Exercises for the Introduction

◇ **0E-1.** The following mappings are defined by stating $f(x)$, the domain S, and the target T. For $A \subset S$ and $B \subset T$, as given, compute $f(A)$ and $f^{-1}(B)$.

(a) $f(x) = x^2$, $S = \{-1, 0, 1\}$, $T =$ all real numbers,
$A = \{-1, 1\}$, $B = \{0, 1\}$.

(b) $f(x) = \begin{cases} x^2, & \text{if } x \geq 0 \\ -x^2, & \text{if } x < 0 \end{cases}$
$S = T =$ all real numbers,
$A = \{x \in \text{ real numbers } \mid x > 0\}$, $B = \{0\}$.

(c) $f(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0 \end{cases}$
$S = T =$ all real numbers,
$A = B = \{x \in \text{ real numbers } \mid -2 < x < 1\}$.

Answer. (a) $f(A) = \{1\}$, $f^{-1}(B) = S$.
(b) $f(A) = A$, $f^{-1}(B) = B$.
(c) $f(A) = \{1, 0, -1\}$, $f^{-1}(B) = \{x \mid x \leq 0\}$.

◇

Solution. (a) $f(A) = \{f(-1), f(1)\} = \{1, 1\} = \{1\}$. Since $f(-1) = f(1) = 1$, and $f(0) = 0$, we have $f^{-1}(\{0, 1\}) = \{-1, 0, 1\}$.

(b) If $x > 0$, then $x^2 > 0$, and if $y > 0$, then $\sqrt{y} > 0$. Thus f takes A onto A and $f(A) = A$. Since $f(x) = 0$ if and only if $x = 0$, we have $f^{-1}(B) = f^{-1}(\{0\}) = \{0\}$.

(c) Since the interval $A =]-2, 1[$ contains positive numbers, negative numbers, and 0, we have $f(A) = \{-1, 0, 1\}$. The interval $B =]-2, 1[$ contains -1 and 0, but not 1. So $f^{-1}(B)$ contains all of the negative numbers and 0, but not the positive numbers. So $f^{-1}(B) = \{x \in \mathbb{R} \mid x \leq 0\}$.

♦

◇ **0E-2.** Determine whether the functions listed in Exercise 0E-1 are one-to-one or onto (or both) for the given domains and targets.

Answer. (a) Neither one-to-one nor onto.
(b) One-to-one and onto.
(c) Neither one-to-one nor onto.

◇

Solution. (a) Since $f(-1) = f(1) = 1$, the function f is not one-to-one. Since $f(S) = \{0, 1\}$, the function f does not map S onto $\mathbb{R}$.

(b) To see that f is one-to-one, suppose that $f(s) = f(t) = y$. If $y > 0$, then $y = s^2 = t^2$ with $s \geq 0$ and $t \geq 0$. So $0 = s^2 - t^2 = (s-t)(s+t)$. Since s and t are nonnegative, the only way for $s + t$ to be 0 is for $s = t = 0$. If $s = t$ is not 0, then $s - t$ must be 0, so again $s = t$. If $y < 0$, then $-s^2 = y = -t^2$ with $s < 0$ and $t < 0$. Thus $s + t < 0$. Again we have $0 = s^2 - t^2 = (s+t)(s-t)$, and, since $s + t$ is not 0, we have $s - t = 0$. So again, $f(s) = f(t)$ implies that $s = t$. To see that f takes $\mathbb{R}$ onto $\mathbb{R}$, let $y \in \mathbb{R}$. If $y \geq 0$, then y has a nonnegative square root. Let $x = \sqrt{y} \geq 0$. Then $f(x) = (\sqrt{y})^2 = y$. If $y < 0$, then $-y > 0$ and has a positive square root. Let $x = -\sqrt{-y} < 0$. Then $f(x) = -(-\sqrt{-y})^2 = -(-y) = y$. So for each y in $\mathbb{R}$ there is an x in $\mathbb{R}$ with $f(x) = y$. That is, f maps $\mathbb{R}$ onto $\mathbb{R}$.

(c) Since the range of f is the set $\{-1, 0, 1\}$, f certainly does not map $\mathbb{R}$ onto $\mathbb{R}$. In particular, there is no x in $\mathbb{R}$ with $f(x) = 3$. Since $f(2) = f(1) = 1$, the function f is not one-to-one.

♦

◇ **0E-3.** Let $f : S \to T$ be a function, $C_1, C_2 \subset T$, and $D_1, D_2 \subset S$. Prove

(a) $f^{-1}(C_1 \cup C_2) = f^{-1}(C_1) \cup f^{-1}(C_2)$.
(b) $f(D_1 \cup D_2) = f(D_1) \cup f(D_2)$.
(c) $f^{-1}(C_1 \cap C_2) = f^{-1}(C_1) \cap f^{-1}(C_2)$.
(d) $f(D_1 \cap D_2) \subset f(D_1) \cap f(D_2)$.

Answer. In each case use the definitions: For example, for (a), $x \in f^{-1}(C_1 \cup C_2)$ iff $f(x) \in C_1 \cup C_2$ iff ($f(x) \in C_1$ or $f(x) \in C_2$) iff $x \in f^{-1}(C_1) \cup f^{-1}(C_2)$. One need not have equality in (d). ◇

Solution. (a)

$$\begin{aligned} x \in f^{-1}(C_1 \cup C_2) &\iff f(x) \in C_1 \cup C_2 \iff f(x) \in C_1 \text{ or } f(x) \in C_2 \\ &\iff x \in f^{-1}(C_1) \text{ or } x \in f^{-1}(C_2) \iff x \in f^{-1}(C_1) \cup f^{-1}(C_2) \end{aligned}$$

The two sets have the same elements and must be equal.

(b) If $y \in f(D_1 \cup D_2)$ then there is an x in $D_1 \cup D_2$ with $f(x) = y$. x must be in D_1 or in D_2. If $x \in D_1$, then $f(x) \in f(D_1)$ and so $f(x) \in f(D_1) \cup f(D_2)$. If $x \in D_2$, then $f(x) \in f(D_2)$ and so $y = f(x) \in f(D_1) \cup f(D_2)$ Thus $f(D_1 \cup D_2) \subseteq f(D_1) \cup f(D_2)$
Conversely, if $y \in f(D_1) \cup f(D_2)$, then $y \in f(D_1)$ or $y \in f(D_2)$. In the first case there is an x in D_1 with $f(x) = y$ and in the second there is an x in D_2 with $y = f(x)$. In either case we have an x in $D_1 \cup D_2$ with $y = f(x)$, so $y \in f(D_1 \cup D_2)$. Thus $f(D_1) \cup f(D_2) \subseteq f(D_1 \cup D_2)$.

We have inclusion in both directions so the sets are equal.

(c)

$$x \in f^{-1}(C_1 \cap C_2) \iff f(x) \in C_1 \cap C_2 \iff f(x) \in C_1 \text{ and } f(x) \in C_2$$
$$\iff x \in f^{-1}(C_1) \text{ and } x \in f^{-1}(C_2) \iff x \in f^{-1}(C_1) \cap f^{-1}(C_2)$$

The sets have the same elements, so they are equal.

(d) If $y \in f(D_1 \cap D_2)$, then there is an x in $D_1 \cap D_2$ with $f(x) = y$. x is in D_1, so $y = f(x) \in f(D_1)$, and $x \in D_2$, so $y = f(x) \in f(D_2)$. Thus y is in both $f(D_1)$ and $f(D_2)$m so $y \in f(D_1) \cap f(D_2)$. Thus $f(D_1 \cap D_2) \subseteq f(D_1) \cap f(D_2)$ as claimed.

♦

◇ **0E-4.** Verify the relations (a) through (d) in Exercise 0E-3 for each of the functions (a) through (c) in Exercise 0E-1 and the following sets; use the sets in part (a) below for the function in Exercise 0E-1 (a), the sets in part (b) for the function in Exercise 0E-1 (b), and the sets in part (c) for the functions in Exercise 0E-1 (c):

(a) $C_1 =$ all $x > 0$, $D_1 = \{-1, 1\}$, $C_2 =$ all $x \le 0$, $D_2 = \{0, 1\}$;
(b) $C_1 =$ all $x \ge 0$, $D_1 =$ all $x > 0$, $C_2 =$ all $x \le 2$, $D_2 =$ all $x \ge -1$;
(c) $C_1 =$ all $x \ge 0$, $D_1 =$ all x, $C_2 =$ all $x > -1$, $D_2 =$ all $x > 0$.

Solution. (a) $f(x) = x^2$; $C_1 = \{x \in \mathbb{R} \mid x > 0\}$; $C_2 = \{x \in \mathbb{R} \mid x \le 0\}$; $D_1 = \{-1, 1\}$; and $D_2 = \{0, 1\}$. So

$$\begin{aligned}
f^{-1}(C_1 \cup C_2) &= f^{-1}(\mathbb{R}) = \mathbb{R} \\
&= \{x \in \mathbb{R} \mid x \ne 0\} \cup \{0\} = f^{-1}(C_1) \cup f^{-1}(C_2). \\
f(D_1 \cup D_2) &= f(\{-1, 0, 1\}) = \{0, 1\} \\
&= \{1\} \cup \{0, 1\} = f(D_1) \cup f(D_2). \\
f^{-1}(C_1 \cap C_2) &= f^{-1}(\emptyset) = \emptyset \\
&= \{x \in \mathbb{R} \mid x \ne 0\} \cap \{0\} = f^{-1}(C_1) \cap f^{-1}(C_2). \\
f(D_1 \cap D_2) &= f(\{1\}) = \{1\} \subseteq \{1\} \cap \{0, 1\} \\
&= f(D_1) \cap f(D_2).
\end{aligned}$$

(b)

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

$$C_1 = \{x \in \mathbb{R} \mid x \ge 0\}; C_2 = \{x \in \mathbb{R} \mid x \le 2\}$$
$$D_1 = \{x \in \mathbb{R} \mid x > 0\}; \text{ and } D_2 = \{x \in \mathbb{R} \mid x \ge -1\}.$$

$$\begin{aligned} f^{-1}(C_1 \cup C_2) &= f^{-1}(\mathbb{R}) = \mathbb{R} \\ &= \{x \in \mathbb{R} \mid x \geq 0\} \cup \{x \in \mathbb{R} \mid x \leq \sqrt{2}\} \\ &= f^{-1}(C_1) \cup f^{-1}(C_2). \end{aligned}$$

$$\begin{aligned} f(D_1 \cup D_2) &= f(\{x \in \mathbb{R} \mid x \geq -1\}) = \{x \in \mathbb{R} \mid x \geq -1\} \\ &= \{x \in \mathbb{R} \mid x \geq -1\} \cup \{x \in \mathbb{R} \mid x \geq 0\} \\ &= f(D_1) \cup f(D_2). \end{aligned}$$

$$\begin{aligned} f^{-1}(C_1 \cap C_2) &= f^{-1}([0, 2]) = [0, \sqrt{2}] \\ &= \{x \in \mathbb{R} \mid x \geq 0\} \cap \{x \in \mathbb{R} \mid x \leq \sqrt{2}\} \\ &= f^{-1}(C_1) \cap f^{-1}(C_2). \end{aligned}$$

$$\begin{aligned} f(D_1 \cap D_2) &= f(D_1) = \{x \in \mathbb{R} \mid x > 0\} \\ &= \{x \in \mathbb{R} \mid x > 0\} \cap \{x \in \mathbb{R} \mid x > -1\} \\ &= f(D_1) \cap f(D_2). \end{aligned}$$

(c)

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

$$C_1 = \{x \in \mathbb{R} \mid x \geq 0\}; C_2 = \{x \in \mathbb{R} \mid x \geq -1\}$$

$$D_1 = \mathbb{R}; \text{ and } D_2 = \{x \in \mathbb{R} \mid x > 0\}.$$

$$\begin{aligned} f^{-1}(C_1 \cup C_2) &= f^{-1}(\{x \in \mathbb{R} \mid x > -1\}) = \{x \in \mathbb{R} \mid x \geq 0\} \\ &= \{x \in \mathbb{R} \mid x \geq 0\} \cup \{x \in \mathbb{R} \mid x \geq 0\} \\ &= f^{-1}(C_1) \cup f^{-1}(C_2). \end{aligned}$$

$$f(D_1 \cup D_2) = f(\mathbb{R}) = \{-1, 0, 1\} = \{-1, 0, 1\} \cup \{1\} = f(D_1) \cup f(D_2).$$

$$\begin{aligned} f^{-1}(C_1 \cap C_2) &= f^{-1}(\{x \in \mathbb{R} \mid x \geq 0\}) = \{x \in \mathbb{R} \mid x \geq 0\} \\ &= \{x \in \mathbb{R} \mid x \geq 0\} \cap \{x \in \mathbb{R} \mid x \geq 0\} \\ &= f^{-1}(C_1) \cap f^{-1}(C_2). \end{aligned}$$

$$f(D_1 \cap D_2) = f(D_2) = \{1\} = \{-1, 0, 1\} \cap \{1\} = f(D_1) \cap f(D_2).$$

♦

◇ **0E-5.** If $f : S \to T$ is a function from S into T, show that the following are equivalent. (Each implies the other two.)

(a) f is one-to-one.
(b) For every y in T, the set $f^{-1}(\{y\})$ contains at most one point.
(c) $f(D_1 \cap D_2) = f(D_1) \cap f(D_2)$ for all subsets D_1 and D_2 of S.

Develop similar criteria for "ontoness."

Answer. For example, assume $f(D_1 \cap D_2) = f(D_1) \cap f(D_2)$. If $x_1 \neq x_2$ and $f(x_1) = f(x_2) = y$, let $D_1 = \{x_1\}$ and $D_2 = \{x_2\}$ to deduce a contradiction. So f is one-to-one. Some possible equivalents to "onto" are: for all $y \in T, f^{-1}(\{y\}) \neq \emptyset$, or, for all $R \subseteq T, R = f(f^{-1}(R))$. Other answers are possible. ◇

Solution. (a) $\implies$ (b): If s and t are in $f^{-1}(\{y\})$, then $f(s) = f(t) = y$. Since f is one-to-one, we must have $s = y$. Thus the set $f^{-1}(\{y\})$ contains at most one point.

(b) $\implies$ (c): Suppose (b) holds and D_1 and D_2 are subsets of S. We know from problem 0E-3 (d) that $f(D_1 \cap d_2) \subseteq f(D_1) \cap f(D_2)$. Now suppose $y \in f(D_1) \cap f(D_2)$. Then $y \in f(D_1)$ and $y \in f(D_2)$. So there are points s in D_1 and t in D_2 such that $f(s) = y$ and $f(t) = y$. So s and t are in $f^{-1}(\{y\})$. Hypothesis (b) implies that $s = t \in D_1 \cap D_2$. So $y \in f(D_1 \cap D_2)$. Thus $f(D_1) \cap f(D_2) \subseteq f(D_1 \cap D_2)$. We have inclusion in both directions, so the sets are equal.

(c) $\implies$ (a): Suppose x and s are in S and $f(x) = f(s) = y \in T$. Let $D_1 = \{x\}$ and $D_2 = \{s\}$. Using hypothesis (c), we compute $\{y\} = f(D_1) \cap f(D_2) = f(D_1 \cap D_2)$. But $D_1 \cap D_2 = \emptyset$ unless $x = s$. Since it is not empty, we must have $x = s$. ♦

◇ **0E-6.** Show that the set of positive integers $\mathbb{N} = \{1, 2, 3, \ldots\}$ has as many elements as there are integers, by setting up a one-to-one correspondence between the set $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ and the set $\mathbb{N}$. Conclude that $\mathbb{Z}$ is countable.

Solution. The pairing between $\mathbb{N}$ and $\mathbb{Z}$ may be established according to the following pattern

$$\begin{array}{cccccccccc} \mathbb{N} & : & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \ldots \\ \mathbb{Z} & : & 0 & 1 & -1 & 2 & -2 & 3 & -3 & \ldots \end{array}$$

That is

$$\begin{aligned} \mathbb{N} &\to \mathbb{Z} \\ 1 &\mapsto 0 \\ 2k &\mapsto k \qquad \text{for } k = 1, 2, 3, \ldots \\ 2k+1 &\mapsto -k \qquad \text{for } k = 1, 2, 3, \ldots \end{aligned}$$

♦

◇ **0E-7.** Let A be a finite set with N elements, and let $\mathcal{P}(A)$ denote the collection of all subsets of A, including the empty set. Prove that $\mathcal{P}(A)$ has 2^N elements.

Sketch. A subset B of A is determined by choosing for each $x \in A$ either $x \in B$ or $x \notin B$. Apply a standard argument for counting a sequence of choices or use induction. ◇

Solution. If $N = 0$, then $A = \emptyset$. The power set is $\mathcal{P}(\emptyset) = \{\emptyset\}$. This has $1 = 2^0$ elements.

If $N = 1$, put $A = \{x\}$. The power set is $\mathcal{P}(\{x\}) = \{\emptyset, \{x\}\}$. This has $2 = 2^1$ elements.

Suppose inductively that we know that every set with exactly $N = k$ elements has exactly 2^k subsets. Suppose A has $N = k+1$ elements. Let $A = \{x_1, x_2, \ldots, x_k, x_{k+1}\}$ and $B = \{x_1, x_2, \ldots, x_k\}$. For each subset S of B, we get two distinct subsets, S and $S \cup \{x_{k+1}\}$, of A. Furthermore, if S and T are different subsets of B, then the resulting 4 subsets of A are all different. Finally, all subsets of A are obtained in this way. Thus $\mathcal{P}(A)$ has exactly twice as many elements as $\mathcal{P}(B)$. By the induction hypothesis, $\mathcal{P}(B)$ has 2^k elements, so $\mathcal{P}(A)$ has 2^{k+1} elements. This is exactly the desired conclusion for $N = k+1$, so our proposition follows by the principle of mathematical induction. ♦

◇ **0E-8.** (a) Let $N = \{0, 1, 2, 3, \ldots\}$. Define $\varphi : N \times N \to N$ by $\varphi(i, j) = j + \frac{1}{2}k(k+1)$ where $k = i + j$. Show that φ is a bijection and that it has something to do with Figure 0.9.

(b) Show that if $A_1, A_2, \ldots$ are countable sets, so is $A_1 \cup A_2 \cup \cdots$.

Solution. (a) $\underline{One-to-one}$: Suppose $\varphi(i, j) = \varphi(I, J)$. Put $k = i+j$ and $K = I + J$. Then $2j + k(k+1) = 2J + K(K+1)$, so $2(j - J) = (K+k+1)(K-k)$. If $K = k$, then we must have $j - J = 0$, so $j = J$ and also $i = I$. Now we show that we must have $K = k$. If they were not equal, then one would be larger. Say $K > k$. If $K = k+1$, our equation would become $2(j - J) = 2K$, so $j = J + K$. This would

Figure 0.9.

imply that $i = k - j = K - 1 - (J + K) = -(J + 1) < 0$. This is not possible since $i \in N$. If $K \geq k+2$, we would also have a contradiction since then

$$\begin{aligned}(K + k + 1)(K - k) &\geq 2(K + k + 1) = 2(J + j + I + i + 1) \\ &\geq 2(J + j + 1) > 2(j - J)\end{aligned}$$

So $K > k$ is not possible. The alternative $k > K$ may be eliminated by a similar argument. The only possibility left is that $K = k$ and we have seen that this forces $J = j$ and $I = i$. So the map φ is one-to-one.

Onto : First note that $\varphi(0, 0) = 0$, $\varphi(1, 0) = 1$, and $\varphi(0, 1) = 2$. Suppose m is an integer with $m \geq 3$, and let k be the largest integer such that $k(k+1)/2 \leq m$. Notice that $k \geq 2$. Then $(k+1)(k+2)/2 > m$, so $k^2 + 3k + 2 > 2m$, and $3k - 2m + k^2 > -2$. Thus if we put $j = m - k(k + 1)/2$ and $i = k - j$, we have

$$i = k - m + \frac{k(k + 1)}{2} = \frac{3k - 2m + k^2}{2} > -1$$

Since i is an integer, we have $i \geq 0$. Thus i and j are in N, and $\varphi(i, j) = m$. Thus φ takes $N \times N$ onto N. The diagram shows a table of the values of φ with $\varphi(i, j)$ in the square at row i column j.

(b) The idea is first to observe that every subset of a countable set is countable. (This is actually a bit tricky.) With this is hand, we might as well assume that the sets A_k are denumerably infinite and pairwise disjoint. Now establish pairings between each of the sets A_k and a copy of N. Their union is then naturally paired with $N \times N$. But this may then be paired with N by the map established in part (a). ♦

◇ **0E-9.** Let $\mathcal{A}$ be a family of subsets of a set S. Write $\cup\mathcal{A}$ for the union of all members of $\mathcal{A}$ and similarly define $\cap\mathcal{A}$. Suppose $\mathcal{B} \supset \mathcal{A}$. Show that $\cup\mathcal{A} \subset \cup\mathcal{B}$ and $\cap\mathcal{B} \subset \cap\mathcal{A}$.

Answer. If $x \in \bigcup\mathcal{A}$, then $x \in A$ for some $A \in \mathcal{A}$. Since $\mathcal{A} \subseteq \mathcal{B}$, $A \in \mathcal{B}$, and so $x \in \bigcup\mathcal{B}$. Thus $\bigcup\mathcal{A} \subseteq \bigcup\mathcal{B}$. The second part is similar. ◊

Solution. If $x \in \bigcup\mathcal{A}$, then $x \in A$ for some $A \in \mathcal{A}$. Since $\mathcal{A} \subseteq \mathcal{B}$, $A \in \mathcal{B}$, and so $x \in \bigcup\mathcal{B}$. Thus $\bigcup\mathcal{A} \subseteq \bigcup\mathcal{B}$.

For the second assertion, suppose that $x \in \bigcap\mathcal{B}$. Then $x \in B$ for every set B in the collection $\mathcal{B}$. If A is one of the sets in $\mathcal{A}$, then $A \in \mathcal{B}$, so $x \in A$. Since this holds for every such x, we have $x \in \bigcap\mathcal{A}$. Thus $\bigcap\mathcal{B} \subseteq \bigcap\mathcal{A}$. ♦

◇ **0E-10.** Let $f : S \to T,\ g : T \to U$, and $h : U \to V$ be mappings. Prove that $h \circ (g \circ f) = (h \circ g) \circ f$ (that is, that *composition is associative*).

Solution. If $x \in S$, then

$$(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x)))$$

and

$$((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x)))$$

So $(h\circ(g\circ f))(x) = ((h\circ g)\circ f)(x)$ for every x in S. Thus $h\circ(g\circ f) = (h\circ g)\circ f$. ♦

◇ **0E-11.** Let $f : S \to T,\ g : T \to U$ be given mappings. Show that for $C \subset U,\ (g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$.

Solution. $x \in (g \circ f)^{-1}(C) \iff (g \circ f)(x) \in C \iff g(f(x)) \in C \iff f(x) \in g^{-1}(C) \iff x \in f^{-1}(g^{-1}(C))$. ♦

◇ **0E-12.** Let $\mathcal{A}$ be a collection of subsets of a set S and $\mathcal{B}$ the collection of complementary sets; that is, $B \in \mathcal{B}$ iff $S\backslash B \in \mathcal{A}$. Prove ***de Morgan's laws:***

(a) $S\backslash \cup \mathcal{A} = \cap\mathcal{B}$.
(b) $S\backslash \cap \mathcal{A} = \cup\mathcal{B}$.

Here $\cup\mathcal{A}$ denotes the union of all sets in $\mathcal{A}$.

For example, if $\mathcal{A} = \{A_1, A_2\}$, then

(a) reads $S\backslash(A_1 \cup A_2) = (S\backslash A_1) \cap (S\backslash A_2)$ and
(b) reads $S\backslash(A_1 \cap A_2) = (S\backslash A_1) \cup (S\backslash A_2)$.

Solution. (a) Suppose $x \in S \setminus (\bigcup \mathcal{A})$. Then x is not in $\bigcup \mathcal{A}$, and so it is not in any of the sets A in the collection $\mathcal{A}$. It is thus in each complementary set $B = S \setminus A$. That is, $x \in \bigcap \mathcal{B}$. Thus $S \setminus (\bigcup \mathcal{A}) \subseteq \bigcap \mathcal{B}$.
For the opposite inclusion, suppose $x \in \bigcap \mathcal{B}$, then x is in each of the sets $B = S \setminus A$. Thus x is in none of the sets A, so x is not in $\bigcup \mathcal{A}$, and x is in $S \setminus (\bigcup \mathcal{A})$. Thus $\bigcap \mathcal{B} \subseteq S \setminus (\bigcup \mathcal{A})$. We have inclusion in both directions, so the sets are equal.

(b) Suppose $x \in S \setminus (\bigcap \mathcal{A})$. Then x is not in $\bigcap \mathcal{A}$, and so there is at least one of the sets A with x not in A. So $x \in B = S \setminus A$ for that A. Thus $x \in \bigcup \mathcal{B}$. Thus $S \setminus (\bigcap \mathcal{A}) \subseteq \bigcup \mathcal{B}$.
For the opposite inclusion, suppose $x \in \bigcup \mathcal{B}$, then x is in at least one of the sets $B = S \setminus A$. Thus x is not in that A, so x is not in $\bigcap \mathcal{A}$, and x is in $S \setminus (\bigcap \mathcal{A})$. Thus $\bigcup \mathcal{B} \subseteq S \setminus (\bigcap \mathcal{A})$. We have inclusion in both directions, so the sets are equal.

♦

◇ **0E-13.** Let $A, B \subset S$. Show that

$$A \times B = \emptyset \;\Leftrightarrow\; A = \emptyset \text{ or } B = \emptyset.$$

Solution. $A, B \neq \emptyset \iff (\exists a \in A \text{ and } \exists b \in B) \iff (a, b) \in A \times B$ and so $A \times B \neq \emptyset$. ♦

◇ **0E-14.** Show that

(a) $(A \times B) \cup (A' \times B) = (A \cup A') \times B$.
(b) $(A \times B) \cap (A' \times B') = (A \cap A') \times (B \cap B')$.

Solution. (a)

$$\begin{aligned}
(x, y) \in (A \times B) \cup (A' \times B) &\iff (x, y) \in A \times B \text{ or } (x, y) \in A' \times B \\
&\iff (x \in A \text{ and } y \in B) \text{ or } (x \in A' \text{ and } y \in B) \\
&\iff (x \in A \text{ or } x \in A') \text{ and } (y \in B) \\
&\iff (x \in A \cup A') \text{ and } (y \in B) \\
&\iff (x, y) \in (A \cup A') \times B.
\end{aligned}$$

So $(A \times B) \cup (A' \times B) = (A \cup A') \times B$.

(b)

$$\begin{aligned}(x,y) \in (A \times B) \cap (A' \times B') &\iff (x,y) \in A \times B \text{ and } (x,y) \in A' \times B' \\ &\iff (x \in A \text{ and } y \in B) \text{ and } (x \in A' \text{ and } y \in B') \\ &\iff (x \in A \text{ and } x \in A') \text{ and } (y \in B \text{ and } y \in B') \\ &\iff (x \in A \cap A') \text{ and } (y \in B \cap B') \\ &\iff (x,y) \in (A \cap A') \times (B \cap B').\end{aligned}$$

So $(A \times B) \cap (A' \times B') = (A \cap A') \times (B \cap B')$.

♦

◇ **0E-15.** Show that

(a) $f : S \to T$ is one-to-one iff there is a function $g : T \to S$ such that $g \circ f = I_S$; we call g a ***left inverse*** of f.
(b) $f : S \to T$ is onto iff there is a function $h : T \to S$ such that $f \circ h = I_T$; we call h a ***right inverse*** of f.
(c) A map $f : S \to T$ is a bijection iff there is a map $g : T \to S$ such that $f \circ g = I_T$ and $g \circ f = I_S$. Show also that $g = f^{-1}$ and is uniquely determined.

Sketch. (a) If f is one-to-one, let $g(y) = x$ if $f(x) = y$ and let $g(y)$ be anything if there is no such x.
(b) If f is onto and $y \in T$, choose some x such that $f(x) = y$ and let $h(y) = x$. (Note the use of the Axiom of Choice.)

◇

Solution. (a) Suppose $f : S \to T$ is one-to-one and let $s_0 \in S$. To define the map g from T to S, let $t \in T$. If t is not in $f(S)$, simply put $g(t) = s_0$. If t is in $f(S)$, then there is exactly one s in S with $f(s) = t$ since f is one-to-one. Put $g(t) = s$. If s is in S, then$(g \circ f)(s) = g(f(s)) = s$. So $g \circ f = I_S$.
For the converse, suppose there is such a function g and that $f(s_1) = f(s_2)$. Then

$$s_1 = (g \circ f)(s_1) = g(f(s_1)) = g(f(s_2)) = (g \circ f)(s_2) = s_2.$$

Thus $f(s_1) = f(s_2) \implies s_1 = s_2$., so f is one-to-one.
(b) Suppose $f : S \to T$ takes S onto T. For each t in T, then set $f^{-1}(\{t\})$, is nonempty, so, by the Axiom of Choice, we can select an element $s_t \in f^{-1}(\{t\})$ from each. Put $h(t) = s_t$. We then have $(f \circ h)(t) = f(h(t)) = f(s_t) = t$. So $f \circ h = I_T$.
For the converse, suppose there is such a function h and that $t \in T$. Put $s = h(t)$. Then $f(s) = f(h(t)) = (f \circ h)(t) = t$. For each t in T there is at least one s in S with $f(s) = t$, so f maps S onto T.

(c) If there is such a map g, then by part (a) the function f is one-to-one, and by part (b) it takes S onto T. So f is a bijection.
If f is a bijection, then it is one-to-one and onto, so both constructions can be carried out as in (a) and (b). Since f takes S onto T, the first part of the construction in (a) is not needed. Since f is one-to-one, there is actually no freedom of choice in the construction of (b). Thus the map h constructed in (b) must actually be the same as the map g constructed in part (a). So g satisfies the criteria required for part (c).

♦

◇ **0E-16.** Let $f : S \to T$ and $g : T \to U$ be bijections. Show that $g \circ f$ is a bijection and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. [Hint: Use Exercise 0E-15 (c).]

Solution. Suppose $u \in U$. Since g takes T onto U, there is a t in T with $g(t) = u$. Since f takes S onto T, there is an s in S with $f(s) = t$. Then $(g \circ f)(s) = g(f(s)) = g(t) = u$. Thus $g \circ f$ maps S onto U.

Suppose s_1 and s_2 are in S and that $(g \circ f)(s_1) = (g \circ f)(s_2)$. Then $g(f(s_1)) = g(f(s_2))$. Since g is one-to-one, we must have $f(s_1) = f(s_2)$. Since f is one-to-one, we must have $s_1 = s_2$. Thus $g \circ f$ is one-to-one.

Since $g \circ f$ is a bijection, we have an inverse function $(g \circ f)^{-1}$ mapping U one-to-one onto S as in exercise 0E-15(c). Furthermore, the method of solution for that problem indicates that there is exactly one choice for that inverse function. Computing directly with the function $f^{-1} \circ g^{-1}$, we find

$$\begin{aligned}(f^{-1} \circ g^{-1})(g \circ f)(s) &= (f^{-1} \circ g^{-1})(g(f(s))) \\ &= f^{-1}(g^{-1}(g(f(s)))) \\ &= f^{-1}(f(s)) = s\end{aligned}$$

So $f^{-1} \circ g^{-1}$ is a left inverse for $g \circ f$.

$$\begin{aligned}(g \circ f)(f^{-1} \circ g^{-1})(u) &= (g \circ f)(f^{-1}(g^{-1}(u))) \\ &= g(f(f^{-1}(g^{-1}(u)))) \\ &= g(g^{-1}(u)) = u\end{aligned}$$

So $f^{-1} \circ g^{-1}$ is a right inverse for $g \circ f$. Since the constructions of exercise 0E-15 indicate that the inverse function of a bijection is unique, we must have $f^{-1} \circ g^{-1} = (g \circ f)^{-1}$. ♦

◇ **0E-17.** (For this problem you may wish to review some linear algebra.) Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

be an $m \times n$ matrix where the a_{ij} are real numbers. Use Exercise 0E-15 to show that A has rank m if and only if there is a matrix B such that AB is the $m \times m$ identity matrix.

Sketch. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation associated with A. We learn in linear algebra that T is onto, i.e., $Tx = y$ is solvable for $x \in \mathbb{R}^n$ for any $y \in \mathbb{R}^m$, if and only if T has rank m. If B exists, T is onto by Exercise 0E-15(b). Conversely, if T is onto, choose a complementary subspace $W \subseteq \mathbb{R}^n$ of dimension m to $\ker(T)$. Then $T|W : W \to \mathbb{R}^m$ is an isomorphism. Let U be its inverse. Then $T \circ U =$ identity . If B is the matrix of U, we get $AB =$ identity . ♢

Solution. We know that the matrix A corresponds to a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ given for a vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ by

$$Tx = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.$$

If B is an $n \times m$ matrix

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix},$$

then B corresponds to a linear transformation $S : \mathbb{R}^m \to \mathbb{R}^n$ given for $y = (y_1, y_2, \ldots, y_m) \in \mathbb{R}^m$ by

$$Sy = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} b_{11}y_1 + b_{12}y_2 + \cdots + b_{1m}y_m \\ b_{21}y_1 + b_{22}y_2 + \cdots + b_{2m}y_m \\ \vdots \\ b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{nm}y_m \end{pmatrix}.$$

The product matrix

$$C = AB = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1m} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{nm} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1} & \cdots & & \cdots & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{k1}b_{11} + a_{k2}b_{21} + \cdots + a_{kn}b_{n1} & \cdots & a_{k1}b_{1j} + a_{k2}b_{2j} + \cdots + a_{kn}b_{nj} & \cdots & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} + a_{m2}b_{21} + \cdots + a_{mn}b_{n1} & \cdots & & \cdots & \end{pmatrix}$$

with $C_{kj} = \sum_{i=1}^{n} a_{ki}b_{ij}$, is the matrix corresponding to the composition operator $R = T \circ S : \mathbb{R}^m \to \mathbb{R}^m$.

Exercise 0E-15 tells us that a right inverse S exists, at least as a function, if and only if the operator T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$.

If the right inverse matrix B exists, then the right inverse function S exists. The operator T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$. Thus

$$\operatorname{rank}(T) = \dim(\operatorname{range}(T)) = \dim(\mathbb{R}^m) = m.$$

For the converse, suppose $\operatorname{rank}(T) = m$. Then $\operatorname{range}(T)$ has full dimension in $\mathbb{R}^m$ and so must be all of $\mathbb{R}^m$. A right inverse function $S : \mathbb{R}^m \to \mathbb{R}^n$ exists by exercise 0E-15. The operator T cannot also be one-to-one unless $m = n$, so the function S does not have to be linear. If $y \in \mathbb{R}^n$ and $Tx = y$, then x is a possible choice for $S(y)$. But so is $x+\hat{x}$ for every $\hat{x}$ with $T\hat{x} = 0$. The point is that the values of S can be selected in such a way as to make S linear. Let W be the orthogonal complement of the kernel (or null space) of T.

$$W = (\ker(T))^{\perp}.$$

The restriction, $T|_W$, takes W one-to-one onto the range of T. (Why?). Since in our case the range is all of $\mathbb{R}^m$, there is a two-sided inverse function $S : \mathbb{R}^m \to W \subseteq \mathbb{R}^n$.

$$S \circ (T|_W) = (T|_W) \circ S = (I|_W).$$

This S must be linear. Suppose y_1 and y_2 are in $\mathbb{R}^m$ and $\lambda \in \mathbb{R}$. Then there are unique x_1 and x_2 in W with $Tx_1 = y_1$ and $Tx_2 = y_2$. Since T is linear we have

$$S(y_1 + y_2) = S(Tx_1 + Tx_2) = S(T(x_1 + x_2)) = x_1 + x_2 = Sy_1 + Sy_2.$$

and

$$S(\lambda y_1) = S(\lambda T x_1) = S(T(\lambda x_1)) = \lambda x_1 = \lambda S y_1$$

The matrix B for the transformation S is then a right inverse for the matrix A. ♦

1

The Real Line and Euclidean Space

1.1 Ordered Fields and the Number Systems

◇ **1.1-1.** In an ordered field prove that $(a+b)^2 = a^2 + 2ab + b^2$.

Sketch. $(a+b)^2 = (a+b)(a+b) = (a+b)a + (a+b)b = a^2 + ba + ab + b^2$ (by the distributive law twice). Now use commutativity. ◊

Solution. Here is a fairly complete proof from the basic properties of a field. This includes more details than we will probably ever want to put down again.

$$
\begin{aligned}
(a+b)^2 &= (a+b)(a+b) && \text{by definition of squaring} \\
&= (a+b)a + (a+b)b && \text{by the distributive law} \\
&= a(a+b) + b(a+b) && \text{by commutativity of mult.} \\
&= (aa+ab) + (ba+bb) && \text{by the distributive law} \\
&= (a^2+ab) + (ba+b^2) && \text{by definition of squaring} \\
&= (a^2+ab) + (ab+b^2) && \text{by commutativity of mult.} \\
&= [(a^2+ab)+ab] + b^2 && \text{by associativity of addition}
\end{aligned}
$$

$$
\begin{aligned}
&= [a^2 + (ab + ab)] + b^2 && \text{by associativity of addition} \\
&= [a^2 + (1 \cdot (ab) + 1 \cdot (ab))] + b^2 && \\
&= [a^2 + ((ab) \cdot 1 + (ab) \cdot 1)] + b^2 && \text{by commutativity of mult.} \\
&= [a^2 + (ab(1 + 1))] + b^2 && \text{by the distributive law} \\
&= [a^2 + (ab(2))] + b^2 && \\
&= [a^2 + 2ab] + b^2 && \text{by commutativity of mult.} \\
&= a^2 + 2ab + b^2 &&
\end{aligned}
$$

♦

◇ **1.1-2.** In a field, show that if $b \neq 0$ and $d \neq 0$, then

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}.$$

Solution. Recall that division is defined by $\frac{x}{y} = xy^{-1}$. We compute:

$$
\begin{aligned}
\frac{a}{b} + \frac{c}{d} &= \left(\frac{a}{b} + \frac{c}{d}\right)\left((bd)(bd)^{-1}\right) && \text{multiply by one} \\
&= \left(\left(\frac{a}{b} + \frac{c}{d}\right)(bd)\right)(bd)^{-1} && \text{associative law} \\
&= ((ab^{-1} + cd^{-1})(bd))(bd)^{-1} && \text{definition of division} \\
&= ((ab^{-1})(bd) + (cd^{-1})(bd))(bd)^{-1} && \text{distributive law} \\
&= ((ab^{-1})b)d + (cd^{-1})d)b)(bd)^{-1} && \text{assoc. and comm. laws} \\
&= ((a(b^{-1}b))d + (c(d^{-1}d))b)(bd)^{-1} && \text{associative law} \\
&= ((a(1))d + (c(1))b)(bd)^{-1} && \text{reciprocals} \\
&= (ad + cb)(bd)^{-1} = (ad + bc)(bd)^{-1} && \text{property of one} \\
&= \frac{ad + bc}{bd} && \text{definition of division}
\end{aligned}
$$

as claimed. ♦

◇ **1.1-3.** In an ordered field, if $a > b$, show that $a^2b < ab^2 + (a^3 - b^3)/3$.

Sketch. Expand $(a - b)^3 > 0$ and rearrange. ◊

Solution. Here is a solution with many, but not all steps. Some expansions and manipulations analogous to the expansion done in Exercise 1.1-1 have been grouped. The reader should be able to fill these in if requested.

$b \le a$	given
$b + (-b) \le a + (-b)$	by order axiom 15
$0 \le a + (-b)$	
$0 \le a - b$	by definition of subtraction
$0 \le (a-b)^3$	by order axiom 16 twice
$0 \le a^3 - 3a^2b + 3ab^2 - b^3$	similar to exercise 1
$0(1/3) \le (a^3 - 3a^2b + 3ab^2 - b^3)/3$	by order axiom 16
$0 \le (a^3 - 3a^2b + 3ab^2 - b^3)/3$	by 1.1.2 vi
$0 \le (ab^2 + (a^3 - b^3)/3) - a^2b$	why ?
$0 + a^2b \le ((ab^2 + (a^3 - b^3)/3) - a^2b) + a^2b$	by order axiom 15
$0 + a^2b \le (ab^2 + (a^3 - b^3)/3) + (-a^2b + a^2b)$	why ?
$0 + a^2b \le (ab^2 + (a^3 - b^3)/3) + 0$	why ?
$a^2b \le ab^2 + (a^3 - b^3)/3$	why ?

♦

◇ **1.1-4.** Prove that in an ordered field, if $\sqrt{2}$ is a positive number whose square is 2, then $\sqrt{2} < 3/2$. (Do this without using a numerical approximation for $\sqrt{2}$.)

Solution. It is convenient to set up a lemma first. This lemma corresponds to the fact that the squaring and square root functions are monotone increasing on the positive real numbers.

Lemma. *If $a > 0$ and $b > 0$, then $a^2 \le b^2 \iff a \le b$.*

Proof: Suppose $0 \le a \le b$. If we use Property 1.1.2 xi twice we find $a^2 \le ab$ and $ab \le b^2$. Transitivity of inequality gives $a^2 \le b^2$.

In the other direction, if $a^2 \le b^2$, then

$$0 \le b^2 - a^2 = (b-a)(b+a).$$

If $b + a = 0$, then $a = b = 0$ since both are non-negative. If it is not 0, then it is positive and so is $(b+a)^{-1}$. So

$$\begin{aligned} 0 \cdot (b+a)^{-1} &\le ((b-a)(b+a))(b+a)^{-1} = (b-a)((b+a)(b+a)^{-1}) \\ 0 &\le b - a \\ a &\le b \end{aligned}$$

We have implication in both directions as claimed.

Using this we can get our result. If $\sqrt{2} \geq 3/2$, we would have $2 \geq 9/4$ and so $8 \geq 9$. Subtracting 8 from both sides would give $0 \geq 1$, but we know this is false. Thus $\sqrt{2} \geq 3/2$ has led to a contradiction and cannot be true. We must have $\sqrt{2} < 3/2$ as claimed. ♦

◇ **1.1-5.** Give an example of a field with only three elements. Prove that it cannot be made into an ordered field.

Sketch. Let $\mathbb{F} = \{0, 1, 2\}$ with arithmetic mod 3. For example, $2 \cdot 2 = 1$ and $1 + 2 = 0$. To show it cannot be ordered, get a contradiction from (for example) $1 > 0$ so $1 + 1 = 2 > 0$ so $1 + 2 = 0 > 0$. ◊

Solution. Let $\mathbb{F} = \{0, 1, 2\}$ with arithmetic mod 3. For example, $2 \cdot 2 = 1$ and $1 + 2 = 0$. The commutative associative and distributive properties work for modular arithmetic with any base. When the base is a prime, the result is a field. In particular in arithmetic modulo 3 we have

$$1 \cdot 1 = 1 \quad \text{and} \quad 2 \cdot 2 = 1.$$

Thus 1 and 2 are their own reciprocals. Since they are the only two nonzero elements, we have a field.

To show it cannot be ordered, get a contradiction from (for example) $1 > 0$ so $1+1 = 2 > 0$ so $1+2 = 0 > 0$. We know that $0 \leq 1$ in any ordered field. So $1 \leq 1 + 1 = 2$ by order axiom 15. Transitivity gives $0 \leq 2$. So far there is no problem. But, if we add 1 to the inequality $1 \leq 2$ obtained above, we find $2 = 1 + 1 \leq 2 + 1 = 0$. So $2 \leq 0$. Thus $0 = 2$. If we multiply by 2, we get $0 = 0 \cdot 2 = 2 \cdot 2 = 1$. So $0 = 1$. But we know this is false. ♦

1.2 Completeness and the Real Number System

◇ **1.2-1.** In Example 1.2.10, let $\lambda = \lim_{n\to\infty} x_n$.

(a) Show that λ is a root of $\lambda^2 - \lambda - 2 = 0$.
(b) Find $\lim_{n\to\infty} x_n$.

Sketch. (a) Take limits as $n \to \infty$ in $x_n^2 = 2 + x_{n-1}$.
(b) 2

◊

Solution. (a) Letting $k = n-1$, we see that $k \to \infty$ if and only if $n \to \infty$. So $\lim_{k\to\infty} x_{n-1} = \lim_{k\to\infty} x_k = \lambda$. Since we know that the limit exists as a finite real number, we can use the arithmetic of limits to compute

$$\lambda^2 = \left(\lim_{n\to\infty} x_n\right)^2 = \lim_{n\to\infty} (x_n^2) = \lim_{n\to\infty} (2+x_{n-1}) = 2+\lim_{n\to\infty} x_{n-1} = 2+\lambda$$

So $\lambda^2 - \lambda - 2 = 0$ as claimed.

(b) We know that the limit λ is a solution to the equation $\lambda^2 - \lambda - 2 = 0$. This equation has solutions 2 and -1. Since all the terms of the sequence are positive, the limit cannot be -1 and must be 2.

♦

◇ **1.2-2.** Show that $3^n/n!$ converges to 0.

Solution. Let $a_n = \frac{3^n}{n!}$ If $n \geq 5$ we can compute

$$0 \leq \frac{3^n}{n!} = \frac{3\cdot 3\cdot 3\cdot 3\cdot 3}{1\cdot 2\cdot 3\cdot 4\cdot 5}\cdot\frac{3}{6}\cdot\frac{3}{7}\cdots\frac{3}{n} \leq \frac{243}{120}\cdot\left(\frac{1}{2}\right)^{n-5} = \frac{324}{5}\cdot\frac{1}{2^n}$$

Since $1/2^n$ goes to 0, so does any constant multiple of it. Thus a_n is trapped between 0 and something known to tend to 0, so it must tend to 0 also as claimed. ♦

◇ **1.2-3.** Let $x_n = \sqrt{n^2+1} - n$. Compute $\lim_{n\to\infty} x_n$.

Answer. 0. ◇

Solution. Let $x_n = \sqrt{n^2+1} - n$, and notice that

$$(\sqrt{n^2+1} - n)\cdot(\sqrt{n^2+1} + n) = (n^2+1) - n^2 = 1.$$

Thus we have

$$0 \leq x_n = \sqrt{n^2+1} - n = \frac{1}{\sqrt{n^2+1}+n} \leq \frac{1}{\sqrt{n^2}+n} = \frac{1}{2n} \leq \frac{1}{n}.$$

We know that $1/n \to 0$, so $x_n \to 0$ by the “Sandwich Lemma” 1.2.2. ♦

◇ **1.2-4.** Let x_n be a monotone increasing sequence such that $x_{n+1} - x_n \leq 1/n$. Must x_n converge?

Answer. No, not necessarily. ◇

Solution. If we put $x_1 = 1$ and suppose that $x_{n+1} = x_n + (1/n)$, then

$$\begin{aligned} x_2 &= 1 + \frac{1}{1} \\ x_3 &= x_2 + \frac{1}{2} = 1 + \frac{1}{1} + \frac{1}{2} \\ x_4 &= x_3 + \frac{1}{3} = 1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{3} \\ &\vdots \\ x_{n+1} &= x_n + \frac{1}{n} = 1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{n} \end{aligned}$$

Since we know that the harmonic series diverges to infinity, we see that the sequence $\langle x_n \rangle_1^\infty$ cannot converge. ♦

◇ **1.2-5.** Let $\mathbb{F}$ be an ordered field in which every *strictly* monotone increasing sequence bounded above converges. Prove that $\mathbb{F}$ is complete.

Sketch. From a (nontrivial) monotone sequence $\langle x_n \rangle_1^\infty$, extract a subsequence which is strictly monotone. ◇

Solution. We need to show that every increasing sequence which is bounded above must converge to an element of F whether the increase is strict or not. If the sequence is constant beyond some point, then it certainly converges to that constant. If not, then we can inductively extract a strictly increasing subsequence from it as follows. Let $\langle x_n \rangle_1^\infty$ be the sequence, and net $n(1) = 1$. Let $n(2)$ be the first integer larger than 1 with $x_{n(1)} < x_{n(2)}$. There must be such an index since the sequence never again gets smaller than $x_{n(1)}$ and it is not constantly equal to $x_{n(1)}$ beyond that point. Repeat this process. Having selected $n(1) < n(2) < \cdots < n(k) < \ldots$ with $x_{n(1)} < x_{n(2)} < \cdots < x_{n(k)} \ldots$, let $n(k+1)$ be the first index larger than $n(k)$ with $x_{n(k)} < x_{n(k+1)}$. Such an index exists since the sequence never again is smaller than $x_{(n(k)}$, and it is not constantly equal to $x_{n(k)}$ beyond this point. This inductively produces indices $n(1) < n(2) < n(3) < \ldots$ with $x_{n(1)} < x_{n(2)} < x_{n(3)} < \ldots$. By hypothesis, this strictly increasing sequence must converge to some element λ of F since it is bounded above by the same bound as the original sequence. We claim that the whole sequence must converge to λ. If $\varepsilon > 0$, then there is a J such that $\left| x_{n(k)} - \lambda \right| < \varepsilon$ whenever $k \geq J$. Put $N = n(J)$. If $n \geq N$, there is a $k > J$ with $n(J) \leq n \leq n(k)$. So $x_{n(J)} \leq x_n \leq x_{n(k)} \leq \lambda$. Thus $|x_n - \lambda| = \lambda - x_n \leq \lambda - x_{n(J)} < \varepsilon$. Thus $\lim_{n\to\infty} x_n = \lambda$ as claimed. ♦

1.3 Least Upper Bounds

◇ **1.3-1.** Let $S = \{x \mid x^3 < 1\}$. Find $\sup S$. Is S bounded below?

Answer. $\sup(S) = 1$; S is not bounded below. ♢

Solution. If $x \leq 0$, then $x^3 \leq 0 < 1$, so $x \in S$. If $0 < x < 1$, then $0 < x^3 < 1$, so $x \in S$. If $x \geq 1$, then $x^3 \geq 1$, so $x \notin S$. Thus $S = \{x \in \mathbb{R} \mid x < 1\}$. The number 1 is an upper bound for S, but nothing smaller is, so $\sup S = 1$. Since all negative numbers are in S, the set is not bounded below. ♦

◇ **1.3-2.** Consider an increasing sequence x_n that is bounded above and that converges to x. Let $S = \{x_n \mid n = 1, 2, 3, \dots\}$. Give a plausibility argument that $x = \sup S$.

Sketch. Since the sequence is increasing, its limit, x, must be at least as large as any of the terms of the sequence. So it is an upper bound. If $t < x$, then the terms x_n must eventually be above t since they come arbitrarily close to x. So t cannot be an upper bound. Thus x must be the smallest upper bound. We should have $x = \sup S$. ♢

Solution. Here is one way to make the arguments given above a bit more precise. Let $S = \{x_1, x_2, x_3, \dots\}$. We are assuming that

$$x_1 \leq x_2 \leq x_3 \leq \dots$$

and that

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

Suppose there were an index k with $x_k > x$, then, since the sequence is monotone increasing, we would have

$$x < x_k \leq x_{k+1} \leq x_{k+2} \leq \dots .$$

So that we would have $x < x_k \leq x_n$ for all $n \geq k$. If we put $\varepsilon = (x_k - x)/2$, we would have $|x_n = x| > \varepsilon$ for every n with $n > k$. The sequence could not converge to x. Thus it must be that $x_k \leq x$ for every k. The number x is thus an upper bound for the set S. So $\sup S \leq x$.

