x$ , and  $x + y$  as an arrow at the origin. These three arrows form a triangle, and the assertion is that the lengths of two sides sum to at least the length of the third. (See Figure 2.7.)



**Figure 2.7.** Sides of a triangle

The **full triangle inequality** says that for all  $x, y \in \mathbb{R}^n$ ,

$$||x| - |y|| \leq |x \pm y| \leq |x| + |y|.$$

The proof is Exercise 2.2.7.

A small argument, which can be formalized as induction if one is painstaking, shows that the basic triangle inequality extends from two vectors to any finite number of vectors. For example,

$$|x + y + z| \leq |x + y| + |z| \leq |x| + |y| + |z|.$$

The only obstacle to generalizing the basic triangle inequality in this fashion is notation. The argument can't use the symbol  $n$  to denote the number of vectors, because  $n$  already denotes the dimension of the Euclidean space where we are working; and furthermore, the vectors can't be denoted with subscripts since a subscript denotes a component of an individual vector. Thus, for now we are stuck writing something like

$$|x^{(1)} + \cdots + x^{(k)}| \leq |x^{(1)}| + \cdots + |x^{(k)}| \quad \text{for all } x^{(1)}, \dots, x^{(k)} \in \mathbb{R}^n,$$

or

$$\left| \sum_{i=1}^k x^{(i)} \right| \leq \sum_{i=1}^k |x^{(i)}|, \quad x^{(1)}, \dots, x^{(k)} \in \mathbb{R}^n.$$

As our work with vectors becomes more intrinsic, vector entries will demand less of our attention, and we will be able to denote vectors by subscripts. The notation-change will be implemented in the next section.

For every vector  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ , useful bounds on the modulus  $|x|$  in terms of the scalar absolute values  $|x_i|$  are as follows.

**Proposition 2.2.7 (Size bounds).** *For every  $j \in \{1, \dots, n\}$ ,*

$$|x_j| \leq |x| \leq \sum_{i=1}^n |x_i|.$$

The proof (by quick applications of the Cauchy–Schwarz inequality and the triangle inequality) is Exercise 2.2.8.

The modulus gives rise to a distance function on  $\mathbb{R}^n$  that behaves as distance should. Define

$$d : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$$

by

$$d(x, y) = |y - x|.$$

For example,  $d(e_i, e_j) = \sqrt{2}(1 - \delta_{ij})$ .

**Theorem 2.2.8 (Distance properties).**

- (D1) *Distance is positive:*  $d(x, y) \geq 0$  for all  $x, y \in \mathbb{R}^n$ , and  $d(x, y) = 0$  if and only if  $x = y$ .
- (D2) *Distance is symmetric:*  $d(x, y) = d(y, x)$  for all  $x, y \in \mathbb{R}^n$ .
- (D3) *Triangle inequality:*  $d(x, z) \leq d(x, y) + d(y, z)$  for all  $x, y, z \in \mathbb{R}^n$ .

(D1) and (D2) are clearly desirable as properties of a distance function. Property (D3) says that you can't shorten your trip from  $x$  to  $z$  by making a stop at  $y$ .

*Proof.* Exercise 2.2.9. □

The Cauchy–Schwarz inequality also lets us define the angle between two nonzero vectors in terms of the inner product. If  $x$  and  $y$  are nonzero vectors in  $\mathbb{R}^n$ , define their angle  $\theta_{x,y}$  by the condition

$$\cos \theta_{x,y} = \frac{\langle x, y \rangle}{|x||y|}, \quad 0 \leq \theta_{x,y} \leq \pi. \quad (2.2)$$

The condition is sensible because  $-1 \leq \frac{\langle x, y \rangle}{|x||y|} \leq 1$  by the Cauchy–Schwarz inequality. For example,  $\cos \theta_{(1,0), (1,1)} = 1/\sqrt{2}$ , and so  $\theta_{(1,0), (1,1)} = \pi/4$ . In particular, two nonzero vectors  $x$  and  $y$  are **orthogonal** when  $\langle x, y \rangle = 0$ . Naturally, we would like  $\theta_{x,y}$  to correspond to the usual notion of angle, at least in  $\mathbb{R}^2$ , and indeed it does—see Exercise 2.2.10. For convenience, define *any* two vectors  $x$  and  $y$  to be orthogonal if  $\langle x, y \rangle = 0$ , thus making  $\mathbf{0}$  orthogonal to all vectors.

Rephrasing geometry in terms of intrinsic vector algebra not only extends the geometric notions of length and angle uniformly to any dimension, it also makes some low-dimensional geometry easier. For example, vectors show in a natural way that the three altitudes of every triangle must meet. Let  $x$  and  $y$  denote two sides of the triangle, making the third side  $x - y$  by the *head minus tail* mnemonic. Let  $q$  be the point where the altitudes to  $x$  and  $y$  meet. (See Figure 2.8, which also shows the third altitude.) Thus

$$q - y \perp x \quad \text{and} \quad q - x \perp y.$$

We want to show that  $q$  also lies on the third altitude, i.e., that

$$q \perp x - y.$$

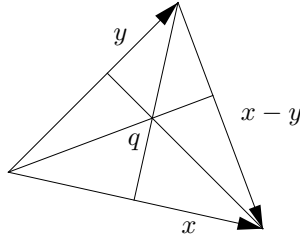
To rephrase matters in terms of inner products, we want to show that

$$\left\{ \begin{array}{l} \langle q - y, x \rangle = 0 \\ \langle q - x, y \rangle = 0 \end{array} \right\} \implies \langle q, x - y \rangle = 0.$$

Since the inner product is linear in each of its arguments, a further rephrasing is that we want to show that

$$\left\{ \begin{array}{l} \langle q, x \rangle = \langle y, x \rangle \\ \langle q, y \rangle = \langle x, y \rangle \end{array} \right\} \implies \langle q, x \rangle = \langle q, y \rangle.$$

And this is immediate because the inner product is symmetric:  $\langle q, x \rangle$  and  $\langle q, y \rangle$  both equal  $\langle x, y \rangle$ , and so they equal each other as desired. The point  $q$  where the three altitudes meet is called the **orthocenter** of the triangle. In general, the orthocenter of a triangle is not the geometric center that we considered in the previous section.



**Figure 2.8.** Three altitudes of a triangle

### Exercises

**2.2.1.** Let  $x = (\frac{\sqrt{3}}{2}, -\frac{1}{2}, 0)$ ,  $y = (\frac{1}{2}, \frac{\sqrt{3}}{2}, 1)$ ,  $z = (1, 1, 1)$ . Compute  $\langle x, x \rangle$ ,  $\langle x, y \rangle$ ,  $\langle y, z \rangle$ ,  $|x|$ ,  $|y|$ ,  $|z|$ ,  $\theta_{x,y}$ ,  $\theta_{y,e_1}$ ,  $\theta_{z,e_2}$ .

**2.2.2.** Show that the points  $x = (2, -1, 3, 1)$ ,  $y = (4, 2, 1, 4)$ ,  $z = (1, 3, 6, 1)$  form the vertices of a triangle in  $\mathbb{R}^4$  with two equal angles.

**2.2.3.** Explain why for all  $x \in \mathbb{R}^n$ ,  $x = \sum_{j=1}^n \langle x, e_j \rangle e_j$ .

**2.2.4.** Prove the inner product properties.

**2.2.5.** Use the inner product properties and the definition of the modulus in terms of the inner product to prove the modulus properties.

**2.2.6.** In the text, the modulus is defined in terms of the inner product. Prove that this can be turned around by showing that for every  $x, y \in \mathbb{R}^n$ ,

$$\langle x, y \rangle = \frac{|x+y|^2 - |x-y|^2}{4}.$$

**2.2.7.** Prove the full triangle inequality: for every  $x, y \in \mathbb{R}^n$ ,

$$||x| - |y|| \leq |x \pm y| \leq |x| + |y|.$$

Do not do this by writing three more variants of the proof of the triangle inequality, but by substituting suitably into the basic triangle inequality, which is already proved.

**2.2.8.** Let  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ . Prove the size bounds: for every  $j \in \{1, \dots, n\}$ ,

$$|x_j| \leq |x| \leq \sum_{i=1}^n |x_i|.$$

(One approach is to start by noting that  $x_j = \langle x, e_j \rangle$  and recalling equation (2.1).) When can each “ $\leq$ ” be an “ $=$ ”?