On the other hand, if $\varepsilon > 0$, then we must have $|x_n - x| < \varepsilon$ for large enough n. That is,

$$-\varepsilon < x_n - x < \varepsilon.$$

So

$$x - \varepsilon < x_n < x + \varepsilon$$

for large enough n. In particular, the number $x - \varepsilon$ is not an upper bound for S. No number smaller than x can be an upper bound, so $x \le \sup S$.

We have inequality in both directions, so $\sup S = x$ as claimed. ◆

◇ **1.3-3.** If $P \subset Q \subset \mathbb{R}$, $P \neq \varnothing$, and P and Q are bounded above, show that $\sup P \le \sup Q$.

Sketch. $\sup(Q)$ is an upper bound for P, so $\sup(Q) \ge \sup(P)$. ◊

Solution. Let $x \in P$. Then $x \in Q$ since $P \subseteq Q$. So $x \le \sup Q$ since $\sup Q$ is an upper bound for the set Q. Thus $x \le \sup Q$ for every x in P. Thus $\sup Q$ is an upper bound for the set P. Since $\sup P$ is the least upper bond for P, we must have $\sup P \le \sup Q$ as claimed. ◆

◇ **1.3-4.** Let $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ be bounded below and define $A + B = \{x + y \mid x \in A \text{ and } y \in B\}$. Is it true that $\inf(A + B) = \inf A + \inf B$?

Answer. Yes. ◊

Solution. First suppose $z \in A + B$, then there are points $x \in A$ and $y \in B$ with $z = x + y$. Certainly $\inf A \le x$ and $\inf B \le y$. So

$$\inf A + \inf B \le x + y = z.$$

Thus $\inf A + \inf B$ is a lower bound for the set $A + B$. So $\inf A + \inf B \le \inf(A + B)$.

To get the opposite inequality, let $\varepsilon \ge 0$. There must be points $x \in A$ and $y \in B$ with

$$\inf A \le x < \inf A + \frac{\varepsilon}{2} \qquad \text{and} \qquad \inf B \le y < \inf B + \frac{\varepsilon}{2}.$$

Since $x + y \in A + B$ we must have

$$\inf(A + B) \le x + y \le \inf A + \frac{\varepsilon}{2} + \inf B + \frac{\varepsilon}{2} = \inf A + \inf B + \varepsilon.$$

Since this holds for every $\varepsilon > 0$, we must have $\inf(A + B) \le \inf A + \inf B$.

We have inequality in both directions, so $\inf(A+B) = \inf A + \inf B$. ◆

◇ **1.3-5.** Let $S \subset [0, 1]$ consist of all infinite decimal expansions $x = 0.a_1a_2a_3\cdots$ where all but finitely many digits are 5 or 6. Find $\sup S$.

Answer. $\sup(S) = 1$. ◊

Solution. The numbers $x_n = 0.99999\ldots9995555555\ldots$ consisting of n 9's followed by infinitely many 5's are all in S. Since these come as close to 1 as we want, we must have $\sup S = 1$. ◆

1.4 Cauchy Sequences

◇ **1.4-1.** Let x_n satisfy $|x_n - x_{n+1}| < 1/3^n$. Show that x_n converges.

Sketch. $|x_n - x_{n+k}| \le \sum_{i-n}^{n+k-1} |x_i - x_{i+1}| \le \sum_{i=n}^{\infty}(1/3^i) = 2/3^{n-1}$. So, for all $\varepsilon > 0$, we can choose N such that $2/3^{n-1} < \varepsilon$. We get a Cauchy sequence, so $\langle x_n \rangle_1^\infty$ converges. ♢

Solution. Let's simply expand the computation sketched above somewhat:

$$\begin{aligned}
|x_n - x_{n+k}| &= |x_n - x_{n+1} + x_{n+1} - x_{n+2} + x_{n+2} - + \cdots + x_{n+1} - x_{n+k}| \\
&\le |x_n - x_{n+1}| + |x_{n+1} - x_{n+2}| + \cdots + |x_{n+k-1} - x_{n+k}| \\
&\le \frac{1}{3^n} + \frac{1}{3^{n+1}} + \cdots + \frac{1}{3^{n+k-1}} \\
&\le \frac{1}{3^n}\left(1 + \frac{1}{3} + \frac{1}{3^2} + \cdots + \frac{1}{3^{k-1}}\right) \\
&\le \frac{1}{3^n}\left(1 + \frac{1}{3} + \frac{1}{3^2} + \dots\right) \\
&\le \frac{1}{3^n}\,\frac{1}{1-(1/3)} = \frac{2}{3^{n-1}}
\end{aligned}$$

We know this tends to 0, so if $\varepsilon > 0$, we can pick N so that $2/3^{N-1} < \varepsilon$. If $n \ge N$ and $k > 0$, we have

$$|x_n - x_{n+k}| \le \frac{2}{3^{n-1}} \le \frac{2}{3^{N-1}} < \varepsilon.$$

This shows that our sequence is a Cauchy sequence in $\mathbb{R}$. So, by completeness, there must be a $\lambda \in \mathbb{R}$ such that $\lim_{n\to\infty} x_n = \lambda$. The sequence converges to λ. ♦

◇ **1.4-2.** Show that the sequence $x_n = e^{\sin(5n)}$ has a convergent subsequence.

Solution. We know that $-1 \le \sin(5n) \le 1$, so $1/e \le x_n \le e$ Thus the sequence is bounded. By Theorem 1.4.3, it must have a convergent subsequence.

Notice that we are able to conclude that there must be a convergent subsequence, but we do not obtain a way actually to get our hands explicitly on such a subsequence. ♦

◇ **1.4-3.** Find a bounded sequence with three subsequences converging to three different numbers.

Sketch. A possibility is $1, 0, -1, 1, 0, -1, 1, 0, -1, \ldots$. ◊

Discussion. We simply need to pick any three sequences converging to different limits and then interleave these sequences as subsequences of one big sequence. That is, a solution is produced by picking three different numbers a, b, and c, and sequences $\langle a_n\rangle_1^\infty$, $\langle b_n\rangle_1^\infty$, and $\langle c_n\rangle_1^\infty$ converging to them.

$$a_n \to a \quad ; \qquad b_n \to b \quad ; \text{ and } \qquad c_n \to c\,.$$

Finally, interleave them to form one sequence.

$$a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3, c_3, \ldots .$$

Such a presentation is probably clear enough, but, if required, one could write a specification of these terms.

$$x_n = \begin{cases} a_k & , \text{ if } n = 3k+1 \\ b_k & , \text{ if } n = 3k+2 \\ c_k & , \text{ if } n = 3k+3 \end{cases} \qquad \text{with } k = 0, 1, 2, 3, \ldots .$$

The example given is about as simple as possible with each of the three subsequences being a constant sequence. ◊

◇ **1.4-4.** Let x_n be a Cauchy sequence. Suppose that for every $\varepsilon > 0$ there is some $n > 1/\varepsilon$ such that $|x_n| < \varepsilon$. Prove that $x_n \to 0$.

Discussion. The assumption that $\langle x_n\rangle_1^\infty$ is a Cauchy sequence says that far out in the sequence, all of the terms are close to each other. The second assumption, that for every $\varepsilon > 0$ there is an $n > 1/\varepsilon$ such that $|x_n| < \varepsilon$, says, more or less, that no matter how far out we go in the sequence there will be at least one term out beyond that point which is small. Combining these two produces the proof. If some of the points far out in the sequence are small and all of the points far out in the sequence are close together, then all of the terms far out in the sequence must be small. The technical tool used to merge the two assumptions is the triangle inequality. ◊

Solution. Let $\varepsilon > 0$. Since the sequence is a Cauchy sequence, there is an N_1 such that $|x_n - x_k| < \varepsilon/2$ whenever $n \ge N_1$ and $k \ge N_1$.

Pick $N_2 > N_1$ large enough so that $1/N_2 < \varepsilon/2$. By hypothesis there is at least one index $n > 1/(1/N_2) = N_2$ with $|x_n| < 1/N_2$.

If $k \ge N_1$, then both k and n are at least as large as N_1 and we can compute

$$|x_k| = |x_k - x_n + x_n| \le |x_k - x_n| + |x_n| < \frac{\varepsilon}{2} + \frac{1}{N_2} < \varepsilon$$

Thus $x_k \to 0$ as claimed. ♦

◇ **1.4-5.** True or false: If x_n is a Cauchy sequence, then for n and m large enough, $d(x_{n+1}, x_{m+1}) \leq d(x_n, x_m)$.

Answer. False. ♢

Solution. If this were true, it would hold in particular with m selected as $m = n + 1$. That is, we should have $d(x_{n+1}, x_{n+2}) \leq d(x_n, x_{n+1})$ for large enough n. But this need not be true. Consider the sequence

$$1, 0, 0, \frac{1}{2}, 0, 0, \frac{1}{3}, 0, 0, \frac{1}{4}, 0, 0, \frac{1}{5}, \ldots .$$

Then

$$\begin{aligned}
d(x_1, x_2) &= 1 \\
d(x_2, x_3) &= 0 \\
d(x_3, x_4) &= \frac{1}{2} \\
d(x_4, x_5) &= \frac{1}{2} \\
d(x_5, x_6) &= 0 \\
d(x_6, x_7) &= \frac{1}{3} \\
d(x_7, x_8) &= \frac{1}{3} \\
d(x_8, x_9) &= 0 \\
&\vdots
\end{aligned}$$

This sequence converges to 0 and is certainly a Cauchy sequence. But the differences of succeeding terms keep dropping to 0 and then coming back up a bit. ♦

1.5 Cluster Points; lim inf and lim sup

◇ **1.5-1.** Let $x_n = 3 + (-1)^n(1 + 1/n)$. Calculate $\liminf x_n$ and $\limsup x_n$.

Answer. $\liminf(x_n) = 2$. $\limsup(x_n) = 4$. ♢

Solution. If n is even, then $x_n = 3 + 1 + \frac{1}{n} = 4 + \frac{1}{n}$. These converge to 4. That is, the subsequence $x_2, x_4, x_6, \ldots$ converges to 4. If n is odd, then $x_n = 3 - \left(1 + \frac{1}{n}\right) = 2 - \frac{1}{n}$. These converge to 2. That is, the subsequence $x_1, x_3, x_5, \ldots$ converges to 2. All terms of the sequence are involved in these two subsequences, so there are none left over to form a subsequence converging to anything else. The only two cluster points are 2 and 4. So $\liminf x_n = 2$ and $\limsup x_n = 4$. ♦

◇ **1.5-2.** Find a sequence x_n with $\limsup x_n = 5$ and $\liminf x_n = -3$.

Solution. We need a sequence with at least two cluster points. The smallest is to be -3 and the largest is to be 5. So there must be a subsequence converging to -3 and another converging to 5. The easiest way to accomplish this is to interweave two constant sequences and use the resulting sequence

$$-3, 5, -3, 5, -3, 5, -3, 5, \dots \quad .$$

If you really want to write a formula for this, you could do something like

$$x_n = \begin{cases} -3, & \text{if } n \text{ is odd} \\ 5, & \text{if } n \text{ is even} \end{cases} \quad .$$

♦

◇ **1.5-3.** Let x_n be a sequence with $\limsup x_n = b \in \mathbb{R}$ and $\liminf x_n = a \in \mathbb{R}$. Show that x_n has subsequences u_n and l_n with $u_n \to b$ and $l_n \to a$.

Sketch. Use Proposition 1.5.5 to show that there are points $x_{N(n)}$ within $1/n$ of a (or b). Make sure you end up with a subsequence. ◇

Solution. If b is a finite real number and $b = \limsup x_n$, then for each N and for each $\varepsilon > 0$ there is an index n with $n > N$ and $b - \varepsilon < x_n \le b + \varepsilon$. Use this repeatedly to generate the desired subsequence.

Step One: There is an index $n(1)$ such that $\left|b - x_{n(1)}\right| < 1$.

Step Two: There is an index $n(2)$ such that $n(2) > n(1)$ and $\left|b - x_{n(2)}\right| < 1/2$.

. . . and so forth.

Induction Step: Having selected indices

$$n(1) < n(2) < n(3) < \cdots < n(k)$$

with $\left|b - x_{n(j)}\right| < 1/j$. for $1 \le j \le k$, there is an index $n(k+1)$ with $n(k+1) > n(k)$ and $\left|b - x_{n(k+1)}\right| < 1/(k+1)$

By induction, this gives a subsequence converging to b.

The inductive definition of a subsequence converging to the limit inferior is similar. Notice that the tricky part is to make the selection of indices dependent on the earlier ones so that the indices used are increasing. This is what makes the selection of terms a subsequence. They remain in the same order as they appeared in the original sequence. ♦

◇ **1.5-4.** Let $\limsup x_n = 2$. True or false: If n is large enough, then $x_n > 1.99$.

Answer. False. ♢

Solution. The largest cluster point is 2, but this need not be the only cluster point. Consider for example the sequence

$$1, 2, 1, 2, 1, 2, 1, 2, \ldots,$$

where

$$x_n = \begin{cases} 1, & \text{if } n \text{ is odd} \\ 2, & \text{if } n \text{ is even} \end{cases} .$$

In this example we have $x_n < 1.99$ for all odd n. ♦

◇ **1.5-5.** True or false: If $\limsup x_n = b$, then for n large enough, $x_n \leq b$.

Answer. False. ♢

Solution. A counterexample is easy to give. For example, the sequence

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots$$

converges to a limit of 0. So the lim sup and the lim inf are both equal to 0. All terms of the sequence are larger than the lim sup. ♦

1.6 Euclidean Space

◇ **1.6-1.** If $\| x + y \| = \| x \| + \| y \|$, show that x and y lie on the same ray from the origin.

Sketch. If equality holds in the Cauchy-Schwarz or triangle inequality, then x and y are parallel. Expand $\| x + y \|^2 = (\| x \| + \| y \|)^2$ to obtain $\| x \| \cdot \| y \| = \langle x, y \rangle = \| x \| \cdot \| y \| \cdot \cos(\vartheta)$. Use this directly or let $u = y / \| y \|$ and $z = x - \langle x, u \rangle u$. Check that $\langle z, u \rangle = 0$ and that $\| x \|^2 = \| z \|^2 + |\langle x, u \rangle|^2 = \| x \|^2$. Conclude that $z = 0$ and $x = (\langle x, y \rangle / \| y \|^2) y$. ♢

Solution. If either x or y is 0, then they are certainly on the same ray through the origin. If y is not the zero vector, we can let u be the unit vector $u = (1/ \| y \|)\, y$. The projection of x in the direction of y is then

$v = \langle x, u \rangle \, u = (\langle x, y \rangle / \| y \|^2) \, y$. We claim that this is equal to x. Let $z = x - v$. So $x = z + v$. But z and v are orthogonal.

$$\begin{aligned} \langle z, v \rangle &= \langle x - v, v \rangle = \langle x, v \rangle - \langle v, v \rangle \\ &= \langle x, u \rangle^2 - \langle x, u \rangle^2 \, \langle u, u \rangle = 0 \end{aligned}$$

Since

$$\| x + y \|^2 = (\| x \| + \| y \|)^2 = \| x \|^2 + 2 \, \| x \| \, \| y \| + \| y \|^2 ,$$

we have

$$\langle x, x \rangle + 2 \langle x, y \rangle + \langle y, y \rangle = \langle x, x \rangle + 2 \, \| x \| \, \| y \| + \langle y, y \rangle$$

and

$$\langle x, y \rangle = \| x \| \, \| y \| .$$

So $\langle x, u \rangle = \| x \|$.

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

We must have $\| z \| = 0$. So $z = 0$ and $x = v$ as claimed.

Notice that we have written the proof for a real inner product space. Exercise: what happens with a complex inner product ? ♦

◇ **1.6-2.** What is the angle between $(3, 2, 2)$ and $(0, 1, 0)$?

Solution. Let $v = (3, 2, 2)$ and $w = (0, 1, 0)$, Then $\| v \| = \sqrt{9 + 4 + 4} = \sqrt{17}$ and $\| w \| = \sqrt{0 + 1 + 0} = 1$. So

$$v \cdot w = \| v \| \, \| w \| \cos \vartheta = \sqrt{17} \cos \vartheta$$

where ϑ is the angle between v and w. But we also know that $v \cdot w = (3)(0) + (2)(1) + (2)(0) = 2$. So

$$2 = \sqrt{17} \cos \vartheta .$$

Thus $\vartheta = \arccos (2/\sqrt{17}) = \arccos (0.48507) \approx 1.06435$ radians. ♦

◇ **1.6-3.** Find the orthogonal complement of the plane spanned by $(3, 2, 2)$ and $(0, 1, 0)$ in $\mathbb{R}^3$.

Answer. $\{\lambda \cdot (-2, 0, 3) \in \mathbb{R}^3 \mid \lambda \in \mathbb{R}\}$. ◇

Solution. If (x, y, z) is to be in the orthogonal complement, it must be perpendicular to $(3, 2, 2)$ and to $(0, 1, 0)$. Setting both dot products equal to 0 gives two equations.

$$3x + 2y + 2z = 0 \qquad \text{and} \qquad y = 0$$

So $y = 0$ and $z = (-3/2)x$. Any scalar multiple of $(-2, 0, 3)$ will do. The orthogonal complement is the straight line (one dimensional subspace)

$$\{\lambda \cdot (-2, 0, 3) \in \mathbb{R}^3 \mid \lambda \in \mathbb{R}\}.$$

♦

◇ **1.6-4.** Describe the sets $B = \{x \in \mathbb{R}^3 \mid \| x \| \leq 3\}$ and $Q = \{x \in \mathbb{R}^3 \mid \| x \| < 3\}$.

Solution. The set $B = \{x \in \mathbb{R}^3 \mid \| x \| \leq 3\}$ is a solid ball of radius 3 centered at the origin together with the sphere which forms its surface. The set $Q = \{x \in \mathbb{R}^3 \mid \| x \| < 3\}$ is the same solid ball, but not including the surface sphere. B is referred to as a *closed* ball, and Q is called an *open* ball. The corresponding sets in $\mathbb{R}^2$ are called closed disks and open disks and are illustrated in the figure. ♦

◇ **1.6-5.** Find the equation of the line through $(1, 1, 1)$ and $(2, 3, 4)$. Is this line a linear subspace?

Answer. $x = (y + 1)/2 = (z + 2)/3$. Or $P(t) = (1 + t, 1 + 2t, 1 + 3t)$. This line is not a linear subspace since $(0, 0, 0)$ is not on it. ◇

Solution. Let $A = (1, 1, 1)$ and $B = (2, 3, 4)$. The vector from A to B is $v = B - A =< 1, 2, 3 >$. For each real t, the point $P(t) = A + tv = (1 + t, 1 + 2t, 1 + 3t)$ will lie on the line through A and B. This is a parameterization of that line. Note that $P(0) = A$ and $P(1) = B$. If t is between 0 and 1, then $P(t)$ is on the straight line segment between A and B.

Relations among x, y, and z along the line can be derived from the parameterization.

$$y = 1 + 2t = 1 + t + t = x + (x - 1) = 2x - 1.$$

$$z = 1 + 3t = 1 + t + 2t = x + 2(x - 1) = 3x - 2.$$

or

$$(y + 1)/2 = x = (z + 2)/3.$$

♦

1.7 Norms, Inner Products, and Metrics

◇ **1.7-1.** In $\mathcal{C}([0,1])$ find $d(f,g)$ where $f(x)=1$ and $g(x)=x$ for both the sup norm and the norm given in Example 1.7.7.

Answer. The distance from the sup norm is 1. That from Example 1.7.7 is $1/\sqrt{3}$. ♢

Solution. The distance between f and g as defined by the sup norm is

$$d_\infty(f,g) = \| f-g \|_\infty = \sup\{|f(x)-g(x)| \mid 0 \le x \le 1\} = 1.$$

That defined by the inner product in Example 1.7.7 is

$$\begin{aligned} d_2(f,g) = \| f-g \|_2 &= \sqrt{\langle f-g, f-g \rangle} \\ &= \left(\int_0^1 |f(x)-g(x)|^2 \, dx \right)^{1/2} \\ &= \left(\int_0^1 |1-x|^2 \, dx \right)^{1/2} = \frac{1}{\sqrt{3}}. \end{aligned}$$

♦

◇ **1.7-2.** Identify the space of polynomials of degree $n-1$ in a real variable with $\mathbb{R}^n$. In the corresponding Euclidean metric, calculate $d(f,g)$ where $f(x)=1$ and $g(x)=x$.

Answer. $d(f,g)=\sqrt{2}$. ♢

Solution. The sense of the word "identify" here is that we are to show that the two spaces have essentially the same structure by establishing a correspondence between the vectors in the spaces which is also a correspondence between the operations. the technical word for this in abstract algebra is "isomorphism". We say that the spaces are isomorphic.

Let $\mathcal{V}$ be the space of all polynomials of degree no more than $n-1$. A polynomial p in $\mathcal{V}$ can be represented uniquely in the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}$$

Looked at in the other direction we have a function $\varphi : \mathbb{R}^n \to \mathcal{V}$ defined by

$$\varphi : (a_0, a_1, \ldots, a_{n-1}) \mapsto p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}$$

Since two polynomials are equal if and only if they have the same coefficients, we see that the function φ takes $\mathbb{R}^n$ one-to-one onto $\mathcal{V}$. Furthermore, if v and w are in $\mathbb{R}^n$ and λ is a number, we evidently have

$$\varphi(v+w) = \varphi(v) + \varphi(w) \qquad \text{and} \qquad \varphi(\lambda v) = \lambda \varphi(v).$$

The function f is the correspondence or "identification" requested. It is said to be an "isomorphism" from $\mathbb{R}^n$ onto $\mathcal{V}$.

Under that correspondence, the polynomial $f(x) = 1$ corresponds to the vector $v = (1, 0, 0, 0, \dots, 0)$ while $g(x) = x$ corresponds to $w = (0, 1, 0, 0, \dots, 0)$. So the "corresponding Euclidean metric" distance requested is

$$d(f, g) = d(v, w) = \sqrt{1 + 1 + 0 + \cdots + 0} = \sqrt{2}.$$

♦

◇ **1.7-3.** Put the inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$ on $\mathcal{C}([0, 1])$. Verify the Cauchy-Schwarz inequality for $f(x) = 1$ and $g(x) = x$.

Sketch. $\sqrt{\langle f, f \rangle} = 1$, $\sqrt{\langle g, g \rangle} = 1/\sqrt{3}$, and $\langle f, g \rangle = 1/2$. So $|\langle f, g \rangle| \le \sqrt{\langle f, f \rangle} \cdot \sqrt{\langle g, g \rangle}$ is true. ◇

Solution. $\sqrt{\langle f, f \rangle} = \left(\int_0^1 (f(x))^2\, dx\right)^{1/2} = \left(\int_0^1 1\, dx\right)^{1/2} = 1,$

$\sqrt{\langle g, g \rangle} = \left(\int_0^1 (g(x))^2\, dx\right)^{1/2} = \left(\int_0^1 x^2\, dx\right)^{1/2} = 1/\sqrt{3},$

and $\langle f, g \rangle = \int_0^1 f(x)g(x)\, dx = \int_0^1 x\, dx = 1/2$.

So $|\langle f, g \rangle| \le \sqrt{\langle f, f \rangle} \cdot \sqrt{\langle g, g \rangle}$ is true. ♦

◇ **1.7-4.** Using the inner product in Exercise 1.7-3, verify the triangle inequality for $f(x) = x$ and $g(x) = x^2$.

Solution. Direct computation gives

$$\begin{aligned} \| f + g \|^2 &= \int_0^1 (x + x^2)^2\, dx = \int_0^1 (x^2 + 2x^3 + x^4)\, dx \\ &= \frac{1}{3} + \frac{1}{2} + \frac{1}{5} = \frac{10 + 15 + 6}{30} = \frac{31}{30}. \end{aligned}$$

So $\| f + g \| = \sqrt{31/30}$. Also

$$\| f \|^2 = \int_0^1 x^2\, dx = \frac{1}{3} \qquad \text{and} \qquad \| g \|^2 = \int_0^1 x^4\, dx = \frac{1}{5}.$$

So $\| f \| + \| g \| = 1/\sqrt{3} + 1/\sqrt{5} = (\sqrt{5} + \sqrt{3})/\sqrt{15}$. The triangle inequality becomes

$$\sqrt{\frac{31}{30}} \le \frac{\sqrt{5} + \sqrt{3}}{\sqrt{15}} = \frac{\sqrt{10} + \sqrt{6}}{\sqrt{30}}$$

This is equivalent to $\sqrt{31} \le \sqrt{10} + \sqrt{6}$ or, squaring, to $31 \le 10 + 2\sqrt{60} + 6 = 16 + 4\sqrt{15}$. By subtracting 16, we see that this is equivalent to $15 \le 4\sqrt{15}$ which is equivalent to $\sqrt{15} \le 4$ or $15 \le 16$. This is certainly true, so the triangle inequality does indeed hold for these two functions. ♦

◇ **1.7-5.** Show that $\|\cdot\|_\infty$ is not the norm defined by the inner product in Example 1.7.7.

Sketch. For $f(x) = x$, $\| f \|_\infty = 1$, but $\langle f, f \rangle^{1/2} = 1/\sqrt{3}$. So these norms are different. ◇

Solution. All we have to do to show that the norms are different is to display a function to which they assign different numbers. Let $f(x) = x$. The norm defined by the inner product of Example 1.7.7 assigns to f the number

$$\| f \|_2 = \left(\int_0^1 |f|^2 \, dx \right)^{1/2} = \left(\int_0^1 x^2 \, dx \right)^{1/2} = \frac{1}{\sqrt{3}}$$

But

$$\| f \|_\infty = \sup\{|f(x)| \mid x \in [0,1]\} = \sup\{x \mid x \in [0,1]\} = 1.$$

Since $1/\sqrt{3}$ is not equal to 1, these norms are different. ♦

1.8 The Complex Numbers

◇ **1.8-1.** Express the following complex numbers in the form $a + ib$:

(a) $(2 + 3i) + (4 + i)$
(b) $(2 + 3i)/(4 + i)$
(c) $1/i \ + \ 3/(1 + i)$

Answer. (a) $6 + 4i$
(b) $(11/5) + 2i$
(c) $(3/2) - (5/2)i$

◇

Solution. (a) $(2 + 3i) + (4 + i) = (2 + 4) + (3 + 1)i = 6 + 4i$.
(b) $\dfrac{2+3i}{4+i} = \dfrac{2+3i}{4+i}\,\dfrac{4-i}{4-i} = \dfrac{8+12i-2i+3}{4+1} = \dfrac{11+10i}{5}$
(c) $\dfrac{1}{i} + \dfrac{3}{1+i} = \dfrac{1}{i}\,\dfrac{-i}{-i} + \dfrac{3}{1+i}\,\dfrac{1-i}{1-i} = \dfrac{-i}{1} + \dfrac{3-3i}{1+1} = \dfrac{3}{2} - \dfrac{5}{2}\,i$

♦

◇ **1.8-2.** Find the real and imaginary parts of the following, where $z = x + iy$:

(a) $(z + 1)/(2z - 5)$

(b) z^3

Answer. (a)

$$\operatorname{Re}\left(\frac{z+1}{2z-5}\right) = \frac{(x+1)(2x-5)+2y^2}{(2x-5)^2+4y^2} \quad \text{and}$$
$$\operatorname{Im}\left(\frac{z+1}{2z-5}\right) = \frac{y(2x-5)-2y(x+1)}{(2x-5)^2+4y^2}$$

(b) $\operatorname{Re}\left(z^3\right) = x^3 - 3xy^2$ and $\operatorname{Im}\left(z^3\right) = 3x^2y - y^3$.

◊

Solution. (a)

$$\begin{aligned}\frac{z+1}{2z-5} &= \frac{x+1+yi}{2x-5+2yi} = \frac{x+1+yi}{2x-5+2yi}\,\frac{2x-5-2yi}{2x-5-2yi} \\ &= \frac{(x+1)(2x-5)+2y^2+\left(y(2x-5)-2y(x+1)\right)i}{(2x-5)^2+4y^2} \\ &= \frac{(x+1)(2x-5)+2y^2}{(2x-5)^2+4y^2} + \frac{y(2x-5)-2y(x+1)}{(2x-5)^2+4y^2}\,i\end{aligned}$$

So

$$\operatorname{Re}\left(\frac{z+1}{2z-5}\right) = \frac{(x+1)(2x-5)+2y^2}{(2x-5)^2+4y^2} \quad \text{and}$$
$$\operatorname{Im}\left(\frac{z+1}{2z-5}\right) = \frac{y(2x-5)-2y(x+1)}{(2x-5)^2+4y^2}$$

(b) $z^3 = (x+iy)^3 = x^3 + 3x^2yi + 3x(yi)^2 + (yi)^3 = x^3 - 3xy^2 + (3x^2y - y^3)\,i$
So $\operatorname{Re}\left(z^3\right) = x^3 - 3xy^2$ and $\operatorname{Im}\left(z^3\right) = 3x^2y - y^3$.

⧫

◇ **1.8-3.** Is it true that $\operatorname{Re}(zw) = (\operatorname{Re} z)(\operatorname{Re} w)$?

Answer. No, not always. ◊

Solution. Consider the example $z = w = i$. Then $\operatorname{Re}(zw) = \operatorname{Re}(-1) = -1$, but $\operatorname{Re}(z)\operatorname{Re}(w) = 0 \cdot 0 = 0$. But -1 is not equal to 0. ⧫

◇ **1.8-4.** What is the complex conjugate of $(8-2i)^{10}/(4+6i)^5$?

Answer. $(8+2i)^{10}/(4-6i)^5$ ◊

Solution. Using Proposition 1.8.1 parts v, vi, and vii, we obtain $(8 + 2i)^{10}/(4 - 6i)^5$ as the complex conjugate of $(8 - 2i)^{10}/(4 + 6i)^5$. ♦

◇ **1.8-5.** Does $z^2 = |z|^2$? If so, prove this equality. If not, for what z is it true?

Answer. No; true iff z is real. ◊

Solution. Let $z = x + iy$. Then $z^2 = x^2 - y^2 + 2xyi$ while $|z|^2 = x^2 + y^2$. For these to be equal we must have $x^2 - y^2 + 2xyi = x^2 + y^2$. Since two complex numbers are equal if and only if their real and imaginary parts are equal, we must have

$$x^2 - y^2 = x^2 + y^2 \quad \text{and} \quad 2xy = 0.$$

This is true if and only if $y = 0$. That is, if and only if z is real. ♦

◇ **1.8-6.** Assuming either $|z| = 1$ or $|w| = 1$ and $\overline{z}w \neq 1$, prove that

$$\left|\frac{z - w}{1 - \overline{z}w}\right| = 1.$$

Solution. Notice that for complex numbers ζ we have $|\zeta| = 1$ if and only if $1 = |\zeta|^2 = \zeta\overline{\zeta}$. We apply this to the complex number $\zeta = \dfrac{z - w}{1 - \overline{z}w}$, using the properties of complex conjugation listed in Proposition 1.8.1 to compute $\zeta\overline{\zeta}$.

$$\begin{aligned}
\left|\frac{z - w}{1 - \overline{z}w}\right|^2 &= \left(\frac{z - w}{1 - \overline{z}w}\right)\overline{\left(\frac{z - w}{1 - \overline{z}w}\right)} \\
&= \left(\frac{z - w}{1 - \overline{z}w}\right)\left(\frac{\overline{z} - \overline{w}}{1 - z\overline{w}}\right) \\
&= \frac{z\overline{z} - w\overline{z} - z\overline{w} + w\overline{w}}{1 - \overline{z}w - z\overline{w} + \overline{z}x\overline{w}w} \\
&= \frac{|z|^2 - w\overline{z} - z\overline{w} + |w|^2}{1 - w\overline{z} - z\overline{w} + |z|^2\,|w|^2}
\end{aligned}$$

Since one of z or w has absolute value 1, the numerator and denominator are identical, and the value of this fraction is 1 as desired. ♦

◇ **1.8-7.** Prove that

(a) $e^{z_1} \cdot e^{z_2} = e^{z_1 + z_2}$.
(b) $e^z \neq 0$ for any complex number z.

(c) $|e^{i\theta}| = 1$ for each real number θ.
(d) $(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$.

Sketch. (a) Use trig identities and the properties of real exponents.
(b) $|e^z| = e^{Re(z)} \neq 0$. So $e^z \neq 0$.
(c) $\cos^2\theta + \sin^2\theta = 1$.
(d) Use induction.

◊

Solution. (a) Let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$. Then

$$\begin{aligned} e^{z_1} \cdot e^{z_2} &= e^{x_1}(\cos y_1 + i\sin y_1) \cdot e^{x_2}(\cos y_2 + i\sin y_2) \\ &= e^{x_1}e^{x_2}\left((\cos y_1\cos y_2 - \sin y_1\sin y_2) + i\,(\sin y_1\cos y_2 + \cos y_1\sin y_2)\right) \\ &= e^{x_1+x_2}\left(\cos(y_1+y_2) + i\,\sin(y_1+y_2)\right) \\ &= e^{x_1+x_2+i\,(y_1+y_2)} = e^{z_1+z_2}. \end{aligned}$$

(b) Using part (a), we find that

$$e^z \cdot e^{-z} = e^{z-z} = e^0 = 1.$$

So e^z cannot be 0.
(c) $\left|e^{i\theta}\right|^2 = |\cos\theta + i\,\sin\theta|^2 = \cos^2\theta + \sin^2\theta = 1$. So $\left|e^{i\theta}\right| = 1$.
(d) Here is a method which uses part (a).

$$(\cos\theta + i\sin\theta)^n = \left(e^{i\theta}\right)^n = e^{in\theta} = \cos n\theta + i\,\sin n\theta$$

(The middle equality uses part (a) and mathematical induction.)

♦

◇ **1.8-8.** Show that

(a) $e^z = 1$ iff $z = k2\pi i$ for some integer k.
(b) $e^{z_1} = e^{z_2}$ iff $z_1 - z_2 = k2\pi i$ for some integer k.

Solution. (a) Let $z = x + iy$ with x and y real. Then $e^z = e^x\cos y + ie^x\sin y$. So $|e^z|^2 = e^{2x}\cos^2 y + e^{2x}\sin^2 y = e^{2x}$, and $|e^z| = e^x$. If $e^z = 1$, this forces $e^x = 1$, so $x = 0$, and $1 = \cos y + i\,\sin y$. This means we must have $\sin y = 0$, so $y = n\pi$. for some integer n. This leaves $1 = \cos n\pi = (-1)^n$. So n must be even. Thus $z = x + iy = 2k\pi i$ for some integer k. Conversely, if $z = x + iy = 2k\pi i$ for some integer k, then $x = 0$ and $y = 2k\pi$. Direct computation gives $e^z = e^0(\cos 0 + i\sin 2k\pi) = 1$.

(b) We can use part (a) and Exercise 1.8-7.

$$\begin{aligned} e^{z_1} = e^{z_2} &\iff e^{z_1}e^{-z_2} = e^{z_2}e^{-z_2} \\ &\iff e^{z_1 - z_2} = e^{z_2 - z_2} = e^0 = 1 \\ &\iff z_1 - z_2 = 2k\pi i \qquad \text{for some integer } k. \end{aligned}$$

♦

◇ **1.8-9.** For what z does the sequence $z_n = nz^n$ converge?

Sketch. $|z| < 1$; then $z_n \to 0$ as $n \to \infty$. ◊

Solution. If $|z| \geq 1$, then $|nz^n| = n\,|z|^n \to \infty$ as $n \to \infty$. The sequence cannot converge in these cases. If $|z| < 1$, then $n\,|z|^n \to 0$ as $n \to \infty$. (One way to see this is to notice that the series $\sum_{n=1}^{\infty} n\,|z|^n$ converges by the ratio test.) So in these cases, the sequence converges to 0. ♦

Exercises for Chapter 1

◇ **1E-1.** For each of the following sets S, find $\sup(S)$ and $\inf(S)$ if they exist:

(a) $\{x \in \mathbb{R} \mid x^2 < 5\}$
(b) $\{x \in \mathbb{R} \mid x^2 > 7\}$
(c) $\{1/n \mid n$, an integer, $n > 0\}$
(d) $\{-1/n \mid n$ an integer, $n > 0\}$
(e) $\{.3, .33, .333, \dots\}$

Answer. (a) $\sup(S) = \sqrt{5}$; $\inf(S) = -\sqrt{5}$.
(b) Neither $\sup(S)$ nor $\inf(S)$ exist (except as $\pm\infty$).
(c) $\sup(S) = 1$; $\inf(S) = 0$.
(d) $\sup(S) = 0$; $\inf(S) = -1$.
(e) $\sup(S) = 1/3$; $\inf(S) = 0.3$.

◊

Solution. (a) $A = \{x \in \mathbb{R} \mid x^2 < 5\} =]-\sqrt{5}, \sqrt{5}[$, so $\inf A = -\sqrt{5}$ and $\sup A = \sqrt{5}$.
(b) $B = \{x \in \mathbb{R} \mid x^2 > 7\} =]-\infty, -\sqrt{7}[\cup]\sqrt{7}, \infty[$. The set is bounded neither above nor below, so $\sup B = \infty$, and $\inf B = -\infty$.

(c) $C = \{1/n \mid n \in \mathbb{Z}, n > 0\} = \{1, 1/2, 1/3, 1/4. \dots\}$. All elements are positive, so $\inf C \geq 0$, and they come arbitrarily close to 0, so $\inf C = 0$. The number 1 is in the set and nothing larger is in the set, so $\sup C = 1$.

(d) $D = \{-1/n \mid n \in \mathbb{Z}, n > 0\} = \{-1, -1/2, -1/3, -1/4. \dots\}$. All elements are negative, so $\sup D \leq 0$, and they come arbitrarily close to 0, so $\sup C = 0$. The number -1 is in the set and nothing smaller is in the set, so $\inf C = -1$.

(e) $E = \{0.3, 0.33, 0.333, \dots\}$. The number 0.3 is in the set and nothing smaller is in the set, so $\inf E = 0.3$. The elements of the set are all smaller than $1/3$, so $\sup E \leq 1/3$, and they come arbitrarily close to $1/3$, so $\sup E = 1/3$.

♦

◇ **1E-2.** Review the proof that $\sqrt{2}$ is irrational. Generalize this to $\sqrt{k}$ for k a positive integer that is not a perfect square.

Suggestion. Write k as mn where m is a perfect square and n has no repeated prime factors. Show that n cannot have a square root which is a fraction of integers "in lowest terms" (numerator and denominator having no common factor). ◇

Solution. Suppose k is a positive integer which is not a perfect square. Then $k = mn$ where m is a perfect square and n has no repeated prime factors. $n = p_1 p_2 p_3 \cdot \dots \cdot p_J$ where the p_j are all prime and all different from one another. Since $\sqrt{n} = \sqrt{k}/\sqrt{m}$, we see that $\sqrt{k}$ is rational if and only if $\sqrt{n}$ is rational. We will obtain a contradiction from the supposition that $\sqrt{n}$ is rational. If $\sqrt{n}$ were rational, then there would be integers a and b having no factors in common such that $a/b = \sqrt{n}$. We would have $a^2 = p_1 p_2 \cdot \dots \cdot p_J b^2$. So p_1 would appear in the list of prime factors of the integer a^2. But this is simply the list of prime factors of a written twice. So p_1 must be a factor of a and appear twice in the list of factors of $a^2 = p_1 p_2 \cdot \dots \cdot p_J b^2$. Since the primes p_j are all different, p_1 must be a factor of b^2, and, since it is prime, of b. Thus a and b would both have a factor of p_1. This contradicts their selection as having no common factor. Thus no such representation of $\sqrt{n}$ as a fraction in "lowest terms" can exist. $\sqrt{n}$ must be irrational, and so $\sqrt{k}$ must be also. ♦

◇ **1E-3.** (a) Let $x \geq 0$ be a real number such that for any $\varepsilon > 0$, $x \leq \varepsilon$. Show that $x = 0$.

(b) Let $S =]0, 1[$. Show that for each $\varepsilon > 0$ there exists an $x \in S$ such that $x < \varepsilon$.

Suggestion. (a) Suppose $x > 0$ and consider $\varepsilon = x/2$.

(b) Let $x = \min\{\varepsilon/2, 1/2\}$.

$\Diamond$

Solution. (a) Suppose $x \geq 0$ and that $x \leq \varepsilon$ for every $\varepsilon > 0$. If $x > 0$, then we would have $x/2 > 0$ so that $0 < x \leq x/2$ by hypothesis. Multiplication by the positive number 2 gives $0 \leq 2x \leq x$. Subtraction of x gives $-x \leq x \leq 0$. We are left with $x \geq 0$ and $x \leq 0$, so that $x = 0$ Thus the assumption that x is strictly greater than 0 fails, and we must have $x = 0$.

Notice that the argument given for part (a) shows that if $x \geq 0$, then $0 \leq x/2 \leq x$ and if $x > 0$, then $0 < x/2 < x$.

(b) Suppose $S =]0, 1[$, and $\varepsilon > 0$. If $\varepsilon \geq 1$, let $x = 1/2$. Then $x \in S$, and $x < \varepsilon$. If $\varepsilon < 1$, let $x = \varepsilon > 2$. Then $0 < x = \varepsilon/2 < \varepsilon < 1$. So $x \in S$, and $x < \varepsilon$. So in either case we have an x with $x \in S$, and $x < \varepsilon$.

$\blacklozenge$

$\diamond$ **1E-4.** Show that $d = \inf(S)$ iff d is a lower bound for S and for any $\varepsilon > 0$ there is an $x \in S$ such that $d \geq x - \varepsilon$.

Solution. Suppose $d \in \mathbb{R}$ and that $d = \inf S$, then d is a lower bound for S by definition, and d is a greatest lower bound. Let $\varepsilon > 0$. If there were no x in S with $d > x - \varepsilon$, then $d \leq x - \varepsilon$, so $x \geq d + \varepsilon$ for every x in S. Thus $d + \varepsilon$ would be a lower bound for S. Since d is a greatest lower bound, this would mean that $d + \varepsilon \leq d$, so $\varepsilon \leq 0$. Since this is false, there must be an x in S with $d > x - \varepsilon$.

Conversely, suppose d is a lower bound for the set S and for each $\varepsilon > 0$ there is an x in S with $d \geq x - \varepsilon$. If $c > d$, then $c = d + \varepsilon$ where $\varepsilon = c - d > 0$. By hypothesis, there is an x in S with $d \geq x - \varepsilon/2 = x - (c - d) + \varepsilon/2$. This gives $c > c - \varepsilon/2 \geq x$. So c is not a lower bound for S. Thus $c \leq d$ for every lower bound c. That is, d is a greatest lower bound. So $d = \inf S$.

$\blacklozenge$

$\diamond$ **1E-5.** Let x_n be a monotone increasing sequence bounded above and consider the set $S = \{x_1, x_2, \dots\}$. Show that x_n converges to $\sup(S)$. Make a similar statement for decreasing sequences.

Sketch. If $\varepsilon > 0$, use Proposition 1.3.2 to get N with $\sup(S) - \varepsilon < x_N \leq \sup(S)$. If $n \geq N$, then $\sup(S) - \varepsilon < x_N \leq x_n \leq \sup(S)$, so $|x_n - \sup(S)| < \varepsilon$. If $\langle y_n \rangle_1^\infty$ is monotone decreasing and $R = \{y_1, y_2, y_3, \dots\}$, then $y_n \to \inf(R)$. $\Diamond$

Solution. Since $S = \{x_1, x_2, x_3, \dots\}$ and the sequence $\langle x_n \rangle_1^\infty$ is bounded above, the set S is nonempty and bounded above. If we assume Theorem

1.3.4(ii), then $b = \sup S$ exists as an element of $\mathbb{R}$. Then b is an upper bound for S, so $x_n \le b$ for every n. If $\varepsilon > 0$, then $b - \varepsilon < b$, so $b - \varepsilon$ is not an upper bound for S. There must be an N such that $b - \varepsilon < x_N$. If $n \ge N$, we have $b - \varepsilon < x_N \le x_n \le b$ since the sequence is monotone increasing. Thus $|b - x_n| < \varepsilon$ for all such n. This shows that $\lim_{n\to\infty} = b$ as claimed.

If we do not wish to assume Theorem 1.3.4, we can proceed as follows. Since the sequence $\langle x_n \rangle_1^\infty$ is monotone increasing and bounded above in $\mathbb{R}$, there is a b in $\mathbb{R}$ to which the sequence converges by completeness. Let $\varepsilon > 0$. Then there is an index N such that $b - \varepsilon < x_n < b + \varepsilon$ whenever $n \ge N$. In particular, $x_N < b+\varepsilon$. Since the sequence is monotone increasing, we also have $x_k < b + \varepsilon$ for all k with $k \le N$. Thus $x_n < b + \varepsilon$ for every n. So $b + \varepsilon$ is an upper bound for the set S for every $\varepsilon > 0$. If $x \in S$, then $x < b + \varepsilon$ for every $\varepsilon > 0$. We must have $x \le b$. So b is an upper bound for S. On the other hand, if $\varepsilon > 0$, there is an n with $b - \varepsilon < x_N \le b$, so $b - \varepsilon$ is not an upper bound for S. Thus b is the least upper bound for S. That is, $b = \sup S$. ♦

◇ **1E-6.** Let A and B be two nonempty sets of real numbers with the property that $x \le y$ for all $x \in A$, $y \in B$. Show that there exists a number $c \in \mathbb{R}$ such that $x \le c \le y$ for all $x \in A$, $y \in B$. Give a counterexample to this statement for rational numbers (it is, in fact, equivalent to the completeness axiom and is the basis for another way of formulating the completeness axiom known as ***Dedekind cuts***).

Solution. Let y_0 be in B. Then $x \le y_0$ for every x in A. So A is bounded above. Since we are given that it is not empty, $r = \sup A$ exists in $\mathbb{R}$. Since r is an upper bound for A, we have $x \le r$ for every x in A. If $y \in B$, then $y \ge x$ for every x in A, so y is an upper bound for A, and $r \le y$. Thus we have $x \le r \le y$ for every x in A and every y in B.

In $\mathbb{Q}$ this fails. Let $A = \{x \in \mathbb{Q} \mid x < 0 \text{ or } x^2 < 2\}$ and $B = \{y \in \mathbb{Q} \mid y > 0 \text{ and } y^2 > 0\}$. The $x \le y$ for every x in A and every y in B. But there is no rational number r with $r^2 = 2$. So $A \cap B = \emptyset$, $A \cup B = \mathbb{Q}$, and there is no r in $\mathbb{Q}$ with $x \le r \le y$ for every x in A and y in B. ♦

◇ **1E-7.** For nonempty sets $A, B \subset \mathbb{R}$, let $A + B = \{x + y \mid x \in A \text{ and } y \in B\}$. Show that $\sup(A + B) = \sup(A) + \sup(B)$.

Sketch. Since $\sup(A) + \sup(B)$ is an upper bound for $A + B$ (Why?), $\sup(A + B) \le \sup(A) + \sup(B)$. Next show that elements of $A + B$ get arbitrarily close to $\sup(A) + \sup(B)$. ♢

Solution. Suppose $w \in A + B$. Then there are x in A and y in B with $w = x + y$. So

$$w = x + y \le x + \sup B \le \sup A + \sup B.$$

The number $\sup A + \sup B$ is an upper bound for the set $A+B$, so $\sup(A+B) \le \sup A + \sup B$.

Now let $\varepsilon > 0$. Then there are x in A and y in B with

$$\sup A - \varepsilon/2 < x \quad \text{and} \quad \sup B - \varepsilon/2 < y.$$

Adding these inequalities gives

$$\sup A + \sup B - \varepsilon < x + y \in A + B.$$

So $\sup A + \sup B - \varepsilon$ is not an upper bound for the set $A + B$. Thus $\sup(A+B) > \sup A + \sup B - \varepsilon$. This is true for all $\varepsilon > 0$, so $\sup(A+B) \ge \sup A + \sup B$.

We have inequality in both directions, so $\sup(A + B) = \sup A + \sup B$ as claimed. ♦

◇ **1E-8.** For nonempty sets $A, B \subset \mathbb{R}$, determine which of the following statements are true. Prove the true statements and give a counterexample for those that are false:

(a) $\sup(A \cap B) \le \inf\{\sup(A), \sup(B)\}$.
(b) $\sup(A \cap B) = \inf\{\sup(A), \sup(B)\}$.
(c) $\sup(A \cup B) \ge \sup\{\sup(A), \sup(B)\}$.
(d) $\sup(A \cup B) = \sup\{\sup(A), \sup(B)\}$.

Solution. (a) $\sup(A \cap B) \le \inf\{\sup A, \sup B\}$.

This is true if the intersection is not empty.

If $x \in A \cap B$, then $x \in A$, so $x \le \sup A$. Also, $x \in B$, so $x \le \sup B$. Thus x is no larger than the smaller of the two numbers $\sup A$, and $\sup B$. That is, $x \le \inf\{\sup A, \sup B\}$ for every x in $A \cap B$. Thus $\inf\{\sup A, \sup B\}$ is an upper bound for $A \cap B$. So $\sup(A \cap B) \le \inf\{\sup A, \sup B\}$.

Here is another argument using Proposition 1.3.3. We notice that $\sup(A \cap B) \subseteq A$, so by Proposition 1.3.3 we should have $\sup(A \cap B) \le \sup A$. Similarly, $\sup(A \cap B) \le \sup B$. So $\sup(A \cap B)$ is no larger than the smaller of the two numbers $\sup A$ and $\sup B$. That is

$$\sup(A \cap B) \le \inf\{\sup A, \sup B\}.$$

There is a problem if the intersection is empty. We have defined $\sup(\emptyset)$ to be $+\infty$, and this is likely to be larger than $\inf\{\sup A, \sup B\}$.

One can make a reasonable argument for defining $\sup(\emptyset) = -\infty$. Since any real number is an upper bound for the empty set, and the supremum is to be smaller than any other upper bound, we should take $\sup(\emptyset) = -\infty$. (Also, since any real number is a lower bound for $\emptyset$, and the infimum is to be larger than any other lower bound, we could set $\inf(\emptyset) = +\infty$). If we did this then the inequality would be true even if the intersection were empty.

(b) $\sup(A \cap B) = \inf\{\sup A, \sup B\}$

This one need not be true even if the intersection is not empty. Consider the two element sets $A = \{1, 2\}$ and $B = \{1, 3\}$, Then $A \cap B = \{1\}$. So $\sup(A \cap B) = 1$. But $\sup A = 2$ and $\sup B = 3$. So $\inf\{\sup A, \sup B\} = 2$ We do have $1 \leq 2$, but they are certainly not equal.

We know that $\sup(A \cap B) \leq \inf\{\sup A, \sup B\}$. Can we get the opposite inequality?

(c) $\sup(A \cup B) \geq \sup\{\sup A, \sup B\}$

(d) $\sup(A \cup B) = \sup\{\sup A, \sup B\}$

Neither A nor B nor the union is empty, so we will not be troubled with problems in the definition of the supremum of the empty set.

If x is in $a \cup B$, then $x \in A$ or $x \in B$. If $x \in A$, then $x \leq \sup A$. If $x \in B$, then $x \leq \sup B$. In either case it is no larger than the larger of the two numbers $\sup A$ and $\sup B$. So $x \leq \sup\{\sup A, \sup B\}$ for every x in the union. Thus $\sup\{\sup A, \sup B\}$ is an upper bound for $A\cup$. So

$$\sup(A \cup B) \leq \sup\{\sup A, \sup B\}$$

In the opposite direction, we note that $A \subseteq A \cup B$, so by Proposition 1.3.3, we have $\sup A \leq \sup(A \cup B)$. Similarly $\sup B \leq \sup(A \cup B)$. So $\sup(A \cup B)$ is at least as large as the larger of the two numbers $\sup A$ and $\sup B$. That is

$$\sup(A \cup B) \geq \sup\{\sup A, \sup B\}$$

We have inequality in both directions, so in fact equality must hold.

♦

◇ **1E-9.** Let x_n be a bounded sequence of real numbers and $y_n = (-1)^n x_n$. Show that $\limsup y_n \leq \limsup |x_n|$. Need we have equality? Make up a similar inequality for lim inf.

Sketch. Use $y_n \leq |x_n|$. Equality need not hold. For example, let $x_n = (-1)^{n+1}$. Then $\limsup y_n = -1$ and $\limsup |x_n| = 1$. Similarly, $\liminf y_n \geq \liminf(-|x_n|)$. ◊

Solution. Since $y_n = (-1)^n x_n$, we see that $y_n = \pm |x_n|$, and $y_n \leq |x_n|$ for every n. So $\langle y_n \rangle_1^\infty$ is a bounded sequence in $\mathbb{R}$. Let $b = \limsup y_n$. Then there is a subsequence y_{n_k} converging to b. If $\varepsilon > 0$, then $b - \varepsilon < y_{n_k} \leq |x_{n_k}|$ for large enough k. The sequence $|a_{n_1}|, |a_{n_2}|, |a_{n_3}|, \ldots$ is bounded and must have a cluster point, and they must all be as large as $b - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, they must all be at least as large as b. So $\limsup y_n = b \leq \limsup |x_n|$.

Let $x_n = (-1)^{n+1}$. Then $y_n = -1$ for all n, and $|x_n| = 1$ for all n. So $\limsup y_n = -1$, and $\limsup |x_n| = 1$. Thus equality need not hold.

Since $y_n \geq -|x_n|$ for all n, a similar argument establishes the inequality

$$\liminf y_n \geq \liminf(-|x_n|).$$

♦

◇ **1E-10.** Verify that the bounded metric in Example 1.7.2(d) is indeed a metric.

Sketch. Use the basic properties of a metric for $d(x, y)$ to establish those properties for $\rho(x, y)$. For the triangle inequality, work backwards from what you want to discover a proof. ◊

Solution. Suppose d is a metric on a set M, and ρ is defined by $\rho(x,y) = \frac{d(x,y)}{1+d(x,y)}$. We are to show that this is a metric on M and that it is bounded by 1.

(i) Positivity: Since $d(x, y) \geq 0$, we have $0 \leq d(x, y) < 1 + d(x, y)$. So

$$0 \leq \frac{d(x,y)}{1+d(x,y)} < 1 \quad \text{for every } x \text{ and } y \text{ in } M.$$

Notice that this also shows that ρ is bounded by 1.

(ii) Nondegeneracy: If $x = y$, then $d(x, y) = 0$, so $\rho(x, y) = 0$ also. In the other direction, we know that $1 + d(x, y)$ is never 0, so

$$\begin{aligned}\rho(x,y) = 0 &\iff \frac{d(x,y)}{1+d(x,y)} = 0\\ &\iff d(x,y) = 0\\ &\iff x = y\end{aligned}$$

(iii) Symmetry: Since $d(y, x) = d(x, y)$, we have

$$\rho(y,x) = \frac{d(y,x)}{1+d(y,x)} = \frac{d(x,y)}{1+d(x,y)} = \rho(x,y).$$

(iv) Triangle Inequality: Suppose x, y, and z are in M. Then

$$\begin{aligned}
&\rho(x,y) \leq \rho(x,z) + \rho(z,y) \iff \frac{d(x,y)}{1+d(x,y)} \leq \frac{d(x,z)}{1+d(x,z)} + \frac{d(z,y)}{1+d(z,y)}\\
&\iff d(x,y)(1+d(x,z))(1+d(z,y)) \leq d(x,z)(1+d(x,y))(1+d(z,y))\\
&\qquad + d(z,y)(1+d(x,y))(1+d(x,z))\\
&\iff d(x,y) + d(x,y)d(x,z) + d(x,y)d(z,y) + d(x,y)d(x,z)d(z,y)\\
&\qquad \leq d(x,z) + d(x,z)d(x,y) + d(x,z)d(z,y) + d(x,y)d(x,z)d(z,y)\\
&\qquad + d(z,y) + d(z,y)d(x,y) + d(z,y)d(x,z) + d(x,y)d(x,z)d(z,y)\\
&\iff d(x,y) \leq d(x,z) + d(z,y) + 2(d(x,z)d(z,y) + d(x,y)d(x,z)d(z,y))
\end{aligned}$$

This last line is true since $d(x, y) \le d(y, z) + d(z, y)$ by the triangle inequality for d and the last term is non-negative.

The function ρ satisfies all of the defining properties of a metric, so it is a metric on the space M.

What makes this example particularly interesting is not merely the fact that there are such things as bounded metrics, but that we can build one starting from any metric, even the usual Euclidean metric on $\mathbb{R}^n$. Even this would not be particularly exciting were it not for the fact that the new metric makes the same convergent sequences and the same limits as the original one. In chapter 2 we will study such things in more detail, but the basic idea is essentially the same as for convergence in $\mathbb{R}$. We simply replace the distance between two real numbers, $|x - y|$, by the metric, $d(x, y)$, which is supposed to capture the idea of how "different" the points x and y in the metric space M are. The definition of convergence of a sequence is then a straight translation.

Definition. *A sequence $x_1, x_2, x_3, \dots$ in a metric space M is said to converge to the limit x in M (with respect to the metric d) if for every $\varepsilon > 0$ there is an index N such that $d(x_n, x) < \varepsilon$ whenever $n \ge N$.*

In other words, $x_n \to x$ with respect to the metric d if and only if the sequence of real numbers $d(x_n, x)$ converges to 0 as $n \to \infty$. Since $0 \le \rho(x_n, x) \le d(x_n, x)$, we certainly have $d(x_n, x) \to 0 \implies \rho(x_n, x) \to 0$. If we solve for d in terms of ρ, we find that $d(x, y) = \rho(x, y)/(1 - \rho(x, y))$. If $\rho(x_n, x) \to 0$, then for large enough n we have $\rho(x_n, x) < 1/2$, and $0 \le d(x_n, x) < 2\rho(x_n, x)$. So $\rho(x_n, x) \to 0 \implies d(x_n, x) \to 0$. So we have implication in both directions as claimed.

Proposition. *If d is a metric on a set M and $\rho(x, y)$ is defined for x and y in M by*

$$\rho(x, y) = \frac{d(x, y)}{1 + d(x, y)},$$

then ρ is also a metric on M and $\rho(x, y) < 1$ for every x and y in M. Furthermore, if $\langle x_n \rangle_1^\infty$ is a sequence in M, then x_n converges to x in M with respect to the metric d if and only if $x_n \to x$ with respect to the metric ρ.

♦

◇ **1E-11.** Show that (i) and (ii) of Theorem 1.3.4 both imply the completeness axiom for an ordered field.

Sketch. (i) $\implies$ completeness: Let $\langle x_n \rangle_1^\infty$ be increasing and bounded above. By (i), $\langle x_n \rangle_1^\infty$ has a *lub*, say S. Check that $x_n \to S$. See Exercise 1E-5. The proof that (ii) $\implies$ completeness is done by showing that $x_n \to -\inf\{-x_n\}$. ◊

Solution. Part (i) says: *If S is a nonempty subset of $\mathbb{R}$ which is bounded above, then S has a least upper bound in $\mathbb{R}$.*

Suppose this is true and that $\langle x_n \rangle_1^\infty$ is a monotonically increasing sequence in $\mathbb{R}$ which is bounded above by a number B. Let $S = \{x_1, x_2, x_3, \dots\}$. Then S is nonempty and bounded above by B, so S has a least upper bound b in $\mathbb{R}$ by hypothesis. The sequence $\langle x_n \rangle_1^\infty$ must converge to b. (See the solution to Exercise 1E-5.)

Part (ii) says: *If T is a nonempty subset of $\mathbb{R}$ which is bounded below, then T has a greatest lower bound in $\mathbb{R}$.*

Suppose this is true and that S is a nonempty subset of $\mathbb{R}$ which is bounded above by B. Let $T = \{x \in \mathbb{R} \mid -x \in S\}$. If $x \in T$, then $-x \in S$, so $-s \le B$ and $x \ge -B$. Thus T is bounded below by $-B$. It is not empty since S is not empty. By hypothesis, T has a greatest lower bound a in $\mathbb{R}$. If $x \in S$, then $-x \in T$, so $a \le -x$, and $x \le -a$. So $-a$ is an upper bound for S. If $\varepsilon > 0$, then $a + \varepsilon$ is not a lower bound for T. There is an x in T with $a \le x < a + \varepsilon$. So $-a - \varepsilon < -x < -a$, and $-x \in S$. So $-a - \varepsilon$ is not an upper bound for S. Thus $-a$ is a least upper bound for S. Every nonempty subset of $\mathbb{R}$ which is bounded above has a least upper bound in $\mathbb{R}$. Thus (ii) implies (i).

Since we have (ii) implies (i) and (i) implies completeness, we conclude that (ii) implies completeness.

Alternatively, one could proceed more directly. Let $\langle x_n \rangle_1^\infty$ be a monotonically increasing sequence in $\mathbb{R}$ bounded above by a number B. Then $\langle -x_n \rangle_1^\infty$ is monotonically decreasing and bounded below by $-B$. The set $T = \{-x_1, -x_2, -x_3, \dots\}$ is nonempty and bounded below. By hypothesis it has a greatest lower bound a in $\mathbb{R}$. So $a \le -x_n$ for every n. Modify the argument from Exercise 1E-5 to show that $-x_n$ must converge to a. Then x_n must converge to $-a$. (See Exercise 1E-18.) ♦

◇ **1E-12.** In an inner product space show that

(a) $2\,\| x \|^2 + 2\,\| y \|^2 = \| x + y \|^2 + \| x - y \|^2$ (***parallelogram law***).

(b) $\| x + y \| \;\; \| x - y \| \le \| x \|^2 + \| y \|^2$.

(c) $4\langle x, y \rangle = \| x + y \|^2 - \| x - y \|^2$ (***polarization identity***).

Interpret these results geometrically in terms of the parallelogram formed by x and y.

Solution. In all of these, the key is to use the relationship between the inner product and the norm derived from it: $\| f \|^2 = \langle f, f \rangle$.

(a) We compute

$$\begin{aligned}
\|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\
&= \langle x+y, x\rangle + \langle x+y, y\rangle + \langle x-y, x\rangle + \langle x-y, -y\rangle \\
&= \langle x, x+y\rangle + \langle y, x+y\rangle + \langle x, x-y\rangle + \langle -y, x-y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\qquad + \langle x, x\rangle + \langle x, -y\rangle + \langle -y, x\rangle + \langle -y, -y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\qquad + \langle x, x\rangle + \langle x, -y\rangle - \langle y, x\rangle - \langle y, -y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\qquad + \langle x, x\rangle + \langle -y, x\rangle - \langle y, x\rangle - \langle -y, y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\qquad + \langle x, x\rangle - \langle y, x\rangle - \langle y, x\rangle + \langle y, y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\qquad + \langle x, x\rangle - \langle x, y\rangle - \langle y, x\rangle + \langle y, y\rangle \\
&= \langle x, x\rangle + \langle y, y\rangle + \langle x, x\rangle + \langle y, y\rangle \\
&= 2\|x\|^2 + 2\|y\|^2
\end{aligned}$$

as desired.

(b) Using the result of part (a), we compute

$$\begin{aligned}
0 &\leq (\|x+y\| - \|x-y\|)^2 \\
&\leq \|x+y\|^2 - 2\|x+y\|\,\|x-y\| + \|x-y\|^2 \\
&\leq 2\|x\|^2 - 2\|x+y\|\,\|x-y\| + 2\|y\|^2
\end{aligned}$$

So $\|x+y\|\,\|x-y\| \leq \|x\|^2 + \|y\|^2$ as desired.

(c) The computation is like that of part (a), but with some sign changes.

$$\begin{aligned}
\|x+y\|^2 - \|x-y\|^2 &= \langle x+y, x+y\rangle - \langle x-y, x-y\rangle \\
&= \langle x+y, x\rangle + \langle x+y, y\rangle - \langle x-y, x\rangle - \langle x-y, -y\rangle \\
&= \langle x, x+y\rangle + \langle y, x+y\rangle - \langle x, x-y\rangle - \langle -y, x-y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\qquad - \langle x, x\rangle - \langle x, -y\rangle - \langle -y, x\rangle - \langle -y, -y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\qquad - \langle x, x\rangle - \langle x, -y\rangle + \langle y, x\rangle + \langle y, -y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\qquad - \langle x, x\rangle - \langle -y, x\rangle + \langle y, x\rangle + \langle -y, y\rangle
\end{aligned}$$

$$\begin{aligned}
&= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\
&\qquad\qquad - \langle x, x \rangle + \langle y, x \rangle + \langle y, x \rangle - \langle y, y \rangle \\
&= \langle x, x \rangle + \langle x, y \rangle + \langle x, y \rangle + \langle y, y \rangle \\
&\qquad\qquad - \langle x, x \rangle + \langle x, y \rangle + \langle x, y \rangle - \langle y, y \rangle \\
&= 4 \langle x, y \rangle
\end{aligned}$$

as desired.

Part (a) has the most immediate geometric interpretation. It says that the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of the four sides. If x and y are vectors forming two adjacent sides of a parallelogram, then the two diagonals are $x + y$ and $x - y$, and the four sides are x, y, x, and y. ♦

◇ **1E-13.** What is the orthogonal complement in $\mathbb{R}^4$ of the space spanned by $(1, 0, 1, 1)$ and $(-1, 2, 0, 0)$?

Answer. One possible answer is the space spanned by $\{(2, 1, -2, 0), (2, 1, 0, -2)\}$. Here is a solution that produces a different form of the same answer. ◇

Solution. Suppose $v = (x, y, z, t)$ is in the orthogonal complement, $M^\perp$ of the space M spanned by $u = (1, 0, 1, 1)$ and $v = (-1, 2, 0, 0)$. Then

$$\begin{aligned}
(x, y, z, t) \cdot (1, 0, 1, 1) = 0 \quad &\text{and} \quad (x, y, z, t) \cdot (-1, 2, 0, 0) = 0 \\
x + z + t = 0 \quad &\text{and} \quad -x + 2y = 0
\end{aligned}$$

So $y = x/2$ and $z + t = -x$. Thus

$$\begin{aligned}
(x, y, z, t) &= (x, x/2, z, -x - z) = x(1, 1/2, 0, -1) + z(0, 0, 1, -1) \\
&= s(2, 1, 0, -2) + t(0, 0, 1, -1)
\end{aligned}$$

where $s = x/2$ and $t = z$.

In the other direction, if $w = s(2, 1, 0, -2) + t(0, 0, 1, -1)$ for real s and t, then a direct computation shows $w \cdot u = 0$ and $w \cdot v = 0$. (Do it.) So w is orthogonal to both u and v and is in $M^\perp$. We conclude

$$\begin{aligned}
M^\perp &= \{s(2, 1, 0, -2) + t(0, 0, 1, -1) \mid s \in \mathbb{R} \text{ and } t \in \mathbb{R}\} \\
&= \text{ the space spanned by } (2, 1, 0, -2) \text{ and } (0, 0, 1, -1).
\end{aligned}$$

Different ways of manipulating the algebra may produce different vectors spanning the same space, for example, $(2, 1, -2, 0) = (2, 1, 0, -2) - 2(0, 0, 1, -1)$. ♦

◇ **1E-14.** (a) Prove ***Lagrange's identity***

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 = \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) - \sum_{1\le i<j\le n} (x_i y_j - x_j y_i)^2$$

and use this to give another proof of the Cauchy-Schwarz inequality.

(b) Show that

$$\left(\sum_{i=1}^{n}(x_i + y_i)^2\right)^{1/2} \le \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} + \left(\sum_{i=1}^{n} y_i^2\right)^{1/2}.$$

Sketch. You can get a good idea of what is going on by looking at a simplified version of the problem. Try it with only two terms.

(a) The Lagrange Identity

$$\begin{aligned}
(x_1y_1 + x_2y_2)^2 &= (x_1y_1 + x_2y_2)(x_1y_1 + x_2y_2)\\
&= x_1y_1x_1y_1 + x_1y_1x_2y_2 + x_2y_2x_1y_1 + x_2y_2x_2y_2\\
&= (x_1^2y_1^2 + x_2^2y_2^2) + (x_1y_1x_2y_2 + x_2y_2x_1y_1)\\
&= (x_1^2y_1^2 + x_2^2y_2^2) + (x_1^2y_2^2 + x_2^2y_1^2)\\
&\qquad - (x_1^2y_2^2 + x_2^2y_1^2) + (x_1y_1x_2y_2 + x_2y_2x_1y_1)\\
&= (x_1^2 + x_2^2)(y_1^2 + y_2^2) - (x_1y_2 - x_2y_1)^2.
\end{aligned}$$

This implies the Cauchy-Schwarz inequality for the usual inner product in $\mathbb{R}^2$.

$$\begin{aligned}
((x_1, x_2)\cdot(y_1, y_2))^2 &= (x_1y_1 + x_2y_2)^2\\
&= (x_1^2 + x_2^2)(y_1^2 + y_2^2) - (x_1y_2 - x_2y_1)^2\\
&\le (x_1^2 + x_2^2)(y_1^2 + y_2^2) = \|\,(x_1, x_2)\,\|^2\ \|\,(y_1, y_2)\,\|^2
\end{aligned}$$

Taking square roots on both sides gives the Cauchy - Schwarz inequality.

(b) The Minkowski Inequality.

$$\begin{aligned}
&((x_1 + y_1)^2 + (x_2 + y_2)^2)^{1/2} \le \left(x_1^2 + x_2^2\right)^{1/2} + \left(y_1^2 + y_2^2\right)^{1/2}\\
&\iff (x_1 + y_1)^2 + (x_2 + y_2)^2\\
&\qquad \le \left(x_1^2 + x_2^2\right) + 2\left(x_1^2 + x_2^2\right)^{1/2}\left(y_1^2 + y_2^2\right)^{1/2} + \left(y_1^2 + y_2^2\right)\\
&\iff x_1^2 + 2x_1y_1 + y_1^2 + x_2^2 + 2x_2y_2 + y_2^2\\
&\qquad \le \left(x_1^2 + x_2^2\right) + 2\left(x_1^2 + x_2^2\right)^{1/2}\left(y_1^2 + y_2^2\right)^{1/2} + \left(y_1^2 + y_2^2\right)\\
&\iff x_1y_1 + x_2y_2 \le \left(x_1^2 + x_2^2\right)^{1/2}\left(y_1^2 + y_2^2\right)^{1/2}
\end{aligned}$$

But this is exactly the Cauchy-Schwarz Inequality. ◊

Solution. To get solutions to the exercise as stated we just need to figure out a reasonable notation for general n for the same computation done above for $n = 2$.

(a) The Lagrange Identity:

$$\begin{aligned}
\left(\sum_{i=1}^{n} x_i y_i\right)^2 &= \left(\sum_{i=1}^{n} x_i y_i\right)\left(\sum_{j=1}^{n} x_j y_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_i x_j y_j \\
&= \sum_{i=1}^{n} x_i^2 y_i^2 + \sum_{i\neq j} x_i y_i x_j y_j \\
&= \sum_{i=1}^{n} x_i^2 y_i^2 + \sum_{i\neq j} x_i^2 y_j^2 - \sum_{i\neq j} x_i^2 y_j^2 + \sum_{i\neq j} x_i y_i x_j y_j \\
&= \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{j=1}^{n} y_j^2\right) + \sum_{i\neq j}\left(x_i y_i x_j y_j - x_i^2 y_j^2\right) \\
&= \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{j=1}^{n} y_j^2\right) - \sum_{i\neq j}\left(x_i^2 y_j^2 - x_i y_i x_j y_j\right) \\
&= \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{j=1}^{n} y_j^2\right) \\
&\qquad - \sum_{i<j}\left(x_i^2 y_j^2 - x_i y_i x_j y_j - x_j y_j x_i y_i + x_j^2 y_i^2\right) \\
&= \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{j=1}^{n} y_j^2\right) - \sum_{i<j}\left(x_i y_j - x_j y_i\right)^2
\end{aligned}$$

Now we prove the Cauchy-Schwarz Inequality for the usual inner product in $\mathbb{R}^n$.

$$
\begin{aligned}
((x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n))^2 &= \left(\sum_{i=1}^{n} x_i y_i \right)^2 \\
&= \left(\sum_{i=1}^{n} x_i^2 \right) \left(\sum_{j=1}^{n} y_j^2 \right) - \sum_{i<j} (x_i y_j - x_j y_i)^2 \\
&\leq \left(\sum_{i=1}^{n} x_i^2 \right) \left(\sum_{j=1}^{n} y_j^2 \right) = \| (x_1, \ldots, x_n) \|^2 \ \| (y_1, \ldots, y_n) \|^2
\end{aligned}
$$

Taking square roots on both sides gives the Cauchy-Schwarz Inequality.

(b) The Minkowski Inequality:

$$
\begin{aligned}
&\left(\sum_{i=1}^{n} (x_i + y_i)^2 \right)^{1/2} \leq \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} + \left(\sum_{i=1}^{n} y_i^2 \right)^{1/2} \\
&\iff \sum_{i=1}^{n} (x_i + y_i)^2 \leq \sum_{i=1}^{n} x_i^2 + 2 \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \left(\sum_{i=1}^{n} y_i^2 \right)^{1/2} + \sum_{i=1}^{n} y_i^2 \\
&\iff \sum_{i=1}^{n} x_i^2 + 2 \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 \\
&\qquad \leq \sum_{i=1}^{n} x_i^2 + 2 \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \left(\sum_{i=1}^{n} y_i^2 \right)^{1/2} + \sum_{i=1}^{n} y_i^2 \\
&\iff \sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \left(\sum_{i=1}^{n} y_i^2 \right)^{1/2}
\end{aligned}
$$

But this is exactly the Cauchy-Schwarz inequality which we established in part (a). ♦

◇ **1E-15.** Let x_n be a sequence in $\mathbb{R}$ such that $d(x_n, x_{n+1}) \leq d(x_{n-1}, x_n)/2$. Show that x_n is a Cauchy sequence.

Sketch. First show that $d(x_n, x_{n+1}) \leq d(x_1, x_2)/2^{n-1}$, then that $d(x_n, x_{n+k})/2^{n-2}$. ◇

Solution. There is nothing particularly special about the number 1/2 in this problem. Any constant c with $0 \le c < 1$ will do as well. The key is the convergence of the geometric series with ratio c. $\sum_{k=0}^{\infty} c^k = 1/(1-c)$.

Proposition. *If $\langle x_n \rangle_1^\infty$ is a sequence in $\mathbb{R}$ such that there is a constant c with $0 \le c < 1$ and $d(x_n, x_{n+1}) \le c d(x_{n-1}, x_n)$ for every index n, then $\langle x_n \rangle_1^\infty$ is a Cauchy sequence.*

Proof: First we show that $d(x_m, x_{m+1}) \le c^{m-1} d(x_1, x_2)$ for each $m = 1, 2, 3, \ldots$. by an informal induction

$$\begin{aligned} d(x_m, x_{m+1}) &\le c d(x_{m-1}, x_m) \\ &\le c^2 d(x_{m-2}, x_{m-1}) \\ &\ \ \vdots \\ &\le c^{m-1} d(x_1, x_2) \end{aligned}$$

Now we use this to establish that $\langle x_n \rangle_1^\infty$ is a Cauchy sequence in the form: For each $\varepsilon > 0$ there is an N such that $d(x_n, x_{n+p}) < \varepsilon$ whenever $n \ge N$ and $p > 0$.

The first step is a repeated use of the triangle inequality to get an expression in which we can use the inequality just established.

$$\begin{aligned} d(x_n, x_{n+p}) &\le d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots d(x_{n+p-1}, x_{n+p}) \\ &\le \left(c^{n-1} + c^n + c^{n+1} + \cdots c^{n+p-2}\right) d(x_1, x_2) \\ &\le c^{n-1}\left(1 + c + c^2 + \cdots + c^{p-1}\right) d(x_1, x_2) \\ &\le \frac{c^{n-1}}{1-c} d(x_1, x_2). \end{aligned}$$

Since $0 \le c < 1$, we know that $c^n \to 0$ as $n \to \infty$. If $\varepsilon > 0$, we can select N large enough so that $\frac{c^{N-1}}{1-c} d(x_1, x_2) < \varepsilon$. If $n \ge N$ and $p > 0$, we conclude that

$$d(x_n, x_{n+p}) \le \frac{c^{n-1}}{1-c} d(x_1, x_2) \le \frac{c^{N-1}}{1-c} d(x_1, x_2) < \varepsilon.$$

Thus $\langle x_n \rangle_1^\infty$ is a Cauchy sequence as claimed.

This solution has been written in terms of the metric $d(x, y)$ between x and y instead of $|x - y|$ to emphasize that it works perfectly well for a sequence $\langle x_n \rangle_1^\infty$ in any metric space provided there is a constant c satisfying the hypothesized inequality. In this form the exercise forms the key part of the proof of a very important theorem about complete metric spaces called the *Banach Fixed Point Theorem* or the *Contraction Mapping Principle*. We will study this theorem and some of its consequences in Chapter 5. ♦

◇ **1E-16.** Prove Theorem 1.6.4. In fact, for vector spaces $\mathcal{V}_1, \dots, \mathcal{V}_n$, show that $\mathcal{V} = \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ is a vector space.

Sketch. If $\mathcal{V}_1, \dots, \mathcal{V}_n$ are vector spaces, then the Cartesian product $\mathcal{V} = \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ is the set of ordered n–tuples $v = (v_1, v_2, \dots v_n)$ where each component v_j is a vector in the vector space $\mathcal{V}_j$ just as $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ is the set of ordered n–tuples $v = (x_1, x_2, \dots x_n)$ where each x_j is a real number. Operations in $\mathcal{V}$ are defined componentwise just as they are in $\mathbb{R}^n$. Use the fact that you know each $\mathcal{V}_j$ is a vector space to check that the defining properties of a vector space work in $\mathcal{V}$. (The dimension of $\mathbb{R}^n$ is a separate part of the question.) ◇

Solution. If $\mathcal{V}_1, \dots, \mathcal{V}_n$ are vector spaces, then the Cartesian product $\mathcal{V} = \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ is the set of ordered n–tuples $v = (v_1, v_2, \dots v_n)$ where each component v_j is a vector in the vector space $\mathcal{V}_j$. Operations are defined componentwise in $\mathcal{V}$. If $v = (v_1, v_2, \dots v_n)$ and $w = (w_1, w_2, \dots, w_n)$ are in $\mathcal{V}$ and $\mu \in \mathbb{R}$, then define

$$\begin{aligned} v + w &= (v_1 + w_1, v_2 + w_2, \dots v_n + w_n) \\ \mu v &= (\mu v_1, \mu v_2, \dots, \mu v_n) \\ -v &= (-v_1, -v_2, \dots, -v_n). \end{aligned}$$

Each of the quantities $v_j + w_j$ and μv_j makes sense as a vector in $\mathcal{V}_j$ since $\mathcal{V}_j$ is a vector space. If 0_j is the zero vector for the vector space $\mathcal{V}_j$, let

$$0_{\mathcal{V}} = (0_1, 0_2, \dots 0_n) \in \mathcal{V}.$$

We claim that with these definitions, $\mathcal{V}$ is a vector space with $0_{\mathcal{V}}$ as the zero vector. To establish this we need to check the defining properties of a vector space ((i) through (viii) in Definition 1.6.2). To accomplish this we write everything in components and use the fact that we know these properties hold for the vectors and operations in each $\mathcal{V}_j$. Suppose $u = (u_1, u_2, \dots, u_n)$, $v = (v_1, \dots, v_n)$, and $w = (w_1, \dots, w_n)$ are in $\mathcal{V}$ and λ and μ are in $\mathbb{R}$.

(i) commutativity:

$$\begin{aligned} v + w &= (v_1, v_2, \dots v_n) + (w_1, w_2, \dots, w_n) \\ &= (v_1 + w_1, v_2 + w_2, \dots v_n + w_n) \\ &= (w_1 + v_1, w_2 + v_2, \dots w_n + v_n) \\ &= w + v \end{aligned}$$

(ii) associativity:

$$\begin{aligned}
(v+u)+w &= ((v_1, v_2, \ldots, v_n) + (u_1, u_2, \ldots u_n)) + (w_1, w_2, \ldots w_n) \\
&= (v_1 + u_1, \ldots, v_n + u_n) + (w_1, \ldots, w_n) \\
&= ((v_1 + u_1) + w_1, \ldots (v_n + u_n) + w_n) \\
&= (v_1 + (u_1 + w_1), \ldots v_n + (u_n + w_n)) \\
&= (v_1, \ldots, v_n) + (u_1 + w_1, \ldots u_n + w_n) \\
&= v + (u + w)
\end{aligned}$$

(iii) zero vector:

$$\begin{aligned}
v + 0_{\mathcal{V}} &= (v_1, \ldots, v_n) + (0_1, \ldots 0_n) \\
&= (v_1 + 0_1, \ldots v_n + 0_n) = (v_1, \ldots, v_n) = v
\end{aligned}$$

(iv) negatives:

$$\begin{aligned}
v + (-v) &= (v_1, \ldots, v_n) + (-v_1, \ldots, -v_n) \\
&= (v_1 + (-v_1), \ldots v_n + (-v_n)) \\
&= (0_1, \ldots, 0_n) = 0_{\mathcal{V}}
\end{aligned}$$

(v) distributivity:

$$\begin{aligned}
\lambda(v + w) &= \lambda\left((v_1, \ldots, v_n) + (w_1, \ldots, w_n)\right) \\
&= \lambda(v_1 + w_1, \ldots, v_n + w_n) \\
&= (\lambda(v_1 + w_1), \ldots, \lambda(v_n + w_n)) \\
&= (\lambda v_1 + \lambda w_1, \ldots, \lambda v_n + \lambda w_n) \\
&= (\lambda v_1, \ldots, \lambda v_n) + (\lambda w_1, \ldots, \lambda w_n) \\
&= \lambda(v_1, \ldots, v_n) + \lambda(w_1, \ldots, w_n) \\
&= \lambda v + \lambda w
\end{aligned}$$

(vi) associativity:

$$\begin{aligned}
\lambda(\mu v) &= \lambda\left(\mu(v_1, \ldots, v_n)\right) \\
&= \lambda(\mu v_1, \ldots, \mu v_n) \\
&= (\lambda(\mu v_1), \ldots, \lambda(\mu v_n)) \\
&= ((\lambda\mu) v_1, \ldots, (\lambda\mu) v_n) \\
&= (\lambda\mu)\,(v_1, \ldots, v_n) \\
&= (\lambda\mu) v
\end{aligned}$$

(vii) distributivity:

$$\begin{aligned}(\lambda+\mu)v &= (\lambda+\mu)\,(v_1,\dots,v_n)\\ &= ((\lambda+\mu)v_1,\dots,(\lambda+\mu)v_n)\\ &= (\lambda v_1+\mu v_1,\dots,\lambda v_n+\mu v_n)\\ &= (\lambda v_1,\dots\lambda v_n)+(\mu v_1,\dots,\mu v_n)\\ &= \lambda\,(v_1,\dots,v_n)+\mu\,(v_1,\dots,v_n)\\ &= \lambda v+\mu v\end{aligned}$$

(viii) multiplicative identity:

$$1v = 1\,(v_1,\dots,v_n) = (1\,v_1,\dots,1\,v_n) = (v_1,\dots,v_n) = v$$

The set $\mathcal{V}=\mathcal{V}_1\times\cdots\times\mathcal{V}_n$ with the operations and special elements specified satisfies all the requirements of a vector space listed in the definition. So it is a vector space. In the special case of $\mathbb{R}^n$, each of the spaces $\mathcal{V}_j$ is just a copy of $\mathbb{R}$, and the "vector operations" in it are just the usual operations with real numbers.

DIMENSION: In $\mathbb{R}^n$, let e_1, e_2, ... , e_n be the "standard basis vectors":

$$\begin{aligned}e_1 &= (1,0,0,\dots,0)\\ e_2 &= (0,1,0,\dots,0)\\ &\ \ \vdots\\ e_n &= (0,0,\dots,0,1).\end{aligned}$$

If $v=(v_1,\dots,v_n)$ is in $\mathbb{R}^n$, then $v=v_1e_1+v_2e_2+\cdots+v_ne_n$. So these vectors span $\mathbb{R}^n$. On the other hand, if $v_1e_1+v_2e_2+\cdots+v_ne_n=\vec{0}$, then $(v_1,v_2,\dots,v_n)=(0,0,\dots,0)$, so $v_j=0$ for each j. Thus the vectors e_j are linearly independent. Since they are linearly independent and span $\mathbb{R}^n$, they are a basis for $\mathbb{R}^n$. Since there are exactly n of them, the dimension of $\mathbb{R}^n$ is n.

Challenge: With $\mathcal{V}=\mathcal{V}_1\times\cdots\times\mathcal{V}_n$ as above, show that the dimension of $\mathcal{V}$ is the sum of the dimensions of the spaces $\mathcal{V}_j$. ♦

◇ **1E-17.** Let $S\subset\mathbb{R}$ be bounded below and nonempty. Show that $\inf(S) = \sup\{x\in\mathbb{R}\mid x \text{ is a lower bound for } S\}$.

Sketch. If L is the set of lower bounds for S and $y\in L$, then $y\le\inf(S)$ (why?). So $\sup(L)\le\inf(S)$. But $\inf(S)\in L$. So $\inf(S)\le\sup(L)$. ◇

Solution. Let $L=\{x\in\mathbb{R}\mid x \text{ is a lower bound for } S\}$. Since S is bounded below, there is at least one lower bound a for S. So $a\in L$ and L is not empty. Since S is not empty, there is at least one point b in S. Then if

$y \in L$, we must have $y \le b < b+1$. So $b+1$ is an upper bound for L. Thus L is a nonempty subset of $\mathbb{R}$ which is bounded above. The least upper bound $b_0 = \sup L$ thus exists as a finite real number. Also, if $y \in L$, then y is a lower bound for S, so $y \le \inf S$ because $\inf S$ is the greatest lower bound. Thus $\inf S$ is an upper bound for L. So $\sup L \le \inf S$ since $\sup L$ is the least upper bound. On the other hand, $\inf S$ is a lower bound for S, so $\inf S \in L$. Thus we also have $\inf S \le \sup L$. We have inequality in both directions, so $\inf S = \sup L$ as claimed. ♦

◇ **1E-18.** Show that in $\mathbb{R}$, $x_n \to x$ iff $-x_n \to -x$. Hence, prove that the completeness axiom is equivalent to the statement that every decreasing sequence $x_1 \ge x_2 \ge x_3 \cdots$ bounded below converges. Prove that the limit of the sequence is $\inf\{x_1, x_2, \dots\}$.

Sketch. For the first part observe that $|(-x_n) - (-x)| = |x_n - x|$ and use the definition of convergence. For the second part observe that the sequence $\langle x_n\rangle_1^\infty$ is decreasing and bounded below if and only if the sequence $\langle -x_n\rangle_1^\infty$ is increasing and bounded above. Then use the result of part one to go back and forth between the two assertions:

A. *If $\langle x_n\rangle_1^\infty$ is an increasing sequence which is bounded above in $\mathbb{R}$, then there is an x in $\mathbb{R}$ such that $x_n \to x$.*

B. *If $\langle y_n\rangle_1^\infty$ is a decreasing sequence which is bounded below in $\mathbb{R}$, then there is a y in $\mathbb{R}$ such that $y_n \to y$.*

The last part of the exercise can be interpreted in two ways:

I. Suppose that we know that the following is true:

C. *If S is a nonempty subset of $\mathbb{R}$ which is bounded below in $\mathbb{R}$, then S must have a greatest lower bound in $\mathbb{R}$.*

Use this to show that our S has a greatest lower bound $\inf S$ in $\mathbb{R}$ and then show that $x_n \to \inf S$.

II. Show directly that the limit a guaranteed by part (ii) must be a greatest lower bound for S. ◇

Solution. (i) Suppose $x_n \to x$ and let $\varepsilon > 0$. Then there is an N such that $|x_n - x| < \varepsilon$ whenever $n \ge N$. Thus if $n \ge N$ we have $|(-x_n) - (-x)| = |(-1)(x_n - x)| = |-1|\,|x_n - x| = |x_n - x| < \varepsilon$. For each $\varepsilon > 0$ there is an N such that $n \ge N \implies |(-x_n) - (-x)| < \varepsilon$. Thus $-x_n \to -x$ as claimed.

(ii.) Now consider the two statements:

A. *If $\langle x_n\rangle_1^\infty$ is an increasing sequence which is bounded above in $\mathbb{R}$, then there is an x in $\mathbb{R}$ such that $x_n \to x$.*

B. *If $\langle y_n\rangle_1^\infty$ is a decreasing sequence which is bounded below in $\mathbb{R}$, then there is a y in $\mathbb{R}$ such that $y_n \to y$.*

Our problem is to show that starting from either of these we can prove the other. First suppose that statement A holds and that $\langle y_n\rangle_1^\infty$ is a decreasing sequence bounded below in $\mathbb{R}$.

$$y_1 \geq y_2 \geq y_3 \geq \cdots \qquad \text{and} \qquad y_k \geq a \quad \text{for every } k.$$

Then

$$-y_1 \leq -y_2 \leq -y_3 \leq \cdots \qquad \text{and} \qquad -y_k \leq -a \quad \text{for every } k.$$

So $\langle -y_n \rangle_1^\infty$ is an increasing sequence bounded above in $\mathbb{R}$. Since statement A is assumed to be true, there must be an x in $\mathbb{R}$ with $-y_n \to x$. Let $y = -x$. By part (i) we have

$$y_n = -(-y_n) \to -x = -(-y) = y.$$

On the assumption of statement A we have shown that every decreasing sequence bounded below in $\mathbb{R}$ must converge to a number in $\mathbb{R}$. That is, statement A implies statement B.

The proof that statement B implies statement A is essentially the same. Assume that statement B is true and suppose $\langle x_n \rangle_1^\infty$ is an increasing sequence bounded above in $\mathbb{R}$.

$$x_1 \leq x_2 \leq x_3 \leq \cdots \qquad \text{and} \qquad x_k \leq b \quad \text{for every } k.$$

Then

$$-x_1 \geq -x_2 \geq -x_3 \geq \cdots \qquad \text{and} \qquad -x_k \geq -b \quad \text{for every } k.$$

So $\langle -x_n \rangle_1^\infty$ is a decreasing sequence bounded below in $\mathbb{R}$. Since statement B has been assumed to be true, there must be a number y such that $-x_n \to y$. By part (i) we have

$$x_n = -(-x_n) \to -y = -(-x) = x.$$

On the assumption of statement B we have shown that every increasing sequence bounded above in $\mathbb{R}$ must converge to a number in $\mathbb{R}$. That is, statement B implies statement A.

(iii) Suppose $\langle x_n \rangle_1^\infty$ is a decreasing sequence bounded below in $\mathbb{R}$.

$$x_1 \geq x_2 \geq x_3 \geq \cdots \qquad \text{and} \qquad x_k \geq c \quad \text{for every } k.$$

Then from part (ii) there must be a real number a with $x_n \to a$. Let $S = \{x_1, x_2, x_3, \dots\}$. Then $x \geq c$ for every x in S.

This part of the exercise can be interpreted in two ways:

I. Suppose that we know that the following is true:

C. *If S is a nonempty subset of $\mathbb{R}$ which is bounded below in $\mathbb{R}$, then S must have a greatest lower bound in $\mathbb{R}$.*

Use this to show that our S has a greatest lower bound $\inf S$ in $\mathbb{R}$ and then show that $x_n \to \inf S$.

II. Show directly that the limit a guaranteed by part (ii) must be a greatest lower bound for S.

Solution I: The set S is not empty since $x_1 \in S$. and it is bounded below since $c \leq x$ for every $x \in S$. The assumption of statement C implies that S has a greatest lower bound $\inf S$ in $\mathbb{R}$. Let $\varepsilon > 0$. Since $\inf S$ is a greatest lower bound for S, the number $\inf S + \varepsilon$ is not a lower bound. There must be an index N with $\inf S \leq x_N < \inf S + \varepsilon$. The sequence is decreasing, so if $n \geq N$ we can conclude

$$\inf S \leq x_n \leq x_N < \inf S + \varepsilon.$$

So $|x_n - \inf S| < \varepsilon$ whenever $n \geq N$. Thus $x_n \to \inf S$ as claimed.

Solution II: With this interpretation of the problem, we do not know already that S has a greatest lower bound. rather we are going to prove that the limit a which is guaranteed by part (ii) must be a greatest lower bound for S.

Claim One: a is a lower bound for S:

Suppose there were an index n with $x_n < a$. Let $\varepsilon = (a - x_n)/2$. For every k larger than n we would have $x_k \leq x_n < a$. So $|x_k - a| = a - x_k \geq a - x_n > \varepsilon$.. The sequence could not converge to a, but it does. Thus no such index n can exist and we must have $x_n \geq a$ for every n. So a is a lower bound for the set S.

Claim Two: If c is a lower bound for S, then $c \leq a$:

Let c be a lower bound for S. Then $x_n \geq c$ for every n. If $c > a$, let $\varepsilon = (c - a)/2$. For each index n we would have $x_n \geq c > a$, so that $|x_n - a| = x_n - a \geq c - a > \varepsilon$. The sequence could not converge to a. But it does. We conclude that $c > a$ is impossible, so $c \leq a$ as claimed.

The assertions of the two claims together say that the number a is a greatest lower bound for S. So $\inf S$ exists and is equal to the limit a of the sequence as claimed. ♦

◇ **1E-19.** Let $x = (1, 1, 1) \in \mathbb{R}^3$ be written $x = y_1 f_1 + y_2 f_2 + y_3 f_3$, where $f_1 = (1, 0, 1)$, $f_2 = (0, 1, 1)$, and $f_3 = (1, 1, 0)$. Compute the components y_i.

Answer. $y_1 = y_2 = y_3 = 1/2$. ◇

Solution. If $x = (1, 1, 1) = y_1 f_1 + y_2 f_2 + y_3 f_3$ where $f_1 = (1, 0, 1)$,$f_2 = (0, 1, 1)$, and $f_3 = (1, 1, 0)$, then

$$\begin{aligned}
(1, 1, 1) &= y_1(1, 0, 1) + y_2(0, 1, 1) + y_3(1, 1, 0) \\
&= (y_1, 0, y_1) + (0, y_2, y_2) + (y_3, y_3, 0) \\
&= (y_1 + y_3, y_2 + y_3, y_1 + y_2).
\end{aligned}$$

This is the vector form of a system of three equations:

$$\begin{aligned} y_1 + y_3 &= 1 \\ y_2 + y_3 &= 1 \\ y_1 + y_2 &= 1. \end{aligned}$$

Subtracting the second from the first gives $y_1 - y_2 = 0$, so that $y_2 = y_1$. Subtracting the second from the third gives $y_1 - y_3 = 0$, so that $y_3 = y_1$. Thus $y_1 = y_2 = y_3$. All three are equal. Putting this fact into any one of the three equations gives $y_1 = y_2 = y_3 = 1/2$. ♦

◇ **1E-20.** Let S and T be nonzero orthogonal subspaces of $\mathbb{R}^n$. Prove that if S and T are orthogonal complements (that is, S and T span all of $\mathbb{R}^n$), then $S \cap T = \{0\}$ and $\dim(S) + \dim(T) = n$, where $\dim(S)$ denotes dimension of S. Give examples in $\mathbb{R}^3$ of nonzero orthogonal subspaces for which the condition $\dim(S) + \dim(T) = n$ holds and examples where it fails. Can it fail in $\mathbb{R}^2$?

Suggestion. First show that $S \cap T = \{0\}$ by observing that anything in the intersection is orthogonal to itself and computing what that means. Having done that, check that a basis for $\mathbb{R}^n$ can be formed as the union of a basis for S and a basis for T. ◇

Solution. To say that S and T are nonzero subspaces is to say that there is at least one nonzero vector v in S and at least one nonzero vector w in T. The statement "S and T are orthogonal subspaces" means that they are orthogonal to each other. That is, $v \perp w$ whenever $v \in S$ and $w \in T$. This as abbreviated as $S \perp T$. To say that they are complementary subspaces means that they span the whole space in the sense that if u is any vector, then there are vectors v in S and w in T with $v+w = u$. This is abbreviated by $S + T = \mathbb{R}^n$. The solution can be broken up into useful smaller pieces.

Proposition (1). *If S and T are subspaces, then* $\dim(S+T) \le \dim S + \dim T$.

Proof: Suppose $\dim S = s$ and $\dim T = t$. Let $B_T = \{v_1, v_2, \ldots, v_s\}$ be a basis for S and $B_T = \{w_1, w_2, \ldots, w_t\}$ be a basis for T. Let $B = B_S \cup B_T$. If $u \in S+T$, then there are vectors v in S and w in T with $u = v+w$. There are numbers λ_k and μ_j, $1 \le k \le s$ and $1 \le j \le t$ such that $v = \sum_k \lambda_k v_k$ and $w = \sum_j \mu_j w_j$. So $u = \sum_k \lambda_k v_k + \sum_j \mu_j w_j$. Thus the set B spans the space $S + T$. We must have $\dim(S + T) \le s + t = \dim S + \dim T$.

Proposition (2). *If $S \cap T = \{0\}$, then* $\dim(S+T) = \dim S + \dim T$.

Proof: Let B_S, B_T, and B be as above. Now suppose $\sum_k \alpha_k v_k + \sum_j \beta_j w_j = 0$. Put $v = \sum_k \alpha_k v_k$ and $w = \sum_j \beta_j w_j$. Then $v \in S$ and $w \in T$. But also,

$v + w = 0$, so $v = -w \in T$ and $w = -v \in S$. Thus v and w are in $S \cap T$. So both must be 0. So $\sum_k \alpha_k v_k = 0$ and $\sum_j \beta_j w_j = 0$. Since B_S and B_T are both linearly independent sets, this means that $v_k = 0$ for every k and $w_j = 0$ for every j. Thus B is a linearly independent set. Since it has $s + t$ vectors in it, we have $\dim(S+T) \geq s+t = \dim S + \dim T$. With Proposition 1 we have inequality in both directions, so we have equality as claimed.

Proposition (3). *If $S \perp T$, then $S \cap T = \{0\}$.*

Proof: Since S and T are both subspaces, the zero vector is in both and we have $\{0\} \subseteq S \cap T$. On the other hand, if $v \in S \cap T$, then $v \in S$ and $v \in T$, so $v \perp v$. But we know that $v \perp w \iff \langle v, w \rangle = 0$. So $\| v \|^2 = \langle v, v \rangle = 0$. This forces $\| v \| = 0$ so that $v = 0$. Thus $S \cap T \subseteq \{0\}$. We have inclusion in both directions, so the sets are equal as claimed.

Now suppose S and T are subspaces which are orthogonal complements of each other. Then $S \perp T$, so $S \cap T = \{0\}$ by Proposition 3. Since $S + T = \mathbb{R}^n$, we then use proposition 2 to conclude that

$$n = \dim \mathbb{R}^n = \dim(S + T) = \dim S + \dim T$$

as claimed.

If S and T are nonzero subspaces of $\mathbb{R}^3$, then each has dimension either 1, 2, or 3. For example, let $v_0 = (1, 0, 0)$ and $w_0 = (0, 1, 0)$. Let $S = \{\lambda v_0 \mid \lambda \in \mathbb{R}\}$ and $T = \{\mu w_0 \mid \mu \in \mathbb{R}\}$. Then S and T are subspaces of $\mathbb{R}^3$ with $\dim S = \dim T = 1$. If $v \in S$ and $w \in T$, then there are numbers λ and μ such that $\langle v, w \rangle = \langle \lambda v_0, \mu w_0 \rangle = \lambda\mu \langle v_0, w_0 \rangle = \lambda\mu(1 \cdot 0 + 0 \cdot 1 + 0 \cdot 0) = 0$. Thus $S \perp T$. But $\dim S + \dim T = 2$ not 3.

If S and T are nonzero subspaces of $\mathbb{R}^2$, then each has dimension either 1 or 2. If they are mutually orthogonal, then neither can be all of $\mathbb{R}^2$. (The other would then have to be just the zero vector.) So we must have $\dim S = \dim T = 1$. So $\dim S + \dim T = 2 = \dim \mathbb{R}^2$. ♦

◇ **1E-21.** Show that the sequence in Worked Example 1WE-2 can be chosen to be increasing.

Sketch. If $x \in S$, let $x_k = x$. Otherwise, $x_k < x$ for all k. Let $y_1 = x_1$, y_2 be the first of $x_2, x_3, \ldots$ with $y_1 \leq y_2$, etc. ◇

Solution. We need to show that if S is a nonempty subset of $\mathbb{R}$ which is bounded above, then there is an increasing sequence $\langle x_n \rangle_1^\infty$ in S which converges to $\sup S$. If S is such a set, then we know that S has a least upper bound $x = \sup S \in \mathbb{R}$. If $x \in S$, we may simply take $x_n = x$ for all n. We certainly would have $x_n \to x$, and since they are all the same, we also have

$$x_1 \leq x_2 \leq x_3 \leq \cdots .$$

If x is not in S, we have to do a bit more work. Since we know S is not empty, we can select any point x_1 in S to begin the sequence. Then $x_1 < x$. Set $y_1 = (x_1 + x)/2$. Since y_1 is strictly less than x and x is the least upper bound for S, y_1 is not an upper bound. We can select a point x_2 in S with $x_1 < y_1 < x_2 < x$. The last inequality is strict since x is not in S. Note also that $|x - x_2| \le |x - x_1|/2$. We can now repeat this process inductively to generate the desired sequence. Suppose points $x_1, x_2, \dots, x_k$ have been selected in S in such a way that

$$x_1 < x_2 < x_3 < \cdots < x_k < x \qquad \text{and} \qquad |x - x_k| \le |x - x_1|/2^{k-1}.$$

Set $y_k = (x_k + x)/2$ so that $x_k < y_k < x$. Since x is the least upper bound of S, the number y_k is not an upper bound and we can select a point x_{k+1} in S with $x_k < y_k \le x_{k+1} < x$. The last inequality is strict since x is not in S. We then have

$$x_1 < x_2 < x_3 < \cdots < x_k < x_{k+1} < x$$

and

$$|x - x_{k+1}| = x - x_{k+1} \le (x - x_k)/2 \le |x - x_1|/2^k.$$

The sequence thus generated is strictly increasing. Since $|x - x_1|/2^k \to 0$ as $k \to \infty$, it converges to x. ♦

◇ **1E-22.** (a) If x_n and y_n are bounded sequences in $\mathbb{R}$, prove that

$$\limsup(x_n + y_n) \le \limsup x_n + \limsup y_n.$$

(b) Is the product rule true for lim sups?