**2.2.9.** Prove the distance properties.

**2.2.10.** Working in  $\mathbb{R}^2$ , depict the nonzero vectors  $x$  and  $y$  as arrows from the origin and depict  $x - y$  as an arrow from the endpoint of  $y$  to the endpoint of  $x$ . Let  $\theta$  denote the angle (in the usual geometric sense) between  $x$  and  $y$ . Use the law of cosines to show that

$$\cos \theta = \frac{\langle x, y \rangle}{|x||y|},$$

so that our notion of angle agrees with the geometric one, at least in  $\mathbb{R}^2$ .

**2.2.11.** Prove that for every nonzero  $x \in \mathbb{R}^n$ ,  $\sum_{i=1}^n \cos^2 \theta_{x, e_i} = 1$ .

**2.2.12.** Prove that two nonzero vectors  $x, y$  are orthogonal if and only if  $|x + y|^2 = |x|^2 + |y|^2$ .

**2.2.13.** Use vectors in  $\mathbb{R}^2$  to show that the diagonals of a parallelogram are perpendicular if and only if the parallelogram is a rhombus.

**2.2.14.** Use vectors to show that every angle inscribed in a semicircle is right.

**2.2.15.** Let  $x$  and  $y$  be vectors, with  $x$  nonzero. Define the **parallel** component of  $y$  along  $x$  and the **normal** component of  $y$  to  $x$  to be

$$y_{(\parallel x)} = \frac{\langle x, y \rangle}{|x|^2} x \quad \text{and} \quad y_{(\perp x)} = y - y_{(\parallel x)}.$$

(a) Show that  $y = y_{(\parallel x)} + y_{(\perp x)}$ ; show that  $y_{(\parallel x)}$  is a scalar multiple of  $x$ ; show that  $y_{(\perp x)}$  is orthogonal to  $x$ . Show that the decomposition of  $y$  as a sum of vectors parallel and perpendicular to  $x$  is unique. Draw an illustration.

(b) Show that

$$|y|^2 = |y_{(\parallel x)}|^2 + |y_{(\perp x)}|^2.$$

What theorem from classical geometry does this encompass?

(c) Explain why it follows from (b) that

$$|y_{(\parallel x)}| \leq |y|,$$

with equality if and only if  $y$  is a scalar multiple of  $x$ . Use this inequality to give another proof of the Cauchy–Schwarz inequality. This argument gives the geometric content of Cauchy–Schwarz: *the parallel component of one vector along another is at most as long as the original vector*.

(d) The proof of the Cauchy–Schwarz inequality in part (c) refers to parts (a) and (b), part (a) refers to orthogonality, orthogonality refers to an angle, and as explained in the text, the fact that angles make sense depends on the Cauchy–Schwarz inequality. And so the proof in part (c) apparently relies on circular logic. Explain why the logic is in fact not circular.

**2.2.16.** Given nonzero vectors  $x_1, x_2, \dots, x_n$  in  $\mathbb{R}^n$ , the **Gram–Schmidt process** is to set

$$\begin{aligned} x'_1 &= x_1 \\ x'_2 &= x_2 - (x_2)_{(\|x'_1\|)} \\ x'_3 &= x_3 - (x_3)_{(\|x'_2\|)} - (x_3)_{(\|x'_1\|)} \\ &\vdots \\ x'_n &= x_n - (x_n)_{(\|x'_{n-1}\|)} - \cdots - (x_n)_{(\|x'_1\|)}. \end{aligned}$$

(a) What is the result of applying the Gram–Schmidt process to the vectors  $x_1 = (1, 0, 0)$ ,  $x_2 = (1, 1, 0)$ , and  $x_3 = (1, 1, 1)$ ?

(b) Returning to the general case, show that  $x'_1, \dots, x'_n$  are pairwise orthogonal and that each  $x'_j$  has the form

$$x'_j = a_{j1}x_1 + a_{j2}x_2 + \cdots + a_{j,j-1}x_{j-1} + x_j.$$

Thus every linear combination of the new  $\{x'_j\}$  is also a linear combination of the original  $\{x_j\}$ . The converse is also true and will be shown in Exercise 3.3.13.

## 2.3 Analysis: Continuous Mappings

A **mapping** from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  is some rule that assigns to each point  $x$  in  $\mathbb{R}^n$  a point in  $\mathbb{R}^m$ . Generally, mappings will be denoted by letters such as  $f$ ,  $g$ ,  $h$ . When  $m = 1$ , we usually say *function* instead of mapping.

For example, the mapping

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

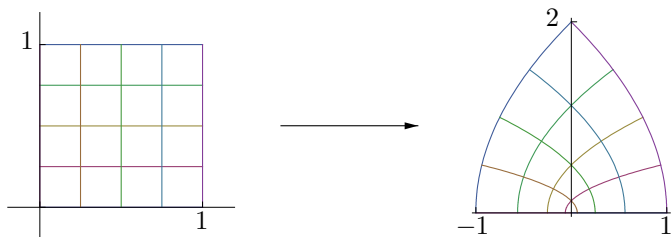
defined by

$$f(x, y) = (x^2 - y^2, 2xy)$$

takes the real and imaginary parts of a complex number  $z = x + iy$  and returns the real and imaginary parts of  $z^2$ . By the nature of multiplication of complex numbers, this means that each output point has modulus equal to the square of the modulus of the input point and has angle equal to twice the angle of the input point. Make sure that you see how this is shown in Figure 2.9.

Mappings expressed by formulas may be undefined at certain points (e.g.,  $f(x) = 1/|x|$  is undefined at  $\mathbf{0}$ ), so we need to restrict their domains. For a given dimension  $n$ , a given set  $A \subset \mathbb{R}^n$ , and a second dimension  $m$ , let  $\mathcal{M}(A, \mathbb{R}^m)$  denote the set of all mappings  $f : A \longrightarrow \mathbb{R}^m$ . This set forms a vector space over  $\mathbb{R}$  (whose *points* are functions) under the operations

$$+ : \mathcal{M}(A, \mathbb{R}^m) \times \mathcal{M}(A, \mathbb{R}^m) \longrightarrow \mathcal{M}(A, \mathbb{R}^m),$$



**Figure 2.9.** The complex square as a mapping from  $\mathbb{R}^2$  to  $\mathbb{R}^2$

defined by

$$(f + g)(x) = f(x) + g(x) \quad \text{for all } x \in A,$$

and

$$\cdot : \mathbb{R} \times \mathcal{M}(A, \mathbb{R}^m) \longrightarrow \mathcal{M}(A, \mathbb{R}^m),$$

defined by

$$(a \cdot f)(x) = a \cdot f(x) \quad \text{for all } x \in A.$$

As usual, “+” and “ $\cdot$ ” are overloaded: on the left they denote operations on  $\mathcal{M}(A, \mathbb{R}^m)$ , while on the right they denote the operations on  $\mathbb{R}^m$  defined in Section 2.1. Also as usual, the “ $\cdot$ ” is generally omitted. The origin in  $\mathcal{M}(A, \mathbb{R}^m)$  is the **zero mapping**,  $\mathbf{0} : A \longrightarrow \mathbb{R}^m$ , defined by

$$\mathbf{0}(x) = \mathbf{0}_m \quad \text{for all } x \in A.$$

For example, to verify that  $\mathcal{M}(A, \mathbb{R}^m)$  satisfies (A1), consider any mappings  $f, g, h \in \mathcal{M}(A, \mathbb{R}^m)$ . For every  $x \in A$ ,

$$\begin{aligned} ((f + g) + h)(x) &= (f + g)(x) + h(x) && \text{by definition of “+” in } \mathcal{M}(A, \mathbb{R}^m) \\ &= (f(x) + g(x)) + h(x) && \text{by definition of “+” in } \mathcal{M}(A, \mathbb{R}^m) \\ &= f(x) + (g(x) + h(x)) && \text{by associativity of “+” in } \mathbb{R}^m \\ &= f(x) + (g + h)(x) && \text{by definition of “+” in } \mathcal{M}(A, \mathbb{R}^m) \\ &= (f + (g + h))(x) && \text{by definition of “+” in } \mathcal{M}(A, \mathbb{R}^m). \end{aligned}$$

Since  $x$  is arbitrary,  $(f + g) + h = f + (g + h)$ .

Let  $A$  be a subset of  $\mathbb{R}^n$ . A **sequence in  $A$**  is an infinite list of vectors  $\{x_1, x_2, x_3, \dots\}$  in  $A$ , often written  $\{x_\nu\}$ . (The symbol  $n$  is already in use, so its Greek counterpart  $\nu$ —pronounced nu—is used as the index-counter.) Since a vector has  $n$  entries, each vector  $x_\nu$  in the sequence takes the form  $(x_{1,\nu}, \dots, x_{n,\nu})$ .

**Definition 2.3.1 (Null Sequence).** *The sequence  $\{x_\nu\}$  in  $\mathbb{R}^n$  is **null** if for every  $\varepsilon > 0$  there exists some  $\nu_0$  such that*

if  $\nu > \nu_0$  then  $|x_\nu| < \varepsilon$ .

That is, a sequence is null if for every  $\varepsilon > 0$ , all but finitely many terms of the sequence lie within distance  $\varepsilon$  of  $\mathbf{0}_n$ .

Quickly from the definition, if  $\{x_\nu\}$  is a null sequence in  $\mathbb{R}^n$  and  $\{y_\nu\}$  is a sequence in  $\mathbb{R}^n$  such that  $|y_\nu| \leq |x_\nu|$  for all  $\nu$  then also  $\{y_\nu\}$  is null.

Let  $\{x_\nu\}$  and  $\{y_\nu\}$  be null sequences in  $\mathbb{R}^n$ , and let  $c$  be a scalar. Then the sequence  $\{x_\nu + y_\nu\}$  is null because  $|x_\nu + y_\nu| \leq |x_\nu| + |y_\nu|$  for each  $\nu$ , and the sequence  $\{cx_\nu\}$  is null because  $|cx_\nu| = |c||x_\nu|$  for each  $\nu$ . These two results show that the set of null sequences in  $\mathbb{R}^n$  forms a vector space.

For every vector  $x \in \mathbb{R}^n$  the absolute value  $|x|$  is a nonnegative scalar, and so no further effect is produced by taking the scalar absolute value in turn,

$$||x|| = |x|, \quad x \in \mathbb{R}^n,$$

and so a vector sequence  $\{x_\nu\}$  is null if and only if the scalar sequence  $\{|x_\nu|\}$  is null.

**Lemma 2.3.2 (Componentwise nature of nullness).** *The vector sequence  $\{(x_{1,\nu}, \dots, x_{n,\nu})\}$  is null if and only if each of its component scalar sequences  $\{x_{j,\nu}\}$  ( $j \in \{1, \dots, n\}$ ) is null.*

*Proof.* By the observation just before the lemma, it suffices to show that  $\{|(x_{1,\nu}, \dots, x_{n,\nu})|\}$  is null if and only if each  $\{|x_{j,\nu}|\}$  is null. The size bounds give for every  $j \in \{1, \dots, n\}$  and every  $\nu$ ,

$$|x_{j,\nu}| \leq |(x_{1,\nu}, \dots, x_{n,\nu})| \leq \sum_{i=1}^n |x_{i,\nu}|.$$

If  $\{|(x_{1,\nu}, \dots, x_{n,\nu})|\}$  is null then by the first inequality, so is each  $\{|x_{j,\nu}|\}$ . On the other hand, if each  $\{|x_{j,\nu}|\}$  is null then so is  $\{\sum_{i=1}^n |x_{i,\nu}|\}$ , and thus by the second inequality,  $\{|(x_{1,\nu}, \dots, x_{n,\nu})|\}$  is null as well.  $\square$

We define the convergence of vector sequences in terms of null sequences.

**Definition 2.3.3 (Sequence convergence, sequence limit).** *Let  $A$  be a subset of  $\mathbb{R}^n$ . Consider a sequence  $\{x_\nu\}$  in  $A$  and a point  $p \in \mathbb{R}^n$ . The sequence  $\{x_\nu\}$  **converges to  $p$**  (or **has limit  $p$** ), written  $\{x_\nu\} \rightarrow p$ , if the sequence  $\{x_\nu - p\}$  is null. When the limit  $p$  is a point of  $A$ , the sequence  $\{x_\nu\}$  **converges in  $A$** .*

If a sequence  $\{x_\nu\}$  converges to  $p$  and also converges to  $p'$  then the constant sequence  $\{p' - p\}$  is the difference of the null sequences  $\{x_\nu - p\}$  and  $\{x_\nu - p'\}$ , hence null, forcing  $p' = p$ . Thus a sequence cannot converge to two distinct values.

Many texts define convergence directly rather than by reference to nullness, the key part of the definition being

if  $\nu > \nu_0$  then  $|x_\nu - p| < \varepsilon$ .

In particular, a null sequence is a sequence that converges to  $\mathbf{0}_n$ . However, in contrast to the situation for null sequences, for  $p \neq \mathbf{0}_n$  it is emphatically false that if  $\{|x_\nu|\}$  converges to  $|p|$  then necessarily  $\{x_\nu\}$  converges to  $p$  or even converges at all. Also, for every nonzero  $p$ , the sequences that converge to  $p$  do not form a vector space.

Vector versions of the sum rule and the constant multiple rule for convergent sequences follow immediately from the vector space properties of null sequences:

**Proposition 2.3.4 (Linearity of convergence).** *Let  $\{x_\nu\}$  be a sequence in  $\mathbb{R}^n$  converging to  $p$ , let  $\{y_\nu\}$  be a sequence in  $\mathbb{R}^n$  converging to  $q$ , and let  $c$  be a scalar. Then the sequence  $\{x_\nu + y_\nu\}$  converges to  $p + q$ , and the sequence  $\{cx_\nu\}$  converges to  $cp$ .*

Similarly, since a sequence  $\{x_\nu\}$  converges to  $p$  if and only if  $\{x_\nu - p\}$  is null, we have the following corollary in consequence of the componentwise nature of nullness (Exercise 2.3.5):

**Proposition 2.3.5 (Componentwise nature of convergence).** *The vector sequence  $\{(x_{1,\nu}, \dots, x_{n,\nu})\}$  converges to the vector  $(p_1, \dots, p_n)$  if and only if each component scalar sequence  $\{x_{j,\nu}\}$  ( $j = 1, \dots, n$ ) converges to the scalar  $p_j$ .*

Continuity, like convergence, is typographically indistinguishable in  $\mathbb{R}$  and  $\mathbb{R}^n$ .

**Definition 2.3.6 (Continuity).** *Let  $A$  be a subset of  $\mathbb{R}^n$ , let  $f : A \rightarrow \mathbb{R}^m$  be a mapping, and let  $p$  be a point of  $A$ . Then  $f$  is **continuous at  $p$**  if for every sequence  $\{x_\nu\}$  in  $A$  converging to  $p$ , the sequence  $\{f(x_\nu)\}$  converges to  $f(p)$ . The mapping  $f$  is **continuous on  $A$**  (or just **continuous** when  $A$  is clearly established) if it is continuous at each point  $p \in A$ .*

For example, the modulus function

$$| \cdot | : \mathbb{R}^n \rightarrow \mathbb{R}$$

is continuous on  $\mathbb{R}^n$ . To see this, consider any point  $p \in \mathbb{R}^n$  and consider any sequence  $\{x_\nu\}$  in  $\mathbb{R}^n$  that converges to  $p$ . We need to show that the sequence  $\{|x_\nu|\}$  in  $\mathbb{R}$  converges to  $|p|$ . But by the full triangle inequality,

$$||x_\nu| - |p|| \leq |x_\nu - p|.$$

Since the right side is the  $\nu$ th term of a null sequence, so is the left, giving the result.

For another example, let  $a \in \mathbb{R}^n$  be any fixed vector and consider the function defined by taking the inner product of this vector with other vectors,

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad T(x) = \langle a, x \rangle.$$

This function is also continuous on  $\mathbb{R}^n$ . To see this, again consider any  $p \in \mathbb{R}^n$  and any sequence  $\{x_\nu\}$  in  $\mathbb{R}^n$  converging to  $p$ . Then the definition of  $T$ , the bilinearity of the inner product, and the Cauchy–Schwarz inequality combine to show that

$$|T(x_\nu) - T(p)| = |\langle a, x_\nu \rangle - \langle a, p \rangle| = |\langle a, x_\nu - p \rangle| \leq |a| |x_\nu - p|.$$

Since  $|a|$  is a constant, the right side is the  $\nu$ th term of a null sequence, whence so is the left, and the proof is complete. We will refer to this example in Section 3.1. Also, note that as a special case of this example we may take any  $j \in \{1, \dots, n\}$  and set the fixed vector  $a$  to  $e_j$ , showing that the  **$j$ th coordinate function map**,

$$\pi_j : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad \pi_j(x_1, \dots, x_n) = x_j,$$

is continuous.