Suggestion. Show that if $\varepsilon > 0$, then $x_n + y_n < \limsup x_n + \limsup y_n + \varepsilon$ for large enough n. ◇

Solution. (a) Let $A = \limsup x_n$ and $B = \limsup y_n$. Then A and B are both finite real numbers since $\langle x_n \rangle_1^\infty$ and $\langle y_n \rangle_1^\infty$ are bounded sequences. Suppose $\varepsilon > 0$. There are indices N_1 and N_2 such that $x_n < A + \varepsilon/2$ whenever $n \ge N_1$ and $y_n < B + \varepsilon/2$ whenever $n \ge N_2$. Let $N = \max(N_1, N_2)$. If $n \ge N$, we have

$$x_n + y_n < x_n + B + \varepsilon/2 < A + B + \varepsilon.$$

Since this can be done for any $\varepsilon > 0$, we conclude that $x_n + y_n$ can have no cluster points larger than A+B. Thus

$$\limsup(x_n + y_n) \le A + B = \limsup x_n + \limsup y_n$$

as claimed.

We cannot guarantee equality. The inequality just established could be strict. For example, consider the sequence defined by $x_n = (-1)^n$ and $y_n = (-1)^{n+1}$. Then $\limsup x_n = \limsup y_n = 1$, but $x_n + y_n = 0$ for each n, so $\limsup(x_n + y_n) = 0$.

(b) There does not seem to be any reasonable relationship which can be guaranteed between the two quantities $\limsup(x_n y_n)$ and $(\limsup x_n) \cdot (\limsup y_n)$. Each of the possibilities "$<$", "$=$", and "$>$" can occur:
Example 1. Let $\langle x_n \rangle_1^\infty$ be the sequence $\langle 1, 0, 1, 0, 1, 0, \dots \rangle$ and $\langle y_n \rangle_1^\infty$ be the sequence $\langle 0, 1, 0, 1, 0, 1, \dots \rangle$. Then $\limsup x_n = \limsup y_n = 1$. But $x_n y_n = 0$ for all n. We have

$$\limsup(x_n y_n) = 0 < 1 = (\limsup x_n) \cdot (\limsup y_n).$$

Example 2. Let $x_n = y_n = 1$ for each n. Then $\limsup x_n = \limsup y_n = 1$. But $x_n y_n = 1$ for all n. We have

$$\limsup(x_n y_n) = 1 = (\limsup x_n) \cdot (\limsup y_n).$$

Example 3. Let $\langle x_n \rangle_1^\infty$ be the sequence $\langle 0, -1, 0, -1, 0, -1, 0, \dots \rangle$ and $\langle y_n \rangle_1^\infty$ be the sequence $\langle 0, -1, 0, -1, 0, -1, \dots \rangle$. Then $\limsup x_n = \limsup y_n = 0$. But $x_n y_n = 0$ is the sequence $\langle 0, 1, 0, 1, 0, 1, \dots \rangle$. We have

$$\limsup(x_n y_n) = 1 > 0 = (\limsup x_n) \cdot (\limsup y_n).$$

♦

◇ **1E-23.** Let $P \subset \mathbb{R}$ be a set such that $x \geq 0$ for all $x \in P$ and for each integer k there is an $x_k \in P$ such that $k x_k \leq 1$. Prove that $0 = \inf(P)$.

Sketch. Since 0 is a lower bound for P, $0 \leq \inf(P)$. Use the given condition to rule out $0 < \inf(P)$. ◇

Solution. With $k = 1$ the hypothesis gives an element x_1 of the set P with $x_1 \leq 1$. In particular, P is not empty. We have assumed that $x \geq 0$ for every x in P. So 0 is a lower bound for P. We conclude that $\inf P$ exists as a finite real number and that $0 \leq \inf P$.

Now let $\varepsilon > 0$. By the Archimedean Principle there is an integer k with $0 < 1/k < \varepsilon$. By hypothesis, there is an element x_k of P with $k x_k \leq 1$. So $0 \leq x_k \leq 1/k < \varepsilon$. So ε is not an lower bound for P and $\inf P \leq \varepsilon$. This holds for every $\varepsilon > 0$, so $\inf P \leq 0$. We have inequality in both directions, so $\inf P = 0$ as claimed. ♦

◇ **1E-24.** If $\sup(P) = \sup(Q)$ and $\inf(P) = \inf(Q)$, does $P = Q$?

Solution. No, P and Q need not be the same. More or less, they have the same top and bottom ends, but in between anything can happen. For example, let P be the two point set $\{0,1\}$, and let Q be the closed interval $[0,1]$. Then $\inf P = \inf Q = 0$ and $\sup P = \sup Q = 1$, but the two sets are certainly not the same. ♦

◇ **1E-25.** We say that $P \le Q$ if for each $x \in P$, there is a $y \in Q$ with $x \le y$.

(a) If $P \le Q$, then show that $\sup(P) \le \sup(Q)$.
(b) Is it true that $\inf(P) \le \inf(Q)$?
(c) If $P \le Q$ and $Q \le P$, does $P = Q$?

Solution. (a) If $x \in P$, then there is a $y \in Q$ with $x \le y \le \sup Q$. So $\sup Q$ is an upper bound for P. Thus $\sup P \le \sup Q$ as claimed.

(b) Not necessarily. It is possible that $\inf P > \inf Q$. Consider $P = [0,1]$ and $Q = [-1,1]$. Then $P \le Q$ but $\inf P > \inf Q$.

(c) Again, not necessarily. It is possible that P and Q are different sets even if $P \le Q$ and $Q \le P$. Consider, for example, $P = \mathbb{R} \setminus \mathbb{Q}$, and $Q = \mathbb{Q}$. Here is another: $P = \{1\}$ and $Q = \{0,1\}$. Then $P \le Q$ and $Q \le P$, but $P \ne Q$.

♦

◇ **1E-26.** Assume that $A = \{a_{m,n} \mid m = 1,2,3,\dots \text{ and } n = 1,2,3,\dots\}$ is a bounded set and that $a_{m,n} \ge a_{p,q}$ whenever $m \ge p$ and $n \ge q$. Show that

$$\lim_{n\to\infty} a_{n,n} = \sup A.$$

Solution. Since A is a bounded, nonempty subset of $\mathbb{R}$, we know that $c = \sup A$ exists as a finite real number and that $a_{j,k} \le c$ for all j and k. For convenience, let $b_n = a_{n,n}$. If $n \le k$, we have $b_n = a_{n,n} \le a_{k,k} = b_k$. So the sequence $\langle b_n \rangle_1^\infty$ is increasing and bounded above by c. So $b = \lim_{n\to\infty} b_n$ exists and $b \le c$. Let $d < c$, then there is a $a_{k,j}$ in A with $d < a_{k,j} \le c$. If $n \ge \max(k,j)$, then

$$d < a_{k,j} \le a_{n,n} = b_n \le b_{n+1} \le \dots .$$

So $b = \lim b_n > d$. This is true for every $d < c$, so $b \ge c$. We have inequality in both directions, so

$$\lim_{n\to\infty} a_{n,n} = \lim_{n\to\infty} b_n = b = c = \sup A$$

as claimed. ♦

◇ **1E-27.** Let $S = \{(x, y) \in \mathbb{R}^2 \mid xy > 1\}$ and $B = \{d((x, y), (0, 0)) \mid (x, y) \in S\}$. Find $\inf(B)$.

Answer. $\inf(B) = \sqrt{2}$. ◇

Solution. The set S is the region "outside" the unit hyperbola illustrated in Figure 1.1. The number D is the minimum distance from that hyperbola to the origin. It is reasonably geometrically clear from the figure that this distance is $\sqrt{2}$ and that it is attained at the points $(1, 1)$ and $(-1, -1)$.

FIGURE 1.1. The distance from a hyperbola to the origin.

There are several ways to check this analytically. Here is one. Note that if $\varepsilon > 0$, then the point $P = (1 + \varepsilon, 1 + \varepsilon)$ is in S, and that $d(P, (0, 0)) = (1 + \varepsilon)\sqrt{2} = \sqrt{2} + \varepsilon\sqrt{2}$. So $D \leq \sqrt{2} + \varepsilon\sqrt{2}$. This holds for every $\varepsilon > 0$, so $D \leq \sqrt{2}$.

On the other hand, if $(x, y) \in S$, then $xy > 1$, so $d((x, y), (0, 0))^2 = x^2 + y^2 \geq 2xy > 2$, so $d((x, y), (0, 0)) > \sqrt{2}$. (We use here the fact that $x^2 + y^2 \geq 2xy$ which is established by expanding the inequality $(x - y)^2 \geq 0$.) Since this holds for all points in S, we have $D \geq \sqrt{2}$.

We have inequality in both directions, so $D = \sqrt{2}$. ♦

◇ **1E-28.** Let x_n be a convergent sequence in $\mathbb{R}$ and define $A_n = \sup\{x_n, x_{n+1}, \dots\}$ and $B_n = \inf\{x_n, x_{n+1}, \dots\}$. Prove that A_n converges to the same limit as B_n, which in turn is the same as the limit of x_n.

Solution. Suppose $x_n \to a \in \mathbb{R}$. Let $\varepsilon > 0$. There is an index N such that $a-\varepsilon < x_k < a+\varepsilon$ whenever $k \geq N$. Thus if $n \geq N$ we have $a-\varepsilon < x < a+\varepsilon$ for all x in the set $S_n = \{x_n, x_{n+1}, x_{n+2}, \dots\}$. That is, $a-\varepsilon$ is a lower bound for S_n, and $a+\varepsilon$ is an upper bound. Hence

$$a - \varepsilon \leq B_n = \inf S_n \leq \sup S_n = A_n \leq a + \varepsilon.$$

So

$$|A_n - a| \leq \varepsilon \qquad \text{and} \qquad |B_n - a| \leq \varepsilon$$

whenever $n \geq N$. We conclude that

$$\lim_{n\to\infty} A_n = a = \lim_{n\to\infty} B_n$$

as claimed. ♦

◇ **1E-29.** For each $x \in \mathbb{R}$ satisfying $x \geq 0$, prove the existence of $y \in \mathbb{R}$ such that $y^2 = x$.

Sketch. Let $y = \sup\{u \in \mathbb{R} \mid u^2 \leq x\}$. Obtain contradictions from $y^2 < x$ and from $y^2 > x$. ◊

Solution. There are several ways to establish the existence of a real square root for every nonnegative real number. Whatever method is selected must use the completeness of the real line or something very much like it, since we know that it simply is not true if we only have rational numbers. We seek a proof of the following.

Theorem. *If $c \in \mathbb{R}$ and $c \geq 0$, then there is exactly one b in $\mathbb{R}$ with $b \geq 0$ and $b^2 = c$. Each positive real number has a unique positive real square root.*

It is convenient to establish the following elementary lemma.

Lemma. *If x and y are nonnegative real numbers, then $x^2 \leq y^2$ if and only if $x \leq y$.*

Proof: First suppose $0 \leq x \leq y$. Then multiplication by either x or y preserves the order of inequalities. We can compute

$$x^2 \leq xy \qquad \text{and} \qquad xy \leq y^2.$$

We conclude that $x^2 \leq y^2$ as desired. For the converse implication, assume that $x \geq 0$, $y \geq 0$, and that $x^2 \leq y^2$. Then

$$0 \leq y^2 - x^2 = (y - x)(y + x).$$

If y and x are both 0, then we certainly have $x \leq y$. If they are not both 0, then $x + y > 0$ since they are both nonnegative. Multiplication of the last inequality by the positive number $(x + y)^{-1}$ gives $0 \leq y - x$. So $x \leq y$ as claimed.

Proof of the Theorem: Let $S = \{x \in \mathbb{R} \mid x^2 \leq c\}$. The intuition is that $S = \{x \in \mathbb{R} \mid -\sqrt{c} \leq x \leq \sqrt{c}\}$ and that $\sup S = \sqrt{c}$ is the number we want. However, since the point of the exercise is to show that there is such a thing as $\sqrt{c}$, we can't just write this down. The plan is to show first that S is bounded and not empty so that it has a least upper bound, and then to show that this least upper bound is a square root for c. Since $0^2 = 0 \leq c$, we see that $0 \in S$ so that S is not empty. Now suppose $x > c + 1$. Using the lemma we can compute

$$x^2 \geq (c + 1)^2 = c^2 + 2c + 1 > c.$$

So x is not in S. Thus $c+1$ is an upper bound for S. The set S is a nonempty subset of $\mathbb{R}$ which is bounded above. Since $\mathbb{R}$ is complete, $b = \sup S$ exists as a finite real number. We want to show that $b^2 = c$. To do this we obtain contradictions from the other two possibilities.

Suppose $b^2 < c$. In this case we will see that we can get a slightly larger number whose square is also less than c. This number would then be in S contradicting the choice of b as an upper bound for S. To do this, simply select a number d small enough so that $2bd + d^2 < c - b^2$. Then

$$(b + d)^2 = b^2 + 2bd + d^2 < b^2 + (c - b^2) = c.$$

Since b is an upper bound for S, the inequality $b^2 < c$ must be false.

Now suppose $b^2 > c$. In this case we will be able to find a number slightly smaller than b whose square is also larger than c. By the lemma, this number would be an upper bound for S. But that would contradict the selection of b as the least upper bound. So this inequality must also be false. To accomplish this, select d small enough so that $0 < d < b$ and $2bd - d^2 < b^2 - c$. Then

$$(b - d)^2 = b^2 - 2bd + d^2 = b^2 - (2bd - d^2) > b^2 - (b^2 - c) = c.$$

We have $b - d > 0$ and $(b - d)^2 > c$. By the lemma, if $x > b - d$, then $x^2 \geq (b - d)^2 > c$, so x is not in S. Thus $b - d$ would be an upper bound for S contradicting the selection of b as the least upper bound for S.

Both inequalities have led to contradiction, so we must have $b^2 = c$.

We have shown that c has one positive square root. Now we want to show that it has only one. If we also had $B^2 = c$, then we would have

$$0 = c - c = B^2 - b^2 = (B - b)(B + b).$$

So either $B - b = 0$ or $B + b = 0$. Thus $B = \pm b$. So b is the only nonnegative square root for c. The only other real number whose square is c is $-b$. We conclude that each nonnegative real number has a unique nonnegative real square root. ◆

◇ **1E-30.** Let $\mathcal{V}$ be the vector space $\mathcal{C}([0,1])$ with the norm $\|f\|_\infty = \sup\{|f(x)| \mid x \in [0,1]\}$. Show that the parallelogram law fails and conclude that this norm does not come from any inner product. (Refer to Exercise 1E-12.)

Solution. Let $f(x) = x$ and $g(x) = 1 - x$. Both of these functions are in $\mathcal{C}([0,1])$, and $\| f \|_\infty = \sup\{x \mid x \in [0,1]\} = 1$ and $\| g \|_\infty = \sup\{1 - x \mid x \in [0,1]\} = 1$. For the sum and difference we have, $(f+g)(x) = 1$ and $(f-g)(x) = 2x - 1$. So

$$\| f + g \|_\infty + \| f - g \|_\infty = 1 + 1 = 2$$

while

$$2 \| f \|_\infty^2 + 2 \| g \|_\infty^2 = 2 \cdot 1^2 + 2 \cdot 1^2 = 4.$$

Since 2 and 4 are not the same, we see that this norm does not satisfy the parallelogram law. If there were any way to define an inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{C}([0,1])$ in such a way that $\| h \|_\infty^2 = \langle h, h \rangle$ for each h in the space, then the parallelogram law would have to hold by the work of Exercise 1E-12(a). Since it does not, there can be no such inner product.

Another example which is useful for other problems also is shown in the figure. We let f and g be "tent functions" based on intervals which do not overlap. For example we can use the subintervals $[0, 1/2]$ and $[1/2, 1]$ of the unit interval as illustrated in the figure.

FIGURE 1.2. "Tent functions" on disjoint intervals.

For these choices of f and g, we have

$$\| f \|_\infty = \| g \|_\infty = \| f + g \|_\infty = \| f - g \|_\infty = 1.$$

So

$$\| f + g \|_\infty{}^2 + \| f - g \|_\infty{}^2 = 2,$$

while

$$2\| f \|_\infty{}^2 + 2\| g \|_\infty{}^2 = 4.$$

Since 2 is not equal to 4, the parallelogram law fails for these functions also. ♦

◇ **1E-31.** Let $A, B \subset \mathbb{R}$ and $f : A \times B \to \mathbb{R}$ be bounded. Is it true that

$$\sup\{\, f(x,y) \mid (x,y) \in A \times B\} = \sup\{\sup\{\, f(x,y) \mid x \in A\} \mid y \in B\}$$

or, the same thing in different notation,

$$\sup_{(x,y)\in A\times B} f(x,y) = \sup_{y\in B}\left(\sup_{x\in A}\ f(x,y)\right)?$$

Answer. True. ◇

Solution. First suppose (s,t) is a point in $A \times B$. Then $s \in A$ and $t \in B$, so

$$\begin{aligned} f(s,t) &\le \sup_{x\in A} f(x,t) \\ &\le \sup_{y\in B}\left(\sup_{x\in A} f(x,y)\right) \end{aligned}$$

Since this is true for every point (s,t) in $A \times B$, we can conclude that

$$\sup_{(x,y)\in A\times B} f(x,y) \le \sup_{y\in B}\left(\sup_{x\in A} f(x,y)\right).$$

For the opposite inequality, let $\varepsilon > 0$. There is a $t \in B$ such that

$$\sup_{x\in A} f(x,t) > \sup_{y\in B}\left(\sup_{x\in A} f(x,y)\right) - \varepsilon/2.$$

So there must be an $s \in A$ with

$$f(s,t) > \sup_{y\in B}\left(\sup_{x\in A} f(x,y)\right) - \varepsilon/2 - \varepsilon/2.$$

So

$$\sup_{(x,y)\in A\times B} f(x,y) \ge \sup_{y\in B}\left(\sup_{x\in A} f(x,y)\right) - \varepsilon.$$

Since this holds for every $\varepsilon > 0$, we must have

$$\sup_{(x,y)\in A\times B} f(x,y) \ge \sup_{y\in B}\left(\sup_{x\in A} f(x,y)\right).$$

We have inequality in both directions, and thus equality as claimed. ♦

◇ **1E-32.** (a) Give a reasonable definition for what $\lim_{n\to\infty} x_n = \infty$ should mean.

(b) Let $x_1 = 1$ and define inductively $x_{n+1} = (x_1 + \cdots + x_n)/2$. Prove that $x_n \to \infty$.

Solution. (a) **Definition.** *We say the sequence* $\langle x_n \rangle_1^\infty$ *tends to infinity and write* $\lim_{n\to\infty} x_n = \infty$ *if for each* $B > 0$ *there is an* N *such that* $x_n > B$ *whenever* $n \geq N$.

(b) Computation of the first few terms of the sequence shows that

$$x_1 = 1 \qquad x_2 = 1/2 \qquad x_3 = 3/4 \qquad x_4 = 9/8 \qquad \ldots$$

If x_1, x_2, x_3, ... , x_n are all positive, then $x_{n+1} = (x_1 + \cdots + x_n)/2$ must be positive also. Since the first few terms are positive, it follows by induction that all terms of the sequence are positive. With this in hand we can get a lower estimate for the terms. If $n \geq 4$, then

$$x_{n+1} = \frac{x_1 + x_2 + \cdots + x_n}{2} > \frac{x_1 + x_4}{2} > 1.$$

If we feed this information back into the formula we find

$$\begin{aligned} x_{n+1} &= \frac{x_1 + x_2 + \cdots + x_n}{2} > \frac{1 + 1/2 + 3/4 + \overbrace{1 + 1 + 1 + \cdots + 1}^{n-3 \text{ terms}}}{2} \\ &> \frac{n-1}{2}. \end{aligned}$$

Thus $x_n > (n/2) - 1$ for $n = 5, 6, 7, \ldots$. If $B > 0$ and $n > 2B + 2$, then $x_n > B$. Thus $\lim_{n\to\infty} = \infty$ as claimed.

Method Two: Observation of the first few terms might lead to the conjecture that $x_n = 3^{n-2}/2^{n-1} = (1/2)(3/2)^{n-2}$ for $n = 2, 3, 4, \ldots$. We can confirm this by induction. It is true by direct computation for $n =$ 2, 3, and 4. If it is true for $1 \leq k \leq n$, then

$$\begin{aligned} x_{n+1} &= \frac{x_1 + x_2 + \cdots + x_n}{2} \\ &= \frac{1}{2}\left(1 + \frac{1}{2} + \frac{1}{2}\left(\frac{3}{2}\right) + \frac{1}{2}\left(\frac{3}{2}\right)^2 + \cdots + \frac{1}{2}\left(\frac{3}{2}\right)^{n-2}\right) \\ &= \frac{1}{2}\left(1 + \frac{1}{2} \cdot \frac{1 - (3/2)^{n-1}}{1 - (3/2)}\right) \\ &= \frac{1}{2}\left(1 + \frac{1 - (3/2)^{n-1}}{2 - 3}\right) = \frac{1}{2}\left(\frac{3}{2}\right)^{n-1} \end{aligned}$$

This is exactly the desired form for the $(n+1)^{st}$ term. Since $x_n = (1/2)(3/2)^{n-2}$, a straightforward computation shows that $x_n > B$ whenever $n > \frac{\log(2B)}{\log(3/2)} + 2$. Again we can conclude that $\lim_{n\to\infty} x_n = \infty$.

Method Three: A bit of clever manipulation with the formula first along with a much easier induction produces the same formula for the n^th term.

$$\begin{aligned} x_{n+1} &= \frac{x_1 + \cdots + x_n}{2} = \frac{x_1 + \cdots + x_{n-1}}{2} + \frac{x_n}{2} \\ &= x_n + \frac{x_n}{2} = \frac{3}{2} x_n \quad \text{for } n = 2, 3, \ldots \end{aligned}$$

With this and the first two terms, a much simpler induction produces the same general formula for x_n as in Method Two.

♦

◇ **1E-33.** (a) Show that $(\log x)/x \to 0$ as $x \to \infty$. (You may consult your calculus text and use, for example, l'Hôpital's rule.)

(b) Show that $n^{1/n} \to 1$ as $n \to \infty$.

Sketch. (a) Use the ∞/∞ form of L'Hôpital's rule.

(b) Take logarithms and use (a).

◇

Solution. (a) Since both x and $\log x$ tend to infinity with x, the expression $(\log x)/x$ has the "indeterminate form" "∞/∞" there. We may try L'Hôpital's Rule. We examine the limit of the ratio of derivatives

$$\lim_{x\to\infty} \frac{1/x}{1} = \lim_{x\to\infty} \frac{1}{x} = 0.$$

By L'Hôpital's Rule, $\lim_{x\to\infty} \dfrac{\log x}{x}$ also exists and is 0.

(b) Let $a_n = n^{1/n}$ for $n = 1, 2, 3, \ldots$. Set $b_n = \log a_n = (1/n)\log(n)$. From part (a) we know that $(1/x)\log(x)$ is near 0 for large x. So it certainly is near 0 if x happens to be a large positive integer. Thus

$$\lim_{n\to\infty} \log a_n = \lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{\log n}{n} = 0.$$

Since $\log a_n \to 0$, we can conclude by the continuity of the exponential function, that

$$n^{1/n} = a_n = e^{\log a_n} \to e^0 = 1.$$

(We will study continuity and its relationship to sequences more fully in Chapters 3 and 4, and the exponential function in particular in Chapter 5.)

♦

◇ **1E-34.** Express the following complex numbers in the form $a + bi$:

(a) $(2+3i)(4+i)$
(b) $(8+6i)^2$
(c) $(1+3/(1+i))^2$

Answer. (a) $(2+3i)(4+i) = 5+14i$
(b) $(8+6i)^2 = 28+96i$
(c) $(1+3/(1+i))^2 = 6-(5/2)i$

♢

Solution. (a) $(2+3i)(4+i) = 2(4+i)+3i(4+i) = 8+2i+12i+3i^2 = 8+14i-3 = 5+14i$
(b) $(8+6i)^2 = 64+96i+36i^2 = 64-36+96i = 28+96i$
(c)

$$\begin{aligned}\left(1+\frac{3}{1+i}\right)^2 &= \left(1+\frac{3}{1+i}\,\frac{1-i}{1-i}\right)^2 = \left(1+\frac{3-3i}{1+1}\right)^2\\ &= \left(\frac{5}{2}-\frac{1}{2}\,i\right)^2 = \frac{1}{4}\,(5-i)^2\\ &= \frac{1}{4}\,(25-10i+i^2) = \frac{24-10i}{4} = 6-\frac{5}{2}\,i\end{aligned}$$

♦

◇ **1E-35.** What is the complex conjugate of $(3+8i)^4/(1+i)^{10}$?

Answer. $(3-8i)^4/(1-i)^{10}$. ♢

Solution. Use properties (v), (vi), and (vii) from Proposition 1.8.1.

$$\overline{\left(\frac{(3+8i)^4}{(1+i)^{10}}\right)} = \frac{\overline{(3+8i)^4}}{\overline{(1+i)^{10}}} = \frac{\left(\overline{(3+8i)}\right)^4}{\left(\overline{(1+i)}\right)^{10}} = \frac{(3-8i)^4}{(1-i)^{10}}$$

♦

◇ **1E-36.** Find the solutions to:

(a) $(z+1)^2 = 3+4i$.
(b) $z^4 - i = 0$.

Answer. (a) Our solutions are $z = x+iy = -3-i$ and $z = x+iy = 1+i$.

(b) $z = e^{\pi i/8 + k\pi i/2}$, for $k = 0, 1, 2, 3$.

$\Diamond$

Solution. (a) Let $z = x + iy$ with x and y real. We want

$$\begin{aligned} 3 + 4i &= (z+1)^2 \\ &= (x + 1 + iy)^2 \\ &= (x+1)^2 - y^2 + 2(x+1)y\, i. \end{aligned}$$

For convenience change variables to $s = x + 1$. Comparing real and imaginary parts, we require

$$s^2 - y^2 = 3 \qquad \text{and} \qquad 2sy = 4.$$

This gives $s^2 - (2/s)^2 = 3$ so $s^4 - 3s^2 - 4 = 0$. This is a quadratic in s^2 with solutions $s^2 = 4$ and $s^2 = -1$. Since s is real, the second is impossible and the first gives $s = \pm 2$. The corresponding values of x and y are $x = s - 1 = -1 \pm 2$ and $y = 2/s = \pm 1$. Our solutions are $z = x + iy = -3 - i$ and $z = x + iy = 1 + i$.

(b) This problem is much more easily expressed and solved in polar coordinates. We want $z^4 = i = e^{\pi i/2} = e^{\pi i/2 + 2k\pi i}$ where k can be any integer. So solutions are given by $z = e^{\pi i/8 + k\pi i/2}$. The parameter k can be any integer, but due to the periodicity of the exponential, only four different values of z are produced corresponding to $k = 0, 1, 2, 3$. These are four points evenly spaced around the unit circle as indicated in Figure 1.3.

FIGURE 1.3. Solutions of $z^4 = i$.

$\blacklozenge$

$\diamond$ **1E-37.** Find the solutions to $z^2 = 3 - 4i$.

Answer. $z = \pm(2 - i)$. $\diamondsuit$

Solution. Let $z = x + iy$ with x and y real. We want

$$3 - 4i = z^2 = (x + iy)^2 = x^2 - y^2 + 2xy\, i.$$

Comparing real and imaginary parts, we require

$$x^2 - y^2 = 3 \qquad \text{and} \qquad 2xy = -4.$$

This gives $x^2 - (-2/x)^2 = 3$ or $x^4 - 3x^2 - 4 = 0$. This is a quadratic in x^2 with solutions $x^2 = 4$ and $x^2 = -1$. Since x is to be real, the second is impossible and the first gives $x = \pm 2$. The corresponding values of y are $y = \mp 1$. Our solutions are $z = 2 - i$ and $z = -2 + i$. That is, $z = \pm(2 - i)$. $\blacklozenge$

$\diamond$ **1E-38.** If a is real and z is complex, prove that $\mathrm{Re}(az) = a\ \mathrm{Re}\ z$ and that $\mathrm{Im}(az) = a\ \mathrm{Im}\ z$. Generally, show that $\mathrm{Re} : \mathbb{C} \to \mathbb{R}$ is a real linear map; that is, that $\mathrm{Re}\,(az + bw) = a\ \mathrm{Re}\ z + b\ \mathrm{Re}\ w$ for a, b real and z, w complex.

Solution. Let a be a real number and $z = x + iy$ a complex number with x and y real. Then since ax and ay are real, we have

$$\mathrm{Re}\,(az) = \mathrm{Re}\,(a(x + iy)) = \mathrm{Re}\,(ax + iay) = ax = a\ \mathrm{Re}\,(z)$$

and

$$\mathrm{Im}\,(az) = \mathrm{Im}\,(a(x + iy)) = \mathrm{Im}\,(ax + iay) = ay = a\ \mathrm{Im}\,(z)\,.$$

Now let b be another real number and $w = u + iv$ another complex number with u and v real. Then

$$\begin{aligned} \mathrm{Re}\,(az + bw) &= \mathrm{Re}\,(a(x + iy) + b(u + iv)) = \mathrm{Re}\,((ax + bu) + i(ay + bv)) \\ &= az + bu = a\ \mathrm{Re}\,(z) + b\ \mathrm{Re}\,(w)\,. \end{aligned}$$

$\blacklozenge$

$\diamond$ **1E-39.** Find the real and imaginary parts of the following, where $z = x + iy$:

(a) $1/z^2$
(b) $1/(3z + 2)$

Answer. (a) $\operatorname{Re}\left(1/z^2\right) = (x^2 - y^2)/(x^2 + y^2)^2$
$\operatorname{Im}\left(1/z^2\right) = -2xy/(x^2 + y^2)^2$.
(b) $\operatorname{Re}(1/(3z + 2)) = (3x + 2)/((3x + 2)^2 + 9y^2)$
$\operatorname{Im}(1/(3z + 2)) = -3y/((3x + 2)^2 + 9y^2)$.

$\diamondsuit$

Solution. (a) Let $z = x + iy$ with x and y real. Then

$$\begin{aligned}\frac{1}{z^2} = \frac{1}{x^2 - y^2 + 2xyi} &= \frac{1}{x^2 - y^2 + 2xyi} \cdot \frac{x^2 - y^2 - 2xyi}{x^2 - y^2 - 2xyi} \\ &= \frac{x^2 - y^2 - 2xyi}{(x^2 - y^2)^2 + 4x^2y^2} = \frac{x^2 - y^2 - 2xyi}{(x^2 + y^2)^2}.\end{aligned}$$

So

$$\operatorname{Re}\left(\frac{1}{z^2}\right) = \frac{x^2 - y^2}{(x^2 + y^2)^2} \qquad \text{and} \qquad \operatorname{Im}\left(\frac{1}{z^2}\right) = \frac{-2xy}{(x^2 + y^2)^2}.$$

(b)

$$\frac{1}{3z + 2} = \frac{1}{(3x + 2) + 3yi} \cdot \frac{(3x + 2) - 3yi}{(3x + 2) - 3yi} = \frac{(3x + 2) - 3yi}{(3x + 2)^2 + 9y^2}.$$

Thus

$$\begin{aligned}\operatorname{Re}\left(\frac{1}{3z + 2}\right) &= \frac{3x + 2}{(3x + 2)^2 + 9y^2} \qquad \text{and} \qquad (1.8.1)\\ \operatorname{Im}\left(\frac{1}{3z + 2}\right) &= \frac{-3y}{(3x + 2)^2 + 9y^2}.\end{aligned}$$

$\blacklozenge$

$\diamond$ **1E-40.** (a) Fix a complex number $z = x + iy$ and consider the linear mapping $\varphi_z : \mathbb{R}^2 \to \mathbb{R}^2$ (that is, of $\mathbb{C} \to \mathbb{C}$) defined by $\varphi_z(w) = z \cdot w$ (that is, multiplication by z). Prove that the matrix of φ_z in the standard basis $(1, 0), (0, 1)$ of $\mathbb{R}^2$ is given by

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}.$$

(b) Show that $\varphi_{z_1 z_2} = \varphi_{z_1} \circ \varphi_{z_2}$.

Solution. (a) Recall a bit of linear algebra. If T is a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ and $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, then there are real numbers a, b, c, and d such that

$$Te_1 = \begin{pmatrix} a \\ c \end{pmatrix} \qquad \text{and} \qquad Te_2 = \begin{pmatrix} b \\ d \end{pmatrix}.$$

Then T is represented by the matrix $\operatorname{Mat}(T) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in the sense that if $v = \begin{pmatrix} x \\ y \end{pmatrix}$ is any vector in $\mathbb{R}^2$, then

$$\begin{aligned} Tv = T(xe_1 + ye_2) = xTe_1 + yTe_2 &= \begin{pmatrix} ax \\ cx \end{pmatrix} + \begin{pmatrix} by \\ dy \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} \\ &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \operatorname{Mat}(T)\, v. \end{aligned}$$

A direct computation shows that this assignment of a matrix to a linear transformation preserves algebra. If S and T are linear transformations and α and β are numbers, then

$$\begin{gathered} \operatorname{Mat}(\alpha S + \beta T) = \alpha \operatorname{Mat}(S) + \beta \operatorname{Mat}(T) \\ \operatorname{Mat}(S \circ T) = \operatorname{Mat}(S) \cdot \operatorname{Mat}(T). \end{gathered}$$

Notice that the columns of $\operatorname{Mat}(T)$ are the images Te_1 and Te_2 of the basis vectors.

A complex number z can be "identified" with a vector in $\mathbb{R}^2$ by associating $z = x + iy$ with the vector $\begin{pmatrix} x \\ y \end{pmatrix}$. With this identification, the complex number 1 is associated with the vector e_1, and the complex number i is associated with e_2. If $z \in \mathbb{C}$, then the operation φ_z of multiplication by z takes $\mathbb{C}$ to $\mathbb{C}$ by $\phi_z(w) = zw$. This corresponds to a linear transformation of $\mathbb{R}^2$ to $\mathbb{R}^2$ since if α and β are real and w and ζ are complex, then

$$\varphi_z(\alpha w + \beta\zeta) = z(\alpha w + \beta\zeta) = \alpha z w + \beta z \zeta = \alpha\varphi_z(w) + \beta\varphi_z(\zeta).$$

The images of the basis vectors are

$$\begin{aligned} \operatorname{Mat}(\phi_z)(e_1) &\sim \varphi_z(1) = z = x + iy \sim \begin{pmatrix} x \\ y \end{pmatrix} \\ \operatorname{Mat}(\phi_z)(e_2) &\sim \varphi_z(i) = zi = -y + ix \sim \begin{pmatrix} -y \\ x \end{pmatrix}. \end{aligned}$$

Thus

$$\operatorname{Mat}(\phi_z) = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

as claimed.

(b) The assertion about compositions can be handled either in terms of the complex numbers or the matrices. Suppose z_1, z_2, and w are in $\mathbb{C}$. Then

$$\begin{aligned} \varphi_{z_1 z_2}(w) &= (z_1 z_2) w = z_1(z_2 w) = \varphi_{z_1}(z_2 w) \\ &= \varphi_{z_1}(\varphi_{z_2}(w)) = (\varphi_{z_1} \circ \varphi_{z_2})(w) \end{aligned}$$

This holds for all w in $\mathbb{C}$. So $\varphi_{z_1 z_2} = \varphi_{z_1} \circ \varphi_{z_2}$ as claimed.
In terms of matrices, we would first note that

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + (x_1 y_2 + y_1 x_2)i$$

and then compute

$$\begin{aligned}
\mathrm{Mat}(\varphi_{z_1 z_2}) &= \begin{pmatrix} x_1 x_2 - y_1 y_2 & -x_1 y_2 - y_1 x_2 \\ x_1 y_2 + y_1 x_2 & x_1 x_2 - y_1 y_2 \end{pmatrix} \\
&= \begin{pmatrix} x_1 & -y_1 \\ y_1 & x_1 \end{pmatrix} \cdot \begin{pmatrix} x_2 & -y_2 \\ y_2 & x_2 \end{pmatrix} \\
&= \mathrm{Mat}(\varphi_{z_1}) \cdot \mathrm{Mat}(\varphi_{z_2}) \\
&= \mathrm{Mat}(\varphi_{z_1} \circ \varphi_{z_2}).
\end{aligned}$$

The transformations $\varphi_{z_1 z_2}$ and $\varphi_{z_1} \circ \varphi_{z_2}$ have the same matrices with respect to the standard basis vectors, so they must be equal as claimed.

Discussion: A Matrix Representation of the Complex Numbers. The work done in the solution of Exercise 1E-40 goes a long way toward supplying a different way of thinking about the complex numbers. The identification of the complex numbers with $\mathbb{R}^2$ shows a way to think of them as a two-dimensional vector space over $\mathbb{R}$. But there are other ways. The complex numbers, $\mathbb{C}$, are a two dimensional space in which we also know how to multiply. Such a structure is called a (real) algebra. More precisely:

Definition. *An algebra $\mathcal{A}$ over $\mathbb{R}$ is a vector space over $\mathbb{R}$ in which there is a multiplication defined satisfying*

1. $(ab)c = a(bc)$ *for all* a, b, *and* c *in* $\mathcal{A}$,
2. $\alpha(ab) = (\alpha a)b$ *for all* a *and* b *in* $\mathcal{A}$ *and all* α *in* $\mathbb{R}$,
3. $\alpha(\beta a) = (\alpha\beta)a$ *for all* a *in* $\mathcal{A}$ *and* α *and* β *in* $\mathbb{R}$.

If $ab = ba$ for all a and b in $\mathcal{A}$, then $\mathcal{A}$ is called a commutative algebra. If there is an element e in $\mathcal{A}$ such that $ae = ea = a$ for every a in $\mathcal{A}$, then $\mathcal{A}$ is an algebra with identity. If for each a in $\mathcal{A}$ except 0 *there is an element a^{-1} such that $aa^{-1} = a^{-1}a = e$, then $\mathcal{A}$ is called a division algebra.*

Notice in particular that a commutative division algebra is a field. There are many examples of algebras:

1. The set of all $n \times n$ matrices with real entries using the usual addition and multiplication of matrices is an algebra over $\mathbb{R}$ with dimension n^2.

2. The set of all polynomials with real coefficients with the usual addition and multiplication of polynomials is an algebra over $\mathbb{R}$. Its dimension is infinite.

3. The set of all continuous real-valued functions on the unit interval is an algebra over $\mathbb{R}$. Here addition and multiplication are defined pointwise. That is, $(f+g)(x) = f(x)+g(x)$ and $(fg)(x) = f(x)g(x)$ for all x in $[0,1]$.

4. The complex numbers are a commutative division algebra over $\mathbb{R}$ with dimension 2.

The work of Exercise 1E-40 shows a way to think of a complex number as a 2×2 matrix. The set of all 2×2 matrices has dimension 4 and is certainly not a division algebra. There are many non-invertible matrices. Nor is it commutative. However, we do not want to use all 2×2 matrices. To represent the complex number z, we basically want to use the matrix $\mathrm{Mat}(\varphi_z)$ studied above. These all have the form $a = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$ where x and y are real. So we let $\mathcal{A}$ be the collection of all such matrices. Since we already know that addition and multiplication of 2×2 matrices satisfy all the required associative, commutative, and distributive laws, all that we need to do to show that $\mathcal{A}$ is an algebra is to show the following:

1. $a + b \in \mathcal{A}$ whenever a and b are in $\mathcal{A}$,
2. $ab \in \mathcal{A}$ whenever a and b are in $\mathcal{A}$,
3. $\alpha a \in \mathcal{A}$ whenever $a \in \mathcal{A}$ and $\alpha \in \mathbb{R}$.

These computations are straightforward. Having done them we note that

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix} = x\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + y\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

so that $\mathcal{A}$ has dimension 2 over $\mathbb{R}$ with the matrices $e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ as basis. The matrix e is certainly a multiplicative identity. There is a copy of $\mathbb{R}$ embedded in $\mathcal{A}$ by $\alpha \sim \alpha e$. Furthermore, a direct computation shows that $i^2 = -e \sim -1$, so that in this sense, i is a square root for -1. Finally we need to check that $\mathcal{A}$ is a division algebra. If $a = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \in \mathcal{A}$, then $\det a = x^2 + y^2$. So unless a is the zero matrix, it is invertible. Since $a^{-1} = \dfrac{1}{x^2+y^2}\begin{pmatrix} x & y \\ -y & x \end{pmatrix}$, it is also in $\mathcal{A}$

In summary, $\mathcal{A}$ is a two-dimensional division algebra over $\mathbb{R}$ in which the negative of the identity has a square root. It supplies a representation of the complex numbers $\mathbb{C}$. ♦

◇ **1E-41.** Show that $\mathrm{Re}(iz) = -\mathrm{Im}(z)$ and that $\mathrm{Im}(iz) = \mathrm{Re}(z)$ for all complex numbers z.

Sketch. If $z = x + iy$, then $\operatorname{Re}(iz) = \operatorname{Re}(ix - y) = -y = -\operatorname{Im}(z)$, and $\operatorname{Im}(iz) = \operatorname{Im}(ix - y) = x = \operatorname{Re}(z)$. ♢

Solution. Suppose $z \in \mathbb{C}$ and let $z = x + iy$ with x and y in $\mathbb{R}$. Then

$$\operatorname{Re}(iz) = \operatorname{Re}(i(x + iy)) = \operatorname{Re}(-y + ix) = -y = -\operatorname{Im}(z)$$

and

$$\operatorname{Im}(iz) = \operatorname{Im}(i(x + iy)) = \operatorname{Im}(-y + ix) = x = \operatorname{Re}(z)$$

as claimed. ♦

◇ **1E-42.** Letting $z = x + iy$, prove that $|x| + |y| \le \sqrt{2}|z|$.

Solution. Here is a solution written in such a way as to indicate how it might have been discovered by reasonable experimentation with the computation. One should be able to supply reasons at each line.

$$\begin{aligned}
|x| + |y| \le \sqrt{2}\ |z| &\iff |x|^2 + 2\,|x|\ |y| + |y|^2 \le 2\ |z|^2 \\
&\iff |x|^2 + 2\,|x|\ |y| + |y|^2 \le 2\ |x|^2 + 2\ |y|^2 \\
&\iff 2\,|x|\ |y| \le |x|^2 + |y|^2 \\
&\iff |x|^2 - 2\ |x|\ |y| + |y|^2 \ge 0 \\
&\iff (|x| - |y|)^2 \ge 0
\end{aligned}$$

The last line is true since the square of any real number is nonnegative. Working our way back up along the line of equivalences, we get a proof of our assertion. ♦

◇ **1E-43.** If $a,\ b \in \mathbb{C}$, prove the ***parallelogram identity:***

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

Sketch. Use $|z|^2 = z \cdot \bar{z}$ and $\overline{z + w} = \bar{z} + \bar{w}$ to expand $|a - b|^2 + |a + b|^2$. ♢

Solution. If a and b are complex numbers, we can use the properties listed in Proposition 1.8.1 to compute

$$\begin{aligned}
|a - b|^2 + |a + b|^2 &= (a - b) \cdot \overline{(a - b)} + (a + b) \cdot \overline{(a + b)} \\
&= (a - b) \cdot (\overline{a} - \overline{b}) + (a + b) \cdot (\overline{a} + \overline{b}) \\
&= a\overline{a} - b\overline{a} - a\overline{b} + b\overline{b} + a\overline{a} + b\overline{a} + a\overline{b} + b\overline{b} \\
&= 2a\overline{a} + 2b\overline{b} \\
&= |a|^2 + |b|^2
\end{aligned}$$

as claimed.

Compare this to the work and discussion of Exercise 1E-12(a). ♦

◇ **1E-44.** Prove Lagrange's identity for complex numbers:

$$\left|\sum_{k=1}^{n} z_k w_k\right|^2 = \left(\sum_{k=1}^{n} |z_k|^2\right)\left(\sum_{k=1}^{n} |w_k|^2\right) - \left(\sum_{k<j} |z_k \overline{w}_j - z_j \overline{w}_k|^2\right).$$

Deduce the Cauchy inequality from your proof.

Sketch. This is very similar to the computation for the "real" Lagrange Identity carried out in Exercise 1E-14(a). The x_k become z_k, the y_k become w_k, and we need to introduce some complex conjugations. Change the first line to

$$\left|\sum_{k=1}^{n} z_k w_k\right|^2 = \left(\sum_{k=1}^{n} z_k w_k\right)\overline{\left(\sum_{j=1}^{n} z_j w_j\right)} = \left(\sum_{k=1}^{n} z_k w_k\right)\left(\sum_{j=1}^{n} \overline{z}_j \overline{w}_j\right)$$

and continue accordingly. ◊

◇ **1E-45.** Show that if $|z| > 1$ then $\lim\limits_{n\to\infty} z^n/n = \infty$.

Sketch. $|z|^n/n > R \iff \log|z| > (\log n + \log R)/n$. The left side is positive and the right side tends to 0 by L'Hôpital's rule. ◊

Solution. Suppose $|z| > 1$ and $R > 1$ and n is a positive integer. Then

$$\begin{aligned}\left|\frac{z^n}{n}\right| \geq R \iff \frac{|z|^n}{n} \geq R &\iff |z|^n \geq nR \\ &\iff n\log|z| \geq \log n + \log R \\ &\iff \frac{\log n + \log R}{n} \leq \log z\end{aligned}$$

The right side of the last inequality is larger than zero since $|z| > 1$. Use of L'Hôpital's Rule (see Exercise 1E-33) shows that the left side tends to 0 as $n \to \infty$. Thus there is an N such that the last inequality holds whenever $n \geq N$. For such n we have $|z^n/n| \geq R$. Since this can be done for any $R > 1$, we conclude that $\lim_{n\to\infty} z^n/n = \infty$ as claimed. ♦

◇ **1E-46.** Prove that each nonempty set S of $\mathbb{R}$ that is bounded above has a least upper bound as follows: Choose $x_0 \in S$ and M_0 an upper bound. Let $a_0 = (x_0 + M_0)/2$. If a_0 is an upper bound, let $M_1 = a_0$ and $x_1 = x_0$; otherwise let $M_1 = M_0$ and $x_1 > a_0$, $x_1 \in S$. Repeat, generating sequences x_n and M_n. Prove that they both converge to $\sup(S)$.

Solution. (Thanks to Maria Korolov) Since S is bounded, there is an M_0 with $x \leq M_0$ for all x in S. Since S is not empty, we can select x_0 in S and let $a_0 = (x_0 + M_0)/2$. If this is an upper bound, let $M_1 = a_0$ and $x_1 = M_0$. If not, then there is an x in S with $x > a_0$. Let $x_1 = x$ and $M_1 = M_0$. Repeat with $a_1 = (x_1 + M_1)/2$ and $a_2 = (x_2 + M_2)/2$, and so forth. Each M_{k+1} is either M_k or $(x_k + M_k)/2$. Since M_k is an upper bound and $x_k \in S$, we have $x_k \leq M_k$, and $M_{k+1} \leq M_k$. The M_k form a monotone decreasing sequence and since every term is greater than some x in S, it is bounded below and must converge to some number M in $\mathbb{R}$ by completeness.

Similarly, the x_k form a monotone increasing sequence bounded above by any of the M's. So they must converge to something. We claim that they converge to M and that M is a least upper bound for S.

Let $d_0 = M_0 - x_0 = d(M_0, x_0)$. For each k we have that $M_k - x_k = d(M_k, x_k)$ is equal either to $x_{k-1} + M_{k-1})/2 - x_{k-1}$ or $M_{k-1} - (x_{k-1} + M_{k-1})/2$. Both of these are equal to $(M_{k-1} - x_{k-1})/2 = d(M_{k-1}, x_{k-1})/2$. Inductively we obtain $d(M_n, x_n) = d_0/2^n$ for each n. This tends to 0 as $n \to \infty$, so the limits of the two convergent sequences must be the same.

To show that M is an upper bound for S, suppose that there were a point x in S with $x > M$. Select k with $M_k - M < x - M$. This would force $M_k < x$ contradicting the fact that M_k is an upper bound for S. To show that M is a least upper bound for S, suppose that b were an upper bound with $b < M$. Since the points x_k converge to M, there is a k with $|M - x_k| < M - b$. This forces $b < x_k$ contradicting the supposition that it was an upper bound. ♦

2
The Topology of Euclidean Space

2.1 Open Sets

◇ **2.1-1.** Show that $\mathbb{R}^2\backslash\{(0,0)\}$ is open in $\mathbb{R}^2$.

Sketch. If $v = (a, b) \neq (0, 0)$, then $v \in U = D(v, \| v \|) \subseteq \mathbb{R}^2 \setminus \{(0, 0)\}$. (Give detail.) ◊

Solution. Suppose $v = (a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$. Then v is not $(0, 0)$, so $\| v \| \neq 0$. Let $r = \| v \|$. We will show that $D(v, r) \subseteq \mathbb{R}^2 \setminus \{(0, 0)\}$. Suppose $w \in D(v, r)$, then $\| w - v \| < r = \| v \|$. If w were equal to $(0, 0)$, we would wave $\| v \| = r > \| 0 - v \| = \| v \|$ which is nonsense. So $w \neq (0, 0)$. Since this is true for every w in $D(v, r)$, we conclude that $D(v, t) \subseteq \mathbb{R}^2 \setminus \{(0, 0)\}$ as claimed. Since $\mathbb{R}^2 \setminus \{(0, 0)\}$ contains a ball around each of its points, it is open. See Figure 2-1. ♦

◇ **2.1-2.** Let $S = \{(x, y) \in \mathbb{R}^2 \mid xy > 1\}$. Show that S is open.

Solution. The set in question is the set of points "outside" the hyperbola $xy = 1$. We need to show that each point in S is surrounded by some small disk completely contained in S. This seems reasonably clear from Figure 2-2. Any radius shorter than the distance from our point to the closest point on the curve will do.

Figure 2-1. The "punctured plane" is open.

Figure 2-2. The region "outside" a hyperbola.

One could actually try to find the closest point on the hyperbola to the origin and its distance using the methods of beginning calculus. This would give the largest radius which will do. Any smaller radius will also work. Here is a geometric argument which shows that such a radius exists. It assumes that the branch of the hyperbola in the first quadrant is the graph of a decreasing function and is concave up. This is easily checked for $f(x) = 1/x$. The argument is based on the figures sketched in Figure 2-3.

Suppose $P = (a, b) \in S$. Draw the horizontal from P to the y–axis. It intersects the curve at the point $A = (1/b, b)$. Draw the vertical from P to the x–axis. It intersects the curve at the point $B = (a, 1/a)$. The hypotenuse of the right triangle APB lies inside S since the curve $xy = 1$ (in the first quadrant) is concave up. Draw the perpendicular from P to this hypotenuse. It meets the hypotenuse at a point Q. The circle through Q with center at P has the hypotenuse AB as tangent at Q. This circle

FIGURE 2-3. Geometric construction for Exercise 2.1-2.

intersects the legs AP and BP at right angles at points D and C. The circle and its interior lie entirely inside the region above and to the right of the vertical through A, the hypotenuse AB, and the horizontal through B. But this region lies inside S since the boundary curve is the graph of a decreasing function of x. Thus this supplies an open disk centered at P and contained within S. The argument for a point in the portion of S in the third quadrant is symmetrical. Since this can be done for every point in S, the set S is open.