**Proposition 2.3.7 (Vector space properties of continuity).** *Let  $A$  be a subset of  $\mathbb{R}^n$ , let  $f, g : A \longrightarrow \mathbb{R}^m$  be continuous mappings, and let  $c \in \mathbb{R}$ . Then the sum and the scalar multiple mappings*

$$f + g, \quad cf : A \longrightarrow \mathbb{R}^m$$

*are continuous. Thus the set of continuous mappings from  $A$  to  $\mathbb{R}^m$  forms a vector subspace of  $\mathcal{M}(A, \mathbb{R}^m)$ .*

The vector space properties of continuity follow immediately from the linearity of convergence and from the definition of continuity. Another consequence of the definition of continuity is as follows.

**Proposition 2.3.8 (Persistence of continuity under composition).** *Let  $A$  be a subset of  $\mathbb{R}^n$ , and let  $f : A \longrightarrow \mathbb{R}^m$  be a continuous mapping. Let  $B$  be a superset of  $f(A)$  in  $\mathbb{R}^m$ , and let  $g : B \longrightarrow \mathbb{R}^\ell$  be a continuous mapping. Then the composition mapping*

$$g \circ f : A \longrightarrow \mathbb{R}^\ell$$

*is continuous.*

The proof is Exercise 2.3.7.

Let  $A$  be a subset of  $\mathbb{R}^n$ . Every mapping  $f : A \longrightarrow \mathbb{R}^m$  decomposes as  $m$  functions  $f_1, \dots, f_m$ , with each  $f_i : A \longrightarrow \mathbb{R}$ , by the formula

$$f(x) = (f_1(x), \dots, f_m(x)).$$

For example, if  $f(x, y) = (x^2 - y^2, 2xy)$  then  $f_1(x, y) = x^2 - y^2$  and  $f_2(x, y) = 2xy$ . The decomposition of  $f$  can also be written

$$f(x) = \sum_{i=1}^m f_i(x)e_i,$$

or equivalently, the functions  $f_i$  are defined by the condition

$$f_i(x) = f(x)_i \quad \text{for } i = 1, \dots, m.$$

Conversely, given  $m$  functions  $f_1, \dots, f_m$  from  $A$  to  $\mathbb{R}$ , each of the preceding three displayed formulas assembles a mapping  $f : A \rightarrow \mathbb{R}^m$ . Thus, each mapping  $f$  determines and is determined by its **component functions**  $f_1, \dots, f_m$ . Conveniently, to check continuity of the vector-valued mapping  $f$  we only need to check its scalar-valued component functions.

**Theorem 2.3.9 (Componentwise nature of continuity).** *Let  $A \subset \mathbb{R}^n$ , let  $f : A \rightarrow \mathbb{R}^m$  have component functions  $f_1, \dots, f_m$ , and let  $p$  be a point in  $A$ . Then*

$$f \text{ is continuous at } p \iff \text{each } f_i \text{ is continuous at } p.$$

The componentwise nature of continuity follows from the componentwise nature of convergence and is left as Exercise 2.3.6.

Let  $A$  be a subset of  $\mathbb{R}^n$ , let  $f$  and  $g$  be continuous functions from  $A$  to  $\mathbb{R}$ , and let  $c \in \mathbb{R}$ . Then the familiar sum rule, constant multiple rule, product rule, and quotient rule for continuous functions hold. That is, the sum  $f + g$ , the constant multiple  $cf$ , the product  $fg$ , and the quotient  $f/g$  (at points  $p \in A$  such that  $g(p) \neq 0$ ) are again continuous. The first two of these facts are special cases of the vector space properties of continuity. The proofs of the other two are typographically identical to their one-variable counterparts. With the various continuity results obtained thus far in hand, it is clear that a function such as

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}, \quad f(x, y, z) = \frac{\sin(\sqrt{x^2 + y^2 + z^2})}{e^{xy+z}}$$

is continuous. The continuity of such functions, and of mappings with such functions as their components, will go without comment from now on.

However, the continuity of functions of  $n$  variables also has new, subtle features when  $n > 1$ . In  $\mathbb{R}$ , a sequence  $\{x_\nu\}$  can approach the point  $p$  in only two essential ways: from the left and from the right. But in  $\mathbb{R}^n$  for  $n \geq 2$ ,  $\{x_\nu\}$  can approach  $p$  along a line from infinitely many directions, or not approach along a line at all, and so the convergence of  $\{f(x_\nu)\}$  can be trickier. For example, consider the function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  defined by

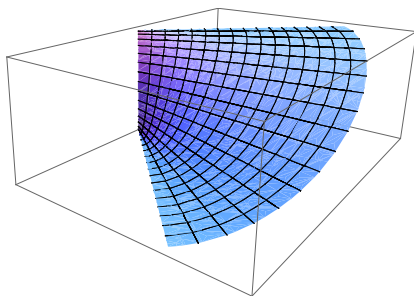
$$f(x, y) = \begin{cases} \frac{2xy}{x^2 + y^2} & \text{if } (x, y) \neq \mathbf{0}, \\ b & \text{if } (x, y) = \mathbf{0}. \end{cases}$$

Can the constant  $b$  be specified to make  $f$  continuous at  $\mathbf{0}$ ?

It can't. Take a sequence  $\{(x_\nu, y_\nu)\}$  approaching  $\mathbf{0}$  along the line  $y = mx$  of slope  $m$ . For every point  $(x_\nu, y_\nu)$  of this sequence,

$$f(x_\nu, y_\nu) = f(x_\nu, mx_\nu) = \frac{2x_\nu mx_\nu}{x_\nu^2 + m^2 x_\nu^2} = \frac{2mx_\nu^2}{(1 + m^2)x_\nu^2} = \frac{2m}{1 + m^2}.$$

Thus, as the sequence of inputs  $\{(x_\nu, y_\nu)\}$  approaches  $\mathbf{0}$  along the line of slope  $m$ , the corresponding sequence of outputs  $\{f(x_\nu, y_\nu)\}$  holds steady at  $2m/(1 + m^2)$ , and so  $f(\mathbf{0})$  needs to take this value for continuity. Taking input sequences  $\{(x_\nu, y_\nu)\}$  that approach  $\mathbf{0}$  along lines of different slope shows that  $f(\mathbf{0})$  needs to take different values for continuity, and hence  $f$  cannot be made continuous at  $\mathbf{0}$ . The graph of  $f$  away from  $\mathbf{0}$  is a sort of spiral staircase, and no height over  $\mathbf{0}$  is compatible with all the stairs. (See Figure 2.10. The figure displays only the portion of the graph for slopes between 0 and 1 in the input plane.) The reader who wants to work a virtually identical example can replace the formula  $2xy/(x^2 + y^2)$  in  $f$  by  $(x^2 - y^2)/(x^2 + y^2)$  and run the same procedure as in this paragraph.



**Figure 2.10.** A spiral staircase

The previous example was actually fairly simple in that we only needed to study  $f(x, y)$  as  $(x, y)$  approached  $\mathbf{0}$  along straight lines. Consider the function  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  defined by

$$g(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & \text{if } (x, y) \neq \mathbf{0}, \\ b & \text{if } (x, y) = \mathbf{0}. \end{cases}$$

For a nonzero slope  $m$ , take a sequence  $\{(x_\nu, y_\nu)\}$  approaching  $\mathbf{0}$  along the line  $y = mx$ . Compute that for each point of this sequence,

$$g(x_\nu, y_\nu) = g(x_\nu, mx_\nu) = \frac{mx_\nu^3}{x_\nu^4 + m^2 x_\nu^2} = \frac{mx_\nu}{x_\nu^2 + m^2}.$$

This quantity tends to 0 as  $x_\nu$  goes to 0. That is, as the sequence of inputs  $\{(x_\nu, y_\nu)\}$  approaches  $\mathbf{0}$  along the line of slope  $m$ , the corresponding sequence of outputs  $\{g(x_\nu, y_\nu)\}$  approaches 0, and so  $g(\mathbf{0})$  needs to take the value 0 for continuity. Since  $g$  is 0 at the nonzero points of either axis in the  $(x, y)$ -plane, this requirement extends to the cases that  $\{(x_\nu, y_\nu)\}$  approaches  $\mathbf{0}$  along a horizontal or vertical line. However, next consider a sequence  $\{(x_\nu, y_\nu)\}$  approaching  $\mathbf{0}$  along the parabola  $y = x^2$ . For each point of this sequence,

$$g(x_\nu, y_\nu) = g(x_\nu, x_\nu^2) = \frac{x_\nu^4}{x_\nu^4 + x_\nu^4} = \frac{1}{2}.$$

Thus, as the sequence of inputs  $\{(x_\nu, y_\nu)\}$  approaches  $\mathbf{0}$  along the parabola, the corresponding sequence of outputs  $\{g(x_\nu, y_\nu)\}$  holds steady at  $1/2$ , and so  $g(\mathbf{0})$  needs to be  $1/2$  for continuity as well. Thus  $g$  cannot be made continuous at  $\mathbf{0}$ , even though approaching  $\mathbf{0}$  only along lines suggests that it can. The reader who wants to work a virtually identical example can replace the formula  $x^2y/(x^4 + y^2)$  in  $g$  by  $x^3y/(x^6 + y^2)$  and run the same procedure as in this paragraph but using the curve  $y = x^3$ .

Thus, given a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , letting  $\{(x_\nu, y_\nu)\}$  approach  $\mathbf{0}$  along lines can disprove continuity at  $\mathbf{0}$ , but it can only suggest continuity at  $\mathbf{0}$ , not prove it. To prove continuity, the size bounds may be helpful. For example, let

$$h(x, y) = \begin{cases} \frac{x^3}{x^2 + y^2} & \text{if } (x, y) \neq \mathbf{0}, \\ b & \text{if } (x, y) = \mathbf{0}. \end{cases}$$

Can  $b$  be specified to make  $h$  continuous at  $\mathbf{0}$ ? The estimate  $|x| \leq |(x, y)|$  gives for every  $(x, y) \neq \mathbf{0}$ ,

$$0 \leq |h(x, y)| = \frac{|x^3|}{x^2 + y^2} = \frac{|x|^3}{|(x, y)|^2} \leq \frac{|(x, y)|^3}{|(x, y)|^2} = |(x, y)|,$$

so as a sequence  $\{(x_\nu, y_\nu)\}$  of nonzero input vectors converges to  $\mathbf{0}$ , the corresponding sequence of outputs  $\{h(x_\nu, y_\nu)\}$  is squeezed to 0 in absolute value and hence converges to 0. Setting  $b = 0$  makes  $h$  continuous at  $\mathbf{0}$ . The reader who wants to work a virtually identical example can replace the formula  $x^3/(x^2 + y^2)$  in  $h$  by  $x^2y^2/(x^4 + y^2)$  and run the same procedure as in this paragraph but applying the size bounds to vectors  $(x_\nu^2, y_\nu)$ .

Returning to the spiral staircase example,

$$f(x, y) = \begin{cases} \frac{2xy}{x^2 + y^2} & \text{if } (x, y) \neq \mathbf{0}, \\ b & \text{if } (x, y) = \mathbf{0}, \end{cases}$$

the size bounds show that that for every  $(x, y) \neq \mathbf{0}$ ,

$$0 \leq |f(x, y)| = \frac{2|x||y|}{|(x, y)|^2} \leq \frac{2|(x, y)|^2}{|(x, y)|^2} = 2.$$

The display tells us only that as a sequence of inputs  $\{(x_\nu, y_\nu)\}$  approaches  $\mathbf{0}$ , the sequence of outputs  $\{f(x_\nu, y_\nu)\}$  might converge to some limit between  $-2$  and  $2$ . The outputs needn't converge to  $0$  (or converge at all), but according to this diagnostic they possibly could. Thus the size bounds tell us only that  $f$  *could* be discontinuous at  $(0, 0)$ , but they give no conclusive information.

In sum, these examples illustrate three ideas.

- The straight line test can prove that a limit does not exist, or it can determine the only candidate for the value of the limit, but it cannot prove that the candidate value is the limit.
- When the straight line test determines a candidate value of the limit, approaching along a curve can further support the candidate, or it can prove that the limit does not exist by determining a different candidate as well.
- The size bounds can prove that a limit does exist, but they can only suggest that a limit does not exist.

The next proposition is a handy encoding of an intuitively plausible property of continuous mappings. The result is so natural that it often is tacitly taken for granted, but it is worth stating and proving carefully.

**Proposition 2.3.10 (Persistence of inequality).** *Let  $A$  be a subset of  $\mathbb{R}^n$  and let  $f : A \rightarrow \mathbb{R}^m$  be a continuous mapping. Let  $p$  be a point of  $A$ , let  $b$  be a point of  $\mathbb{R}^m$ , and suppose that  $f(p) \neq b$ . Then there exists some  $\varepsilon > 0$  such that*

$$\text{for all } x \in A \text{ such that } |x - p| < \varepsilon, f(x) \neq b.$$

*Proof.* Assume that the displayed statement in the proposition fails for every  $\varepsilon > 0$ . Then in particular, it fails for  $\varepsilon = 1/\nu$  for  $\nu = 1, 2, 3, \dots$ . So there is a sequence  $\{x_\nu\}$  in  $A$  such that

$$|x_\nu - p| < 1/\nu \quad \text{and} \quad f(x_\nu) = b, \quad \nu = 1, 2, 3, \dots$$

Since  $f$  is continuous at  $p$ , this condition shows that  $f(p) = b$ . But in fact  $f(p) \neq b$ , and so our assumption that the displayed statement in the proposition fails for every  $\varepsilon > 0$  leads to a contradiction. Therefore the statement holds for some  $\varepsilon > 0$ , as desired.  $\square$

## Exercises

**2.3.1.** For  $A \subset \mathbb{R}^n$ , partially verify that  $\mathcal{M}(A, \mathbb{R}^m)$  is a vector space over  $\mathbb{R}$  by showing that it satisfies vector space axioms (A4) and (D1).

**2.3.2.** Define multiplication  $*$  :  $\mathcal{M}(A, \mathbb{R}) \times \mathcal{M}(A, \mathbb{R}) \rightarrow \mathcal{M}(A, \mathbb{R})$ . Is  $\mathcal{M}(A, \mathbb{R})$  a field with “+” from the section and this multiplication? Does it have a sub-space that is a field?

**2.3.3.** For  $A \subset \mathbb{R}^n$  and  $m \in \mathbb{Z}^+$  define a subspace of the space of mappings from  $A$  to  $\mathbb{R}^m$ ,

$$\mathcal{C}(A, \mathbb{R}^m) = \{f \in \mathcal{M}(A, \mathbb{R}^m) : f \text{ is continuous on } A\}.$$

Briefly explain how this section has shown that  $\mathcal{C}(A, \mathbb{R}^m)$  is a vector space.

**2.3.4.** Define an inner product and a modulus on  $\mathcal{C}([0, 1], \mathbb{R})$  by

$$\langle f, g \rangle = \int_0^1 f(t)g(t) dt, \quad |f| = \sqrt{\langle f, f \rangle}.$$

Do the inner product properties (IP1), (IP2), and (IP3) (see Proposition 2.2.2) hold for this inner product on  $\mathcal{C}([0, 1], \mathbb{R})$ ? How much of the material from Section 2.2 on the inner product and modulus in  $\mathbb{R}^n$  carries over to  $\mathcal{C}([0, 1], \mathbb{R})$ ? Express the Cauchy-Schwarz inequality as a relation between integrals.

**2.3.5.** Use the definition of convergence and the componentwise nature of nullness to prove the componentwise nature of convergence. (The argument is short.)

**2.3.6.** Use the definition of continuity and the componentwise nature of convergence to prove the componentwise nature of continuity.

**2.3.7.** Prove the persistence of continuity under composition.

**2.3.8.** Define  $f : \mathbb{Q} \rightarrow \mathbb{R}$  by the rule

$$f(x) = \begin{cases} 1 & \text{if } x^2 < 2, \\ 0 & \text{if } x^2 > 2. \end{cases}$$

Is  $f$  continuous?

**2.3.9.** Which of the following functions on  $\mathbb{R}^2$  can be defined continuously at  $\mathbf{0}$ ?

$$\begin{aligned} f(x, y) &= \begin{cases} \frac{x^4 - y^4}{(x^2 + y^2)^2} & \text{if } (x, y) \neq \mathbf{0}, \\ b & \text{if } (x, y) = \mathbf{0}, \end{cases} & g(x, y) &= \begin{cases} \frac{x^2 - y^3}{x^2 + y^2} & \text{if } (x, y) \neq \mathbf{0}, \\ b & \text{if } (x, y) = \mathbf{0}, \end{cases} \\ h(x, y) &= \begin{cases} \frac{x^3 - y^3}{x^2 + y^2} & \text{if } (x, y) \neq \mathbf{0}, \\ b & \text{if } (x, y) = \mathbf{0}, \end{cases} & k(x, y) &= \begin{cases} \frac{xy^2}{x^2 + y^6} & \text{if } (x, y) \neq \mathbf{0}, \\ b & \text{if } (x, y) = \mathbf{0}. \end{cases} \end{aligned}$$

**2.3.10.** Let  $f(x, y) = g(xy)$ , where  $g : \mathbb{R} \rightarrow \mathbb{R}$  is continuous. Is  $f$  continuous?

**2.3.11.** Let  $f, g \in \mathcal{M}(\mathbb{R}^n, \mathbb{R})$  be such that  $f + g$  and  $fg$  are continuous. Are  $f$  and  $g$  necessarily continuous?