It is interesting to note that we can actually compute the length of the radius PQ. If we let that radius be r, then since the triangle AQP is similar to triangle APB, we have

$$\frac{r}{AP} = \frac{BP}{AB} \qquad \text{so} \qquad r = \frac{AP \cdot BP}{AB}$$

But

$$AP = |a - (1/b)| \qquad \text{and} \qquad BP = |b - (1/a)|$$
$$\text{and} \qquad AB = \sqrt{(a - (1/b)^2 + (b - (1/a)^2}.$$

So

$$r = \frac{|a - (1/b)| \cdot |b - (1/a)|}{\sqrt{(a - (1/b)^2 + (b - (1/a)^2}} = \frac{ab - 1}{\sqrt{a^2 + b^2}}.$$

(Do some algebra and remember that $ab > 1$.) ♦

◇ **2.1-3.** Let $A \subset \mathbb{R}$ be open and $B \subset \mathbb{R}^2$ be defined by $B = \{(x, y) \in \mathbb{R}^2 \mid x \in A\}$. Show that B is open.

Sketch. If $v_0 = (x_0, y_0) \in B$, there is an $r > 0$ with $|x - x_0| < r \implies x \in A$. (Why ?) Show $\| v - v_0 \| < r \implies v \in B$. $\Diamond$

Solution. Let $v_0 = (x_0, y_0) \in B$. Then $x_0 \in A$. Since A is an open subset of $\mathbb{R}$, there is a $\delta > 0$ such that the open interval; $J =]x_0 - \delta, x_0 + \delta[$ is contained in A. Let $r = \delta$ and $B = D(v_0, r) = \{v \in \mathbb{R}^2 \mid \| v - v_0 \| < r\}$. If $v = (x, y) \in B$, then

$$(x - x_0)^2 \le (x - x_0)^2 + (y - y_0)^2 = \| v - v_0 \|^2 < r^2.$$

So $|x - x_0| < r = \delta$. This implies that x is in J and so in A. In turn, this implies that $v = (x, y) \in B$. For each v in B, there is an $r > 0$ such that $D(v, r) \subseteq B$. Therefore B is an open subset of $\mathbb{R}^2$ as claimed. See Figure 2-4. $\blacklozenge$

FIGURE 2-4. An open set in $\mathbb{R}$ "pulls back" to an open set in $\mathbb{R}^2$.

$\diamond$ **2.1-4.** Let $B \subset \mathbb{R}^n$ be any set. Define $C = \{x \in \mathbb{R}^n \mid d(x, y) < 1$ for some $y \in B\}$. Show that C is open.

Suggestion. The set C is a union of open balls. $\Diamond$

Solution. Let $A = \bigcup_{y \in B} D(y, 1)$.Since each of the disks $D(y, 1)$ is an open subset of $\mathbb{R}^n$ and A is their union, we know that A is an open set by Proposition 2.1.3(ii). We will show that C is open by showing that $C = A$. If $y \in B$ and x is any point in $\mathbb{R}^n$ with $\| x - y \| < 1$, then $x \in C$ by the definition of the set C. So $D(y, 1) \subseteq C$ for every y in B. Thus

$A = \bigcup_{y \in B} D(y,1) \subseteq C$. In the other direction, if $x \in C$, then there is a point y in B with $d(x,y) = \| x - y \| < 1$. So $x \in D(y,1) \subseteq A$. Thus $C \subseteq A$. We have containment in both directions, so $C = A$. Thus C is an open set as claimed. ♦

◇ **2.1-5.** Let $A \subset \mathbb{R}$ be open and $B \subset \mathbb{R}$. Define $AB = \{xy \in \mathbb{R} \mid x \in A$ and $y \in B\}$. Is AB necessarily open?

Answer. "No" in general; "yes" if B is open or $0 \notin B$. ◊

Solution. If A is any open subset of $\mathbb{R}$ and $B = \{0\}$, then $AB = \{0\}$ and is not open. So the answer is "No" in general.

To find conditions under which AB will be open, we define for each $y \in \mathbb{R}$ the set

$$yA = \{y\}A = A\{y\} = \{xy \in \mathbb{R} \mid x \in A\}.$$

Then $AB = \bigcup_{y \in B} yA$.

Lemma. *If A is open and $y \neq 0$, then yA is an open subset of $\mathbb{R}$.*

Proof: Suppose $z_0 \in yA$. Then there is an $x_0 \in A$ with $z = yx_0$. . Since A is open there is a $\delta > 0$ with $J =]x_0 - \delta, x_0 + \delta[\subseteq A$.

First suppose that $y > 0$. If $z_0 - y\delta < z < z_0 + y\delta$, then let $x = z/y$. We have

$$x_0 - \delta = (z_0 - y\delta)/y < z/y = x < (z_0 + y\delta)/y = x_0 + \delta.$$

This implies that x is in J and so in A and hence that $z = xy \in yA$. For each z_0 in yA there is an open interval around z_0 contained in yA, so A is open.

The argument for negative y is similar and should be supplied by the reader. (By the way, where is the hole in the argument if $y = 0$?)

If 0 is not in B, then, from the lemma, each of the sets yA for y in B is open. Since $AB = \bigcup_{y \in B} yA$, we see that A is open by Proposition 2.1.3(ii).

Now suppose B is an open subset of $\mathbb{R}$. If $0 \notin B$, then we already know that AB is open. So suppose $0 \in B$. Let $B_0 = B \setminus \{0\}$. Then B_0 is open ($B_0 = B \cap (\mathbb{R} \setminus \{0\})$, and

$$AB = A(B_0 \cup \{0\}) = AB_0 \cup A\{0\} = AB_0 \cup \{0\}.$$

If $z \in AB$ and $z \neq 0$, then $z \in AB_0$. We know from above that AB_0 is an open set, so there is an open interval J_z with $z \in J_z \subseteq AB_0 \subseteq AB$. To show that AB is open, it remains to show that it contains an interval around 0. Since A is open, it contains a nonzero element a. Since B is open and contains 0, there is a $\delta > 0$ such that B contains the interval $J = \{y \mid |y| < \delta\}$. If $|z| < |a|\,\delta$, then $|\pm z/a| < \delta$. This implies that $\pm z/a$ are in J and so in B. Since $a \in A$, this puts z and $-z$ in AB. Thus AB contains the interval $\{z \mid |z| < |a|\,\delta\}$ around zero. Since AB contains an interval around each of its points, it is an open subset of $\mathbb{R}$ as claimed. ♦

◇ **2.1-6.** Show that $\mathbb{R}^2$ with the taxicab metric has the same open sets as it does with the standard metric.

Sketch. The key idea is that each "taxicab disk" centered at a point P contains a "Euclidean disk" centered at P and also the reverse. ◊

Solution. Recall that the "taxicab metric" on $\mathbb{R}^2$ is defined for points $P = (a, b)$ and $Q = (x, y)$ by $d_1(Q, P) = |x - a| + |y - b|$. The Euclidean metric is defined by $d_2(Q, P) = \sqrt{(x-a)^2 + (y-b)^2}$. Euclidean disks look like the interior of a usual Euclidean circle, while a "taxicab disk" looks like the inside of a "diamond". The key idea is that each taxicab disk centered at P contains a Euclidean disk centered at P and each Euclidean disk centered at P contains a taxicab disk centered at P. See the figure.

FIGURE 2-5. The "taxicab" and Euclidean distances make the same sets open.

$$\begin{aligned} d_1(Q,P)^2 &= (|x-a| + |y-b|)^2 = (x-a)^2 + 2\,|x-a| \cdot |y-b| + (y-b)^2 \\ &\geq (x-a)^2 + (y-b)^2 = d_2(Q,P)^2 \end{aligned}$$

On the other hand, we know that for any real s and t that $2st \leq s^2 + t^2$ since $0 \leq (s-t)^2 = s^2 - 2st + t^2$. With $s = |x - a|$ and $t = |y - b|$, this gives

$$\begin{aligned} d_1(Q,P)^2 &= (|x-a| + |y-b|)^2 = (x-a)^2 + 2\,|x-a| \cdot |y-b| + (y-b)^2 \\ &\leq 2((x-a)^2 + (y-b)^2) = 2\,d_2(Q,P)^2 \end{aligned}$$

Taking square roots gives

$$d_2(Q,P) \le d_1(Q,P) \le \sqrt{2}\,d(Q,P) \tag{$*$}$$

for every pair of points P and Q in $\mathbb{R}^2$. Denote the Euclidean and taxicab disks around P of radius ρ by

$$\begin{aligned} D_2(P,\rho) &= \{P \in \mathbb{R}^2 \mid d_2(P,Q) < \rho\} \\ D_1(P,\rho) &= \{P \in \mathbb{R}^2 \mid d_1(P,Q) < \rho\}. \end{aligned}$$

If $d_1(P,Q) < \rho$, then by the first part of $(*)$, we also have $d_2(P,Q) < \rho$. So $D_1(P,\rho) \subseteq D_2(P,\rho)$. If $d_2(P,Q) < r$, then by the second part of $(*)$, we also have $d_1(P,Q) < \sqrt{2}\,r$. So $D_2(P,r) \subseteq D_1(P,\sqrt{2}\,r)$. Equivalently $D_2(P,\rho/\sqrt{2}) \subseteq D_1(P,\rho)$ for each $\rho > 0$.

Now Let $S \subseteq \mathbb{R}^2$ and suppose that S is open with respect to the usual Euclidean distance. Let $P \in S$. Then there is an $\rho > 0$ such that $D_2(P,\rho) \subseteq S$. But then $P \in D_1(P,\rho) \subseteq D_2(P,\rho) \subseteq S$. So S contains a taxicab disk around each of its points and must be "taxicab open". On the other hand, if we assume that S is open with respect to the taxicab distance and $P \in S$, then there is an $r > 0$ such that $P \in D_1(P,r) \subseteq S$. The argument above then shows that $P \in D_2(P,r/\sqrt{2}) \subseteq D_1(P,r) \subseteq S$. So S contains a Euclidean disk around each of its points and is open in the usual sense. We have shown

$$S \text{ "Euclidean open"} \implies S \text{ "taxicab open"},$$

and that

$$S \text{ "taxicab open"} \implies S \text{ "Euclidean open"}.$$

So the "open sets" are the same no matter which of the two metrics we choose to use for measuring distance between points in the plane. ♦

2.2 Interior of a Set

◇ **2.2-1.** Let $S = \{(x,y) \in \mathbb{R}^2 \mid xy \ge 1\}$. Find $\operatorname{int}(S)$.

Answer. $\operatorname{int}(S) = \{(x,y) \in \mathbb{R}^2 \mid xy > 1\}$. ◊

Sketch. If we replace "$\ge$" by "$>$", we have the set studied in Exercise 2.1-2. This is open, so all its points are interior points of S. Points on the boundary hyperbola are not interior since any disk around such a point sticks outside of S. Sketch and give a bit more detail. ◊

Solution. We saw in Exercise 2.1-2 that the set $U = \{(x, y) \in \mathbb{R}^2 \mid xy > 1\}$ is open. Since it is a subset of S, all of its points are interior points of S, and we have $U \subseteq \text{int}(S)$. If $(x_0, y_0) \in S \setminus U$, then it is a point on the boundary hyperbola where $x_0 y_0 = 1$. Such a point is not in $\text{int}(S)$. If $r > 0$, there are points in $D((x_0, y_0), r)$ which are not in S. If x_0 and y_0 are greater than 0, then $(x_0 - (r/2), y_0)$ is such a point. If they are less than 0, then $(x_0 + (r/2), y_0)$ is such a point. See Figure 2-6. Thus $(S \setminus U) \cap \text{int}(S) = \emptyset$. So $\text{int}(S) = U$. ♦

FIGURE 2-6. The hyperbola again.

◇ **2.2-2.** Let $S = \{(x, y, z) \in \mathbb{R}^3 \mid 0 < x < 1,\ y^2 + z^2 \leq 1\}$. Find $\text{int}(S)$.

Solution. The region S is the inside and part of the surface of the cylindrical solid of radius 1 with axis along the x–axis. It is that part of the cylinder between the planes $x = 0$ and $x = 1$. It includes the curved cylindrical part of the surface but not the circular disk "end plates". See Figure 2-7. The interior is essentially the same cylindrical solid, but now without the curved sides or the end plates included. This is obtained by changing to a strict inequality.

$$\text{int}(S) = U = \{(x, y, z) \in \mathbb{R}^3 \mid 0 < x < 1, y^2 + z^2 < 1\}.$$

To see this, note that we know already that the half spaces $H_1 = \{(x, y, z) \mid x > 0\}$ and $H_2 = \{(x, y, z) \mid x < 1\}$ are open subsets of $\mathbb{R}^3$, and the the disk $D = \{(y, z) \mid y^2 + z^2 < 1\}$ is an open subset of $\mathbb{R}^2$. An argument like that for Exercise 2.1-3 shows that the cylinder $C = \{(x, y, z) \in \mathbb{R}^3 \mid y^2 + z^2 < 1\}$ is an open subset of $\mathbb{R}^3$. Thus the set $U = \{(x, y, z) \in \mathbb{R}^3 \mid 0 < x < 1, y^2 + z^2 \leq 1\}$ is open in $\mathbb{R}^3$ since

FIGURE 2-7. A cylinder.

$U = H_1 \cap H_2 \cap C$. Since $U \subseteq S$ and U is open, all of its points are interior points of S, and $U \subseteq \text{int}\,(S)$. If $P = (x_0 y_0 z_0)$ is any point on the boundaries of S, the end plates $D_0 = \{(x, y, z) \mid x = 0 \text{ and } y^2 + z^2 \leq 1\}$ and $D_1 = \{(x, y, z) \mid x = 1 \text{ and } y^2 + z^2 \leq 1\}$, or the curved part $B = \{(x, y, z) \mid 0 \leq x \leq 1 \text{ and } y^2 + z^2 = 1\}$, then any ball around P will extend outside S. (Why? See the solution to Exercise 2.2-1.) So these points are not in the interior of S. Thus $\text{int}\,(S) = U$. ♦

⋄ **2.2-3.** If $A \subset B$, is $\text{int}(A) \subset \text{int}(B)$?

Answer. Yes. ◊

Solution. To establish the inclusion, let $x \in \text{int}\,(A)$. Then there is an open set U such that $x \in U \subseteq A$. But we know that $A \subseteq B$. So we have $x \in U \subseteq A \subseteq B$ and hence $x \in U \subseteq B$. Thus $x \in \text{int}\,(B)$. Since x was an arbitrary point in $\text{int}\,(A)$, this shows that $\text{int}\,(A) \subseteq \text{int}\,(B)$. ♦

⋄ **2.2-4.** Is it true that $\text{int}(A) \cap \text{int}(B) = \text{int}(A \cap B)$?

Answer. Yes. ◊

Solution. First suppose $x \in \text{int}\,(A \cap B)$. Then there is an $r > 0$ such that $x \in D(x, r) \subseteq A \cap B$. Since $A \cap B \subseteq A$, we have $x \in D(x, r) \subseteq A$. So $x \in \text{int}\,(A)$. Also $A \cap B \subseteq B$, so $x \in D(x, r) \subseteq B$ and $x \in \text{int}\,(B)$. Thus $x \in \text{int}\,(A) \cap \text{int}\,(B)$. Since x was an arbitrary point in $\text{int}\,(A \cap B)$, this shows that $\mathrm{int}\,(A \cap B) \subseteq \mathrm{int}\,(A) \cap \mathrm{int}\,(B)$. (Notice that we could also employ the result of Exercise 2.2-3 to obtain this inclusion.)

Now suppose that $y \in \mathrm{int}\,(A) \cap \mathrm{int}\,(B)$. Then $y \in \mathrm{int}\,(A)$, so there is an $r_1 > 0$ such that $x \in D(x, r_1) \subseteq A$. Also $y \in \mathrm{int}\,(B)$. So there is an $r_2 > 0$ such that $y \in D(y, r_2) \subseteq B$. Let r be the smaller of r_1 and r_2 so that $r \leq r_1$ and $r \leq r_2$. Then $D(y, r) \subseteq D(y, r_1) \subseteq A$, and $D(x, r) \subseteq D(y, r_2) \subseteq B$. So $y \in D(y, r) \subseteq A \cap B$. Thus $y \in \mathrm{int}\,(A \cap B)$. Since y was an arbitrary point in $\mathrm{int}\,(A) \cap \mathrm{int}\,(B)$, this shows that $\mathrm{int}\,(A) \cap \mathrm{int}\,(B) \subseteq \mathrm{int}\,(A \cap B)$.

We have inclusion in both directions, so $\mathrm{int}\,(A \cap B) = \mathrm{int}\,(A) \cap \mathrm{int}\,(B)$ as claimed. ♦

◇ **2.2-5.** Let (M, d) be a metric space, and let $x_0 \in M$ and $r > 0$. Show that

$$D(x_0, r) \subset \mathrm{int}\{y \in M \mid d(\,y, x_0) \leq r\}.$$

Suggestion. Show more generally that if U is open and $U \subseteq A$, then $U \subseteq \mathrm{int}\,(A)$. (See Exercise 2E-3 at the end of the chapter.) ◊

Solution. Let $A = \{y \in M \mid d(y, x_0) \leq r\}$ and $U = D(x_0, r) = \{y \in M \mid d(y, x_0) < r\}$. Then we certainly have $U \subseteq A$, and we know that U is open from Proposition 2.1.2. So the desired conclusion follows immediately from this observation.

Proposition. *If U is an open set and $U \subseteq A$, then $U \subseteq \mathrm{int}\,(A)$.*

Proof: If x in U, then U is open and $x \in U \subseteq A$. So x satisfies the definition of an interior point of A and $x \in \mathrm{int}\,(A)$. Since x was an arbitrary point in U, this shows that $U \subseteq \mathrm{int}\,(A)$ as claimed. ♦

2.3 Closed Sets

◇ **2.3-1.** Let $S = \{(x, y) \in \mathbb{R}^2 \mid x \geq 1 \text{ and } y \geq 1\}$. Is S closed?

Answer. Yes. ◊

Solution. Let $U = \{(x, y) \in \mathbb{R}^2 \mid x < 1\}$ and $V = \{(x, y) \in \mathbb{R}^2 \mid y < 1\}$. Then $\mathbb{R}^2 \setminus S = U \cup V$. If we knew that U and V were open, then we would also know that this union is an open set. So the set S would be closed since its complement is open. Let's give some detail for the proof that such half planes are open.

Proposition. *Suppose a, b, c and d are real constants and let*

$$\begin{aligned} S &= \{(x, y) \in \mathbb{R}^2 \mid x > a\} \\ U &= \{(x, y) \in \mathbb{R}^2 \mid x < b\} \\ V &= \{(x, y) \in \mathbb{R}^2 \mid y > c\} \\ W &= \{(x, y) \in \mathbb{R}^2 \mid y < d\}. \end{aligned}$$

Then S, U, V, and W are open subsets of $\mathbb{R}^2$.

Proof: The four arguments are quite similar. We give that for S. Suppose $v_0 = (x_0, y_0) \in S$. Then $x_0 > a$ so $x_0 - a > 0$. Let $r = x_0 - a$. We claim that the disk around v_0 of radius r is contained in S. See Figure 2-8.

Figure 2-8. An open half plane.

Let $v = (x, y) \in D(v_0, r)$. Then

$$|x - x_0| = \sqrt{(x - x_0)^2} \leq \sqrt{(x - x_0)^2 + (y - y_0)^2} = \| v - v_0 \| < r = x_0 - a.$$

So

$$a - x_0 = -(x_0 - a) < x - x_0 < x_0 - a.$$

Adding x_0 to the first inequality gives $x > a$, so $v = (x, y) \in S$. Since v was an arbitrary point in $D(v_0, r)$, we conclude that $D(v_0, r) \subseteq S$. The set S contains a disk around each of its points, so it is an open subset of $\mathbb{R}^2$ as claimed. The proofs for U, V, and W are analogous. ♦

◇ **2.3-2.** Let $S = \{(x, y) \in \mathbb{R}^2 \mid x = 0,\ 0 < y < 1\}$. Is S closed?

Answer. No. ◇

Solution. The set S is not closed since its complement is not open. The point $P = (0, 0)$ is in $\mathbb{R}^2 \setminus S$. But if $r > 0$, then the point $(0, r/2)$ is in

$D((0,0),r)\cap S$. Thus $\mathbb{R}^2\setminus S$ contains no nontrivial disk around the origin $(0,0)$ even though that point is in the set. $\mathbb{R}^2\setminus S$ is not open, so S is not closed. See Figure 2-9. ♦

FIGURE 2-9. A line segment in $\mathbb{R}^2$.

◇ **2.3-3.** Redo Example 2.3.5 directly, this time showing that the complement is open.

Sketch. Suppose $S=\{x_1,\dots,x_k\}$ and $v_0\in\mathbb{R}^n\setminus S$. Show that if we let $r=(1/2)\min\{d(v_0,x_1),\dots,d(v_0,x_k)\}$, then $\|\,v-v_0\,\|<r \implies v\in\mathbb{R}^n\setminus S$. ◇

Solution. Recall that Example 2.3.5 asserts that if S is a finite set in $\mathbb{R}^n$, then S is closed. So let $S=\{x_1,x_2,\dots,x_k\}$ be any finite set in $\mathbb{R}^n$. Let v_0 be any point in $\mathbb{R}^n\setminus S$. Then none of the finitely many nonnegative numbers $d(v_0,x_1)$, $d(v_0,x_2)$, $\dots$, $d(v_0,x_k)$ are 0. Some one of them must be smallest. The number $r=(1/2)\min\{d(v_0,x_1),\dots,d(v_0,x_k)\}$ is positive. We have $r>0$ and $r<d(v_0,x_j)$ for $j=1,2,\dots,k$. Suppose $\|\,v-v_0\,\|<r$. Then

$$\begin{aligned}\|\,x_j-v\,\| &= \|\,(x_j-v_0)-(v-v_0)\,\| \geq \|\,x_j-v_0\,\|-\|\,v-v_0\,\|\\ &\geq 2r-\|\,v-v_0\,\|>2r-r=r.\end{aligned}$$

In particular, v is not equal to x_j. This works for each $j=1,2,\dots,k$. So none of the points x_j are in $D(v_0,r)$. This disk is contained in $\mathbb{R}^n\setminus S$. Since $\mathbb{R}^n\setminus S$ contains a nontrivial disk around each of its points, it is open and S is closed. ♦

◇ **2.3-4.** Let $A \subset \mathbb{R}^n$ be arbitrary. Show that $\mathbb{R}^n \setminus \operatorname{int}(A)$ is closed.

Suggestion. Show that the interior of every set is open. ◊

Solution. Since a set is closed if and only if its complement is open, and $\mathbb{R}^n \setminus (\mathbb{R}^n \setminus \operatorname{int}(A)) = \operatorname{int}(A)$, the assertion that $\mathbb{R}^n \setminus \operatorname{int}(A)$ is closed will follow from the observation.

Proposition. *If A is any set, then* $\operatorname{int}(A)$ *is open.*

Proof: Suppose $x \in \operatorname{int}(A)$. Then there is an open set U such that $x \in U \subseteq A$. If y is any other point in U, then $y \in U \subseteq A$, so $y \in \operatorname{int}(A)$. Thus $U \subseteq \operatorname{int}(A)$. Since U is open and $x \in U$, there is an $r > 0$ such that $D(x, r) \subseteq U$. Thus

$$x \in D(x, r) \subseteq U \subseteq \operatorname{int}(A).$$

If $x \in \operatorname{int}(A)$, then there is a nontrivial disk around x which is contained in $\operatorname{int}(A)$. So $\operatorname{int}(A)$ is open. ♦

◇ **2.3-5.** Let $S = \{x \in \mathbb{R} \mid x \text{ is irrational}\}$. Is S closed?

Answer. No. ◊

Solution. The point 0 is in the complement of S, but is $r > 0$, then the open interval $]-r, r[$ around 0 must contain irrational points such as $\sqrt{2}/n$ for large integer n. Thus $\mathbb{R} \setminus S = \mathbb{Q}$ is not open. So S is not closed.

More generally, it is not too difficult to show using the irrationality of $\sqrt{2}$ and the Archimedean Principle, that both the rationals and irrationals are scattered densely along the real line in the following sense.

Proposition. *If a and b are real numbers with $a < b$, then there are a rational number r and an irrational number z such that*

$$a < r < b \qquad \text{and} \qquad a < z < b.$$

Proof: The first is accomplished by noting that $b - a > 0$. By the Archimedean principle there is an integer n such that $0 < 1/n < b - a$. By another version of the Archimedean principle there are positive integers k such that $a < k/n$. By the well-ordering of the positive integers, we may assume that k is the smallest such. Then

$$a < \frac{k}{n} = \frac{k-1}{n} + \frac{1}{n} \leq a + \frac{1}{n} < a + (b - a) = b.$$

So $r = k/n$ is a rational number meeting our requirements.

Now use the argument just given twice to obtain a pair of rational numbers r and s with $a < r < s < b$. Since $\sqrt{2} > 1$, we have

$$a < r < r + \frac{s-r}{\sqrt{2}} < r + (s-r) = s$$

The number $z = r + (s-r)/\sqrt{2}$ cannot be rational, for if it were, then $\sqrt{2} = (s-r)/(z-r)$ would also be rational. But it is not. The number z is an irrational number which meets our requirements. ♦

◇ **2.3-6.** Give an alternative solution of Example 2.3.6 by showing that B is a union of finitely many closed sets.

Solution. Example 2.3.6 asserts that if d is a metric on a set M, and A is a finite subset of M, then the set $B = \{x \in M \mid d(x,y) \le 1 \text{ for some } y \in A\}$ is closed. For each fixed y in M, let $C_y = \{x \in M \mid d(x,y) \le 1\}$. Then $B = \bigcup_{y \in A} C_y$. Since A is a finite set, this expresses B as the union of a finite number of sets of type C_y. If we knew that all of these were closed we would have B closed since it would be a finite union of closed sets.

We proceed to show that for each fixed y the set C_y is closed. The argument is essentially the same as that given for Example 2.3.4 as illustrated in Figure 2.3-3 of the text. We simply need to write it down for a general metric space instead of $\mathbb{R}^2$.

Suppose $z \in M \setminus C_y$. then $d(z,y) > 1$. Set $r = d(z,y) - 1 > 0$. If $w \in M$ and $d(w,z) < r$, then

$$d(z,y) \le d(z,w) + d(w,y) < r + d(w,y) = d(z,y) - 1 + d(w,y),$$

so $1 < d(w,y)$ and $w \in M \setminus C_y$. Thus $D(z,r) \subseteq M \setminus C_y$. The set $M \setminus C_y$ contains a disk around each of its points and so is open. Thus C_y is closed as we needed. ♦

2.4 Accumulation Points

If S is a subset of a metric space M, we let S' denote the set of accumulation points of S, so that $x \in S'$ if and only if every open set U containing x contains points of S not equal to x. Equivalently, for each $\varepsilon > 0$ the set $D(x,\varepsilon) \cap (S \setminus \{x\})$ is not empty. Theorem 2.4.2 becomes

$$A \text{ is closed} \qquad \Longleftrightarrow \qquad A' \subseteq A.$$

◇ **2.4-1.** Find the accumulation points of $A = \{(x,y) \in \mathbb{R}^2 \mid y = 0 \text{ and } 0 < x < 1\}$.

Answer. $A' = \{(x,y) \in \mathbb{R}^2 \mid y = 0 \text{ and } 0 \le x \le 1\}$. ◇

Solution. If $0 < x \le 1$ and $\varepsilon > 0$, then for large enough integer n, the point $(x - (1/n), 0)$ is in A and in $D((x, 0), \varepsilon)$. Since it is not equal to $(x, 0)$ (which is in A), this shows that $(x, 0)$ is an accumulation point of A. Since $(1/n, 0)$ is in $D((0, 0), \varepsilon) \cap (A \setminus (0, 0))$ for large enough n, the origin, $(0, 0)$, is also in A'. Thus the set of accumulation points A' satisfies $\{(x, y) \in \mathbb{R}^2 \mid y = 0 \text{ and } 0 \le x \le 1\} \subseteq A'$.

To get the opposite conclusion, first suppose that y is not). Then no point of A can be closer than $|y|$ to (x, y). So these points are not in A'. If $x < 0$, then no point of A can be closer than $|x|$ to (x, y), and if $x > 0$, then no point of A can be closer than $x - 1$ to (x, y). So these points are not in A' either. Thus $A' \subseteq \{(x, y) \in \mathbb{R}^2 \mid y = 0 \text{ and } 0 \le x \le 1\}$.

We have inclusion in both directions, so $A' = \{(x, y) \in \mathbb{R}^2 \mid y = 0 \text{ and } 0 \le x \le 1\}$. ♦

◇ **2.4-2.** If $A \subset B$ and x is an accumulation point of A, is x an accumulation point of B as well?

Answer. Yes. ♢

Solution. Suppose $A \subseteq B$ and x is an accumulation point of A. $x \in A'$. Let U be any open set with $x \in U$. Then there is a point y in A with y not equal to x and $y \in A \cap U$. But $A \subset B$, so $y \in B$. Thus $y \in B \cap U$ and y is not equal to x. Since this can be done for every open set U containing x, we see that x is an accumulation point of B.

More succinctly, we have the following:

Proposition. *If $A \subseteq B$, then $A' \subseteq B'$.* ♦

◇ **2.4-3.** Find the accumulation points of the following sets in $\mathbb{R}^2$:

(a) $\{(m, n) \mid m, n \text{ integers}\}$
(b) $\{(p, q) \mid p, q \text{ rational}\}$
(c) $\{(m/n, 1/n) \mid m, n \text{ integers}, n \ne 0\}$
(d) $\{(1/n + 1/m, 0) \mid n, m \text{ integers}, n \ne 0, m \ne 0\}$

Answer. (a) $A' = \emptyset$.
(b) $B' = \mathbb{R}^2$.
(c) $C' =$ the $x-$axis.
(d) $D' = \{(1/n, 0) \in \mathbb{R}^2 \mid n \text{ is a nonzero integer }\} \cup \{0\}$. ♢

Solution. (a) All points are isolated. If $(m,n) \in A$, then $D((m,n),1/2) \cap A = \{(m,n)\}$. There can be no accumulation points. $A' = \emptyset$.

(b) As we have seen before, there are rational numbers in every short interval of the real line. So if (x,y) is any point in $\mathbb{R}^2$ and $\varepsilon > 0$, there are rational points closer than ε to (x,y) which are not equal to (x,y). We need only select rational numbers s and t with $0 < |s-x| < \varepsilon/\sqrt{2}$ and $0 < |t-y| < \varepsilon/\sqrt{2}$.

(c) As m and n run over the integers (with n not 0), the fraction m/n runs over all of $\mathbb{Q}$. As n increases, we get more and more points along horizontals coming close to the x – axis. The result is that all points on the x – axis are accumulation points. No others are. Consider Figure 2-10. So $C' = \{(x,0) \mid x \in \mathbb{R}\}$.

FIGURE 2-10. Rational points which accumulate on the horizontal axis.

(d) If we let $m = n$, we see that all of the points $1/2n$ for integer n are in D. Since these converge to 0, this must be an accumulation point. If we hold n fixed and let $m \to \infty$, we find that $1/n + 1/m \to 1/n$. So $1/n$ is an accumulation point. For m and n large, $1/n + 1/m$ moves away from all other points, so there are no other accumulation points. $D' = \{1/n \mid n \text{ is a nonzero integer }\} \cup \{0\}$. ♦

◇ **2.4-4.** Let $A \subset \mathbb{R}$ be nonempty and bounded above and let $x = \sup(A)$. Must x be an accumulation point of A?

Answer. Not necessarily. ♢

Solution. The point $\sup(A)$ might not be an accumulation point of A. It could be what is called an *isolated point* of A. That is, a point x which is in A but which is not an accumulation point of A. Consider, for example, $A = [0,1] \cup \{2\} \subseteq \mathbb{R}$. Then $\sup A = 2$. The interval $]3/2, 5/2[$ around 2 contains no points of A other than 2. So 2 is not an accumulation point of A. ♦

◇ **2.4-5.** Verify Theorem 2.4.2 for the set $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y + 2x = 3\}$.

Sketch. The set A is the parabola $y = 4 - (x + 1)^2$. We need to show that A is closed and then independently that $A' \subseteq A$ to "confirm" Theorem 2.4.2. Each is accomplished by taking a point not on A and getting a disk around that point which misses A. ♢

Solution. Theorem 2.4.2 says that A is closed if and only if $A' \subseteq A$. The set A here is the parabolic curve in $\mathbb{R}^2$ defined by the equation $x^2+y+2x = 3$. That is, $y = 4 - (x + 1)^2$. The exercise is asking us to confirm Theorem 2.4.2, so we are being asked to do two things. First, check that A is closed. Second, check that $A' \subseteq A$. These are accomplished in essentially the same way.

If (a, b) is not on the parabola A, then it lies at a positive distance $r > 0$ from that parabola. There is at least one point (perhaps two) on the parabola at which the distance to (a, b) is minimized, at that minimum is not 0. No point in the disk $D((a, b), r)$ can lies on A. So this disk is contained in the complement of A. This shows that the complement of A is open and hence that A is closed. But it also shows that (a, b) cannot be an accumulation point of A so that $A' \subseteq A$. One can use calculus methods (including solving a cubic equation) to try to find this exact distance, or one can use some geometrical methods to find a shorter distance which will also work. Ideas for these methods are sketched in the figure. The first sketch shows the disk around (a, b) with the largest possible radius r. However, any radius ρ with $0 < \rho < r$ will also serve to make a disk around (a, b) missing A. The peak of the parabola is at $(-1, 4)$, so if $b > 4$, we can simply take $\rho = b - 4$. The second sketch illustrates an argument for the case in which $b < 4$ and (a, b) is outside the parabola. We can draw the horizontal and vertical lines from (a, b) to the parabola and then the tangent to the parabola at one of the intersection points. The result is a right triangle lying outside the parabola (why?). The hypotenuse is part of the tangent line. If we take for radius of a disk centered at (a, b) the perpendicular from (a, b) to the hypotenuse, we get a disk lying completely outside A. The third sketch illustrates the case in which (a, b) is inside the parabola. Again draw the vertical and horizontal from (a, b) to A. The segment connecting the intersection points lies under the parabola since the parabola is "concave down". So we get a right triangle lying inside A. If we take for radius of a disk centered at (a, b) the perpendicular from (a, b) to the hypotenuse, we get a disk lying completely inside A.

Here is another way to show that $A' \subseteq A$ which uses a bit of knowledge about continuous function and convergent sequences. If (a, b) is an accumulation point of A, then there must be points (x_k, y_k) on it with $\| (x - k, y_k) - (a, b) \| < 1/k$. In particular, $|x_k - a| < 1/k$ and $|y_k - b| <$

FIGURE 2-11.

$1/k$. Thus $x_k \to a$ and $y + k \to b$. For each k we have $y_k = 4 - (x_k + 1)^2$, so by continuity we have $b = 4 - (a + 1)^2$. That is (a, b) must be on the parabola A. This shows that $A' \subseteq A$.

It is also interesting to note, though not required for this exercise, that each point in A is an accumulation point of A. If (x, y) satisfies the equation, we can select any sequence of distinct x_k tending to x, put $y_k = 4-(x_k+1)^2$, and obtain a sequence of distinct points (x_k, y_k) on the curve A which tend to (x, y). Thus $(x, y) \in A'$. ♦

◇ **2.4-6.** Let M be a set with the discrete metric and $A \subset M$ be any subset. Find the set of accumulation points of A.

Answer. $A' = \emptyset$. ◇

Solution. If $x_0 \in M$ then the ball of radius $1/2$ around x_0 contains no points other than x_0 since all distances between unequal points are 1. $D(x_0, 1/2) = \{x_0\}$. Nonetheless, this "ball" is an "open set". Thus no set in M can have any accumulation points. ♦

2.5 Closure of a Set

◇ **2.5-1.** Find the closure of $A = \{(x, y) \in \mathbb{R}^2 \mid x > y^2\}$.

Answer. $\mathrm{cl}(A) = \{(x, y) \in \mathbb{R}^2 \mid x \geq y^2\}$. ◇

Solution. The set in question is the region to the right of the parabola $x = y^2$ but not including the curve itself. See Figure 2-11.

The points on the curve are accumulation points of this set since if $P = (x, y)$ satisfies $x = y^2$, then the points $P_n = (x + (1/n), y)$ are in A. Since $\| P - P_n \| = 1/n$, some of these points are to be found in any disk $D(P, \varepsilon)$ around P. There are no accumulation points to the left of the parabola. (See Exercise 2.4-5.) So $\mathrm{cl}(A) = A \cup A' = \{(x, y) \in \mathbb{R}^2 \mid x \geq y^2\}$. ♦

FIGURE 2-11.

$\diamond$ **2.5-2.** Find the closure of $\{1/n \mid n = 1, 2, 3, \dots\}$ in $\mathbb{R}$.

Answer. $\mathrm{cl}(A) = \{0, 1, 1/2, 1/3, \dots\} = A \cup \{0\}$. $\Diamond$

Solution. The point 0 is an accumulation point of A by the Archimedean principle. There are no other accumulation points since the convergent sequence $\langle 1, 1/2, 1/3, 1/4, \dots \rangle$ has only one cluster point. So

$$\mathrm{cl}(A) = A \cup A' = A \cup \{0\} = \{0, 1, 1/2, 1/3, \dots\}.$$ $\blacklozenge$

$\diamond$ **2.5-3.** Let $A = \{(x, y) \in \mathbb{R}^2 \mid x \text{ is rational}\}$. Find $\mathrm{cl}(A)$.

Answer. $\mathbb{R}^2$. $\Diamond$

Solution. As we have seen before, the rational numbers are scattered densely in the real line in the sense that if a and b are real numbers with $a < b$, then there is a rational number r with $a < r < b$. Now suppose $P = (x, y)$ is any point in $\mathbb{R}^2$ and let $\varepsilon > 0$. Find a rational number r with $x < r < x + \varepsilon$. Let $Q = (r, y)$. Then $\| P - Q \| = r - x < \varepsilon$. So $Q \in D(P, \varepsilon)$ and Q is not equal to P and $Q \in A$. Since this can be done for every $\varepsilon > 0$, we see that P is an accumulation point of A. All points of $\mathbb{R}^2$ are accumulation points of A, so $\mathrm{cl}(A) = \mathbb{R}^2$. $\blacklozenge$

$\diamond$ **2.5-4.** (a) For $A \subset \mathbb{R}^n$, show that $\mathrm{cl}(A) \setminus A$ consists entirely of accumulation points of A.
(b) Need it be all of them?

Suggestion. Establish first the following general observation: If A and B are sets then $(A \cup B) \setminus A \subseteq B$. Apply this with $B = A'$. For part (b) the answer is "'No". What might happen if A is a closed set? $\Diamond$

Solution. (a) From Proposition 2.5.2 we know that $\text{cl}(A) = A \cup A'$. So $\text{cl}(A) \setminus A = (A \cup A') \setminus A$.

We establish the following general observation:

Lemma. *If A and B are sets then $(A \cup B) \setminus A \subseteq B$.*

Proof: Suppose $x \in (A \cup B) \setminus A$. Then x must either be in A or B since it is in $A \cup B$. But it is not in A since the points of A have been deleted. Thus x must be in B.

Apply this with $B = A'$ to obtain

$$\text{cl}(A) \setminus A = (A \cup A') \setminus A \subseteq A'$$

as desired.

(b) The answer is "No". Some or all of the accumulation points might be in the set A. If A is closed they all are. For example, if $A = [0, 1] \subseteq \mathbb{R}$, then $A' = A = \text{cl}(A) = [0, 1]$, but $\text{cl}(A) \setminus A = \emptyset$.

Challenge: For what sets is it true that $\text{cl}(A) \setminus A = A'$? ♦

◇ **2.5-5.** In a general metric space M, let $A \subset D(x, r)$ for some $x \in M$ and $r > 0$. Show that $\text{cl}(A) \subset B(x, r) = \{y \in M \mid d(x, y) \leq r\}$.

Sketch. The set $B(x, r) = M \setminus \{y \mid d(y, x) > r\}$ is closed and $A \subseteq B(x, r)$. So $\text{cl}(A) \subseteq B(x, r)$. ◇

Solution. We know that the set $U = \{y \mid d(y, x) > r\}$ is open so that $B = \{y \mid d(y, x) \leq r\}$ is closed. (See the solution to Exercise 2.3-6 or Example 2.3.4 and Figure 2.3-3 of the text for essentially the same thing in $\mathbb{R}^2$.) We certainly have $A \subseteq D(x, r) = \{y \mid d(y, x) < r\} \subseteq B(x, r)$ since $d(y, x) < r \implies d(y, x) \leq r$. So $A \subseteq B$. We have that B is a closed set with $A \subseteq B$, so $\text{cl}(A) \subseteq B$ as claimed. ♦

2.6 Boundary of a Set

◇ **2.6-1.** Find $\text{bd}(A)$ where $A = \{1/n \in \mathbb{R} \mid n = 1, 2, 3, \dots\}$.

Answer. $\text{bd}(A) = \{0\} \cup A$. ◇

Solution. $\text{cl}(A) = A \cup \{0, 1, 1/2, 1/3, \dots\}$ (see Exercise 2.5-2). On the other hand, every disk $D(1/n, \varepsilon)$ contains points not in A. So $1/n \in \text{cl}(\mathbb{R} \setminus A)$ Since this can be done for every $n > 0$, we find that $A \subseteq \text{cl}(\mathbb{R} \setminus A)$. Since $\mathbb{R} \setminus A$ is also contained in $\text{cl}(\mathbb{R} \setminus A)$, we find that $\text{cl}(\mathbb{R} \setminus A) = \mathbb{R}$. So

$$\text{bd}(A) = \text{cl}(A) \cap \text{cl}(\mathbb{R} \setminus A) = \text{cl}(A) \cap \mathbb{R} = \text{cl}(A) = \{0, 1, 1/2.1/3, \dots\}.$$ ♦

◇ **2.6-2.** If $x \in \mathrm{cl}(A) \setminus A$, then show that $x \in \mathrm{bd}(A)$. Is the converse true?

Suggestion. If x is not in A, then it is in the complement of A. ◊

Solution. Let the enclosing metric space be M. Suppose $x \in \mathrm{cl}(A) \setminus A$. Then $x \in \mathrm{cl}(A)$, but x is not in A. So $x \in M \setminus A$. Since $M \setminus A \subseteq \mathrm{cl}(M \setminus A)$, we have $x \in \mathrm{cl}(M \setminus A)$. Thus $x \in \mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A) = \mathrm{bd}(A)$ as desired.

The converse is not true. If A is any nonempty closed set with nonempty boundary, then $A = \mathrm{cl}(A)$, so $\mathrm{cl}(A) \setminus A = \emptyset$. For example, take A to be a single point. $A = \{x\}$. Then $\mathrm{bd}(A) = A = \{x\}$. But $\mathrm{cl}(A) \setminus A = \{x\} \setminus \{x\} = \emptyset$. ♦

◇ **2.6-3.** Find $\mathrm{bd}(A)$ where $A = \{(x, y) \in \mathbb{R}^2 \mid x \leq y\}$.

Answer. $\mathrm{bd}(A) = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$. ◊

Solution. The set A is a closed half plane consisting of those points on or above the main diagonal. So $\mathrm{cl}(A) = A$. (The reader should by now be able to supply detail if required.) The complement, $\mathbb{R}^2 \setminus A$, is the open half plane $\{(x, y) \in \mathbb{R}^2 \mid x > y\}$ consisting of those points below the diagonal. Its closure is $\{(x, y) \in \mathbb{R}^2 \mid x \geq y\}$. The boundary of A is thus

$$\begin{aligned}\mathrm{bd}(A) &= \mathrm{cl}(A) \cap \mathrm{cl}(\mathbb{R}^2 \setminus A) \\ &= \{(x, y) \in \mathbb{R}^2 \mid x \leq y\} \cap \{(x, y) \in \mathbb{R}^2 \mid x \geq y\} \\ &= \{(x, y) \in \mathbb{R}^2 \mid x = y\}.\end{aligned}$$ ♦

◇ **2.6-4.** Is it always true that $\mathrm{bd}(A) = \mathrm{bd}(\mathrm{int}\ A)$?

Answer. Not necessarily. ◊

Solution. Any isolated points of A will be in $\mathrm{bd}(A)$ but will disappear when we take the interior. For example, let $A = [0, 1] \cup \{2\} \subseteq \mathbb{R}$. Then $\mathrm{bd}(A) = \{0, 1, 2\}$. But $\mathrm{int}(A) =]0, 1[$, and $\mathrm{bd}(\mathrm{int}(A)) = \{0, 1\}$. Certainly $\{0, 1, 2\}$ is not the same as $\{0, 1\}$. ♦

◇ **2.6-5.** Let $A \subset \mathbb{R}$ be bounded and nonempty and let $x = \sup(A)$. Is $x \in \mathrm{bd}(A)$?

Answer. Yes. ◊

Solution. Let $\varepsilon > 0$. Since $x = \sup A$, there must be an element y in A with $x - \varepsilon < x \leq x$. Every short interval around x contains points of A, so $x \in \mathrm{cl}(A)$. On the other hand, the upper half of such an interval, $]x, x+\varepsilon[$ consists entirely of points in $\mathbb{R} \setminus A$ since x is an upper bound for A. Thus $x \in \mathrm{cl}(\mathbb{R} \setminus A)$. Thus $x \in \mathrm{cl}(A) \cap \mathrm{cl}(\mathbb{R} \setminus A) = \mathrm{bd}(A)$ as desired. ♦

◇ **2.6-6.** Prove that the boundary of a set in $\mathbb{R}^2$ with the standard metric is the same as it would be with the taxicab metric.

Suggestion. Review Exercise 2.1-6. ◇

Solution. From Exercise 2.1-6, we know that the open sets are the same for these metrics. Since the closed sets are just the complements of the open sets, the closed sets are also the same in the two metrics. The closure of a set is just the intersection of all the closed sets which contain it, so the closure of a set is the same whichever metric we use. The boundary of a set is the intersection of the closure of the set with the closure of its complement. Since the closures are the same, the boundaries are the same. ♦

2.7 Sequences

◇ **2.7-1.** Find the limit of the sequence $\big((\sin n)^n/n, 1/n^2\big)$ in $\mathbb{R}^2$.

Answer. $(0, 0)$. ◇

Solution. Since $|\sin x| \leq 1$ for every x, we have $|(\sin n)^n/n| \leq 1/n$ for all positive integers n. Since $1/n \to 0$, so does $(\sin n)^n/n$. Also, $0 < 1/n^2 \leq 1/n$ for all $n = 1, 2, 3, \ldots$, so $1/n^2 \to 0$ also. Both coordinate sequences tend to 0 in $\mathbb{R}$. So the vectors $((\sin n)^n/n, 1/n^2)$ converge to $(0, 0)$ in $\mathbb{R}^2$. ♦

◇ **2.7-2.** Let $x_n \to x$ in $\mathbb{R}^m$. Show that $A = \{x_n \mid n = 1, 2, \ldots\} \cup \{x\}$ is closed.

Sketch. Let $B = \{x_1, x_2, \ldots\}$, and show that $A = \mathrm{cl}(B)$. ◇

Solution. Let $B = \{x_1, x_2, \ldots\}$. We are given that $x_n \to x$, so for each $\varepsilon > 0$, $x_n \in D(x, \varepsilon)$ for large enough n. Thus $x \in \mathrm{cl}(B)$. On the other hand, if $y \in \mathrm{cl}(B)$, then either $x \in B$ or there is a subsequence which converges to y. To see this suppose y is not in B. There is an index n_1

with $\| y - x_{n_1} \| < 1$. Since y is not among the x_n, there is an index n_2, necessarily larger than n_1, such that

$$\| y - x_{n_2} \| < \min(\| y - x_1 \|, \dots, \| y - x_{n_1} \|)/2.$$

Continuing inductively, we get indices

$$n_1 < n_2 < n_3 < \dots$$

such that

$$\| y - x_{n_k} \| < \min\left(\| y - x_1 \|, \dots, \left\| y - x_{n_{k-1}} \right\| \right)/k \leq \| y - x_1 \| /k \to 0.$$

So this subsequence converges to y. But it is a subsequence of a sequence which converges to x, so we must have $y = x$. Thus the only points which can be in $\mathrm{cl}(B)$ are the points x_n and x. So $\mathrm{cl}(B) = B \cup \{x\} = \{x_n \mid n = 1, 2, 3, \dots\} \cup \{x\}$. Thus this set is closed since the closure of any set is closed. ♦

◊ **2.7-3.** Let $A \subset \mathbb{R}^m$, $x_n \in A$, and $x_n \to x$. Show that $x \in \mathrm{cl}(A)$.

Sketch. Use Proposition 2.7.6(ii). ◊

Solution. This is exactly one direction of the equivalence in Proposition 2.7.6(ii) with $M = \mathbb{R}^m$. ♦

◊ **2.7-4.** Verify Proposition 2.7.6(ii) for the set $B = \{(x, y) \in \mathbb{R}^2 \mid x < y\}$.

Solution. Proposition 2.7.6(ii) says that $v \in \mathrm{cl}(B)$ if and only if there is a sequence $\langle v_n \rangle_1^\infty$ in B which converges to v. The closure of the open half plane $B = \{(x, y) \in \mathbb{R}^2 \mid x < y\}$ is the closed half plane $\mathrm{cl}(B) = \{(x, y) \in \mathbb{R}^2 \mid x \leq y\}$. (Why?) If $v \in B$, we can just take $v_n = v$ for each n to get a sequence converging to v. If $v = (x, x)$ is on the boundary line, we can take $v_n = (x - (1/n), x)$. These are in B, and $\| v - v_n \| = 1/n$, so $v_n \to v$. If $v = (x, y)$ with $y < x$, then v lies at a positive distance away from the boundary line, and no sequence from B can converge to it. ♦

◊ **2.7-5.** Let $S = \{x \in \mathbb{R} \mid x \text{ is rational and } x^2 < 2\}$. Compute $\mathrm{cl}(S)$.

Answer. $\mathrm{cl}(S) = \{x \in \mathbb{R} \mid x^2 \leq 2\} = [-\sqrt{2}, \sqrt{2}]$. ◊

Solution. Rational numbers x with $x^2 \leq 2$ are dense in the open interval $]-\sqrt{2}, \sqrt{2}[$ as we have seen several times before. We can find rationals as close as we wish to any real number in that interval. Their closure is thus the closed interval $[-\sqrt{2}, \sqrt{2}]$ of the real line. ♦

2.8 Completeness

◇ **2.8-1.** Suppose that N is a complete subset of a metric space (M, d). Show that N is closed.

Sketch. If $x_n \in N$ and $x_n \to x$, then $\langle x_n \rangle_1^\infty$ is a Cauchy sequence. Since N is complete, it has a limit in $N \subseteq M$. Since limits in M are unique, this limit must be x. So $x \in N$. This shows that N is closed. (Why?) ♢

Solution. To show that N is a closed subset of M, it suffices by Proposition 2.7.6(i) to show that if $\langle x_n \rangle_1^\infty$ is a sequence of points in N which converges to a limit x in M, then x must be in N. So suppose $\langle x_n \rangle_1^\infty$ is a sequence in N and that $x_n \to x \in M$. Since $\langle x_n \rangle_1^\infty$ is a convergent sequence in the metric space M, it must be a Cauchy sequence. The metric in N is, of course, the same as that of M, and all of the x_n are in N. So $\langle x_n \rangle_1^\infty$ is a Cauchy sequence in N. As a metric space in its own right we have assumed that N is complete. So the sequence $\langle x_n \rangle_1^\infty$ must converge to some point $y \in N$. But $N \subseteq M$, so we have $x_n \to x \in M$ and $x_n \to y \in M$. We know that limits in a metric space are unique, so $y = x$. Thus $x \in N$ as we needed to establish that N is closed. ♦

◇ **2.8-2.** Let (M, d) be a metric space with the property that every bounded sequence has a convergent subsequence. Prove that M is complete.