## 2.4 Topology: Compact Sets and Continuity

The extreme value theorem from one-variable calculus states:

*Let  $I$  be a nonempty closed and bounded interval in  $\mathbb{R}$ , and let  $f : I \rightarrow \mathbb{R}$  be a continuous function. Then  $f$  takes a minimum value and a maximum value on  $I$ .*

This section generalizes the theorem from scalars to vectors. That is, we want a result that if  $A$  is a set in  $\mathbb{R}^n$  with certain properties, and if  $f : A \rightarrow \mathbb{R}^m$  is a continuous mapping, then the output set  $f(A)$  will also have certain properties. The questions are, for what sorts of properties do such statements hold, and when they hold, how do we prove them?

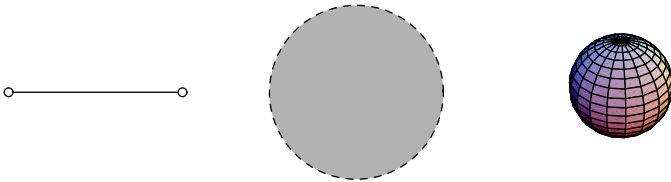
The one-variable theorem hypothesizes two data, the nonempty closed and bounded interval  $I$  and the continuous function  $f$ . Each of these is described in its own terms— $I$  takes the readily recognizable but static form  $[a, b]$  where  $a \leq b$ , while the continuity of  $f$  is a dynamic assertion about convergence of sequences. Because the two data have differently phrased descriptions, a proof of the extreme value theorem doesn't suggest itself immediately: no ideas at hand bear obviously on all the given information. Thus the work of this section is not only to define the sets to appear in the pending theorem, but also to describe them in terms of sequences, compatibly with the sequential description of continuous mappings. The theorem itself will then be easy to prove. Accordingly, most of the section will be spent describing sets in two ways—in terms that are easy to recognize, and in sequential language that dovetails with continuity.

We begin with a little machinery to quantify the intuitive notion of nearness.

**Definition 2.4.1 ( $\varepsilon$ -ball).** *For every point  $p \in \mathbb{R}^n$  and every positive real number  $\varepsilon > 0$ , the  $\varepsilon$ -ball centered at  $p$  is the set*

$$B(p, \varepsilon) = \{x \in \mathbb{R}^n : |x - p| < \varepsilon\}.$$

(See Figure 2.11.)



**Figure 2.11.** Balls in various dimensions

With  $\varepsilon$ -balls it is easy to describe the points that are approached by a set  $A$ .

**Definition 2.4.2 (Limit point).** Let  $A$  be a subset of  $\mathbb{R}^n$ , and let  $p$  be a point of  $\mathbb{R}^n$ . The point  $p$  is a **limit point** of  $A$  if every  $\varepsilon$ -ball centered at  $p$  contains some point  $x \in A$  such that  $x \neq p$ .

A limit point of  $A$  need not belong to  $A$  (Exercise 2.4.2). On the other hand, a point in  $A$  need not be a limit point of  $A$  (Exercise 2.4.2 again); such a point is called an **isolated point** of  $A$ . Equivalently,  $p$  is an isolated point of  $A$  if  $p \in A$  and there exists some  $\varepsilon > 0$  such that  $B(p, \varepsilon) \cap A = \{p\}$ . The next lemma justifies the nomenclature of the previous definition: limit points of  $A$  are precisely the (nontrivial) limits of sequences in  $A$ .

**Lemma 2.4.3 (Sequential characterization of limit points).** Let  $A$  be a subset of  $\mathbb{R}^n$ , and let  $p$  be a point of  $\mathbb{R}^n$ . Then  $p$  is the limit of a sequence  $\{x_\nu\}$  in  $A$  with each  $x_\nu \neq p$  if and only if  $p$  is a limit point of  $A$ .

*Proof.* (  $\implies$  ) If  $p$  is the limit of a sequence  $\{x_\nu\}$  in  $A$  with each  $x_\nu \neq p$  then every  $\varepsilon$ -ball about  $p$  contains an  $x_\nu$  (in fact, infinitely many), so  $p$  is a limit point of  $A$ .

(  $\impliedby$  ) Conversely, if  $p$  is a limit point of  $A$  then  $B(p, 1/2)$  contains some  $x_1 \in A$ ,  $x_1 \neq p$ . Let  $\varepsilon_2 = |x_1 - p|/2$ . The ball  $B(p, \varepsilon_2)$  contains some  $x_2 \in A$ ,  $x_2 \neq p$ . Let  $\varepsilon_3 = |x_2 - p|/2$  and continue defining a sequence  $\{x_\nu\}$  in this fashion with  $|x_\nu - p| < 1/2^\nu$  for all  $\nu$ . This sequence converges to  $p$ , and  $x_\nu \neq p$  for each  $x_\nu$ .  $\square$

The lemma shows that Definition 2.4.2 is more powerful than it appears—every  $\varepsilon$ -ball centered at a limit point of  $A$  contains not only one but *infinitely many* points of  $A$ .

**Definition 2.4.4 (Closed set).** A subset  $A$  of  $\mathbb{R}^n$  is **closed** if it contains all of its limit points.

For example, the  $x_1$ -axis is closed as a subset of  $\mathbb{R}^n$ , because every point off the axis is surrounded by a ball that misses the axis—that is, every point off the axis is not a limit point of the axis, i.e., the axis is not missing any of its limit points, i.e., the axis contains all of its limit points. The interval  $(0, 1)$  is not closed because it does not contain the limit points at its ends. These examples illustrate the fact that with a little practice it becomes easy to recognize quickly whether a set is closed. Loosely speaking, a set is closed when it contains all the points that it seems to want to contain.

**Proposition 2.4.5 (Sequential characterization of closed sets).** Let  $A$  be a subset of  $\mathbb{R}^n$ . Then  $A$  is closed if and only if every sequence in  $A$  that converges in  $\mathbb{R}^n$  in fact converges in  $A$ .

*Proof.* (  $\implies$  ) Suppose that  $A$  is closed, and let  $\{x_\nu\}$  be a sequence in  $A$  converging in  $\mathbb{R}^n$  to  $p$ . If  $x_\nu = p$  for some  $\nu$  then  $p \in A$  because  $x_\nu \in A$ ; and if  $x_\nu \neq p$  for all  $\nu$  then  $p$  is a limit point of  $A$  by “ $\implies$ ” of Lemma 2.4.3, and so  $p \in A$  because  $A$  is closed.

( $\Leftarrow$ ) Conversely, suppose that every convergent sequence in  $A$  has its limit in  $A$ . Then all limit points of  $A$  are in  $A$  by “ $\Leftarrow$ ” of Lemma 2.4.3, and so  $A$  is closed.  $\square$

The proposition equates an easily recognizable condition that we can understand intuitively (a set being closed) with a sequential characterization that we can use in further arguments. Note that the sequential characterization of a closed set  $A$  refers not only to  $A$  but also to the ambient space  $\mathbb{R}^n$  in which  $A$  lies. We will return to this point later in this section.

Closed sets do not necessarily have good properties under continuous mappings. So next we describe another class of sets, the bounded sets. Boundedness is again an easily recognizable condition that also has a characterization in terms of sequences. The sequential characterization will turn out to be complementary to the sequential characterization of closed sets, foreshadowing that the properties of being closed and bounded will work well together.

**Definition 2.4.6 (Bounded set).** A set  $A$  in  $\mathbb{R}^n$  is **bounded** if  $A \subset B(\mathbf{0}, R)$  for some  $R > 0$ .

Thus a bounded set is enclosed in some finite corral centered at the origin, possibly a very big one. For example, every ball  $B(p, \varepsilon)$ , not necessarily centered at the origin, is bounded, by a nice application of the triangle inequality (Exercise 2.4.5). On the other hand, the Archimedean property of the real number system says that  $\mathbb{Z}$  is an unbounded subset of  $\mathbb{R}$ . The size bounds show that a subset of  $\mathbb{R}^n$  is bounded if and only if the  $j$ th coordinates of its points form a bounded subset of  $\mathbb{R}$  for each  $j \in \{1, \dots, n\}$ . The geometric content of this statement is that a set sits inside a ball centered at the origin if and only if it sits inside a box centered at the origin.