Sketch. Use parts (ii) and (iii) of Proposition 2.8.4. ♢

Solution. To show that M is complete we need to show that every Cauchy sequence in M converges to a limit in M. So let $\langle x_n \rangle_1^\infty$ be a Cauchy sequence in M. By Proposition 2.8.4(ii), the sequence is bounded. By hypothesis, this means that it must have a subsequence converging to some point $x \in M$. But then by 2.8.4(iii), the whole sequence converges to x. Thus every Cauchy sequence in M converges to a point in M, so M is complete. ♦

◇ **2.8-3.** Let M be a set with the discrete metric. Is M complete?

Sketch. Yes, every Cauchy sequence is eventually constant. ♢

Solution. Suppose M is a set and d is the discrete metric on M. That is, $d(x, y) = 0$ if $x = y$, and $d(x, y) = 1$ if x and y are different. Suppose $\langle x_n \rangle_1^\infty$ is a Cauchy sequence in M. Then there is an index N such that $d(x_n, x_k) < 1/2$ whenever $n \geq N$ and $k \geq N$. In particular, $d(x_N, x_k) \leq 1/2$ for all $k \geq N$. But this forces $x_k = x_N$ for all $k \geq N$. That is, the sequence is constantly equal to x_N past the N^{th} term. Thus $d(x_k, x_N) = 0$ for $k \geq N$, and the sequence converges to x_N. Every Cauchy sequence in M converges to a point in M, so M is complete. ♦

◇ **2.8-4.** Show that a Cauchy sequence can have at most one cluster point.

Sketch. Use 2.8.7(ii) and 2.8.4(iii) and the uniqueness of limits in a metric space (and possibly 2.8.7(iii)). ◊

Solution. Let $\langle x_n \rangle_1^\infty$ be a Cauchy sequence in a metric space (M, d), and suppose that x and y are cluster points of the sequence. We want to show that $x = y$. By Proposition 2.8.7(ii), there is a subsequence x_{n_k} converging to x and a subsequence x_{n_j} converging to y. By Proposition 2.8.4(iii), the whole sequence converges to x and also to y. But we know that limits in a metric space are unique, so $x = y$ as claimed.

Instead of using the uniqueness of limits directly at the last step, we could also argue as follows after observing that the sequence converges to x. By Proposition 2.8.7(iii) every subsequence converges to x. In particular x_{n_j} should converge to x. But it converges to y. Now apply the uniqueness of limits to conclude that $x = y$. ♦

◇ **2.8-5.** Suppose that a metric space M has the property that every bounded sequence has at least one cluster point. Show that M is complete.

Sketch. If $\langle x_n \rangle_1^\infty$ is a Cauchy sequence it is bounded. It has a cluster points and hence a convergent subsequence. So the whole sequence converges. (Why?) ◊

Solution. To show that M is complete we need to show that every cauchy sequence in M converges to a limit in M. So let $\langle x_n \rangle_1^\infty$ be a Cauchy sequence in M. By Proposition 2.8.4(ii), the sequence is bounded. By hypothesis it must have at least one cluster point x. By 2.8.7(ii) there is a subsequence which converges to x. Finally, by Proposition 2.8.4(iii), the whole sequence converges to x. Every Cauchy sequence in M converges to a point in M, so M is complete. ♦

2.9 Series of Real Numbers and Vectors

◇ **2.9-1.** Determine whether $\displaystyle\sum_{n=1}^{\infty} \left(\frac{(\sin n)^n}{n^2}, \frac{1}{n^2} \right)$ converges.

Answer. Converges. ◊

Solution. An infinite series converges if and only if the sequence of its partial sums converges. From Proposition 2.7.4 we know that a sequence of vectors in $\mathbb{R}^n$ converges to a limit w in $\mathbb{R}^n$ if and only if each of the coordinate sequences converges to the corresponding coordinate of w. The coordinates of the partial sums of an infinite series of vectors are precisely the partial sums of the coordinate series:

$$\sum_{j=1}^{k}(v_j^1,\ldots,v_j^n) = \left(\sum_{j=1}^{k} v_j^1,\ldots,\sum_{j=1}^{k} v_j^n\right)$$

We conclude,

Proposition. *An infinite series of vectors in $\mathbb{R}^n$ converges to a sum w in $\mathbb{R}^n$ if and only if each of the coordinate series converges to the corresponding coordinate of w.*

In our current problem, the coordinate series are $\displaystyle\sum_{n=1}^{\infty}\frac{(\sin n)^n}{n^2}$ and $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n^2}$. The second of these converges by the "p-series test", Proposition 2.9.4(iii). Also, since $|\sin x| \le 1$ for every real x, we have $0 \le \left|(\sin n)^n/n^2\right| \le 1/n^2$ for every n. So the series $\displaystyle\sum_{n=1}^{\infty}\left|\frac{(\sin n)^n}{n^2}\right|$ converges by the comparison test, 2.9.4(ii). Thus $\displaystyle\sum_{n=1}^{\infty}\frac{(\sin n)^n}{n^2}$ is absolutely convergent and so convergent by Theorem 2.9.3. Both of our coordinate series are convergent, so the series of vectors converges in $\mathbb{R}^2$. ♦

◇ **2.9-2.** Show that the series in Example 2.9.6 converges absolutely.

Sketch. Use comparison to a geometric series. ◇

Solution. In Example 2.9.6, we are given that $\|\, x_n \,\| \le 1/2^n$ for each n. But we know that the geometric series $\sum_{n=1}^{\infty}(1/2^n)$ converges as a geometric series with ratio $1/2 < 1$ (Theorem 2.9.4(i)). Since $0 \le \|\, x_n \,\| \le 1/2^n$ for each n, we conclude that $\sum_{n=1}^{\infty}\|\, x_n \,\|$ converges in $\mathbb{R}$ by comparison (2.9.4(ii)). This is exactly the definition of absolute convergence of the series $\sum_{n=1}^{\infty} x_n$. (It implies convergence of the series by Theorem 2.9.3, but this was not actually part of the problem.) ♦

◇ **2.9-3.** Let $\sum x_k$ converge in $\mathbb{R}^n$. Show that $x_k \to 0 = (0,\ldots,0) \in \mathbb{R}^n$.

Sketch. Let $\varepsilon > 0$. By 2.9.2, there is an N such that $\|\, x_k \,\| < \varepsilon$ for every $k \ge N$. So $x_k \to 0$. ◇

Solution. Taking $\mathcal{V} = \mathbb{R}^n$ and $p = 0$ in Theorem 2.9.2 (the Cauchy criterion for convergence of an infinite series), we see that for every $\varepsilon > 0$, there is an N such that $k \geq N$ implies $\| x_k \| < \varepsilon$. This means that $x_k \to \vec{0} = (0, 0, \ldots, 0) \in \mathbb{R}^n$ as claimed. ♦

◇ **2.9-4.** Test for convergence $\displaystyle\sum_{n=3}^{\infty} \frac{2^n + n}{3^n - n}$.

Suggestion. Try the ratio comparison test. ◇

Solution. The idea of the solution is that for large n, the exponentials 2^n and 3^n dominate the n terms and the series should behave much like the series $\sum(2^n/3^n) = \sum(2/3)^n$. We know this last series converges since it is a geometric series with ratio less than 1. So we conjecture that our original series converges. To establish this we will use the ratio comparison test. The dominance mentioned is embodied in the following observation.

Lemma. *If $r > 1$, then* $\displaystyle\lim_{n\to\infty} \frac{n}{r^n} = 0$.

Proof: Let $\varepsilon > 0$. Then

$$\begin{aligned} \frac{n}{r^n} < \varepsilon &\iff n < \varepsilon r^n \\ &\iff \log n < \log \varepsilon + n \log r \\ &\iff \log r > \frac{\log n - \log \varepsilon}{n}. \end{aligned}$$

The left side of the last inequality is positive since $r > 1$, and the right side tends to 0 as $n \to \infty$ by L'Hôpital's Rule. So the inequality is valid for large n, and we have our limit as claimed.

Now to solve the problem, let $a_n = \dfrac{2^n + n}{3^n - n}$ and $b_n = \dfrac{2^n}{3^n}$. Then

$$\frac{a_n}{b_n} = \frac{2^n + n}{3^n - n} \frac{3^n}{2^n} = \frac{1 + (n/2^n)}{1 - (n/3^n)} \to \frac{1+0}{1-0} = 1$$

By the ratio comparison test, the series $\displaystyle\sum_{n=3}^{\infty} \frac{2^n + n}{3^n - n}$ and $\displaystyle\sum_{n=3}^{\infty} \frac{2^n}{3^n} = \sum_{n=3}^{\infty} \left(\frac{2}{3}\right)^n$ either both converge or both diverge. We know the latter converges since it is a geometric series with ratio less than 1. So our series converges also. ♦

◇ **2.9-5.** Test for convergence $\displaystyle\sum_{n=0}^{\infty} \frac{n!}{3^n}$.

Answer. Does not converge. ◊

Solution. We know that $3^n/n!$ converges to 0 as $n \to \infty$ (Exercise 1.2-2). So $n!/3^n \to \infty$. In particular, the terms do not converge to 0. According to Exercise 2.9-3, the series cannot converge.

To avoid reference to §1.2, one can simply note that if $n \geq 3$, then $n!/3^n \geq 2/9$. The terms remain larger than $2/9$ and the partial sums must become arbitrarily large. The sum diverges to $+\infty$. ♦

Exercises for Chapter 2

◇ **2E-1.** Discuss whether the following sets are open or closed:

(a) $]1, 2[$ in $\mathbb{R}^1 = \mathbb{R}$
(b) $[2, 3]$ in $\mathbb{R}$
(c) $\cap_{n=1}^{\infty}[-1, 1/n[$ in$\mathbb{R}$
(d) $\mathbb{R}^n$ in $\mathbb{R}^n$
(e) A hyperplane in $\mathbb{R}^n$
(f) $\{r \in]0, 1[\mid r \text{ is rational}\}$ in$\mathbb{R}$
(g) $\{(x, y) \in \mathbb{R}^2 \mid 0 < x \leq 1\}$ in$\mathbb{R}^2$
(h) $\{x \in \mathbb{R}^n \mid ||x|| = 1\}$ in $\mathbb{R}^n$

Answer. (a) Open.
(b) Closed.
(c) $[-1, 0]$ is closed.
(d) Open and closed.
(e) Closed.
(f) Neither open nor closed.
(g) Neither open nor closed.
(h) Closed. ◊

Solution. (a) Any open interval such as $]1, 2[$ is an open subset of $\mathbb{R}^1 = \mathbb{R}$. This follows from Proposition 2.1.2 since an open interval is a one dimensional "disk" in $\mathbb{R}^1$. In this case the center is at $x_0 = 1.5$ and the radius is $r = 0.5$. $]1, 2[= \{x \in \mathbb{R} \mid |x - 1.5| < 0.5\}$.
A direct argument can be given by letting $x \in]1, 2[$, and setting $\delta = \min(2 - x, x - 1)$. If $y \in]x - \delta, x + \delta[$, then

$$1 = x - (x - 1) \leq x - \delta < y < x + \delta \leq x + (2 - x) = 2.$$

So $]x-\delta, x+\delta[= D(x,\delta) \subseteq]1,2[$. The set $]1,2[$ contains an open interval centered around each of its points, so it is an open set.

(b) If $a \le b \in \mathbb{R}$, then the "closed interval" $[a,b] = \{x \in \mathbb{R} \mid a \le x \le b\}$ is a closed subset of $\mathbb{R}$. One way to see this is to note that its complement is the union of the two "open half-lines" $U = \{x \in \mathbb{R} \mid x < a\}$ and $V = \{x \in \mathbb{R} \mid x > b\}$. Each of these is open by a argument like that of part (a), so their union is open. The complement of that union, $[a,b]$, is thus closed.

(c) $\bigcap_{n=1}^{\infty} [-1, 1/n[= [-1,0]$ since if $-1 \le x \le 0$, then $x \in [-1, 1/n[$ for each positive n, and so is in the intersection. If $x > 0$, then by the Archimedean principle there is an integer $n > 0$ such that $0 < 1/n < x$. So x is not in $]-1, 1/n[$ for that (or larger) n and is not in the intersection. The interval $[-1,0]$ is closed by part (b).

(d) As with any metric space, the entire space $\mathbb{R}^n$ is both open and closed as a subset of itself. Propositions 2.1.3(iii) and 2.3.2(iii).

(e) A hyperplane in $\mathbb{R}^n$ is an $(n-1)$–dimensional linear subspace. (Sometimes the word is also used to include what we have called "affine hyperplanes", that is, translations of such subspaces.) In $\mathbb{R}^2$ these would be straight lines. In $\mathbb{R}^3$ they would be ordinary planes, and so forth. In the jargon of the trade they have "co-dimension one". Such a set is a closed subset since its complement is open. One way to see this is to use a bit of knowledge about the ability to choose bases for vector spaces.

If H is a hyperplane in $\mathbb{R}^n$, then we can select an orthonormal basis $\{e_1, e_2, \dots e_{n-1}\}$ for H. We can then extend this to an orthonormal basis $\{e_1, e_2, \dots e_n\}$ for $\mathbb{R}^n$ by selecting a final unit vector e_n orthogonal to the first $n-1$. If v is not in H, then $v = \sum_{k=1}^{n} \langle v, e_k \rangle e_k$, and $\langle v, e_n \rangle$ is not 0. The vector $Pv = \sum_{k=1}^{n-1} \langle v, e_k \rangle e_k$ is in H and is called the projection of v onto H. Furthermore, if $w \in H$, then $Pv - w \in H$, so $Pv - w$ is perpendicular to e_n and thus to $v - Pv = \langle v, e_n \rangle e_n$. In general we have the "Pythagorean theorem":

Lemma. *If $\langle f, g \rangle = 0$, then $\| f + g \|^2 = \| f \|^2 + \| g \|^2$.*

Proof: Using the assumed orthogonality, we compute

$$\begin{aligned} \| f+g \|^2 &= \langle f+g, f+g \rangle = \langle f,f \rangle + \langle f,g \rangle + \langle g,f \rangle + \langle g,g \rangle \\ &= \| f \|^2 + 0 + 0 + \| g \|^2 = \| f \|^2 + \| g \|^2 \end{aligned}$$

as desired.

Applying this in our current setting, we find

$$\begin{aligned} \| v - w \|^2 &= \| (v - Pv) + (Pv - w) \|^2 = \| v - Pv \|^2 + \| Pv - w \|^2 \\ &= |\langle v, e_n \rangle|^2 + \| Pv - w \|^2 \le |\langle v, e_n \rangle|^2 . \end{aligned}$$

So, if we set $r = |\langle v, e_n \rangle|$, we find that $r > 0$ and $\| v - w \| \ge r$ for every w in H. Thus $D(v, r/2) \subseteq \mathbb{R}^n \setminus H$. So $\mathbb{R}^n \setminus H$ is open and H is closed in $\mathbb{R}^n$. See Figure 2-13.

Figure 2-13. A hyperplane is closed.

(f) The set $F = \{r \in]0, 1[\mid r \in \mathbb{Q}\}$ is not open in $\mathbb{R}$ since, for example, every open interval centered at $1/2$ contains irrational numbers. It is not closed since the complement is not open. Every open interval around 0 contains the rational numbers $1/n$ for large enough positive integers n, and these are in F. Thus F is neither open nor closed.

(g) The set $G = \{(x, y) \in \mathbb{R}^2 \mid 0 < x \leq 1\}$ is neither open nor closed in $\mathbb{R}^2$. As in Example 2.1.5 it is not open since $(1, 0) \in G$, but each disk around $(1, 0)$ contains points $(x, 0)$ not in G. Just take x slightly larger than 1. But, similarly, the complement is not open either. The point $(0, 0)$ is in the complement, but each small disk around $(0, 0)$ contains points $(x, 0)$ which are in G. Just take x slightly larger than 0. The complement is not open, so G is not closed.

(h) The set $H = \{x \in \mathbb{R}^n \mid \| x \| = 1\}$ is a closed subset of $\mathbb{R}^n$. One way to see this is to note that its complement is the union of the two sets $U = \{x \in \mathbb{R}^n \mid \| x \| < 1\}$ and $V = \{x \in \mathbb{R}^n \mid \| x \| > 1\}$. The disk U is open by Proposition 2.1.2. The set V is also open. The argument is basically that given in $\mathbb{R}^2$ in the solution to Example 2.3.4 and illustrated in Figure 2.3-3 of the text. Their union is thus also open and the complement of that union, which is H, is closed. ♦

◇ **2E-2.** Determine the interiors, closures, and boundaries of the sets in Exercise 2E-1.

Answer. (a) If $A =]1, 2[$ in $\mathbb{R}^1 = \mathbb{R}$, then $\mathrm{int}(A) = A =]1, 2[$, $\mathrm{cl}(A) = [1, 2]$, and $\mathrm{bd}(A) = \{1, 2\}$.

(b) If $B = [2, 3]$ in $\mathbb{R}$, then $\mathrm{int}(B) =]2, 3[$, $\mathrm{cl}(B) = B = [2, 3]$, and $\mathrm{bd}(B) = \{2, 3\}$.

(c) If $C = \bigcap_{n=1}^{\infty} [-1, 1/n[$ in $\mathbb{R}$, then $\mathrm{int}(C) =]-1, 0[$, $\mathrm{cl}(C) = [-1, 0]$, and $\mathrm{bd}(C) = \{-1, 0\}$.

(d) If $D = \mathbb{R}^n$ in $\mathbb{R}^n$, then $\mathrm{int}(D) = \mathrm{cl}(D) = D$ and $\mathrm{bd}(D) = \emptyset$.

(e) If E is a hyperplane in $\mathbb{R}^n$, then $\mathrm{int}(E) = \emptyset$ and $\mathrm{cl}(E) = \mathrm{bd}(E) = E$.

(f) If $F = \{r \in]0, 1[\mid r \in \mathbb{Q}\}$ in $\mathbb{R}$, then $\mathrm{int}(F) = \emptyset$ and $\mathrm{cl}(F) = \mathrm{bd}(F) = [0, 1]$.

(g) If $G = \{(x, y) \in \mathbb{R}^2 \mid 0 < x \leq 1\}$ in $\mathbb{R}^2$, then $\mathrm{int}(G) = \{(x, y) \in \mathbb{R}^2 \mid 0 < x < 1\}$, $\mathrm{cl}(G) = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1\}$, and $\mathrm{bd}(G) = \{(x, y) \in \mathbb{R}^2 \mid x = 0 \text{ or } x = 1\}$.

(h) If $H = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ in $\mathbb{R}^n$, then $\mathrm{int}(H) = \emptyset$ and $\mathrm{cl}(H) = \mathrm{bd}(H) = H$. ◊

Solution. (a) If $A =]1, 2[$ in $\mathbb{R}^1 = \mathbb{R}$, then we know from Exercise 2E-1(a) that A is open, so $\mathrm{int}(A) = A$. The endpoints 1 and 2 are certainly accumulation points, so they are in the closure. From Exercise 2E-1(b), we know that closed intervals are closed sets, so $[1, 2]$ is closed. Thus $\mathrm{cl}(A) = [1, 2]$. Similarly $\mathrm{cl}(\mathbb{R} \setminus A) = \{x \mid x \leq 1\} \cup \{x \mid x \geq 2\}$, so $\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(\mathbb{R} \setminus A) = \{1, 2\}$.

(b) If $B = [2, 3]$ in $\mathbb{R}$, then we know from Exercise 2E-1(b) that B is closed, so $\mathrm{cl}(B) = B$. The situation is just like that of part (a). We have a finite interval in $\mathbb{R}$. The interior is the open interval, the closure is the closed interval, and the boundary is the set consisting of the two endpoints. So $\mathrm{int}(B) =]2, 3[$, $\mathrm{cl}(B) = B = [2, 3]$, and $\mathrm{bd}(B) = \{2, 3\}$.

(c) If $C = \bigcap_{n=1}^{\infty} [-1, 1/n[$ in $\mathbb{R}$, then as we saw in Exercise 2E-1(c), $C = [-1, 0]$. Again we have an interval in $\mathbb{R}$ and, as discussed above, we must have $\mathrm{int}(C) =]-1, 0[$, $\mathrm{cl}(C) = [-1, 0]$, and $\mathrm{bd}(C) = \{-1, 0\}$.

(d) If $D = \mathbb{R}^n$ in $\mathbb{R}^n$, then we know that D is both open and closed, so $\mathrm{int}(D) = \mathrm{cl}(D) = D$. Since the complement of D is empty, the closure of the complement is empty, and we must have $\mathrm{bd}(D) = \emptyset$.

(e) Suppose E is a hyperplane in $\mathbb{R}^n$. As in Exercise 2E-1(e), E is closed, so $\mathrm{cl}(E) = E$. Let e_n be a unit vector orthogonal to E, then if $w \in E$, the points $v_k = w + (1/k)e_n$ are in the complement of E and converge to w. So w is in the closure of the complement and is not in the interior of E. This is true for every $w \in E$, so $\mathrm{int}(E) = \emptyset$ and $\mathrm{bd}(E) = \mathrm{cl}(E) \cap \mathrm{cl}(\mathbb{R}^n \setminus E) = E \cap \mathbb{R}^n = E$.

(f) Let $F = \{r \in]0, 1[\mid r \in \mathbb{Q}\}$ in $\mathbb{R}$. If $r \in F$ and $\varepsilon > 0$, then we have seen that the interval $]r - \varepsilon, r + \varepsilon[$ contains irrational numbers. So r cannot be an interior point. So $\mathrm{int}(F) = \emptyset$. On the other hand, if $0 \leq x_0 \leq 1$, then there are rational numbers between $x_0 - \varepsilon$ and x_0 and between x_0 and $x_0 + \varepsilon$. So x_0 is an accumulation point of F. We must have $\mathrm{cl}(F) = [0, 1]$. But for similar reasons, $\mathrm{cl}(\mathbb{R} \setminus F) = \mathbb{R}$, so $\mathrm{bd}(F) = \mathrm{cl}(F) \cap \mathrm{cl}(\mathbb{R} \setminus F) = [0, 1] \cap \mathbb{R} = [0, 1]$.

(g) Consider the sets

$$
\begin{aligned}
G &= \{(x, y) \in \mathbb{R}^2 \mid 0 < x \leq 1\} \\
S &= \{(x, y) \in \mathbb{R}^2 \mid 0 < x < 1\} \\
T &= \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1\} \\
U &= \{(x, y) \in \mathbb{R}^2 \mid x < 0\} \\
V &= \{(x, y) \in \mathbb{R}^2 \mid x > 1\} \\
W &= \{(x, y) \in \mathbb{R}^2 \mid x \geq 1\} \\
P &= \{(x, y) \in \mathbb{R}^2 \mid x \leq 0\}.
\end{aligned}
$$

We certainly have $S \subseteq G \subseteq T$. As in Example 2.1.4 of the text, the set S is open. So $S \subseteq \mathrm{int}(G)$. As in Example 2.1.5, none of the points (x, y) with $\mathrm{int}(H) = S = \{(x, y) \in \mathbb{R}^2 \mid 0, x < 1\}$. The half planes U and V are open by an argument similar to that of Example 2.1.5 which we have seen before. For example, suppose $(a, b) \in U$. Then $a < 0$. If $\| (x, y) - (a, b) \| < |a|$, then

$$|x - a| \leq \sqrt{(x-a)^2 + (y-b)^2} < |a| \,.$$

So

$$a = -|a| < x - a < |a| = -a.$$

Adding a to both sides of the second inequality gives $x < 0$. Thus $(x, y) \in U$. Thus $D((a, b), |a|) \subseteq U$. The set U contains an open disk around each of its points and so is open. The argument for the half plane V is similar. Since U and V are open, so is their union. But that union is the complement of T. So T is closed. Thus $\mathrm{cl}(G) \subseteq T$. On the other hand, the points $(1/n, y)$ are in G for each $n = 1, 2, \ldots$ and converge to $(0, y)$. So $(0, y) \in \mathrm{cl}(G)$. We must have $\mathrm{cl}(G) = T = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1\}$. Finally, W and P are closed since they are the complements of V and U. So their union is closed and must be the closure of the complement of G. So $\mathrm{bd}(G) = T \cap (W \cup P) = \{(x, y) \in \mathbb{R}^2 \mid x = 0 \text{ or } x = 1\}$.

(h) Let $H = \{x \in \mathbb{R}^n \mid \| x \| = 1\}$ in $\mathbb{R}^n$. Thus H is closed (Exercise 2E-1(h)). So $\mathrm{cl}(H) = H$. If $x \in H$, then the points $w_n = (1 + (1/n)x$ are not in H but do converge to x. So x cannot be an interior point and $\mathrm{int}(H) = \emptyset$. This also shows that $x \in \mathrm{cl}(\mathbb{R}^n \setminus H)$. So $H \subseteq \mathrm{cl}(\mathbb{R}^n \setminus H)$, and $\mathrm{bd}(H) = \mathrm{cl}(H) \cap \mathrm{cl}(\mathbb{R}^n \setminus H) = H \cap \mathrm{cl}(\mathbb{R}^n \setminus H) = H$. ♦

◇ **2E-3.** Let U be open in M and $U \subset A$. Show that $U \subset \mathrm{int}(A)$. What is the corresponding statement for closed sets?

Suggestion. Each x in U satisfies the definition of an interior point of A. For closed sets: If $A \subseteq C$ and C is closed, then $\mathrm{cl}(A) \subseteq C$. ◇

Solution. Suppose U is open and $U \subseteq A$. Let $x \in U$, then U itself is an open set such that $x \in U \subseteq A$. So $x \in \text{int}(A)$. Thus $U \subseteq \text{int}(A)$ as claimed.

Here is a similar sort of assertion about closed sets and closures:

Proposition. *If $A \subseteq C$ and C is closed, then* $\text{cl}(A) \subseteq C$.

Proof: The closure of a set A was defined as the intersection of all closed sets which contain A. Since C is one of the sets being intersected, we certainly have

$$\text{cl}(A) = \cap\{D \subseteq M \mid A \subseteq D \text{ and } D \text{ is closed }\} \subseteq C$$

as claimed. ♦

◇ **2E-4.** (a) Show that if $x_n \to x$ in a metric space M, then $x \in \text{cl}\{x_1, x_2, \ldots\}$. When is x an accumulation point?

(b) Can a sequence have more than one accumulation point?

(c) If x is an accumulation point of a set A, prove that there is a sequence of *distinct* points of A converging to x.

Sketch. (a) Use 2.7.6(ii).

(b) Yes.

(c) Choose the sequence inductively with $d(x, x_{n+1}) < d(x, x_n)/2$. ◇

Solution. (a) Let $S = \{x_1, x_2, x_3, \ldots\}$. The set whose elements appear as entries in the sequence. Then we certainly have $\langle x_n \rangle_1^\infty$ is a sequence in S which converges to x. By Proposition 2.7.6(ii), we conclude that $x \in \text{cl}(S)$ as requested. For x to be an accumulation point of S, we need every open set containing x to contain points of the sequence which are not equal to x. This will happen if the sequence contains infinitely many terms which are not equal to each other. In that case we will have: For each $\varepsilon > 0$ and each N there is an index $n > N$ with $d(x, x_n) < \varepsilon$.

(b) A sequence can have more than one accumulation point. For example, if $x_n = (-1)^n + (1/n)$, then 1 and -1 are accumulation points of the sequence $\langle x_n \rangle_1^\infty$.

(c) Suppose x is an accumulation point of a set A. A sequence of distinct (that is, all different) points can be selected from A converging to x by choosing the sequence inductively. That is, the choice of each term in the sequence is made depending on knowledge of the earlier terms.

STEP 0: Since x is an accumulation point of A, there is a point $x_0 \in A$ with $d(x_0, x) < 1$ and x_0 not equal to x.

STEP 1: Since x_0 is not x, we have $d(x_0, x) > 0$. Since x is an accumulation point of A, there is a point x_1 not equal to x in A such that

$d(x_1, x) < d(x_0, x)/2$. Note that this forces x_1 to be different from x_0.

STEP 2: Since x_1 is not x, we have $d(x_1, x) > 0$. Since x is an accumulation point of A, there is a point x_2 not equal to x in A such that $d(x_2, x) < d(x_1, x)/2$. We have x, x_0, x_1, and x_2 are all different and

$$d(x_2, x) < d(x_1, x)/2 < d(x_0, x)/4 < 1/4.$$

$$\vdots$$

STEP $n+1$: Having selected x_0, x_1, x_2, ... , and x_n in A, all different from x and from each other with

$$0 < d(x_n, x) < d(x_{n-1}, x)/2 < \cdots < d(x_1, x)/2^{n-1} < d(x_0, x)/2^n < 1/2^n$$

we can select a point x_{n+1} in A not equal to x such that $d(x_{n+1}, x) < d(x_n, x)/2$. We then have x_{n+1} different from all preceding terms in the sequence and

$$\begin{aligned} 0 < d(x_{n+1}, x) <& d(x_n, x)/2 < d(x_{n-1}, x)/4 < \cdots < d(x_2, x)/2^{n-1} \\ &< d(x_1, x)/2^n < d(x_0, x)/2^{n+1} < 1/2^{n+1}. \end{aligned}$$

$$\vdots$$

This process inductively selects a sequence $\langle x_n \rangle_1^\infty$ of points in A which are all different, none of which are equal to x, and such that $d(x_n, x) < 1/2^n$ for each n. So $x_n \to x$, and they meet the requirements of the problem. ♦

◇ **2E-5.** Show that $x \in \operatorname{int}(A)$ iff there is an $\varepsilon > 0$ such that $D(x, \varepsilon) \subset A$.

Suggestion. There is an open set U with $x \in U \subseteq A$. Use the definition of openness of U. ♢

Solution. First suppose that $x \in \operatorname{int}(A)$. Then there is an open set U such that $x \in U \subseteq A$. Since U is open and $x \in U$, there is an $\varepsilon > 0$ such that $x \in D(x, \varepsilon) \subseteq U$. Putting these together we obtain $x \in D(x, \varepsilon) \subseteq U \subseteq A$ as desired.

For the opposite implication, suppose there is an $\varepsilon > 0$ such that $D(x, \varepsilon) \subseteq A$. Let $U = D(x, \varepsilon)$. Since $d(x, x) = 0$, we certainly have $x \in U \subseteq A$, and by Proposition 2.1.2, we know that U is open. So $x \in \operatorname{int}(A)$.

We have proved the implications in both directions, so $x \in \operatorname{int}(A)$ iff there is an $\varepsilon > 0$ such that $D(x, \varepsilon) \subseteq A$ as claimed. ♦

◇ **2E-6.** Find the limits, if they exist, of these sequences in $\mathbb{R}^2$:

(a) $\left((-1)^n, \dfrac{1}{n}\right)$

(b) $\left(1, \dfrac{1}{n}\right)$

(c) $\left(\left(\dfrac{1}{n}\right)(\cos n\pi), \left(\dfrac{1}{n}\right)\left(\sin\left(n\pi + \dfrac{\pi}{2}\right)\right)\right)$

(d) $\left(\dfrac{1}{n}, n^{-n}\right)$

Answer. (a) Does not converge.
(b) $(1, 0)$.
(c) $(0, 0)$.
(d) $(0, 0)$. ♢

Solution. (a) The limit $\lim_{n\to\infty}((-1)^n, 1/n))$ does not exist since the first coordinate sequence, $-1, 1, -1, \ldots$ does not converge and we know from Proposition 2.7.4 that a sequence of vectors in $\mathbb{R}^2$ converges to a limit vector w if and only if both coordinate sequences converge to the corresponding coordinate of w. We use this same proposition in parts (b), (c), and (d).

(b) In the sequence of vectors given by $v_n = (1, 1/n)$, the first coordinate sequence is the constant sequence $x_n = 1$ for every n. This certainly converge to 1. The second coordinate sequence is given by $y_n = 1/n$ for each n. We know by the Archimedean Principle that this converges to 0. So, by Proposition 2.7.4, $\lim_{n\to\infty} v_n = (1, 0)$.

(c) Now put $v_n = \left(\left(\dfrac{1}{n}\right)(\cos n\pi), \left(\dfrac{1}{n}\right)\left(\sin\left(n\pi + \dfrac{\pi}{2}\right)\right)\right)$. For the coordinate sequences we have

$$|x_n| = \left|\frac{1}{n}\cos n\pi\right| \leq \frac{1}{n} \quad \text{and} \quad |y_n| = \left|\frac{1}{n}\sin\left(n\pi + \frac{\pi}{2}\right)\right| \leq \frac{1}{n}.$$

Since $1/n \to 0$, we conclude that $x_n \to 0$ and $y_n \to 0$ (Sandwich). By Proposition 2.7.4, $\lim_{n\to\infty} v_n = (0, 0)$.

(d) Now put $v_n = \left(\dfrac{1}{n}, n^{-n}\right)$. For the coordinate sequences we have

$$|x_n| = \left|\frac{1}{n}\right| = \frac{1}{n} \quad \text{and} \quad |y_n| = \left|\frac{1}{n^n}\right| \leq \frac{1}{n}.$$

The latter is because $0 < n < n^n$ for $n \geq 1$. Since $1/n \to 0$, we conclude that $x_n \to 0$ and $y_n \to 0$. (Sandwich). By Proposition 2.7.4, $\lim_{n\to\infty} v_n = (0, 0)$. ♦

◇ **2E-7.** Let U be open in a metric space M. Show that $U = \mathrm{cl}(U) \setminus \mathrm{bd}(U)$. Is this true for every set in M?

Answer. It is not true for every subset of M. Try $U = [0, 1] \subseteq \mathbb{R}$. ◇

Solution. Let U be an open subset of a metric space M and suppose that $x \in U$. Since U is open, $M \setminus U$ is closed. So $\mathrm{cl}(M \setminus U) = M \setminus U$. In particular, x is not in $\mathrm{cl}(M \setminus U)$, so it is not in $\mathrm{bd}(U) = \mathrm{cl}(M \setminus U) \cap \mathrm{cl}(U)$. On the other hand, since $U \subseteq \mathrm{cl}(A)$, we do have $x \in \mathrm{cl}(U)$. Thus $x \in \mathrm{cl}(U) \setminus \mathrm{bd}(U)$. This shows that $U \subseteq \mathrm{cl}(U) \setminus \mathrm{bd}(U)$.

Now suppose $x \in \mathrm{cl}(U) \setminus \mathrm{bd}(U)$. We want to show that x must be in U. Since it is in the closure of U but not in the boundary of U, it must not be in the closure of the complement of U. But U is open, so the complement of U is closed and so equal to its closure. Thus x is not in the complement of U and so it must be in U.

We have proved inclusion in both directions, so the sets are equal as claimed.

The equality is not true for every set. Consider for example, $M = \mathbb{R}$ and $U = [0, 1]$. Since U is closed, $\mathrm{cl}(U) = U = [0, 1]$. But the boundary of U is the two point set $\{0, 1\}$. So $\mathrm{cl}(U) \setminus \mathrm{bd}(U)$ is the open interval $]0, 1[$.

A good discussion question: Does this characterize open sets? Is it possible that U is open if and only if $U = \mathrm{cl}(U) \setminus \mathrm{bd}(U)$? ♦

◇ **2E-8.** Let $S \subset \mathbb{R}$ be nonempty, bounded below, and closed. Show that $\inf(S) \in S$.

Suggestion. Show there is a sequence in S converging to $\inf S$. ◇

Solution. Since S is a nonempty subset of $\mathbb{R}$ which is bounded below, it has a greatest lower bound $\lambda = \inf S \in \mathbb{R}$ by Theorem 1.3.4. Since λ is the greatest lower bound, we know that for each $n = 1, 2, 3, \ldots$, that $\lambda + (1/n)$ is not a lower bound. So there must be points x_n in S with

$$\lambda \le x_n < \lambda + \frac{1}{n}.$$

So $|x_n - \lambda| < 1/n$. Since $1/n \to 0$, we conclude that $x_n \to \lambda = \inf S$. Since $\langle x_n \rangle_1^\infty$ is a sequence in S converging to $\inf S$, we know from Proposition 2.7.6 that $\inf S$ must be in $\mathrm{cl}(S)$. ♦

◇ **2E-9.** Show that

(a) $\mathrm{int} B = B \setminus \mathrm{bd} B$, and
(b) $\mathrm{cl}(A) = M \setminus \mathrm{int}(M \setminus A)$.

Answer. Use Proposition 2.6.2 to get (a). Then use (a) to get (b). $\Diamond$

Solution. (a) Let B be a subset of a metric space M, and suppose that $x \in \mathrm{int}(B)$. Then there is an $\varepsilon > 0$ such that $x \in D(x, \varepsilon) \subseteq B$. (See Exercise 2E-5.) This disk is contained in B and so cannot intersect $M \setminus B$. By Proposition 2.6.2, x cannot be in $\mathrm{bd}(B)$. On the other hand, since $\mathrm{int}(B) \subseteq B$, x is certainly in B. Thus $x \in B \setminus \mathrm{bd}(B)$. This shows that $\mathrm{int}(B) \subseteq B \setminus \mathrm{bd}(B)$.

Now suppose that $x \in B \setminus \mathrm{bd}(B)$. We want to show that x must be an interior point of B. Since $x \in B$, every disk centered at x intersects B (at least at x if nowhere else). If all such disks also intersected $M \setminus B$, then x would be in $\mathrm{bd}(B)$ by 2.6.2. But we have assumed that x is not in $\mathrm{bd}(B)$. So there must be an $\varepsilon > 0$ such that $D(x, \varepsilon) \cap (M \setminus B) = \emptyset$. This forces $x \in D(x, \varepsilon) \subseteq B$. So $x \in \mathrm{int}(B)$. (See Exercise 2E-5.) This holds for every such x, so $B \setminus \mathrm{bd}(B) \subseteq \mathrm{int}(B)$.

We have proved inclusion in both directions, so $\mathrm{int}(B) = B \setminus \mathrm{bd}(B)$ as claimed.

(b) Let A be a subset of a metric space M and set $B = M \setminus A$. So $A = M \setminus B$. Using the result of part (a) we can compute:

$$\begin{aligned} M \setminus \mathrm{int}(M \setminus A) &= M \setminus \mathrm{int}(B) = M \setminus (B \setminus \mathrm{bd}(B)) \\ &= M \setminus \big[B \cap (M \setminus \mathrm{bd}(B))\big] \\ &= [M \setminus B] \cup [M \setminus (M \setminus \mathrm{bd}(B))] \\ &= A \cup \mathrm{bd}(B) = A \cup \mathrm{bd}(M \setminus A) = A \cup \mathrm{bd}(A). \end{aligned}$$

Since $\mathrm{bd}(M \setminus A) = \mathrm{cl}(M \setminus A) \cap \mathrm{cl}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A) = \mathrm{bd}(A)$. We are done as soon as we establish the following:

Proposition. *If A is a subset of a metric space M, then* $\mathrm{cl}(A) = A \cup \mathrm{bd}(A)$.

Proof: First of all, $\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A) \subseteq \mathrm{cl}(A)$, and $A \subseteq \mathrm{cl}(A)$. So $A \cup \mathrm{bd}(A) \subseteq \mathrm{cl}(A)$. For the opposite inclusion, suppose that $x \in \mathrm{cl}(A)$. If $x \in A$, then it is certainly in $A \cup \mathrm{bd}(A)$. If x is not in A, then it is in $M \setminus A$ and so in $\mathrm{cl}(M \setminus A)$. Since we have assumed that it is in $\mathrm{cl}(A)$, we have $x \in \mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A) = \mathrm{bd}(A) \subset A \cup \mathrm{bd}(A)$. In either case, $x \in A \cup \mathrm{bd}(A)$. So $\mathrm{cl}(A) \subseteq A \cup \mathrm{bd}(A)$. We have inclusion in both directions, so the sets are equal as claimed. $\blacklozenge$

$\diamond$ **2E-10.** Determine which of the following statements are true.

(a) $\mathrm{int}(\mathrm{cl}(A)) = \mathrm{int}(A)$.
(b) $\mathrm{cl}(A) \cap A = A$.
(c) $\mathrm{cl}(\mathrm{int}(A)) = A$.

(d) $\text{bd}(\text{cl}(A)) = \text{bd}(A)$.
(e) If A is open, then $\text{bd}(A) \subset M \setminus A$.

Answer. (a) May be false.
(b) True.
(c) May be false.
(d) May be false.
(e) True. ♢

Solution. (a) The equality $\text{int}(\text{cl}(A)) = \text{int}(A)$ is not always true. Consider the example $A = [-1,1]\setminus\{0\} \subseteq \mathbb{R}$. Then $\text{cl}(A)$ is the closed interval $[-1,1]$, and $\text{int}(\text{cl}(A))$ is the open interval $]-1,1[$. But $\text{int}(A)$ is the open interval with zero deleted. $\text{int}(A) =]-1,1[\setminus\{0\}$.
(b) True: Since $A \subseteq \text{cl}(A)$, we always have $\text{cl}(A) \cap A = A$.
(c) The proposed equality, $\text{cl}(\text{int}(A)) = A$, is not always true. Consider the example of a one point set with the usual metric on $\mathbb{R}$. Take $A = \{0\} \subseteq \mathbb{R}$. Then $\int A = \emptyset$. So $\text{cl}(\text{int}(A)) = \emptyset$. But A is not empty.
(d) The proposed equality, $\text{bd}(\text{cl}(A)) = \text{bd}(A)$, is not always true. Consider the same example as in part (a). $A = [-1,1] \setminus \{0\} \subseteq \mathbb{R}$. Then $\text{cl}(A)$ is the closed interval $[-1,1]$ and $\text{bd}(\text{cl}(A))$ is the two point set $\{-1,1\}$. But $\text{bd}(A)$ is the three point set $\{-1,0,1\}$.
(e) The proposed inclusion, $\text{bd}(A) \subseteq M \setminus A$, is true if A is an open subset of the metric space M. The set A is open, so its complement, $M \setminus A$, is closed. Thus

$$\text{bd}(A) = \text{cl}(A) \cap \text{cl}(M \setminus A) = \text{cl}(A) \cap (M \setminus A) \subseteq M \setminus A$$

as claimed. ♦

◇ **2E-11.** Show that in a metric space, $x_m \to x$ iff for every $\varepsilon > 0$, there is an N such that $m \geq N$ implies $d(x_m, x) \leq \varepsilon$ (this differs from Proposition 2.7.2 in that here "$< \varepsilon$" is replaced by "$\leq \varepsilon$").

Suggestion. Consider $d(x_m, x) \leq \varepsilon/2 < \varepsilon$. ♢

Solution. One direction is trivial. If $x_m \to x$, then for each $\varepsilon > 0$, there is an N such that $m \geq N$ implies $d(x_m, x) < \varepsilon \leq \varepsilon$.

For the other direction, suppose that for each $\varepsilon > 0$ there is an N such that $m \geq N$ implies $d(x_m, x) \leq \varepsilon$. Let $\varepsilon_0 > 0$. Then also, $\varepsilon_0/2 > 0$. By hypothesis there is an N such that $m \geq N$ implies $d(x_m, x) \leq \varepsilon_0/2 < \varepsilon_0$. This can be done for each $\varepsilon_0 > 0$, so $x_m \to x$.

We have established the implications in both directions, so $x_m \to x$ iff for every $\varepsilon > 0$, there is an N such that $m \geq N$ implies $d(x_m, x) \leq \varepsilon$ as claimed. ♦

◇ **2E-12.** Prove the following properties for subsets A and B of a metric space:

(a) $\operatorname{int}(\operatorname{int}(A)) = \operatorname{int}(A)$.
(b) $\operatorname{int}(A \cup B) \supset \operatorname{int}(A) \cup \operatorname{int}(B)$.
(c) $\operatorname{int}(A \cap B) = \operatorname{int}(A) \cap \operatorname{int}(B)$.

Solution. (a) For any set C we know that $\operatorname{int}(C) \subseteq C$, so, in particular, $\operatorname{int}(\operatorname{int}(A)) \subseteq \operatorname{int}(A)$. For the other direction, suppose $x \in \operatorname{int}(A)$. Then there is an open set U such that $x \in U \subseteq A$. If $y \in U$, then $y \in U \subseteq A$, so $y \in \operatorname{int}(A)$. Thus $U \subseteq \operatorname{int}(A)$. We have $x \in U \subseteq \operatorname{int}(A)$, so $x \in \operatorname{int}(\operatorname{int}(A))$. This shows that $\operatorname{int}(A) \subseteq \operatorname{int}(\operatorname{int}(A))$. We have inclusion in both directions, so $\operatorname{int}(A) = \operatorname{int}(\operatorname{int}(A))$ as claimed.
Notice that in the middle of the argument just given, we established:

Proposition. *If U is open and $U \subseteq A$, then $U \subseteq \operatorname{int}(A)$.*

This was Exercise 2E-3.

(b) Suppose $x \in \operatorname{int}(A) \cup \operatorname{int}(B)$. Then $x \in \operatorname{int}(A)$ or $x \in \operatorname{int}(B)$.

CASE 1: If $x \in \operatorname{int}(A)$, then there is an open set U with $x \in U \subseteq A \subseteq A \cup B$. So $x \in \operatorname{int}(A \cup B)$.

CASE 2: If $x \in \operatorname{int}(B)$, then there is an open set U with $x \in U \subseteq B \subseteq A \cup B$. So $x \in \operatorname{int}(A \cup B)$.

Since at least one of cases (1) or (2) must hold, we know that $x \in \operatorname{int}(A \cup B)$. This holds for every x in $\operatorname{int}(A) \cup \operatorname{int}(B)$, so $\operatorname{int}(A) \cup \operatorname{int}(B) \subseteq \operatorname{int}(A \cup B)$ as claimed.

The inclusion just established could be proper. Take as subsets of the metric space $\mathbb{R}$ the "punctured" open interval $A =]-1, 1[\setminus \{0\}$, and the one point set $B = \{0\}$. Then A is open, so $\operatorname{int}(A) = A$, and $\operatorname{int}(B) = \emptyset$. So $\operatorname{int}(A) \cup \operatorname{int}(B) = A$. But $A \cup B$ is the unbroken open interval $]-1, 1[$. So $\operatorname{int}(A \cup B) =]-1, 1[$. We see that $0 \in \operatorname{int}(A \cup B) \setminus (\operatorname{int}(A) \cup \operatorname{int}(B))$.

(c) This is the same as Exercise 2.2-4. ♦

◇ **2E-13.** Show that $\operatorname{cl}(A) = A \cup \operatorname{bd}(A)$.

Suggestion. Compare with Exercise 2E-9. ◊

Solution. First of all, $\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A) \subseteq \mathrm{cl}(A)$, and $A \subseteq \mathrm{cl}(A)$. So $A \cup \mathrm{bd}(A) \subseteq \mathrm{cl}(A)$. For the opposite inclusion, suppose that $x \in \mathrm{cl}(A)$. If $x \in A$, then it is certainly in $A \cup \mathrm{bd}(A)$. If x is not in A, then it is in $M \setminus A$ and so in $\mathrm{cl}(M \setminus A)$. Since we have assumed that it is in $\mathrm{cl}(A)$, we have $x \in \mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A) = \mathrm{bd}(A) \subset A \cup \mathrm{bd}(A)$. In either case, $x \in A \cup \mathrm{bd}(A)$. So $\mathrm{cl}(A) \subseteq A \cup \mathrm{bd}(A)$. We have inclusion in both directions, so $\mathrm{cl}(A) = A \cup \mathrm{bd}(A)$ as claimed. ♦

◇ **2E-14.** Prove the following for subsets of a metric space M:

(a) $\mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$.
(b) $\mathrm{cl}(A \cup B) = \mathrm{cl}(A) \cup \mathrm{cl}(B)$.
(c) $\mathrm{cl}(A \cap B) \subset \mathrm{cl}(A) \cap \mathrm{cl}(B)$.

Solution. Let's make explicit a few lemmas about closure.

Lemma. *Suppose A, B, and C are subsets of a metric space M.*

(1) $\mathrm{cl}(A)$ *is a closed set.*
(2) $A \subseteq \mathrm{cl}(A)$.
(3) If C is closed and $A \subseteq C$, then $\mathrm{cl}(A) \subseteq C$.
(4) If $A \subseteq B$, then $\mathrm{cl}(A) \subseteq \mathrm{cl}(B)$.
(5) A is closed if and only if $A = \mathrm{cl}(A)$.

Proof: Since $\mathrm{cl}(A)$ is defined as the intersection of all closed subsets of M which contain A and the intersection of any family of closed sets is closed, (Proposition 2.3.2(ii)), we see that $\mathrm{cl}(A)$ must be a closed set. Since A is contained in each of the sets being intersected, it is contained in the intersection which is $\mathrm{cl}(A)$. For part (3), if C is closed and $A \subseteq C$, then C is one of the sets being intersected to obtain $\mathrm{cl}(A)$. So $\mathrm{cl}(A) \subseteq C$.

For part (4), if $A \subseteq B$, then we know from part (2) that $A \subseteq B \subseteq \mathrm{cl}(B)$, and from part (1) that $\mathrm{cl}(B)$ is a closed set. So $\mathrm{cl}(B)$ is a closed set which contains A. By part (3) we conclude that $\mathrm{cl}(A) \subseteq \mathrm{cl}(B)$ as desired.

Finally, if $A = \mathrm{cl}(A)$, then it is closed by part (1). We always have $A \subseteq \mathrm{cl}(A)$ by part (2). Of course, $A \subseteq A$, so if A is closed, then we also have $\mathrm{cl}(A) \subseteq A$ by part (3). With inclusion in both directions, we conclude that $A = \mathrm{cl}(A)$ as claimed.

(a) If A is any subset of a metric space M, then $\mathrm{cl}(A)$ is a closed set by part (1) of the lemma. So $\mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$ by part (5).
(b) If A and B are subsets of a metric space M, then $A \subseteq A \cup B$ and $B \subseteq A \cup B$. So, by part (4) of the lemma, $\mathrm{cl}(A) \subseteq \mathrm{cl}(A \cup B)$ and $\mathrm{cl}(B) \subseteq \mathrm{cl}(A \cup B)$. So $\mathrm{cl}(A) \cup \mathrm{cl}(B) \subseteq \mathrm{cl}(A \cup B)$. On the other hand, using part (2) gives $A \subseteq \mathrm{cl}(A) \subseteq \mathrm{cl}(A) \cup \mathrm{cl}(B)$. and $B \subseteq \mathrm{cl}(B) \subseteq \mathrm{cl}(A) \cup \mathrm{cl}(B)$.

So $A \cup B \subseteq \text{cl}(A) \cup \text{cl}(B)$. Each of the sets $\text{cl}(A)$ and $\text{cl}(B)$ is closed, so their union is also. Thus $\text{cl}(A) \cup \text{cl}(B)$ is a closed set containing $A \cup B$. By part (3), we have $\text{cl}(A \cup B) \subseteq \text{cl}(A) \cup \text{cl}(B)$. We have inclusion in both directions, so $\text{cl}(A \cup B) = \text{cl}(A) \cup \text{cl}(B)$ as claimed.