Blurring the distinction between a sequence and the set of its elements allows the definition of boundedness to apply to sequences. That is, a sequence  $\{x_\nu\}$  is bounded if there is some  $R > 0$  such that  $|x_\nu| < R$  for all  $\nu \in \mathbb{Z}^+$ . The proof of the next fact in  $\mathbb{R}^n$  is symbol-for-symbol the same as in  $\mathbb{R}$  (or in  $\mathbb{C}$ ), so it is only sketched.

**Proposition 2.4.7 (Convergence implies boundedness).** If the sequence  $\{x_\nu\}$  converges in  $\mathbb{R}^n$  then it is bounded.

*Proof.* Let  $\{x_\nu\}$  converge to  $p$ . Then there exists a starting index  $\nu_0$  such that  $x_\nu \in B(p, 1)$  for all  $\nu > \nu_0$ . Consider any real number  $R$  such that

$$R > \max\{|x_1|, \dots, |x_{\nu_0}|, |p| + 1\}.$$

Then clearly  $x_\nu \in B(\mathbf{0}, R)$  for  $\nu = 1, \dots, \nu_0$ , and the triangle inequality shows that also  $x_\nu \in B(\mathbf{0}, R)$  for all  $\nu > \nu_0$ . Thus  $\{x_\nu\} \subset B(\mathbf{0}, R)$  as a set.  $\square$

**Definition 2.4.8 (Subsequence).** A **subsequence** of the sequence  $\{x_\nu\}$  is a sequence consisting of some (possibly all) of the original terms, in ascending order of indices.

Since a subsequence of  $\{x_\nu\}$  consists of terms  $x_\nu$  only for some values of  $\nu$ , it is often written  $\{x_{\nu_k}\}$ , where now  $k$  is the index variable. For example, given the sequence

$$\{x_1, x_2, x_3, x_4, x_5, \dots\},$$

a subsequence is

$$\{x_2, x_3, x_5, x_7, x_{11}, \dots\},$$

with  $\nu_1 = 2$ ,  $\nu_2 = 3$ ,  $\nu_3 = 5$ , and generally  $\nu_k =$  the  $k$ th prime.

**Lemma 2.4.9 (Persistence of convergence).** *Let  $\{x_\nu\}$  converge to  $p$ . Then every subsequence  $\{x_{\nu_k}\}$  also converges to  $p$ .*

*Proof.* The hypothesis that  $\{x_\nu\}$  converges to  $p$  means that for every given  $\varepsilon > 0$ , only finitely many sequence-terms  $x_\nu$  lie outside the ball  $B(p, \varepsilon)$ . Consequently, only finitely many subsequence-terms  $x_{\nu_k}$  lie outside  $B(p, \varepsilon)$ , which is to say that  $\{x_{\nu_k}\}$  converges to  $p$ .  $\square$

The sequence property that characterizes bounded sets is called the **Bolzano–Weierstrass property**. Once it is proved in  $\mathbb{R}$ , the result follows in  $\mathbb{R}^n$  by arguing one component at a time.

**Theorem 2.4.10 (Bolzano–Weierstrass property in  $\mathbb{R}$ ).** *Let  $A$  be a bounded subset of  $\mathbb{R}$ . Then every sequence in  $A$  has a convergent subsequence.*

*Proof.* Let  $\{x_\nu\}$  be a sequence in  $A$ . Call a term  $x_\nu$  of the sequence a max-point if it is at least as big as all later terms, i.e.,  $x_\nu \geq x_\mu$  for all  $\mu > \nu$ . (For visual intuition, draw a graph plotting  $x_\nu$  as a function of  $\nu$ , with line segments connecting consecutive points. A max-point is a peak of the graph at least as high as all points to its right.) If there are infinitely many max-points in  $\{x_\nu\}$  then these form a decreasing sequence. If there are only finitely many max-points then  $\{x_\nu\}$  has an increasing sequence starting after the last max-point—this follows almost immediately from the definition of max-point. In either case,  $\{x_\nu\}$  has a monotonic subsequence that, being bounded, converges because the real number system is complete.  $\square$

**Theorem 2.4.11 (Bolzano–Weierstrass property in  $\mathbb{R}^n$ : sequential characterization of bounded sets).** *Let  $A$  be a subset of  $\mathbb{R}^n$ . Then  $A$  is bounded if and only if every sequence in  $A$  has a subsequence that converges in  $\mathbb{R}^n$ .*

*Proof.* ( $\implies$ ) Suppose that  $A$  is bounded. Consider any sequence  $\{x_\nu\}$  in  $A$ , written as  $\{(x_{1,\nu}, \dots, x_{n,\nu})\}$ . The real sequence  $\{x_{1,\nu}\}$  takes values in a bounded subset of  $\mathbb{R}$  and thus has a convergent subsequence,  $\{x_{1,\nu_k}\}$ . The subscripts are getting out of hand, so keep only the  $\nu_k$ th terms of the original sequence and relabel it. In other words, we may as well assume that the sequence of first components,  $\{x_{1,\nu}\}$ , converges. The real sequence of second components,  $\{x_{2,\nu}\}$ , in turn has a convergent subsequence, and by

Lemma 2.4.9 the corresponding subsequence of first components,  $\{x_{1,\nu}\}$ , converges too. Relabeling again, we may assume that  $\{x_{1,\nu}\}$  and  $\{x_{2,\nu}\}$  both converge. Continuing in this fashion  $n - 2$  more times exhibits a subsequence of  $\{x_\nu\}$  that converges at each component.

( $\Leftarrow$ ) Conversely, suppose that  $A$  is not bounded. Then there is a sequence  $\{x_\nu\}$  in  $A$  with  $|x_\nu| > \nu$  for all  $\nu$ . This sequence has no bounded subsequence, and hence it has no convergent subsequence by Proposition 2.4.7.  $\square$

Note how the sequential characterizations in Proposition 2.4.5 and in the Bolzano–Weierstrass property complement each other. The proposition characterizes every closed set in  $\mathbb{R}^n$  by the fact that if a sequence converges in the ambient space then it converges in the set. The Bolzano–Weierstrass property characterizes every bounded set in  $\mathbb{R}^n$  by the fact that every sequence in the set has a subsequence that converges in the ambient space but not necessarily in the set. Both the sequential characterization of a closed set and the sequential characterization of a bounded set refer to the ambient space  $\mathbb{R}^n$  in which the set lies. We will return to this point once more in this section.

**Definition 2.4.12 (Compact set).** *A subset  $K$  of  $\mathbb{R}^n$  is compact if it is closed and bounded.*

Since the static notions of closed and bounded are reasonably intuitive, we can usually recognize compact sets on sight. But it is not obvious from how compact sets look that they are related to continuity. So our program now has two steps: first, combine Proposition 2.4.5 and the Bolzano–Weierstrass property to characterize compact sets in terms of sequences, and second, use the characterization to prove that compactness is preserved by continuous mappings.

**Theorem 2.4.13 (Sequential characterization of compact sets).** *Let  $K$  be a subset of  $\mathbb{R}^n$ . Then  $K$  is compact if and only if every sequence in  $K$  has a subsequence that converges in  $K$ .*

*Proof.* ( $\Rightarrow$ ) We show that the sequential characterizations of closed and bounded sets together imply the claimed sequential characterization of compact sets. Suppose that  $K$  is compact and  $\{x_\nu\}$  is a sequence in  $K$ . Then  $K$  is bounded, so by “ $\Rightarrow$ ” of the Bolzano–Weierstrass property,  $\{x_\nu\}$  has a convergent subsequence. But  $K$  is also closed, so by “ $\Rightarrow$ ” of Proposition 2.4.5, this subsequence converges in  $K$ .

( $\Leftarrow$ ) Conversely, we show that the claimed sequential characterization of compact sets subsumes the sequential characterizations of closed and bounded sets. Thus, suppose that every sequence in  $K$  has a subsequence that converges in  $K$ . Then in particular, every sequence in  $K$  that converges in  $\mathbb{R}^n$  has a subsequence that converges in  $K$ . By Lemma 2.4.9 the limit of the sequence is the limit of the subsequence, so the sequence converges in  $K$ . That is, every sequence in  $K$  that converges in  $\mathbb{R}^n$  converges in  $K$ , and hence  $K$  is closed

by “ $\Leftarrow$ ” of Proposition 2.4.5. Also in consequence of the claimed sequential property of compact sets, every sequence in  $K$  has a subsequence that converges in  $\mathbb{R}^n$ . Thus  $K$  is bounded by “ $\Leftarrow$ ” of the Bolzano–Weierstrass Property.  $\square$

By contrast to the sequential characterizations of a closed set and of a bounded set, the sequential characterization of a compact set  $K$  makes no reference to the ambient space  $\mathbb{R}^n$  in which  $K$  lies. A set’s property of being compact is innate in a way that a set’s property of being closed or of being bounded is not.