(c) If A and B are subsets of a metric space M, then $A \cap B \subseteq A$ and $A \cap B \subseteq B$. By part (3) of the lemma we have $\text{cl}(A \cap B) \subseteq \text{cl}(A)$ and $\text{cl}(A \cap B) \subseteq \text{cl}(B)$. So $\text{cl}(A \cap B) \subseteq \text{cl}(A) \cap \text{cl}(B)$ as claimed.

Note: The inclusion established in part (b) might be proper. Consider, for example, the open intervals $A =]0,1[$ and $B =]1,2[$ as subsets of $\mathbb{R}$. Then $\text{cl}(A \cap B) = \text{cl}(\emptyset) = \emptyset$. But $\text{cl}(A) \cap \text{cl}(B) = [0,1] \cap [1,2] = \{1\}$. ♦

◇ **2E-15.** Prove the following for subsets of a metric space M:

(a) $\text{bd}(A) = \text{bd}(M \setminus A)$.
(b) $\text{bd}(\text{bd}(A)) \subset \text{bd}(A)$.
(c) $\text{bd}(A \cup B) \subset \text{bd}(A) \cup \text{bd}(B) \subset \text{bd}(A \cup B) \cup A \cup B$.
(d) $\text{bd}(\text{bd}(\text{bd}(A))) = \text{bd}(\text{bd}(A))$.

Sketch. (a) Use $M \setminus (M \setminus A) = A$ and the definition of boundary.

(b) Use the fact that $\text{bd}(A)$ is closed. (Why?)

(c) The facts $\text{cl}(A \cup B) = \text{cl}(A) \cup \text{cl}(B)$ and $\text{cl}(A \cap B) \subseteq \text{cl}(A) \cup \text{cl}(B)$ from Exercise 2E-14 are useful.

(d) One approach is to show that $\text{cl}(M \setminus \text{bd}(\text{bd}(A))) = M$ or, equivalently, that $\text{int}(\text{bd}(\text{bd}(A))) = \emptyset$, and use that to compute $\text{bd}(\text{bd}(\text{bd}(A)))$. ◇

Solution. (a) If A is a subset of a metric space M, we can compute

$$\begin{aligned}\text{bd}(M \setminus A) &= \text{cl}(M \setminus A) \cap \text{cl}(M \setminus (M \setminus A)) = \text{cl}(M \setminus A) \cap \text{cl}(A)\\ &= \text{cl}(A) \cap \text{cl}(M \setminus A) = \text{bd}(A)\end{aligned}$$

as claimed.

(b) Since $\text{bd}(A)$ is the intersection of $\text{cl}(A)$ and $\text{cl}(M \setminus A)$, both of which are closed, it is closed. In particular, $\text{cl}(\text{bd}(A)) = \text{bd}(A)$, and we have

$$\text{bd}(\text{bd}(A)) = \text{cl}(\text{bd}(A)) \cap \text{cl}(M \setminus \text{bd}(A)) = \text{bd}(A) \cap \text{cl}(M \setminus \text{bd}(A)).$$

Since $\text{bd}(\text{bd}(A))$ is the intersection of $\text{bd}(A)$ with something else, we have $\text{bd}(\text{bd}(A)) \subseteq \text{bd}(A)$ as claimed.

(c) A key to part (c) are the observations that $\text{cl}(A \cup B) = \text{cl}(A) \cup \text{cl}(B)$ and $\text{cl}(A \cap B) \subseteq \text{cl}(A) \cup \text{cl}(B)$ for any subsets A and B of a metric space

M. (See Exercise 2E-14.) Using them we can compute

$$\begin{aligned}
\mathrm{bd}(A \cup B) &= \mathrm{cl}(A \cup B) \cap \mathrm{cl}(M \setminus (A \cup B)) \\
&= \mathrm{cl}(A \cup B) \cap \mathrm{cl}((M \setminus A) \cap (M \setminus B)) \\
&\subseteq \mathrm{cl}(A \cup B) \cap \mathrm{cl}(M \setminus A) \cap \mathrm{cl}(M \setminus B) \\
&= \big[\mathrm{cl}(A) \cup \mathrm{cl}(B)\big] \cap \mathrm{cl}(M \setminus A) \cap \mathrm{cl}(M \setminus B) \\
&= \big[\mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A) \cap \mathrm{cl}(M \setminus B)\big] \\
&\qquad \cup \big[\mathrm{cl}(B) \cap \mathrm{cl}(M \setminus A) \cap \mathrm{cl}(M \setminus B)\big] \\
&\subseteq \big[\mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A)\big] \cup \big[\mathrm{cl}(B) \cap \mathrm{cl}(M \setminus B)\big] = \mathrm{bd}(A) \cup \mathrm{bd}(B).
\end{aligned}$$

This is the first inclusion claimed. Now suppose $x \in \mathrm{bd}(A) \cup \mathrm{bd}(B)$. Then $x \in \mathrm{cl}(A)$ or $x \in \mathrm{cl}(B)$. So $x \in \mathrm{cl}(A) \cup \mathrm{cl}(B) = \mathrm{cl}(A \cup B)$. If x is in neither A nor B, then $x \in (M \setminus A) \cap (M \setminus B) = M \setminus (A \cup B) \subseteq \mathrm{cl}((M \setminus (A \cup B))$. Since we also have $x \in \mathrm{cl}(A \cup B)$, this puts x in $\mathrm{bd}(A \cup B)$. So x must be in at least one of the three sets A, B, or $\mathrm{bd}(A \cup B)$. That is, $\mathrm{bd}(A) \cup \mathrm{bd}(B) \subseteq \mathrm{bd}(A \cup B) \cup A \cup B$ as claimed.

(d) From part (a) we know that $\mathrm{bd}(\mathrm{bd}(C)) \subseteq \mathrm{bd}(C)$ for every set C. So, in particular, $\mathrm{bd}(\mathrm{bd}(\mathrm{bd}(A))) \subseteq \mathrm{bd}(\mathrm{bd}(A))$. But now we want equality:

$$\begin{aligned}
\mathrm{bd}(\mathrm{bd}(\mathrm{bd}(A))) &= \mathrm{cl}(\mathrm{bd}(\mathrm{bd}(A))) \cap \mathrm{cl}(M \setminus \mathrm{bd}(\mathrm{bd}(A))) \\
&= \mathrm{bd}(\mathrm{bd}(A)) \cap \mathrm{cl}(M \setminus \mathrm{bd}(\mathrm{bd}(A))).
\end{aligned}$$

But

$$\begin{aligned}
\mathrm{cl}(M \setminus \mathrm{bd}(\mathrm{bd}(A))) &= \mathrm{cl}(M \setminus (\mathrm{cl}(\mathrm{bd}(A)) \cap \mathrm{cl}(M \setminus \mathrm{bd}(A)))) \\
&= \mathrm{cl}(M \setminus (\mathrm{bd}(A) \cap \mathrm{cl}(M \setminus \mathrm{bd}(A)))) \\
&= \mathrm{cl}((M \setminus \mathrm{bd}(A)) \cup (M \setminus \mathrm{cl}(M \setminus \mathrm{bd}(A)))) \\
&= \mathrm{cl}(M \setminus \mathrm{bd}(A)) \cup \mathrm{cl}(M \setminus \mathrm{cl}(M \setminus \mathrm{bd}(A))) \\
&\subseteq \mathrm{cl}(M \setminus \mathrm{bd}(A)) \cup \mathrm{cl}(M \setminus (M \setminus \mathrm{bd}(A))) \\
&\subseteq (M \setminus \mathrm{bd}(A)) \cup (M \setminus (M \setminus \mathrm{bd}(A))) = M.
\end{aligned}$$

Combining the last two displays gives

$$\mathrm{bd}(\mathrm{bd}(\mathrm{bd}(A))) = \mathrm{bd}(\mathrm{bd}(A)) \cap M = \mathrm{bd}(\mathrm{bd}(A))$$

as claimed.

Remark: The identity in the next to last display is equivalent to the assertion that

$$\mathrm{int}(\mathrm{bd}(\mathrm{bd}(A))) = \emptyset. \qquad \blacklozenge$$

◇ **2E-16.** Let $a_1 = \sqrt{2}$, $a_2 = (\sqrt{2}\ )^{a_1}, \ldots, a_{n+1} = (\sqrt{2}\ )^{a_n}$. Show that $a_n \to 2$ as $n \to \infty$. (You may use any relevant facts from calculus.)

Suggestion. Show the sequence is increasing and bounded above by 2. Conclude that the limit λ exists and is a solution to the equation $\lambda = \left(\sqrt{2}\right)^{\lambda}$. Show $\lambda = 2$ and $\lambda = 4$ are the only solutions of this equation. Conclude that the limit is 2. ♢

Solution. First: $a_1 > 1$, and if $a_k > 1$, then $a_{k+1} = \left(\sqrt{2}\right)^{a_k} > \sqrt{2} > 1$ also. By induction we have

$$a_n > 1 \qquad \text{for every } n = 1, 2, 3, \ldots .$$

Second: The sequence is increasing: $a_1 = \sqrt{2}$ and $a_2 = \left(\sqrt{2}\right)^{\sqrt{2}} > \sqrt{2} = a_1$. If $a_{k-1} < a_k$, we can compute

$$\frac{a_{k+1}}{a_k} = \frac{\left(\sqrt{2}\right)^{a_k}}{\left(\sqrt{2}\right)^{a_{k-1}}} = \left(\sqrt{2}\right)^{a_k - a_{k-1}} > 1.$$

So $a_{k+1} > a_k$. By induction we find that $a_{n+1} > a_n$ for each n so the sequence is increasing.

Third: The sequence is bounded above by 2: $a_1 = \sqrt{2} < 2$. If $a_k < 2$, then $a_{k+1} = \left(\sqrt{2}\right)^{a_k} < \left(\sqrt{2}\right)^2 = 2$ also. By induction we conclude that $a_n < 2$ for every n.

Fourth: Since the sequence $\langle a_n \rangle_1^\infty$ is increasing and bounded above by 2, the completeness of $\mathbb{R}$ implies that it must converge to some $\lambda \in \mathbb{R}$.

Fifth: $a_{n+1} \to \lambda$. But also $a_{n+1} = \left(\sqrt{2}\right)^{a_n} \to \left(\sqrt{2}\right)^{\lambda}$. Since limits in $\mathbb{R}$ are unique we must have

$$\lambda = \left(\sqrt{2}\right)^{\lambda} \qquad \text{or equivalently} \qquad \log \lambda = \frac{\log 2}{2}\, \lambda.$$

These equations have the two solutions $\lambda = 2$ and $\lambda = 4$ as may be checked directly. There can be no other solutions since the solutions occur at the intersection of a straight line with an exponential curve (or a logarithm curve in the second equation). The exponential curve is always concave up and the logarithm curve is always concave down. Either can cross a straight line at most twice. Thus the limit must be either 2 or 4. But all terms of the sequence are smaller than 2, so the limit cannot be 4. It must be 2. $\lim_{n\to\infty} a_n = 2$. ♦

◇ **2E-17.** If $\sum x_m$ converges absolutely in $\mathbb{R}^n$, show that $\sum x_m \sin m$ converges.

Sketch. The series $\sum_m \|\, (\sin m) x_m \,\|$ converges by comparison to the series $\sum_m \|\, x_m \,\|$. (Why?) So $\sum_m (\sin m)\, x_m$ converges by Theorem 2.9.3. ♢

Solution. We know that $|\sin t| \le 1$ for every real number t. So

$$0 \le \| (\sin m)x_m \| = |(\sin m)| \, \| x_m \| \le \| x_m \|$$

for $m = 1, 2, 3, \ldots$.

We have assumed that $\sum x_m$ is absolutely convergent in $\mathbb{R}^n$ which simply means that $\sum \| x_m \|$ converges in $\mathbb{R}$. The inequality above then implies by comparison (Proposition 2.9.4(ii)) that $\sum_m \| (\sin m)x_m \|$ also converges in $\mathbb{R}$. That is, $\sum_m (\sin m)x_m$ is absolutely convergent in $\mathbb{R}^n$. But $\mathbb{R}^n$ is complete (Theorem 2.8.5). So $\sum_m (\sin m)x_m$ converges in $\mathbb{R}^n$ by Theorem 2.9.3. ♦

◇ **2E-18.** If $x, y \in M$ and $x \neq y$, then prove that there exist open sets U and V such that $x \in U$, $y \in V$, and $U \cap V = \varnothing$.

Suggestion. Use disks of radius no larger than half the distance between x and y. ◇

Solution. Since x and y are not equal, we know that $d(x, y) > 0$. Let $r = d(x, y)/3$, and set $U = D(x, r)$ and $V = D(y, r)$. Then U and V are open by Proposition 2.1.2, and we certainly have $x \in U$ and $y \in V$. If z were in $D(x, r) \cap D(y, r)$, we would have

$$0 < r = d(x, y) \le d(x, z) + d(z, y) < \frac{r}{3} + \frac{r}{3} = \frac{2}{3}\, r.$$

But this is impossible. A positive number cannot be smaller than 2/3 of itself. Thus there can be no such z, and we must have $U \cap V = D(x, r) \cap D(y, r) = \emptyset$ as desired.

The property proved here is called the Hausdorff separation property. ♦

◇ **2E-19.** Define a ***limit point*** of a set A in a metric space M to be a point $x \in M$ such that $U \cap A \neq \varnothing$ for every neighborhood U of x.

(a) What is the difference between limit points and accumulation points? Give examples.

(b) If x is a limit point of A, then show that there is a sequence $x_n \in A$ with $x_n \to x$.

(c) If x is an accumulation point of A, then show that x is a limit point of A. Is the converse true?

(d) If x is a limit point of A and $x \notin A$, then show that x is an accumulation point.

(e) Prove: A set is closed iff it contains all of its limit points.

Sketch. A limit point either belongs to A or is an accumulation point of A. It might be an isolated point of A; an accumulation point is not. $\Diamond$

Solution. (a) The difference in definitions is at the very end. For x to be a limit point of A, every neighborhood U of x must intersect A. There must be at least one point in the intersection, but that point is allowed to be x. For an accumulation point, there must be something in the intersection which is not equal to x. Thus if x itself is actually in A it is automatically a limit point since $x \in U \cap A$. However, x might not be an accumulation point even if it is in A since it might be the only point in the intersection. This may be the situation if x is what is called an isolated point of A, that is, a point $x \in A$ such that there is an open set U with $U \cap A = \{x\}$. Consider for example $A = [0,1[\cup\{2\} \subseteq \mathbb{R}$. 2 is in A, so it is a limit point, but it is not an accumulation point. $U =]1.5, 2.5[$ is an open set containing 2 but no other points of A. It is an isolated point of A.

(b) If x is a limit point of the subset A of a metric space M, then for each integer $n > 0$, the disk $D(x, 1/n)$ is an open set containing x, so it must have nonempty intersection with A. There is a point x_n in A with $d(x_n, x) < 1/n$. The sequence $\langle x_n \rangle_1^\infty$ is thus a sequence in A converging to x.

(c) An accumulation point must certainly be a limit point since $U \cap (A \setminus \{x\}) \subseteq U \setminus A$. If x is an accumulation point of A, and U is an open set containing x, then the first must be nonempty, so the second, larger set must be also. This says that x is a limit point. The converse is not true. A limit point need not be an accumulation point since it might be the only point in the larger intersection. This was the case in the example given in part a.

(d) Suppose x is an limit point of A and x is not in A. Let U be an open set containing x. Since x is a limit point of A, the intersection $U \cap A$ is not empty. There is a $y \in U \cap A$. But, since x is not in A, y cannot be equal to x. Thus $y \in U \cap (A \setminus \{x\})$. This can be done for every neighborhood U of x, so x is an accumulation point of A.

(e) If A is a closed set in M, we know from Proposition 2.7.6(i) that A contains the limit of every convergent sequence in A. More precisely, $x_n \in A$ for every n and $x_n \to x \in M$ implies that $x \in M$. We know from part (b) that if x is a limit point of A, then there is such a sequence in A converging to x. So we must have $x \in A$. Thus if A is closed then it must contain all of its limit points. Conversely, suppose A is a subset of M which contains all of its limit points. We know that a set is closed if it contains all its accumulation points. This follows from Proposition 2.5.2. If A contains all its accumulation points, then $\mathrm{cl}(A) = A \cup \{$ accumulation points of $A\} = A$, so A is equal to its closure and is closed. But by part (c) every accumulation point is a limit point.

If A contains all its limit points, then it certainly contains all of its accumulation points and so is closed. ♦

◇ **2E-20.** For a set A in a metric space M and $x \in M$, let

$$d(x, A) = \inf\{d(x, y) \mid y \in A\},$$

and for $\varepsilon > 0$, let $D(A, \varepsilon) = \{x \mid d(x, A) < \varepsilon\}$.

(a) Show that $D(A, \varepsilon)$ is open.
(b) Let $A \subset M$ and $N_\varepsilon = \{x \in M \mid d(x, A) \leq \varepsilon\}$, where $\varepsilon > 0$. Show that N_ε is closed and that A is closed iff $A = \cap \{N_\varepsilon \mid \varepsilon > 0\}$.

Suggestion. For (a), show that $D(A, \epsilon)$ is a union of open disks. For the first part of (b), consider convergent sequences in $N(A, \varepsilon)$ and their limits in M. ◇

Solution. (a) First suppose $x \in D(A, \varepsilon)$. Then $d(x, A) = \inf\{d(x, y) \mid y \in A\} = r < \varepsilon$. So there is a point $y \in A$ with $r \leq d(x, y) < \varepsilon$. Thus $x \in D(y, \varepsilon)$. This can be done for each $x \in D(A, \varepsilon)$. We conclude that

$$D(A, \varepsilon) \subseteq \bigcup_{y \in A} D(y, \varepsilon).$$

Conversely, if there is a y in A with $x \in D(y, \varepsilon)$, then $d(x, y) < \varepsilon$. So $d(x, A) = \inf\{d(x, y) \mid y \in A\} < \varepsilon$, and $x \in D(A, \varepsilon)$. This proves inclusion in the other direction. We conclude that

$$D(A, \varepsilon) = \bigcup_{y \in A} D(y, \varepsilon).$$

Each of the disks $D(y, \varepsilon)$ is open by Proposition 2.1.2, so their union, $D(A, \varepsilon)$ is also open by Proposition 2.1.3(ii).

(b) To show that $N(A, \varepsilon)$ is closed we will show that it contains the limits of all convergent sequences in it. Suppose $\langle x_k \rangle_1^\infty$ is a sequence in $N(A, \varepsilon)$ and that $x_k \to x \in M$. Since each $x_k \in N(A, \varepsilon)$, we have $d(x_k, A) \leq \varepsilon < \varepsilon + (1/k)$. So there are points y_k in A with $d(x_k, y_k) < \varepsilon + (1/k)$. Since $x_k \to x$, we know that $d(x, x_k) \to 0$, and can compute

$$d(x, y_k) \leq d(x, x_k) + d(x_k, y_k) < d(x, x_k) + \varepsilon + \frac{1}{k} \to \varepsilon \quad \text{as } k \to \infty.$$

Thus

$$d(x, A) = \inf\{d(x, y) \mid y \in A\} \leq \varepsilon.$$

So $x \in N(A, \varepsilon)$. We have shown that if $\langle x_k \rangle_1^\infty$ is a sequence in $N(A, \varepsilon)$ and $x_k \to x \in M$, then $x \in N(A, \varepsilon)$. So $N(A, \varepsilon)$ is a closed subset of M by Proposition 2.7.6(i).

We have just shown that each of the sets $N(A, \varepsilon)$ is closed, and we know that the intersection of any family of closed subsets of M is closed. So if $A = \bigcap_{\varepsilon > 0} N(A, \varepsilon)$, then A is closed.

Conversely, if A is closed and $y \in M \setminus A$, then there is an $r > 0$ such that $D(y, r) \subset M \setminus A$ since the latter set is open. Thus y is not in $N(A, r/2)$. So y is not in $\bigcap_{\varepsilon>0} N(A, \varepsilon)$. This establishes the opposite inclusion $\bigcap_{\varepsilon>0} N(A, \varepsilon) \subseteq A$.

If A is closed, we have inclusion in both directions, so $A = \bigcap_{\varepsilon>0} N(A, \varepsilon)$ as claimed. ♦

◇ **2E-21.** Prove that a sequence x_k in a normed space is a Cauchy sequence iff for every neighborhood U of 0, there is an N such that $k, l \geq N$ implies $x_k - x_l \in U$.

Suggestion. Each open neighborhood U of 0 contains a disk $D(0, \varepsilon)$. ◇

Solution. Let $\langle x_n \rangle_1^\infty$ be a sequence in a normed space $\mathcal{V}$.

If $\langle x_n \rangle_1^\infty$ is a Cauchy sequence and U is a neighborhood of 0 in $\mathcal{V}$, then there is an $\varepsilon > 0$ such that $D(0, \varepsilon) \subseteq U$. Since the sequence is a Cauchy sequence, there is an index N such that $\| x_k - x_j \| < \varepsilon$ whenever $k \geq N$ and $j \geq N$. Thus for such k and j we have $x_k - x_j \in D(0, \varepsilon) \subseteq U$.

For the converse, suppose that for every neighborhood U of 0, there is an N such that $k, j \geq N$ implies $x_k - x_j \in U$, and let $\varepsilon > 0$. Take $U = D(0, \varepsilon)$. We know that U is open by Proposition 2.1.2, so by hypothesis there is an N such that $k, j \geq N$ implies $x_k - x_j \in U = D(0, \varepsilon)$. For such k and j we have $\| x_k - x_j \| < \varepsilon$. So the sequence is a Cauchy sequence. ♦

◇ **2E-22.** Prove Proposition 2.3.2. (Hint: Use Exercise 0E-12 of the Introduction.)

Suggestion. Take complements in Proposition 2.1.3. Use De Morgan's Laws and the fact that a set is closed if and only if its complement is open. ◇

Solution. Proposition 2.3.2 is the basic theorem about closed sets:

Proposition. *In a metric space* (M, d),

(1) The union of a finite number of closed sets is closed.

(2) The intersection of an arbitrary family of closed sets is closed.

(3) The sets M and $\emptyset$ are closed.

Each piece of this theorem corresponds to a piece of the similar theorem about open sets, Proposition 2.1.3, and is proved from it by taking complements and using the fact that a set is closed if and only if its complement is open.

Proposition. *In a metric space (M, d),*

(1) The intersection of a finite number of open sets is open.
(2) The union of an arbitrary family of open sets is open.
(3) The sets M and $\emptyset$ are open.

For (1), suppose $F_1, F_2, \ldots, F_n$ are closed subsets of M. For each k, let $U_k = M \setminus F_k$. Then each U_k is an open subset of M, and

$$M \setminus \bigcup_{k=1}^{n} F_k = \bigcap_{k=1}^{n} (M \setminus F_k) = \bigcap_{k=1}^{n} U_k.$$

This set is open by part (1) of Proposition 2.1.2. Since its complement is open, the union $\bigcup_{k=1}^{n} F_k$ is closed as claimed.

For (2), suppose $\{F_\alpha \mid \alpha \in A\}$ is any family of closed subsets of M. Then

$$M \setminus \bigcap_{\alpha \in A} F\alpha = \bigcup_{\alpha \in A} (M \setminus F_\alpha) = \bigcup_{\alpha \in A} U_\alpha.$$

This set is open by part (2) of 2.1.2. Since its complement is open, the intersection $\bigcap_{\alpha \in A} F_\alpha$ is closed as claimed.

For part (3), note that $M \setminus M = \emptyset$ and $M \setminus \emptyset = M$ are open by part (3) of 2.1.3. So their complements, M and $\emptyset$, are closed as claimed. ♦

◇ **2E-23.** Prove that the interior of a set $A \subset M$ is the union of all the subsets of A that are open. Deduce that A is open iff $A = \mathrm{int}(A)$. Also, give a direct proof of the latter statement using the definitions.

Sketch. If V is the union of all open subsets of A, then V is open and contained in A. So $V \subseteq \mathrm{int}(A)$. (See Exercise 2E-3.) If $x \in \mathrm{int}(A)$, there is an open U with $x \in U \subseteq A$. So $x \in V$. ◇

Solution. Let V be the union of all open sets U with $U \subseteq A$. Then V is open since it is a union of open sets (Proposition 2.1.3(ii)), and $V \subseteq A$ since each $U \subseteq A$. If $x \in V$, then there is an open U with $x \in U \subseteq A$, so $x \in \mathrm{int}(A)$. Thus $V \subseteq \mathrm{int}(A)$. On the other hand, if $x \in \mathrm{int}(A)$, then there is an open U with $x \in U \subseteq A$. The set U is one of those whose union is V.

So $x \in U \subseteq V$, and $x \in V$. Thus $\text{int}(A) \subseteq V$. We have inclusion in both directions, so $\text{int}(A) = V$ as claimed.

If $A = \text{int}(A)$, then A is open since we have just shown that the interior of A is a union of open sets and so is open. If A is open, then it is one of the sets whose union was shown to be equal to $\text{int}(A)$. So $A \subseteq \text{int}(A)$. On the other hand, we always have $\text{int}(A) \subseteq A$. We have inclusion in both directions, so $A = \text{int}(A)$. Thus A is open if and only if $A = \text{int}(A)$ as claimed.

Proof: Suppose A is open. If $x \in$, let $U = A$. Then U is an open set with $x \in U \subseteq A$. So x is an interior point of A and $x \in \text{int}(A)$. Thus $A \subseteq \text{int}(A)$. We always have $\text{int}(A) \subseteq A$, so $\text{int}(A) = A$. For the converse, suppose $A = \text{int}(A)$ and that $x \in A$. Then there is an open set U with $x \in U \subseteq A$. Since U is open, there is an $r > 0$ with $x \in D(x, r) \subseteq U \subseteq A$. Thus A contains a disk around each of its points so A is open. We have proved both implications, so A is open if and only if $A = \text{int}(A)$ as claimed. ♦

◇ **2E-24.** Identify $\mathbb{R}^{n+m}$ with $\mathbb{R}^n \times \mathbb{R}^m$. Show that $A \subset \mathbb{R}^{n+m}$ is open iff for each $(x, y) \in A$, with $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, there exist open sets $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ with $x \in U$, $y \in V$ such that $U \times V \subset A$. Deduce that the product of open sets is open.

Suggestion. Try drawing the picture in $\mathbb{R}^2$ and writing the norms out in coordinates in $\mathbb{R}^{n+m}$, $\mathbb{R}^n$, and $\mathbb{R}^m$. ◊

Solution. The spaces $\mathbb{R}^{n+m}$ and $\mathbb{R}^n \times \mathbb{R}^m$ are identified as sets by identifying the element $(x, y) = ((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_m))$ of $\mathbb{R}^n \times \mathbb{R}^m$ with the point
$(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m)$ of $\mathbb{R}^{n+m}$. With this identification, we have

$$\| (x, y) \|^2 = x_1^2 + \cdots + x_n^2 + y_1^2 + \cdots + y_m^2 = \| x \|^2 + \| y \|^2 . \qquad (1)$$

Using this we see that if x and s are in $\mathbb{R}^n$ and y and t are in $\mathbb{R}^m$, then

$$\| (s, t) - (x, y) \|^2 = \| s - x \|^2 + \| t - y \|^2 \qquad (2)$$

and this certainly implies that

$$\| s - x \|^2 \le \| (s, t) - (x, y) \|^2 \quad \text{and} \quad \| t - y \|^2 \le \| (s, t) - (x, y) \|^2$$

so that

$$\| s - x \| \le \| (s, t) - (x, y) \| \quad \text{and} \quad \| t - y \| \le \| (s, t) - (x, y) \| . \qquad (3)$$

Now Let $A \subseteq \mathbb{R}^{n+m}$, and let $(x, y) \in A$ with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. See Figure 2-14.

Figure 2-14. An open set in $\mathbb{R}^{n+m}$.

First suppose that A is open. Then there is an $r > 0$ such that $D((x, y), r) \subseteq A$. Let $\rho = \sqrt{r/2}$, and set $U = D(x, \rho) \subseteq \mathbb{R}^n$ and $V = D(y, \rho) \subseteq \mathbb{R}^m$. We want to show that $U \times V \subseteq D((x, y), r) \subseteq A$. Since $x \in D(x, \rho) = U$, and $y \in D(y, \rho) = V$, we certainly have $(x, y) \in U \times V$. If $(s, t) \in U \times V$, then using (2) we have $\| (s, t) - (x, y) \|^2 = \| s - x \|^2 + \| t - y \|^2 < 2\rho^2 = r$. So $(s, t) \in D((x, y), r) \subseteq A$. Thus $U \times V \subseteq D((x, y), r) \subseteq A$ as we wanted.

For the converse, suppose that there are open sets $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ with $(x, y) \in U \times V \subseteq A$. Since U and V are open in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, there are radii $\rho_1 > 0$ and $\rho_2 > 0$ such that $D(x, \rho_1) \subseteq U$ and $D(y, \rho_2) \subseteq V$. So

$$(x, y) \in D(x, \rho_1) \times D(y, \rho_2) \subseteq U \times V \subseteq A.$$

Let $r = \min(\rho_1, \rho_2)$, and suppose that $(s, t) \in D((x, y), r)$, then by (3) we have $\| s - x \| \le \| (s, t) - (x, y) \| < r \le \rho_1$. So $s \in D(x, \rho_1) \subseteq U$. Also $\| t - y \| \le \| (s, t) - (x, y) \| < r \le \rho_2$. So $t \in D(y, \rho_2) \subseteq V$. Thus $(s, t) \in D(x, \rho_1) \times D(y, \rho_2) \subseteq U \times V$. This is true for every such (s, t). So

$$(x, y) \in D((x, y), r) \subseteq D(x, \rho_1) \times D(y, \rho_2) \subseteq U \times V \subseteq A.$$

The set A contains an open disk around each of its points, so A is open.

We have proved both implications as required.

This argument is a bit like that in Exercise 2.1-6 in which we showed that the taxicab metric and the Euclidean metric produce the same open sets in $\mathbb{R}^2$. The reason is partially that a taxicab "disk" is an open rectangle in $\mathbb{R}^2$ which is the product of two open intervals which are open "disks" in $\mathbb{R}^1$. ♦

◇ **2E-25.** Prove that a set $A \subset M$ is open iff we can write A as the union of some family of ε-disks.

Sketch. Since ε–disks are open, so is any union of them. Conversely, if A is open and $x \in A$, there is an $\varepsilon_x > 0$ with $x \in D(x, \varepsilon_x) \subseteq A$. So $A = \bigcup_{x \in A} D(x, \varepsilon_x)$. ◊

Solution. To say that A is a union of ε–disks is to say that there is a set of points $\{x_\beta \mid \beta \in B\} \subseteq A$ and a set of positive radii $\{r_\beta \mid \beta \in B\}$ such that $A = \bigcup_{\beta \in B} D(x_\beta, r_\beta)$. ($B$ is just any convenient index set for listing these things.) We know from Proposition 2.1.2 that each of the disks $D(x_\beta, r_\beta)$ is open. By 2.1.3(ii), the union of any family of open subsets of M is open. So A must be open.

For the converse, suppose A is an open subset of A. Then for each x in A, there is a radius $r_x > 0$ such that $D(x, r_x) \subseteq A$. Since $x \in D(x, r_x) \subseteq A$, we have

$$A = \bigcup_{x \in A} \{x\} \subseteq \bigcup_{x \in A} D(x, r_x) \subseteq A.$$

So we must have $A = \bigcup_{x \in A} D(x, r_x)$, a union of open disks, as required. ♦

◇ **2E-26.** Define the sequence of numbers a_n by

$$a_0 = 1,\ a_1 = 1 + \frac{1}{1 + a_0},\ \ldots,\ a_n = 1 + \frac{1}{1 + a_{n-1}}.$$

Show that a_n is a convergent sequence. Find the limit.

Suggestion. Compute the first few terms to see what is going on. The sequence is not monotone, but the subsequence consisting of the even index terms is, and so is the one consisting of the odd index terms. ◊

Solution. It helps to compute the first few terms to see what is going on.

$$\begin{aligned}
a_0 &= 1 \\
a_1 &= 1 + \frac{1}{1+1} = \frac{3}{2} = 1.5 \\
a_2 &= 1 + \frac{1}{1 + (3/2)} = \frac{7}{5} = 1.4 \\
a_3 &= 1 + \frac{1}{1 + (7/5)} = \frac{17}{12} \approx 1.4167 \\
a_4 &= 1 + \frac{1}{1 + (17/12)} = \frac{41}{29} \approx 1.4137931 \\
&\vdots
\end{aligned}$$

It appears that the even index terms are increasing while the odd index terms are decreasing. They also appear to be converging to the square root of 2. We show that this is in fact the case.

FIRST: $1 \le a_n < 2$ for $n = 0, 1, 2, \ldots$.

This is an easy induction. $a_0 = 1$, and if $1 \le a_k < 2$, then $a_{k+1} = 1 + \frac{1}{1+a_k}$ is also.

SECOND: We compute a_{n+2} in terms of n.

$$a_{n+2} = 1 + \frac{1}{1 + a_{n+1}} = 1 + \frac{1}{1 + 1 + \frac{1}{1+a_n}} = \frac{4 + 3a_n}{3 + 2a_n}.$$

THIRD: We use this to show that the even index terms stay on one side of $\sqrt{2}$, as do the odd index terms:

$$a_{n+2}^2 - 2 = \left(\frac{4 + 3a_n}{3 + 2a_n}\right)^2 - 2 = \frac{a_n^2 - 2}{(3 + 2a_n)^2}.$$

So the sign of $a_{n+2}^2 - 2$ is the same as that of $a_n^2 - 2$. Combining this with our first observation and the fact that $a_0 = 1 < \sqrt{2}$ and $a_1 = 1.5 > \sqrt{2}$, we conclude

$$1 \le a_n \le \sqrt{2} \qquad \text{for every even } n$$

and

$$\sqrt{2} \le a_n < 2 \qquad \text{for every odd } n.$$

FOURTH: We establish the monotonicity of the subsequences:

$$a_{n+2} - a_n = \frac{4 + 3a_n}{3 + 2a_n} - a_n = \frac{2(2 - a_n^2)}{3 + 2a_n}.$$

So $a_{n+2} > a_n$ if $a_n^2 < 2$, and $a_{n+2} < a_n$ if $a_n^2 > 2$. Combined with observation THREE, this gives

$$\begin{aligned} 1 = a_0 < a_2 < a_4 < a_6 < \cdots < a_{2k} < \cdots < \sqrt{2} \\ 1.5 = a_1 > a_3 > a_5 > a_7 > \cdots > a_{2k+1} > \sqrt{2}. \end{aligned}$$

From this and from the completeness of $\mathbb{R}$, we know that the limits $\mu = \lim_{k\to\infty} a_{2k}$ and $\nu = \lim_{k\to\infty} a_{2k+1}$ exist in $\mathbb{R}$. Now that we know they exist, we can compute them:

$$a_{2(k+1)} \to \mu \quad \text{but} \quad a_{2(k+1)} = a_{2k+2)} = \frac{4 + 3a_{2k}}{3 + 2a_{2k}} \to \frac{4 + 3\mu}{3 + 2\mu}$$

and

$$a_{2(k+1)+1} \to \nu \quad \text{but} \quad a_{2(k+1)+1} = a_{(2k+1)+2)} = \frac{4 + 3a_{2k+1}}{3 + 2a_{2k+1}} \to \frac{4 + 3\nu}{3 + 2\nu}.$$

So both μ and ν satisfy the equation

$$\begin{aligned}
x &= \frac{4+3x}{3+2x} \\
3x + 2x^2 &= 4 + 3x \\
x^2 &= 2 \\
x &= \pm\sqrt{2}.
\end{aligned}$$

Since all of the a_n are positive, both μ and ν must be $\sqrt{2}$. Since both these subsequences converge to $\sqrt{2}$, the only possible cluster point of the sequence is $\sqrt{2}$. The sequence must converge, and its limit is $\sqrt{2}$.

Remark: What we have studied here is called the *continued fraction* expansion of $\sqrt{2}$. To see why, write the terms of the sequence in a somewhat different way.

$$a_0 = 1 \qquad a_1 = 1 + \frac{1}{1+1}$$

$$a_2 = 1 + \frac{1}{1+a_1} = 1 + \frac{1}{1+1+\frac{1}{1+1}} = 1 + \frac{1}{2+\frac{1}{2}}$$

$$a_3 = 1 + \frac{1}{1+a_2} = 1 + \frac{1}{1+1+\frac{1}{2+\frac{1}{2}}} = 1 + \frac{1}{2+\frac{1}{2+\frac{1}{2}}}$$

$$a_4 = 1 + \frac{1}{1+a_3} = 1 + \frac{1}{1+1+\frac{1}{2+\frac{1}{2+\frac{1}{2}}}} = 1 + \frac{1}{2+\frac{1}{2+\frac{1}{2+\frac{1}{2}}}}$$

This process continues, and the convergence to $\sqrt{2}$ can be written, more or less, as

$$\sqrt{2} = 1 + \cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{\ddots}}}}}}}.$$

This sort of thing is called a *continued fraction*. They have a lot of interesting properties and uses in analysis and number theory some of which can be found in the book *Continued Fractions* by C. D. Olds published as volume 9 in the New Mathematical Library series by the Mathematical Association of America. ♦

◇ **2E-27.** Suppose $a_n \geq 0$ and $a_n \to 0$ as $n \to \infty$. Given any $\varepsilon > 0$, show that there is a subsequence b_n of a_n such that $\sum_{n=1}^{\infty} b_n < \varepsilon$.

Suggestion. Pick b_n with $b_n = |b_n| < \varepsilon/2^n$. (Why can you do this?) ◇

Solution. Select the subsequence inductively.

STEP ONE: Since $a_n \geq 0$ and $a_n \to 0$, there is an index $n(1)$ such that $0 \leq a_{n(1)} < \varepsilon/2$.

STEP TWO: Since $a_n \geq 0$ and $a_n \to 0$, there is an index $n(2)$ such that $n(2) > n(1)$ and $0 \leq a_{n(2)} < \varepsilon/4$.

STEP THREE: Since $a_n \geq 0$ and $a_n \to 0$, there is an index $n(3)$ such that $n(3) > n(2)$ and $0 \leq a_{n(3)} < \varepsilon/8$.

$$\vdots$$

STEP k+1: Having selected indices $n(1) < n(2) < \cdots < n(k)$ such that $0 \leq a_{n(j)} < \varepsilon/2^j$ for $j = 1, 2, \ldots, k$, we observe that since $a_n \geq 0$ and $a_n \to 0$, there is an index $n(k+1)$ such that $n(k+1) > n(k)$ and $0 \leq a_{n(k+1)} < \varepsilon/2^{k+1}$.

$$\vdots$$

This process inductively generates a subsequence with indices $n(1) < n(2) < n(3) < \cdots$ such that $0 \leq a_{n(k)} < \varepsilon/2^k$ for each k. So $0 \leq \sum_{k=1}^{\infty} a_{n(k)} < \sum_{k=1}^{\infty} \varepsilon/2^k = \varepsilon \sum_{k=1}^{\infty}(1/2^k) = \varepsilon$, as we wanted. ♦

◇ **2E-28.** Give examples of:

(a) An infinite set in $\mathbb{R}$ with no accumulation points
(b) A nonempty subset of $\mathbb{R}$ that is contained in its set of accumulation points
(c) A subset of $\mathbb{R}$ that has infinitely many accumulation points but contains none of them
(d) A set A such that $\mathrm{bd}(A) = \mathrm{cl}(A)$.

Solution. Here are some possible examples.

(a) We need an infinite set all of whose points are isolated. $\mathbb{Z}$ is a typical example of an infinite subset of $\mathbb{R}$ with no accumulation points.
(b) The closure of a set is equal to the union of the set with its accumulation points. So if the set is to be contained in its set of accumulation points, the closure must be equal to the set of accumulation points. Here are some examples:
If $A = \mathbb{R}$, then $A' = \mathbb{R}$, so $A = A'$.
If $A = \mathbb{Q}$, then $A' = \mathbb{R}$, so $A \subseteq A'$.
If $A = \mathbb{R} \setminus \mathbb{Q}$, then $A' = \mathbb{R}$, so $A \subseteq A'$.
If $A = [0,1] \cap \mathbb{Q}$, then $A' = [0,1]$, so $A \subseteq A'$.

(c) Idea: 0 is the only accumulation point of the set $\{1/2, 1/3, 1/4, \ldots\}$, and is not in this set. Put together infinitely many copies of this effect. If $A = \{n + (1/m) \mid n \in \mathbb{Z}, m = 2, 3, 4, \ldots\}$, then $A' = \mathbb{Z}$ and $A \cap \mathbb{Z} = \emptyset$.

(d) Consider $A = [0, 1] \cap \mathbb{Q} \subseteq \mathbb{R}$ again. Then $\mathrm{cl}(A) = [0, 1]$. But $\mathrm{cl}(\mathbb{R} \setminus A) = \mathbb{R}$. So $\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(\mathbb{R} \setminus A) = \mathrm{cl}(A) = [0, 1]$. ⧫

◇ **2E-29.** Let $A, B \subset \mathbb{R}^n$ and x be an accumulation point of $A \cup B$. Must x be an accumulation point of either A or B?

Sketch. Yes. If x is not an accumulation point of A or of B, then there are small disks centered at x which intersect A and B respectively in at most the point x. The smaller of these two disks intersects $A \cup B$ in at most the point x, so x is not an accumulation point of $A \cup B$. ◊

Solution. Let x be an accumulation point of $A \cup B$. Suppose x were not an accumulation point of either A or B. Since it is not an accumulation point of A, there is a radius $r_1 > 0$ such that $D(x, r_1) \cap (A \setminus \{x\}) = \emptyset$. (Equivalently: $D(x, r_1) \cap A \subseteq \{x\}$.) Since it is not an accumulation point of B, there is a radius $r_2 > 0$ such that $D(x, r_2) \cap (B \setminus \{x\}) = \emptyset$. (Equivalently: $D(x, r_2) \cap B \subseteq \{x\}$.) Let $r = \min(r_1, r_2)$. Then $D(x, r) \subseteq D(x, r_1)$ and $D(x, r) \subseteq D(x, r_2)$. So

$$\begin{aligned} D(x, r) \cap ((A \cup B) \setminus \{x\}) &\subseteq [D(x, r) \cap (A \setminus \{x\})] \cup [D(x, r) \cap (B \setminus \{x\})] \\ &= \emptyset \cup \emptyset = \emptyset. \end{aligned}$$

Thus x would not be an accumulation point of A as it was assumed to be. So x must be an accumulation point of at least one of the sets A or B. ⧫

◇ **2E-30.** Show that each open set in $\mathbb{R}$ is a union of disjoint open intervals. Is this sort of result true in $\mathbb{R}^n$ for $n > 1$, where we define an open interval as the Cartesian product of n open intervals, $]a_1, b_1[\times \cdots \times]a_n, b_n[$?

Suggestion. If two open intervals overlap, then their union is an open interval. ◊

Solution. We need to allow "infinite intervals", $]-\infty, b[= \{x \in \mathbb{R} \mid x < b\}$, $]a, \infty[= \{x \in \mathbb{R} \mid a < x\}$, and $]-\infty, \infty[= \mathbb{R}$.

Let U be an open set in $\mathbb{R}$. For $x \in U$, let $a(x) = \inf\{y \in \mathbb{R} \mid]y, x] \subseteq U\}$ and $b(x) = \sup\{y \in \mathbb{R} \mid [a, y[\subseteq U\}$, and let $J(x)$ be the open interval $]a(x), b(x)[$. Either or both of the ends $a(x)$ or $b(x)$ could be infinite, and we know that $a(x) < b(x)$ so that the interval $J(x)$ really exists as a nontrivial interval containing x since the set U is open and does contain some small interval around x. There is an $r > 0$ such that $]x-r, x+r[\subseteq U$. So $a(x) \leq x - r < x + r \leq b(x)$. So

$$J(x) \text{ is an open interval and } x \in J(x). \tag{1}$$

The next fact we need is that this interval is contained in U:

$$J(x) \subseteq U \tag{2}$$

To prove this, let $y \in J(x)$. That is, $a(x) < y < b(x)$. From the definitions of $a(x)$ and $b(x)$, there are numbers y_1 and y_2 such that $a(x) < y_1 < y < y_2 < b(x)$ with $]y_1, x] \subseteq U$ and $[x, y_2[\subseteq U$. So $y \in]y_1, y_2[=]y_1, x[\cup [x, y_2[\subseteq U$. This works for every y in $J(x)$, so $J(x) \subseteq U$ as claimed.

Since we have for each x in U that $x \in J(x) \subseteq U$, we certainly have

$$U = \bigcup_{x \in U} J(x). \tag{3}$$

This expresses U as a union of open intervals. It remains to show that this is actually a union of disjoint intervals. It is not true that the intervals $J(x)$ are all disjoint from each other. It is true that if x and y are in U, then $J(x)$ and $J(y)$ are either disjoint or equal so that the union in (3) really is a union of a family of disjoint intervals with some of them listed many times. This follows from the next two observations.

$$\text{If } I \text{ is an open interval with } x \in I \subseteq U, \text{ then } I \subseteq J(x). \tag{4}$$

$$J(x) \cap (y) \neq \emptyset \quad \implies J(x) = J(y). \tag{5}$$

The first is clear from the definition of $J(x)$, and the second follows from it and the observation that the union of two open intervals which overlap is an open interval.

We cannot get a straightforward generalization of this to $\mathbb{R}^n$ with "interval" replaced by "rectangles" of the form $]a_1, b_1[\times \cdots \times]a_n, b_n[$. Consider the example in $\mathbb{R}^2$ of the open set $U = \{(x, y) \in \mathbb{R}^2 \mid x > 0 \text{ and } -1/x < y < 1/x\}$. Suppose U were a disjoint union of a family of open rectangles and for each $(x, y) \in U$, let $R(x, y)$ denote the rectangle in that family which contains (x, y). For the family to be a collection of pairwise disjoint rectangles we would require that for points (x, y) and (a, b) in U the rectangles $R(x, y)$ and $R(a, b)$ be either disjoint or identical. But this gets us into trouble.

Let $A = \{(x, 0) \mid x \geq 1\}$ be the positive horizontal axis to the right of 1. Then $A \subset U$ and $(1, 0) \in A$.

CLAIM: We would need to have $A \subseteq R((1, 0))$.

Proof: Let $a = \sup\{x > 1 \mid [1, a[\subseteq R((1, 0))\}$. We want to show that

$a = +\infty$. If it were a finite real number, we would have trouble. If it were in $R((1,0))$, then it would be in that open set but nothing larger would be. This cannot happen in an open set. On the other hand, if it were not in $R((1,0))$, then it would be on the boundary of that rectangle. The open rectangle $R((a,1))$ would be a different rectangle but would intersect $R((1,0))$ since it is open and so extends below A. This was not to happen. So we must have $a = +\infty$, and $A \subset R((0,1))$ as claimed.

So far things sound all right. But they are not. If $R((0,1))$ is a rectangle, it has to have a nonzero vertical width. There would be numbers c and d with $c < 0 < d$ such that the vertical segment $I_x = \{(x,y) \mid c < y < d\}$ has $I_x \subseteq R((0,1)) \subseteq U$ for all $x > 1$. But this cannot be true. The set U is very narrow in the vertical direction for large x. The rectangle $R((0,1))$ would have to stick outside of U. See the figure.

FIGURE 2-15.

For some related ideas see Exercise 2E-35 and the remarks following it. ♦

◇ **2E-31.** Let A' denote the set of accumulation points of a set A. Prove that A' is closed. Is $(A')' = A'$ for all A?

Sketch. What needs to be done is to show that an accumulation point of A' must be an accumulation point of A. $(A')'$ need not be equal to A'. Consider $A = \{1/2, 1/3, \dots\}$. ◇

Solution. To show that A' is closed, we show that it contains all of its accumulation points. That is, $(A')' \subseteq A'$. Suppose x is an accumulation point of A', and let U be an open set containing x. Then U contains a point y in A' with y not equal to x. Let $V = U \setminus \{x\}$. Then V is an open set containing y. Since $y \in A'$, there is a point z in $V \cap A$ with z not equal to y. Since x is not in V, we also know that z is not equal to x. Since $V \subseteq U$, we know that $z \in U$. Every neighborhood U of x contains a point z of A which is not equal to x. So x is an accumulation point of A. This works for every x in A'. So $(A')' \subseteq A'$. Since the set A' contains all of its accumulation points, it is closed as claimed.

Although we now know that $(A')' \subseteq A'$ for all subsets of a metric space M, the inclusion might be proper. Consider $A = \{1, 1/2, 1/3 \dots\} \subseteq \mathbb{R}$. Then $A' = \{0\}$. and $(A')' = \emptyset$. ♦

◇ **2E-32.** Let $A \subset \mathbb{R}^n$ be closed and $x_n \in A$ be a Cauchy sequence. Prove that x_n converges to a point in A.

Suggestion. Use completeness to get the limit to exist and closure to get it in A. ◇

Solution. The sequence $\langle x_n \rangle_1^\infty$ is a Cauchy sequence in $\mathbb{R}^n$. So it converges to some point $v \in \mathbb{R}^n$ by the completeness of $\mathbb{R}^n$. (Theorem 2.8.5.) The limit must be in A since A is closed. (Proposition 2.7.6(i).) ♦

◇ **2E-33.** Let s_n be a bounded sequence of real numbers. Assume that $2s_n \leq s_{n-1} + s_{n+1}$. Show that $\lim_{n\to\infty}(s_{n+1} - s_n) = 0$.

Sketch. Show that $\alpha_n = s_{n+1} - s_n$ are increasing and bounded above. Use the telescoping series $s_n = s_0 + \sum_{k=1}^{n-1} \alpha_k$ to show that the limit must be 0. ◇

Solution. Subtract $(s_n + s_{n+1})$ from both sides of the inequality $2s_n \leq s_{n-1}+s_{n+1}$ to obtain $s_n - s_{n+1} \leq s_{n-1} - s_n$, or equivalently, $s_{n+1} - s_n \geq s_n - s_{n-1}$. Let $\alpha_n = s_{n+1} - s_n$. We have just shown that $\langle \alpha_n \rangle_1^\infty$ is an increasing sequence. Also, we know that the numbers s_n are bounded. There is a B such that $|s_n| \leq B$ for every n. So $|\alpha_n| = |s_{n+1} - s_n| \leq |s_{n+1}| + |s_n| \leq 2B$. The sequence $\langle \alpha_n \rangle_1^\infty$ is a sequence in $\mathbb{R}$ which is increasing and bounded above. So it must converge to some $\lambda \in \mathbb{R}$. We want to show that $\lambda = 0$. If not, then either $\lambda > 0$ or $\lambda < 0$. We eliminate the first and leave the second to the reader.

If $\lambda > 0$, then there is an index N such that $\alpha_n > \lambda/2 > 0$ whenever $n \geq N$. But then we would have, for $n > N$,

$$\begin{aligned} s_n &= s_0 + (s_1 - s_0) + (s_2 - s_1) + \cdots + (s_n - s_{n-1}) \\ &= s_0 + \sum_{k=1}^{n-1} \alpha_k = s_0 + \sum_{k=1}^{N-1} \alpha_k + \sum_{k=N}^{n-1} \alpha_k \\ &\geq s_0 + \sum_{k=1}^{N-1} \alpha_k + (n-N)\lambda/2 \quad \to \quad +\infty \text{ as } n \to \infty. \end{aligned}$$

This contradicts the hypothesis that the sequence $\langle s_n \rangle_1^\infty$ is bounded above. The contradiction obtained from $\lambda < 0$ is similar. ♦

◇ **2E-34.** Let $x_n \in \mathbb{R}^k$ and $d(x_{n+1}, x_n) \leq r d(x_n, x_{n-1})$, where $0 \leq r < 1$. Show that x_n converges.