The next theorem is the main result of this section. Now that all of the objects involved are described in the common language of sequences, its proof is natural.

**Theorem 2.4.14 (The continuous image of a compact set is compact).** *Let  $K$  be a compact subset of  $\mathbb{R}^n$  and let  $f : K \rightarrow \mathbb{R}^m$  be continuous. Then  $f(K)$ , the image set of  $K$  under  $f$ , is a compact subset of  $\mathbb{R}^m$ .*

*Proof.* Let  $\{y_\nu\}$  be any sequence in  $f(K)$ ; by “ $\Leftarrow$ ” of Theorem 2.4.13, it suffices to exhibit a subsequence converging in  $f(K)$ . Each  $y_\nu$  has the form  $f(x_\nu)$ , and this defines a sequence  $\{x_\nu\}$  in  $K$ . By “ $\Rightarrow$ ” of Theorem 2.4.13, since  $K$  is compact,  $\{x_\nu\}$  necessarily has a subsequence  $\{x_{\nu_k}\}$  converging in  $K$ , say to  $p$ . By the continuity of  $f$  at  $p$ , the sequence  $\{f(x_{\nu_k})\}$  converges in  $f(K)$  to  $f(p)$ . Since  $\{f(x_{\nu_k})\}$  is a subsequence of  $\{y_\nu\}$ , the proof is complete.  $\square$

Again, the sets in Theorem 2.4.14 are defined with no direct reference to sequences, but the theorem is proved entirely using sequences. The point is that with the theorem proved, we can easily see that it applies in particular contexts without having to think any longer about the sequences that were used to prove it.

A corollary of Theorem 2.4.14 generalizes the theorem that was quoted to begin the section:

**Theorem 2.4.15 (Extreme value theorem).** *Let  $K$  be a nonempty compact subset of  $\mathbb{R}^n$  and let the function  $f : K \rightarrow \mathbb{R}$  be continuous. Then  $f$  takes a minimum and a maximum value on  $K$ .*

*Proof.* By Theorem 2.4.14,  $f(K)$  is a compact subset of  $\mathbb{R}$ . As a nonempty bounded subset of  $\mathbb{R}$ ,  $f(K)$  has a greatest lower bound and a least upper bound by the completeness of the real number system. Each of these bounds is an isolated point or a limit point of  $f(K)$ , since otherwise some  $\varepsilon$ -ball about it would be disjoint from  $f(K)$ , giving rise to greater lower bounds or lesser upper bounds of  $f(K)$ . Because  $f(K)$  is also closed, it contains its limit points, so in particular it contains its greatest lower bound and its least upper bound. This means precisely that  $f$  takes a minimum and a maximum value on  $K$ .  $\square$

Even when  $n = 1$ , Theorem 2.4.15 generalizes the extreme value theorem from the beginning of the section. In the theorem here,  $K$  can be a finite union of closed and bounded intervals in  $\mathbb{R}$  rather than only one interval, or  $K$  can be a more complicated set, provided only that it is compact.

A *topological property* of sets is a property that is preserved under continuity. Theorem 2.4.14 says that compactness is a topological property. Neither the property of being closed nor the property of being bounded is in itself topological. That is, the continuous image of a closed set need not be closed, and the continuous image of a bounded set need not be bounded; for that matter, the continuous image of a closed set need not be bounded, and the continuous image of a bounded set need not be closed (Exercise 2.4.8).

The nomenclature *continuous image* in the slogan-title of Theorem 2.4.14 and in the previous paragraph is, strictly speaking, inaccurate: the image of a mapping is a set, and the notion of a set being continuous doesn't even make sense according to our grammar. As stated correctly in the body of the theorem, *continuous image* is short for *image under a continuous mapping*.

The property that students often have in mind when they call a set continuous is in fact called *connectedness*. Loosely, a set is connected if it has only one piece, so that a better approximating word from everyday language is *contiguous*. To define connectedness accurately, we would have to use methodology exactly opposite that of this section: rather than relate sets to continuous mappings by characterizing the sets in terms of sequences, the idea is to turn the whole business around and characterize continuous mappings in terms of sets, specifically in terms of open balls. However, the process of doing so, and then characterizing compact sets in terms of open balls as well, is trickier than characterizing sets in terms of sequences; and so we omit it because we do not need connectedness. Indeed, the remark after Theorem 2.4.15 points out that connectedness is unnecessary even for the one-variable extreme value theorem.

However, it deserves passing mention that connectedness is also a topological property: again using language loosely, the continuous image of a connected set is connected. This statement generalizes another theorem that underlies one-variable calculus, the intermediate value theorem. For a notion related to connectedness that is easily shown to be a topological property, see Exercise 2.4.10.

The ideas of this section readily extend to broader environments. The first generalization of Euclidean space is a *metric space*, a set with a well-behaved distance function. Even more general is a *topological space*, a set with some of its subsets designated as closed. Continuous functions, compact sets, and connected sets can be defined meaningfully in these environments, and the theorems remain the same: the continuous image of a compact set is compact, and the continuous image of a connected set is connected.

**Exercises**

**2.4.1.** Are the following subsets of  $\mathbb{R}^n$  closed, bounded, compact?

- (a)  $B(\mathbf{0}, 1)$ ,
- (b)  $\{(x, y) \in \mathbb{R}^2 : y - x^2 = 0\}$ ,
- (c)  $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 - 1 = 0\}$ ,
- (d)  $\{x : f(x) = \mathbf{0}_m\}$ , where  $f \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^m)$  is continuous (this generalizes (b) and (c)),
- (e)  $\mathbb{Q}^n$  where  $\mathbb{Q}$  denotes the rational numbers,
- (f)  $\{(x_1, \dots, x_n) : x_1 + \dots + x_n > 0\}$ .

**2.4.2.** Give a set  $A \subset \mathbb{R}^n$  and limit point  $b$  of  $A$  such that  $b \notin A$ . Give a set  $A \subset \mathbb{R}^n$  and a point  $a \in A$  such that  $a$  is not a limit point of  $A$ .

**2.4.3.** Let  $A$  be a closed subset of  $\mathbb{R}^n$  and let  $f \in \mathcal{M}(A, \mathbb{R}^m)$ . Define the **graph** of  $f$  to be

$$G(f) = \{(a, f(a)) : a \in A\},$$

a subset of  $\mathbb{R}^{n+m}$ . Show that if  $f$  is continuous then its graph is closed.

**2.4.4.** Prove the closed set properties: (1) the empty set  $\emptyset$  and the full space  $\mathbb{R}^n$  are closed subsets of  $\mathbb{R}^n$ ; (2) every intersection of closed sets is closed; (3) every finite union of closed sets is closed.

**2.4.5.** Prove that every ball  $B(p, \varepsilon)$  is bounded in  $\mathbb{R}^n$ .

**2.4.6.** Show that  $A$  is a bounded subset of  $\mathbb{R}^n$  if and only if for each  $j \in \{1, \dots, n\}$ , the  $j$ th coordinates of its points form a bounded subset of  $\mathbb{R}$ .

**2.4.7.** Show by example that a closed set need not satisfy the sequential characterization of bounded sets, and that a bounded set need not satisfy the sequential characterization of closed sets.

**2.4.8.** Show by example that the continuous image of a closed set need not be closed, that the continuous image of a closed set need not be bounded, that the continuous image of a bounded set need not be closed, and that the continuous image of a bounded set need not be bounded.

**2.4.9.** A subset  $A$  of  $\mathbb{R}^n$  is called **discrete** if each of its points is isolated. (Recall that the term *isolated* was defined in this section.) Show or take for granted the (perhaps surprising at first) fact that every mapping whose domain is discrete must be continuous. Is discreteness a topological property? That is, need the continuous image of a discrete set be discrete?

**2.4.10.** A subset  $A$  of  $\mathbb{R}^n$  is called **path-connected** if for every two points  $x, y \in A$ , there is a continuous mapping

$$\gamma : [0, 1] \longrightarrow A$$

such that  $\gamma(0) = x$  and  $\gamma(1) = y$ . (This  $\gamma$  is the path that connects  $x$  and  $y$ .) Draw a picture to illustrate the definition of a path-connected set. Prove that path-connectedness is a topological property.

Calculus and Analysis in Euclidean Space

Shurman, J.

2016, XIII, 505 p. 1 illus., Hardcover

ISBN: 978-3-319-49312-1