Sketch. First use the inequality repeatedly to obtain $d(x_{j+1}, x_j) \leq r^j \cdot d(x_1, x_0)$ for each j. Then use this and the triangle inequality to show $d(x_{n+p}, x_n) \leq r^n d(x_1, x_0)/(1-r)$. Now use completeness. ◇

Solution. We will show that the points x_n form a Cauchy sequence in $\mathbb{R}^k$. From this the convergence of the sequence to some point $x \in \mathbb{R}^k$ follows by the completeness of $\mathbb{R}^k$. (Theorem 2.8.5.)

CLAIM: $d(x_{j+1}, x_j) \leq r^j d(x_1, x_0)$ for each j.

This is done by an induction which we present informally:

$$d(x_{j+1}, x_j) \leq r d(x_j, x_{j-1}) \leq r^2 d(x_{j-1}, x_{j-2}) \leq \cdots \leq r^j d(x_1, x_0).$$

Now we use this in a repeated triangle inequality to estimate $d(x_{n+p}, x_n)$:

$$\begin{aligned} d(x_{n+p}, x_n) &\leq d(x_{n+p}, x_{n+p-1}) + d(x_{n+p-1}, x_{n+p-2}) + \cdots + d(x_{n+1}, x_n) \\ &\leq (r^{n+p-1} + r^{n+p-2} + \cdots + r^n)\, d(x_1, x_0) \\ &\leq r^n (1 + r + r^2 + \cdots + r^{p-1})\, d(x_1, x_0) \\ &\leq r^n (1 + r + r^2 + \dots)\, d(x_1, x_0) = \frac{r^n}{1-r}\, d(x_1, x_0) \\ &\to 0 \qquad \text{since } 0 \leq r < 1. \end{aligned}$$

Thus the points x_n form a Cauchy sequence in $\mathbb{R}^k$ and must converge to a point in $\mathbb{R}^k$ as claimed.

The proof as written uses only the properties of a metric and the completeness of the space. It therefore serves to establish the following somewhat more general fact.

Proposition. *If x_n, $n = 0, 1, 2, 3, \ldots$ are points in a metric space M and there is a constant $r \in \mathbb{R}$ with $0 \le r < 1$ such that $d(x_{n+1}, x_n) < rd(x_n, x_{n-1})$ for each $n = 1, 2, 3, \ldots$, then the points form a Cauchy sequence in M. If M is complete with respect to the metric d, then the sequence must converge to some point in M.*

In this form the proposition forms the heart of the proof of one of the very important theorems about complete metric spaces. This is a result called the Contraction Mapping Principle or the Banach Fixed Point Theorem.

Theorem. *If M is a complete metric space and $T : M \to M$ is a function for which there is a real constant r with $0 \le r < 1$ and $d(Tx, Ty) \le rd(x, y)$ for all x and y in M, then there is a unique point $z \in M$ such that $Tz = z$. Furthermore, if x_0 is any point in M, then the sequence x_0, $x_1 = Tx_0$, $x_2 = Tx_1$, $x_3 = Tx_2$, $\ldots$ converges to z.*

We will study this theorem in §5.7 and see how it can be used, among other things, to prove the existence of solutions to a large class of differential equations, while simultaneously giving a method of generating a sequence of approximate solutions. In this context the repetition of the function T in the theorem is exactly the Picard iteration process for approximating the solution to a differential equation. It is often studied in courses on differential equations, numerical analysis, or integral equations. ♦

◇ **2E-35.** Show that any family of disjoint nonempty open sets of real numbers is countable.

Suggestion. Each of the sets must contain a rational number. ◊

Solution. Let $\{U_\alpha \mid \alpha \in A\}$ be any family of disjoint open subsets of $\mathbb{R}$. That is, each U_α is open, and $U_\alpha \cap U_\beta = \emptyset$ if α and β are different. Since each U_α is open, there are open intervals $]a_\alpha, b_\alpha[\subseteq U_\alpha$. We know that the rational numbers are scattered densely in $\mathbb{R}$ in the sense that if $a_\alpha < b_\alpha$ is any such pair of reals there is a rational number r_α with $r_\alpha \in]a_\alpha, b_\alpha[\subseteq U_\alpha$. Since different α and β have $U_\alpha \cap U_\beta = \emptyset$, we must have r_α and r_β different. But there are only countably many different rational numbers. So there are only countably many different sets U_α in the collection.

Remarks: (1) Without too much effort we can see that points $(r_1, r_2, \ldots, r_n)$ with rational coordinates are dense in $\mathbb{R}^n$ in the sense that any open ball $D(v, \epsilon)$ must contain such points. Furthermore, $\mathbb{Q} \times \cdots \times \mathbb{Q} = \mathbb{Q}^n \subseteq \mathbb{R}^n$ is a countable set of points in $\mathbb{R}^n$. With these facts in hand we can conclude in exactly the same way that every family of disjoint open subsets in $\mathbb{R}^n$ must be countable.

(2) If we combine the results of Exercises 2E-30 and 2E-34, we obtain

Proposition. *Each open set in $\mathbb{R}$ is the union of countably many disjoint open intervals.*

(3) This also has a generalization to $\mathbb{R}^n$. We saw that in Exercise 2E-30 that the appropriate sets to look at are not the direct generalization of intervals to "rectangles". We need to loosen this to "connected sets". These are those sets which are "all in one piece". We will find the right way to formulate this in Chapter 3.

Proposition. *Each open subset of $\mathbb{R}^n$ is the union of a countable number of disjoint connected subsets.*

The connection is that a subset of $\mathbb{R}$ is connected (all one piece) if and only if it is an interval. Here the term "interval" is intended to include half lines and the whole line. ♦

◇ **2E-36.** Let $A, B \subset \mathbb{R}^n$ be closed sets. Does $A + B = \{x + y \mid x \in A$ and $y \in B\}$ have to be closed?

Answer*.* No. ◊

Solution*.* Here is a counterexample. Let $A = \{(x, 0) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ be the horizontal axis and $B = \{(t, 1/t) \in \mathbb{R}^2 \mid t > 0\}$ be one branch of a hyperbola. If k is a positive integer, then $u_k = (-k, 0) \in A$, and $v_k = (k, 1/k) \in B$. So $w_k = (0, 1/k) \in A + B$. Since $\| w_k \| = 1/k \to 0$, we have $w_k \to (0, 0)$ and $(0, 0) \in \mathrm{cl}(A+B)$. But $(0, 0)$ itself is not in $A+B$ since there is no single vector v with $v = (a, b) \in A$ and $-v = (-a, -b) \in B$. The second coordinate, b, of v would have to be 0, and $-1/0$ is not a real number. ♦

◇ **2E-37.** For $A \subset M$, a metric space, prove that

$$\mathrm{bd}(A) = [A \ \cap \ \mathrm{cl}(M \setminus A)] \ \cup \ [\mathrm{cl}(A) \setminus A].$$

Sketch*.* $[A \cap \mathrm{cl}(M \setminus A)] \cup [\mathrm{cl}(A) \setminus A] = [A \cup (\mathrm{cl}(A) \setminus A)] \cap [\mathrm{cl}(M \setminus A) \cup (\mathrm{cl}(A) \setminus A)] = \mathrm{cl}(A) \cap [\mathrm{cl}(M \setminus A) \cup (\mathrm{cl}(A) \setminus A)]$. Why is this $\mathrm{bd}(A)$? ◊

Solution*.* If $x \in \mathrm{cl}(A) \setminus A$, then it is not in A and must be in $M \setminus A$. So it is certainly in $\mathrm{cl}(M \setminus A)$. Thus $\mathrm{cl}(A) \setminus A \subseteq \mathrm{cl}(M \setminus A)$, and

$$\mathrm{cl}(M \setminus A) \cup (\mathrm{cl}(A) \setminus A) = \mathrm{cl}(M \setminus A).$$

We compute

$$\begin{aligned}
[A \cap \mathrm{cl}(M \setminus A)] \cup [\mathrm{cl}(A) \setminus A] &= [A \cup (\mathrm{cl}(A) \setminus A)] \cap [\mathrm{cl}(M \setminus A) \cup (\mathrm{cl}(A) \setminus A)] \\
&= \mathrm{cl}(A) \cap [\mathrm{cl}(M \setminus A) \cup (\mathrm{cl}(A) \setminus A)] \\
&= \mathrm{cl}(A) \cap \mathrm{cl}(M \setminus A) \\
&= \mathrm{bd}(A)
\end{aligned}$$

as claimed. ♦

◇ **2E-38.** Let $x_k \in \mathbb{R}^n$ satisfy $||x_k - x_l|| \leq 1/k + 1/l$. Prove that x_k converges.

Suggestion. Show we have a Cauchy sequence by taking k and l larger than N where $1/N < \varepsilon/2$. ◊

Solution. Let $\varepsilon > 0$. Select N large enough so that $0 < 1/N < \varepsilon/2$. If $k \geq N$ and $l \geq N$, then

$$\| x_k - x_l \| \leq \frac{1}{k} + \frac{1}{l} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus $\langle x_k \rangle_1^\infty$ is a Cauchy sequence in $\mathbb{R}^n$. Since $\mathbb{R}^n$ is complete, the sequence must converge to some point in $\mathbb{R}^n$ (Theorem 2.8.5). ♦

◇ **2E-39.** Let $S \subset \mathbb{R}$ be bounded above and below. Prove that $\sup(S) - \inf(S) = \sup\{x - y \mid x \in S \text{ and } y \in S\}$.

Suggestion. First show that $\sup(S) - \inf(S)$ is an upper bound for $\{x - y \mid x \in S \text{ and } y \in S\}$. Then consider x and y in S very close to $\sup(S)$ and $\inf(S)$. ◊

Solution. Let $T = \{x - y \mid x \in S \text{ and } y \in S\}$. If x and y are in S, we know that $x \leq \sup S$ and $\inf S \leq y$. So $-y \leq -\inf S$, and

$$x - y \leq x - \inf S \leq \sup S - \inf S.$$

Thus $\sup S - \inf S$ is an upper bound for T. If $\varepsilon > 0$, there are x and y is S with $\sup S - (\varepsilon/2) < x \leq \sup S$, and $\inf S \leq y < \inf S + (\varepsilon/2)$. So

$$(\sup S - \inf S) - \varepsilon = \sup S - (\varepsilon/2) - (\inf S + (\varepsilon/2)) < x - y \leq \sup S - \inf S.$$

Thus $(\sup S - \inf S) - \varepsilon$ is not an upper bound for T. This holds for every $\varepsilon > 0$, so $\sup S - \inf S$ is the least upper bound for T as claimed. ♦

◇ **2E-40.** Suppose in $\mathbb{R}$ that for all n, $a_n \leq b_n$, $a_n \leq a_{n+1}$, and $b_{n+1} \leq b_n$. Prove that a_n converges.

Sketch. The sequence is increasing and bounded above by b_1 in $\mathbb{R}$, so it must converge. ◊

Solution. We are given that $a_n \le a_{n+1}$ for all n. So $a_1 \le a_2 \le a_3 \ldots$. The sequence is monotonically increasing. We also know that $b_{k+1} \le b_k$ for each k and $a_n \le b_n$ for each n. So

$$a_n \le b_n \le b_{n-1} \le b_{n-2} \le \cdots \le b_2 \le b_1.$$

This is true for each n. So the sequence $\langle a_n \rangle_1^\infty$ is bounded above by $b_1 \in \mathbb{R}$. Our sequence is a monotonically increasing sequence of real numbers bounded above in $\mathbb{R}$. So it must converge to some $\lambda \in \mathbb{R}$ by completeness. ♦

◇ **2E-41.** Let A_n be subsets of a metric space M, $A_{n+1} \subset A_n$, and $A_n \neq \varnothing$, but assume that $\cap_{n=1}^\infty A_n = \varnothing$. Suppose $x \in \cap_{n=1}^\infty \mathrm{cl}(A_n)$. Show that x is an accumulation point of A_1.

Sketch. Let U be a neighborhood of x. There is an n with $x \in \mathrm{cl}(A_n) \setminus A_n$. (Why?) So $U \cap (A_n \setminus \{x\})$ is not empty. ◇

Solution. Let U be an open set containing x. We must show that U contains some point of A_1 not equal to x. Since $\bigcap_{n=1}^\infty A_n = \emptyset$, there is an n_0 such that x is not in A_{n_0}. But $x \in \bigcap_{n=1}^\infty \mathrm{cl}(A_n)$. So $x \in \mathrm{cl}(A_{n_0})$ Since x is in the closure of A_{n_0} but not in A_{n_0}, it must be an accumulation point of A_{n_0}. So there is a y in $U \cap (A_{n_0} \setminus \{x\})$. But $A_{n_0} \subseteq A_{n_0-1} \subseteq \cdots \subseteq A_1$. So $y \in U \cap (A_1 \setminus \{x\})$. This can be done for every open set containing x. So x is an accumulation point of A_1 as claimed. ♦

◇ **2E-42.** Let $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$. Define $d(x, A) = \inf\{d(x, y) \mid y \in A\}$. Must there be a $z \in A$ such that $d(x, A) = d(x, z)$?

Answer. Not necessarily. ◇

Solution. The number $d(x, A) = \inf\{d(x, y) \mid y \in A\}$ is the greatest lower bound of the distances between x and points of A. If A is not empty, then this is a finite real number. It is 0 if and only if there are points of A either equal to x or arbitrarily close to x. For each $n > 0$, there is a point $x_n \in A$ with $d(x, x_n) < 1/n$. Thus $x_n \to x$ and x must be in $\mathrm{cl}(A)$.

$$d(x, A) = 0 \iff x \in \mathrm{cl}(A).$$

If A is open and $x \in \mathrm{bd}(A)$, then $d(x, A) = 0$, but there will be no point y in A with $d(x, y) = 0$. (That would say $x = y$.) For example, in $\mathbb{R}^2$, let $A = D((0,0), 1) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$, and let $w = (1, 0)$. Then $d(w, A) = 0$ since $v_k = (1 - (1/k), 0) \in A$ and $d(v_k, w) = 1/k \to 0$. But w is not in A. So there is no v in A with $d(v, w) = d(w, A) = 0$. ♦

◇ **2E-43.** Let $x_1 = \sqrt{3}, \dots, x_n = \sqrt{3 + x_{n-1}}$. Compute $\lim_{n\to\infty} x_n$.

Answer. $(1+\sqrt{13})/2$. ◇

Solution. First we show that the limit exists by showing that the sequence is increasing and bounded above. Once we know that it exists we can assign it a symbol and use the properties of limits to compute its value.

FIRST: The sequence is increasing.

The first two values are $x_1 = \sqrt{3}$ and $x_2 = \sqrt{3+\sqrt{3}} > \sqrt{3} = x_1$. For an induction, assume that $x_k - x_{k-1} > 0$. Then

$$\begin{aligned} x_{k+1} - x_k &= \sqrt{3+x_k} - \sqrt{3+x_{k-1}} \\ &= \frac{(3+x_k) - (3+x_{k-1})}{\sqrt{3+x_k} + \sqrt{3+x_{k-1}}} \\ &= \frac{x_k - x_{k-1}}{\sqrt{3+x_k} + \sqrt{3+x_{k-1}}}. \end{aligned}$$

So $x_{k+1} - x_k > 0$ also. By induction we have $x_{n+1} - x_n > 0$ for all $n = 1, 2, 3, \dots$. The sequence is increasing.

SECOND: The sequence is bounded above.

We have $x_1 = \sqrt{3} < 3$. If $x_k < 3$, then $x_{k+1} = \sqrt{3+x_k} < \sqrt{6} < 3$ also. By induction we conclude that $x_n < 3$ for all n.

Our sequence is a monotonically increasing sequence in $\mathbb{R}$ which is bounded above. By completeness there is a $\lambda \in \mathbb{R}$ such that $x_n \to \lambda$. Consider what happens to x_{n+1}.

$$x_{n+1} \to \lambda \quad \text{but} \quad x_{n+1} = \sqrt{3+x_n} \to \sqrt{3+\lambda}.$$

But limits in $\mathbb{R}$ are unique. So $\lambda = \sqrt{3+\lambda}$. With a bit of algebra we get $0 = \lambda^2 - \lambda - 3$. This has solutions $\lambda = (1 \pm \sqrt{13})/2$. But the sequence is increasing away from $x_1 = \sqrt{3}$. So $\lambda > \sqrt{3}$. We must have

$$\lim_{n\to\infty} x_n = \lambda = \frac{1+\sqrt{13}}{2}.$$ ♦

◇ **2E-44.** A set $A \subset \mathbb{R}^n$ is said to be ***dense*** in $B \subset \mathbb{R}^n$ if $B \subset \mathrm{cl}(A)$. If A is dense in $\mathbb{R}^n$ and U is open, prove that $A \cap U$ is dense in U. Is this true if U is not open?

Sketch. It need not be true for sets which are not open. ◇

Solution. Let A be dense in $\mathbb{R}^n$ and U is an open subset of $\mathbb{R}^n$. We want to prove that $U \subseteq \mathrm{cl}(A \cap U)$. If $x \in U$, we need to show that every neighborhood of x intersects $A \cap U$. So, suppose V is an open set containing x. Then $x \in V \cap U$ which is an open set. So there is an $r > 0$ such that $x \in D(x, r) \subseteq U \cap V$. Since A is dense in $\mathbb{R}^n$, there is a

$$y \in A \cap D(x, r) \subseteq A \cap (V \cap U) = V \cap (A \cap U).$$

Since this can be done for every neighborhood V of x, we have $x \in \mathrm{cl}(A \cap U)$. Since this is true for every $x \in U$, we have $U \subseteq \mathrm{cl}(A \cap U)$. Thus $A \cap U$ is dense in U as claimed.

If the set U is not open, this can fail. Consider $U = \mathbb{Q} \subseteq \mathbb{R}$ and $A = \mathbb{R} \setminus \mathbb{Q} \subseteq \mathbb{R}$. Each of the sets U and A is dense in $\mathbb{R}$, but $A \cap U = \emptyset$ and is dense nowhere. ♦

◇ **2E-45.** Show that $x^{\log x} = o(e^x)$ as $x \to \infty$ (see Worked Example 2WE-5).

Suggestion. We need $x^{\log x}/e^x \to 0$. Do some manipulation and use L'Hôpital's rule. ◊

Solution. We need to show that $x^{\log x}/e^x \to 0$ as $x \to +\infty$. Let $y = x^{\log x}/e^x$. So $\log y = \log\left(x^{\log x}\right) - \log(e^x) = (\log x)^2 - x$.

CLAIM: $\lim_{x\to\infty} (\log x)^2/x = 0$.

Proof: The limit is of the indeterminate form ∞/∞, so we may try L'Hôpital's Rule. The ratio of derivatives is

$$\frac{2(\log x)(1/x)}{1} = \frac{2\log x}{x}$$

which is again of the form ∞/∞. So we try L'Hôpital's Rule again. The ratio of derivatives is

$$\frac{2/x}{1} = \frac{2}{x} \to 0.$$

Since this goes to 0 as $x \to \infty$, we have $\lim_{x\to\infty} (2\log x)/x = 0$ and $\lim_{x\to\infty} (\log x)^2/x = 0$ by L'Hôpital's Rule.

From the lemma, there must be a $B > 0$ such that

$$\begin{aligned} x > B \implies &\frac{(\log x)^2}{x} < \frac{1}{2} \\ \implies &(\log x)^2 < x/2 \\ \implies &\log y = (\log x)^2 - x < -x/2 \\ \implies &0 < y(x) < e^{-x/2} = 1/e^{x/2}. \end{aligned}$$

As $x \to +\infty$, $1/e^{x/2} \to 0$. So $y(x) \to 0$. That is

$$\lim_{x\to\infty} y(x) = \lim_{x\to\infty} \frac{x^{\log x}}{e^x} = 0.$$

This is exactly what we wanted. $x^{\log x} = o(e^x)$ as $x \to \infty$ as desired. ♦

◇ **2E-46.** (a) If If $f = o(g)$ and if $g(x) \to \infty$ as $x \to \infty$, then show that $e^{f(x)} = o(e^{g(x)})$ as $x \to \infty$.

(b) Show that $\lim_{n\to\infty}(x \log x)/e^{\,x} = 0$ by showing that $x = o(e^{\,x/2})$ and that $\log\, x = o(e^{\,x/2})$.

Solution. (a) Since $f(x) = o(g(x))$ as $x \to \infty$, there is a $B > 0$ such that $|f(x)| < g(x)/2$ for $x > B$. (We can assume $g(x) > 0$ since $g(x) \to +\infty$.) In particular,

$$f(x) - g(x) < \frac{1}{2}\, g(x) - g(x) = -\frac{g(x)}{2} \qquad \text{for } x > B.$$

So $f(x) - g(x) \to -\infty$ as $x \to +\infty$. Thus

$$\frac{e^{f(x)}}{e^{g}(x)} = e^{f(x)-g(x)} \to 0 \qquad \text{as } x \to +\infty.$$

That is, $e^{f(x)} = o(e^{g(x)})$ as claimed.

(b) The method of proof suggested is based on the last part of this lemma.

Lemma. *Let $f(x)$, $g(x)$, $F(x)$, and $G(x)$ be real valued functions defined for large real x. Then*

(1) If $f(x) = o(g(x))$ and $g(x) = O(h(x))$ as $x \to \infty$, then $f(x) = o(h(x))$ as $x \to \infty$.

(2) If $f(x) = O(g(x))$ and $g(x) = o(h(x))$ as $x \to \infty$, then $f(x) = o(h(x))$ as $x \to \infty$.

(3) If $f(x) = o(g(x))$ and $g(x) = o(h(x))$ as $x \to \infty$, then $f(x) = o(h(x))$ as $x \to \infty$.

(4) If $f(x) = o(g(x))$ and $F(x) = o(G(x))$ as $x \to \infty$, then $f(x)F(x) = o(g(x)G(x))$ as $x \to \infty$.

Proof: For (1) we note that since $g(x) = O(h(x))$ as $x \to \infty$, there is a bound $M > 0$ and a number B_1 such that $|g(x)| < Mh(x)$ whenever $x > B_1$. Since $f(x) = o(g(x))$, there is a B_2 such that $|f(x)| < (\varepsilon/M)\,|g(x)|$ whenever $x > B_2$. If $x > \max\,(B_1, B_2)$, we have

$$|f(x)| < (\varepsilon/M)\,|g(x)| < (\varepsilon/M)Mh(x) = \varepsilon h(x).$$

So $f(x) = o(h(x))$ as $x \to \infty$ as claimed.

The proof of (2) is similar to that of (1) with the roles of f and g reversed.

Assertion (3) follows from (1) since $g(x) = o(h(x))$ implies that $g(x) = O(h(x))$.

To prove part (4), let $\varepsilon > 0$. Since $f(x) = o(g(x))$ and $F(x) = o(G(x))$ as $x \to \infty$, there are $B_1 > 0$ and $B_2 > 0$ such that $|f(x)| < \sqrt{\varepsilon}\,|g(x)|$ whenever $x > B_1$ and $|F(x)| < \sqrt{\varepsilon}\,|G(x)|$ whenever $x > B_2$ If $x > \max(B_1, B_2)$, then

$$|f(x)F(x)| < \sqrt{\varepsilon}\,|g(x)| \cdot \sqrt{\varepsilon}\,|G(x)| = \varepsilon\,|g(x)G(x)|\,.$$

So $f(x)F(x) = o(g(x)G(x))$ as $x \to \infty$ as claimed.

To complete the proof of part (b), we apply part (4) of the lemma with $f(x) = x$, $g(x) = e^{x/2}$, $F(x) = \log x$, and $G(x) = e^{x/2}$. We thus will be done as soon as we have two facts.

CLAIM ONE: $x = o(e^{x/2})$ as $x \to +\infty$.

CLAIM TWO: $\log x = o(e^{x/2})$ as $x \to +\infty$.

For the first assertion we need the limit $\lim\limits_{x\to+\infty} \dfrac{x}{e^{x/2}} = 0$. This is a limit of the indeterminate form ∞/∞, so we may try L'Hôpital's Rule. The ratio of derivatives is

$$\frac{1}{(1/2)e^{x/2}} = \frac{1}{2e^{x/2}} \to 0.$$

Since this limit is 0, we have $\lim\limits_{x\to+\infty} \dfrac{x}{e^{x/2}} = 0$ also by L'Hôpital's Rule.

For the second claim we can either use L'Hôpital's Rule again or we can use part (1) of the lemma. We have just shown that $x = o(e^{x/2})$. We also have $\log x = O(x)$ since $0 < \log x < x$ for $x > 1$. Thus $\log x = o(e^{x/2})$ by part (1).

With both of our claims established, the relation $x \log x = o(e^x)$ follows from part (4) of the lemma. So $\lim_{n\to\infty}(x \log x)/e^x = 0$ as claimed. ♦

◇ **2E-47.** Show that $\gamma = \lim_{n\to\infty}(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n)$ exists by using the proof of the integral test (γ is called ***Euler's constant***).

Suggestion. Compare the graph of $1/x$ to inscribed and circumscribed rectangles to show that the sequence $\gamma_n = 1 + (1/2) + \cdots + (1/n) - \log n$ is decreasing and bounded below by 0. ◊

Solution. A clue as to how to proceed comes from the proof of the integral test for the convergence of infinite series. It is exactly this test which allows us to obtain the divergence of the harmonic series from that of the logarithm integral. A comparison of Riemann sums using inscribed and circumscribed rectangles with the exact value of the integral leads to inequalities

$$\frac{1}{2}+\frac{1}{3}+\cdots++\frac{1}{n}<\log n<1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n-1}$$

This is sketched for $n = 8$ in the figure.

FIGURE 2-16. The proof of the integral test gives the method for proving the existence of Euler's constant.

If we let $a_n = 1+\frac{1}{2}+\cdots+\frac{1}{n}-\log n$, then this certainly implies that $a_n > 0$ for each $n = 2, 3, 4, \ldots$. Furthermore, contemplation of the column between $x = n$ and $x = n+1$ gives

$$\frac{1}{n+1}<\log(n+1)-\log n<\frac{1}{n}.$$

Since $a_{n+1}-a_n = \frac{1}{n+1}-\log(n+1)+\log n$, this shows that $a_{n+1}-a_n < 0$ for each n, and the sequence is decreasing. Since the sequence $\langle a_n\rangle_1^\infty$ is monotonically decreasing and bounded below by 0 in $\mathbb{R}$, there must be a real number γ such that $a_n \to \gamma$ by completeness. That is,

$$\lim n\to\infty\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\log n\right)=\gamma \qquad \text{exists in } \mathbb{R}$$

as claimed. ♦

◇ **2E-48.** Prove the following generalizations of the ratio and root tests:

(a) If $a_n > 0$ and $\limsup_{n\to\infty} a_{n+1}/a_n < 1$, then $\sum a_n$ converges, and if $\liminf_{n\to\infty} a_{n+1}/a_n > 1$, then $\sum a_n$ diverges.
(b) If $a_n \geq 0$ and if $\limsup_{n\to\infty} \sqrt[n]{a_n} < 1$ (respectively, > 1), then $\sum a_n$ converges (respectively, diverges).

(c) In the ratio comparison test, can the limits be replaced by lim sup's?

Suggestion. (a) For large n, the series is comparable to an appropriately selected geometric series.

(b) For large n, the series is comparable to an appropriately selected geometric series.

(c) Not quite, but we can replace the limit in the convergence statement by lim sup and the one in the divergence statement by lim inf. ◊

Solution. (a) Suppose each $a_n > 0$ and that $\limsup_{n\to\infty} a_{n+1}/a_n < 1$. Select a number r such that $\limsup_{n\to\infty} a_{n+1}/a_n < r < 1$. Then there is an integer N such that $a_{n+1}/a_n < r$ whenever $n \geq N$. Since the numbers are nonnegative, this gives $0 \leq a_{n+1} \leq ra_n$ for $n = N, N+1, N+2, \ldots$. Apply this repeatedly.

$$\begin{aligned}
&0 \leq a_{N+1} \leq ra_N \\
&0 \leq a_{N+2} \leq ra_{N+1} \leq r^2 a_N \\
&0 \leq a_{N+3} \leq ra_{N+2} \leq r^3 a_N \\
&\qquad\qquad\vdots \\
&0 \leq a_{N+p} \leq ra_{N+p-1} \leq r^p a_N \\
&\qquad\qquad\vdots
\end{aligned}$$

Since $0 \leq r < 1$, we know that the geometric series $\sum_{p=0}^{\infty} r^p a_N$ converges. By comparison, we conclude that $\sum_{k=N}^{\infty} a_k = \sum_{p=0}^{\infty} a_{N+p}$ also converges. Reintroducing the finitely many terms $a_1 + \cdots + a_{N-1}$ does not change the fact of convergence. We conclude that $\sum_{n=1}^{\infty} a_n$ converges.

Now suppose each $a_n > 0$ and that $\liminf_{n\to\infty} a_{n+1}/a_n > 1$. Select a number r such that $1 < r < \liminf_{n\to\infty} a_{n+1}/a_n$ Then there is an integer N such that $a_{n+1}/a_n > r > 1$ whenever $n \geq N$. Since the numbers are nonnegative, this gives $a_{n+1} \geq ra_n \geq a_n > 0$ for $n = N, N+1, N+2, \ldots$. Apply this repeatedly.

$$\begin{aligned}
a_{N+1} &\geq ra_N \geq a_N > 0 \\
a_{N+2} &\geq ra_{N+1} \geq a_N > 0 \\
a_{N+3} &\geq ra_{N+2} \geq a_N > 0 \\
&\vdots \\
a_{N+p} &\geq ra_{N+p-1} \geq a_N > 0 \\
&\vdots
\end{aligned}$$

Since $a_N > 0$, we know that the constant series $\sum_{p=0}^{\infty} a_N$ diverges to $+\infty$. By comparison, we conclude that $\sum_{k=N}^{\infty} a_k = \sum_{p=0}^{\infty} a_{N+p}$ also diverges. Reintroducing the finitely many terms $a_1 + \cdots + a_{N-1}$ does not change the fact of divergence. We conclude that $\sum_{n=1}^{\infty} a_n$ diverges.

(b) Suppose each $a_n \geq 0$ and that $\limsup_{n\to\infty} \sqrt[n]{a_n} < 1$. Select a number r such that $\limsup_{n\to\infty} \sqrt[n]{a_n} < r < 1$ Then there is an integer N such that $0 \leq \sqrt[n]{a_n} < r < 1$ whenever $n \geq N$. Taking n^{th} powers gives $0 \leq a_n \leq r^n$ for $n = N, N+1, N+2, \ldots$. That is, $0 \leq a_{N+p} \leq r^{N+p}$ for $p = 1, 2, 3, \ldots$. Since $0 \leq r < 1$, we know that the geometric series $\sum_{p=0}^{\infty} r^{N+p} = r^N \sum_{p=0}^{\infty} r^p$ converges. By comparison, the series $\sum_{n=N}^{\infty} a_n = \sum_{p=0}^{\infty} a_{N+p}$ also converges. Reintroducing the finitely many terms $a_1 + \cdots + a_{N-1}$ does not change the fact of convergence. We conclude that $\sum_{n=1}^{\infty} a_n$ converges.

Now suppose each $a_n \geq 0$ and that $\limsup_{n\to\infty} \sqrt[n]{a_n} > 1$. Select a number r such that $1 < r < \limsup_{n\to\infty} \sqrt[n]{a_n}$. Then there is an integer N such that $1 < r < \sqrt[n]{a_n}$ whenever $n \geq N$. Taking n^{th} powers gives $1 < r^n <\leq a_n$ for $n = N, N+1, N+2, \ldots$. That is, $1 < a_{N+p}$ for $p = 1, 2, 3, \ldots$. The constant series $\sum_{p=0}^{\infty} 1$ certainly diverges. By comparison, the series $\sum_{n=N}^{\infty} a_n = \sum_{p=0}^{\infty} a_{N+p}$ also diverges. Reintroducing the finitely many terms $a_1 + \cdots + a_{N-1}$ does not change the fact of divergence. We conclude that $\sum_{n=1}^{\infty} a_n$ diverges.

(c) Following the pattern in the ratio test of part (a), we can replace the limits by lim sup in the convergence part of the result and by lim inf in the divergence part.

Proposition. *Let $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ be series with $b_n > 0$ for each n.*

(1) If $|a_n| \leq b_n$ for each n or $\limsup_{n\to\infty} |a_n|/b_n < \infty$, and if $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ also converges (in fact absolutely).

(2) If $a_n \geq b_n$ for each n or $\liminf_{n\to\infty} a_n/b_n > 0$, and if $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ also diverges.

Proof: Suppose each $0 \leq |a_n| \leq b_n$ and that $\sum_{n=1}^{\infty} b_n$ converges. Then $\sum_{n=1}^{\infty} |a_n|$ also converges by comparison. Since $\mathbb{R}$ is complete, absolute convergence implies convergence, and $\sum_{n=1}^{\infty} a_n$ also converges.

Now suppose $\limsup_{n\to\infty} |a_n|/b_n < \infty$ and $\sum_{n=1}^{\infty} b_n$ converges. Select a number $r > 0$ such that $\limsup_{n\to\infty} |a_n|/b_n < r < \infty$. Then there is an integer N such that $|a_n|/b_n < r$ whenever $n \geq N$. Since the numbers b_n are positive, this gives $0 \leq |a_n| \leq rb_n$ for $n = N, N+1, N+2, \ldots$. Since the series $\sum_{n=N}^{\infty} b_n$ converges, it still does after each term is multiplied by the constant r. By comparison, we conclude that $\sum_{n=N}^{\infty} |a_n|$ also converges. Reintroducing the finitely many terms $|a_1| + \cdots + |a_{N-1}|$ does not change the fact of convergence. We conclude that $\sum_{n=1}^{\infty} |a_n|$ converges. We use again the fact that absolute convergence implies convergence to conclude that $\sum_{n=1}^{\infty} a_n$ also converges.

Suppose each $a_n \geq b_n$ for each n and that $\sum_{n=1}^{\infty} b_n$ converges. Then $\sum_{n=1}^{\infty} a_n$ also diverges by comparison.

Now suppose $\liminf_{n\to\infty} a_n/b_n > 0$ and $\sum_{n=1}^{\infty} b_n$ diverges. Select a number $r > 0$ such that $0 < r < \liminf_{n\to\ infty} a_n/b_n$. Then there is an integer N such that $a_n/b_n > r$ whenever $n \geq N$. Since the numbers b_n are positive, this gives $a_n \geq rb_n > 0$ for $n = N, N+1, N+2, \ldots$. Since the series $\sum_{n=N}^{\infty} b_n$ diverges, it still does after each term is multiplied by the constant r. By comparison, we conclude that $\sum_{n=N}^{\infty} a_n$ also diverges. Reintroducing the finitely many terms $a_1 + \cdots + a_{N-1}$ does not change the fact of divergence. We conclude that $\sum_{n=1}^{\infty} a_n$ diverges. ♦

◊ **2E-49.** Prove ***Raabe's test:*** If $a_n > 0$ and if $a_{n+1}/a_n \leq 1 - A/n$ for some fixed constant $A > 1$ and n sufficiently large, then $\sum a_n$ converges. Similarly, show that if $a_{n+1}/a_n \geq 1 - (1/n)$, then $\sum a_n$ diverges.

Use Raabe's test to prove convergence of the ***hypergeometric series*** whose general term is

$$a_n = \frac{\alpha(\alpha+1)\cdots(\alpha+n-1)\beta(\beta+1)\cdots(\beta+n-1)}{1\cdot 2\cdots n\cdot\gamma(\gamma+1)\cdots(\gamma+n-1)}$$

where α, β, and γ are nonnegative integers, $\gamma > \alpha+\beta$. Show that the series diverges if $\gamma < \alpha + \beta$.

Sketch. Show that $a_n = O(n^{-A})$ by considering $P_n = \prod_{k=1}^{n}(1 - A/k)$ and use Exercise 2E-47 to establish that $\log P_n = -A\log n + O(1)$. ◊

Solution. If $a_n > 0$ and if $a_{n+1}/a_n \leq 1 - A/n$ for some fixed constant $A > 1$ and n sufficiently large, then we can select an integer $N > 2A$ such that the inequality holds for $n \geq N$. Repeatedly applying the inequality $a_{n+1} \leq (1 - (A/n))a_n$, we find

$$\begin{aligned}
a_{N+1} &\leq \left(1 - \frac{A}{N}\right) a_N \\
a_{N+2} &\leq \left(1 - \frac{A}{N+1}\right) a_{N+1} \leq \left(1 - \frac{A}{N+1}\right)\left(1 - \frac{A}{N}\right) a_N \\
&\vdots \\
a_{N+k} &\leq \left(1 - \frac{A}{N+k}\right)\cdot\cdots\cdot\left(1 - \frac{A}{N+1}\right)\left(1 - \frac{A}{N}\right) a_N \\
&\vdots
\end{aligned}$$

So $0 < a_{N+k} \le P_k a_N$ where $P_k = \prod_{j=0}^{k}\left(1 - \frac{A}{N+j}\right)$, and

$$\log P_k = \sum_{j=0}^{k} \log\left(1 - \frac{A}{N+j}\right).$$

For $0 \le x \le 1/2$, we have $-(2\log 2)x \le \log(1-x) \le -x$. So

$$\log\left(1 - \frac{A}{N+j}\right) \le -\frac{A}{N+j}.$$

Adding we get

$$\log P_k \le -A\left(\frac{1}{N} + \frac{1}{N+1} + \cdots + \frac{1}{N+k}\right). \tag{1}$$

We know that

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N-1} + \frac{1}{N} + \cdots + \frac{1}{n} - \log n \to \gamma = \text{ Euler's Constant}$$

(Exercise 2E-47). So

$$\frac{1}{N} + \cdots + \frac{1}{N+k} - \log(N+k) \to \gamma - \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N-1}\right) := \Theta_N.$$

So we can select an integer M large enough so that

$$\Theta_N - 1 < \frac{1}{N} + \cdots + \frac{1}{N+k} - \log(N+k) < \Theta_N + 1$$

whenever $k \ge M$. For such k we have

$$\log(N+k) + \Theta_N - 1 < \frac{1}{N} + \cdots + \frac{1}{N+k} < \log(N+k) + \Theta_N + 1.$$

Using this in (1) we obtain

$$\log P_k < -A\left(\log(N+k) + \Theta_N - 1\right).$$

Exponentiating gives

$$0 < P_k \le (N+k)^{-A}\, e^{-A(\Theta_N - 1)} = \frac{C_N}{(N+k)^A}$$

where C_N is a positive constant. So, for $k \ge M$, we have

$$0 < a_{N+k} \le P_k a_N \le \frac{a_N C_N}{(N+k)^A}.$$

Since $A > 1$, the series

$$\sum_{k=M}^{\infty} \frac{a_N C_N}{(N+k)^A} \quad = \quad a_N C_N \sum_{k=M}^{\infty} \frac{1}{(N+k)^A} \quad = \quad a_N C_N \sum_{j=N+M}^{\infty} \frac{1}{j^A}$$

converges. So $\sum_{k=M}^{\infty} a_{N+k} = \sum_{j=N+M}^{\infty} a_j$ also converges by comparison. Reintroducing the first $N+M-1$ terms does not change the fact of convergence.

The divergence part of the test is much easier. If $a_n > 0$ and $a_{n+1}/a_n > (1-(1/n))$ for each n, then we find

$$\begin{aligned}
a_3 &\geq \frac{1}{2}\, a_2 \\
a_4 &\geq \frac{2}{3}\, a_3 \geq \frac{2}{3}\,\frac{1}{2}\, a_2 = \frac{1}{3}\, a_2 \\
a_5 &\geq \frac{3}{4}\, a_4 \geq \frac{3}{4}\,\frac{1}{3}\, a_2 = \frac{1}{4}\, a_2 \\
&\vdots \\
a_{n+1} &\geq \frac{n-1}{n}\, a_n \geq \frac{n-1}{n}\,\frac{1}{n-1}\, a_2 = \frac{1}{n}\, a_2 \\
&\vdots
\end{aligned}$$

So the series $\sum_{k=3}^{\infty} a_k = \sum_{n=2}^{\infty} a_{n+1}$ diverges by comparison to the divergent harmonic series $\displaystyle\sum_{n=2}^{\infty} \frac{a_2}{n}$. Reintroducing the first two terms does not change the fact of divergence.

If

$$a_n = \frac{\alpha(\alpha+1)\cdot \dots \cdot(\alpha+n-1)\beta(\beta+1)\cdot \dots \cdot(\beta+n-1)}{1\cdot 2\cdot \dots \cdot n\cdot\gamma(\gamma+1)\cdot \dots \cdot(\gamma+n-1)},$$

then

$$\begin{aligned}
\frac{a_{n+1}}{a_n} &= \frac{\alpha(\alpha+1)\dots(\alpha+n-1)(\alpha+n)\beta(\beta+1)\dots(\beta+n-1)(\beta+n)}{1\cdot 2\dots n\cdot(n+1)\cdot\gamma(\gamma+1)\dots(\gamma+n-1)(\gamma+n)} \\
&\qquad \cdot \frac{1\cdot 2\dots n\cdot\gamma(\gamma+1)\dots(\gamma+n-1)}{\alpha(\alpha+1)\dots(\alpha+n-1)\beta(\beta+1)\dots(\beta+n-1)} \\
&= \frac{(\alpha+n)(\beta+n)}{(n+1)(\gamma+n)} = 1 - \frac{(1+[\gamma-(\alpha+\beta)])n}{(n+1)(n+\gamma)} + \frac{\alpha\beta-\gamma}{(n+1)(\gamma+n)}.
\end{aligned}$$

If $\gamma > \alpha + \beta$, then $1 + [\gamma - (\alpha + \beta)] \geq 2$ since α, β, and γ are integers. So

$$\begin{aligned}
\frac{a_{n+1}}{a_n} &\leq 1 - \frac{2n}{(n+1)(\gamma+n)} + \frac{\alpha\beta - \gamma}{(n+1)(\gamma+n)} \\
&\leq 1 - \frac{(1.5)n}{(n+1)(\gamma+n)} - \frac{n - 2(\alpha\beta-\gamma)}{2(n+1)(\gamma+n)} \\
&\leq 1 - \frac{1.5}{n} - \frac{n - 2(\alpha\beta-\gamma)}{2(n+1)(\gamma+n)} \\
&\leq 1 - \frac{1.5}{n}
\end{aligned}$$

with the last inequality valid for $n > 2(\alpha\beta - \gamma)$. So the series $\sum_{n=1}^{\infty} a_n$ converges by Raabe's test.

If $\gamma < \alpha + \beta$, then we rewrite the expression above as

$$\frac{a_{n+1}}{a_n} = 1 + \frac{[(\alpha+\beta) - (\gamma+1)]n + (\alpha\beta - \gamma)}{(n+1)(n+\gamma)}.$$

We distinguish two cases:

FIRST: If $\alpha+\beta > \gamma+1$, then the fraction in the last expression is certainly positive for large n, and $a_{n+1}/a_n > 1 > 1-(1/n)$. The series must diverge.

SECOND: If $\alpha + \beta = \gamma + 1$, the situation is a bit more delicate.

$$\begin{aligned}
\frac{a_{n+1}}{a_n} &= 1 + \frac{\alpha\beta - \gamma}{(n+1)(n+\gamma)} = 1 - \frac{1}{n} + \frac{1}{n} + \frac{\alpha\beta-\gamma}{(n+1)(n+\gamma)} \\
&= 1 - \frac{1}{n} + \frac{(n+1)(n+\gamma) + (\alpha\beta-\gamma)n}{n(n+1)(n+\gamma)} \\
&= 1 - \frac{1}{n} + \frac{n^2 + (\alpha\beta+1)n + \gamma}{n(n+1)(n+\gamma)} \\
&\geq 1 - \frac{1}{n}.
\end{aligned}$$

The series must diverge. ♦

◇ **2E-50.** Show that for x sufficiently large, $f(x) = (x\cos^2 x + \sin^2 x)e^{x^2}$ is monotonic and tends to $+\infty$, but that neither the ratio $f(x)/(x^{1/2}e^{x^2})$ nor its reciprocal is bounded.

Sketch. First show that $f'(x) \geq 0$ for $x > 1$ so that $f(x)$ is increasing past that point. Then compute $f(n\pi)$. For the ratios, look at the values at even and odd multiples of $\pi/2$. ◇

Solution. Our function is $f(x) = (x\cos^2 x + \sin^2 x)e^{x^2}$. So

$$\begin{aligned}
f'(x) &= (-2x\cos x\sin x + \cos^2 x + 2\sin x\cos x)e^{x^2} \\
&\qquad + 2xe^{x^2}(x\cos^2 x + \sin^2 x) \\
&= e^{x^2}(\cos^2 x - x\sin(2x) + \sin(2x) + 2x^2\cos^2 x + 2x\sin^2 x).
\end{aligned}$$

If $x > 1$, we have $x^2 > x$. So

$$\begin{aligned}
f'(x) &\geq e^{x^2}(\cos^2 x - x\sin(2x) + \sin(2x) + 2x\cos^2 x + 2x\sin^2 x) \\
&\geq e^{x^2}(\cos^2 x - x\sin(2x) + \sin(2x) + 2x) \\
&\geq e^{x^2}(\cos^2 x + x(1 - \sin(2x)) + (x + \sin(2x))).
\end{aligned}$$

But $1 - \sin(2x) \geq 0$, and with $x > 1$, we have $x + \sin(2x) > 0$. So $f'(x) \geq 0$ for $x > 1$. Thus the function is monotonically increasing beyond that point. At selected points along the way it is easy to compute, $f(n\pi) = n\pi e^{n^2\pi^2}$. These values certainly tend to ∞, and $f(x)$ is monotonically increasing, so $f(x) \to +\infty$ as $x \to +\infty$.

Let

$$R(x) = f(x)/(x^{1/2}e^{x^2}) = x^{1/2}\cos^2 x + x^{-1/2}\sin^2 x.$$

If n is a positive integer, then $R(n\pi) = \sqrt{n\pi} \to +\infty$ as $n \to \infty$. On the other hand, $R((2n+1)\pi/2) = ((2n+1)\pi/2)^{-1/2} = \sqrt{2}/\sqrt{(2n+1)\pi} \to 0$. So $R((2n+1)\pi/2)^{-1} \to +\infty$ as $n \to \infty$. ♦

◊ **2E-51.** (a) If $u_n > 0$, $n = 1, 2, \ldots$, show that

$$\liminf \frac{u_{n+1}}{u_n} \leq \liminf \sqrt[n]{u_n} \leq \limsup \sqrt[n]{u_n} \leq \limsup \frac{u_{n+1}}{u_n}.$$

(b) Deduce that if $\lim(u_{n+1}/u_n) = A$, then $\limsup \sqrt[n]{u_n} = A$.

(c) Show that the converse of part (b) is false by use of the sequence $u_{2n} = u_{2n+1} = 2^{-n}$.

(d) Calculate $\limsup \sqrt[n]{n!}/n$.

Answer. (d) 1/e. ◊

Solution. (a) There are three inequalities to prove:

$$\liminf_{n\to\infty} \frac{u_{n+1}}{u_n} \leq \liminf_{n\to\infty} \sqrt[n]{u_n} \tag{1}$$

$$\liminf_{n\to\infty} \sqrt[n]{u_n} \leq \limsup_{n\to\infty} \sqrt[n]{u_n} \tag{2}$$

$$\limsup_{n\to\infty} \sqrt[n]{u_n} \leq \limsup_{n\to\infty} \frac{u_{n+1}}{u_n}. \tag{3}$$

Of these the second is true for every sequence, so we need only work on the first and third.

Proof of (1): We know that $u_n > 0$ for each n. So $u_{n+1}/u_n > 0$ and $\sqrt[n]{u_n} > 0$. Therefore $\liminf(u_{n+1}/u_n) \geq 0$ and $\liminf \sqrt[n]{u_n} \geq 0$. If $\liminf(u_{n+1}/u_n) = 0$, then inequality (1) is certainly true. So we may assume that $\liminf(u_{n+1}/u_n) > 0$. If $a < \liminf(u_{n+1}/u_n)$, we can select a number r with $a < r < \liminf(u_{n+1}/u_n)$. Then there is an integer N such that $u_{n+1}/u_n > r$ whenever $n \geq N$. Since $u_n > 0$, we have $u_{n+1} > ru_n$ for $n \geq N$. Applying this repeatedly we find

$$\begin{aligned}
u_{N+1} &\geq ru_N \\
u_{N+2} &\geq r_{N+1} \geq r^2 u_N \\
u_{N+3} &\geq r_{N+2} \geq r^3 u_N \\
&\vdots \\
u_{N+k} &\geq r_{N+k-1} \geq r^k u_N \\
&\vdots
\end{aligned}$$

So

$$\sqrt[N+k]{u_{N+k}} > \sqrt[N+k]{r^k u_N} = r^{k/(N+k)} u_N^{1/(N+k)}.$$

Now we need two facts about exponential functions: $r^x \to r$ as $x \to 1$ and $c^x \to 1$ as $x \to 0$ for positive constants r and c. Applying this to the rightmost expression in the last display, we find

$$\sqrt[N+k]{u_{N+k}} > r^{k/(N+k)} u_N^{1/(N+k)} \to r > a \qquad \text{as} \qquad k \to \infty.$$

So $\sqrt[N+k]{u_{N+k}} > a$ for sufficiently large k. That is, $\sqrt[n]{u_n} > a$ for sufficiently large n. Thus $\liminf \sqrt[n]{u_n} \geq a$. This is true for every a smaller than $\liminf u_{n+1}/u_n$. So

$$\liminf u_{n+1}/u_n \leq \liminf \sqrt[n]{u_n}$$

as claimed.

Proof of (3): The argument for the third inequality is similar to that for the first but with the inequalities reversed. If $\limsup(u_{n+1}/u_n) = +\infty$, then inequality (3) is certainly true. Therefore we may assume that $\limsup(u_{n+1}/u_n) < +\infty$. If $b > \limsup(u_{n+1}/u_n)$, we can select a number r with $b > r > \limsup(u_{n+1}/u_n)$. Then there is an integer N such that $u_{n+1}/u_n < r$ whenever $n \geq N$. Since $u_n > 0$, we have

$u_{n+1} < ru_n$ for $n \geq N$. Applying this repeatedly we find

$$\begin{aligned} u_{N+1} &\leq ru_N \\ u_{N+2} &\leq r_{N+1} \leq r^2 u_N \\ u_{N+3} &\leq r_{N+2} \leq r^3 u_N \\ &\vdots \\ u_{N+k} &\leq r_{N+k-1} \leq r^k u_N \\ &\vdots \end{aligned}$$

So

$$\sqrt[N+k]{u_{N+k}} \leq \sqrt[N+k]{r^k u_N} = r^{k/(N+k)} u_N^{1/(N+k)} \to r < b$$

as $k \to \infty$. So $\sqrt[N+k]{u_{N+k}} < b$ for sufficiently large k. That is, $\sqrt[n]{u_n} < b$ for sufficiently large n. Thus $\limsup \sqrt[n]{u_n} \leq b$. This is true for every b larger than $\limsup u_{n+1}/u_n$. So

$$\limsup \sqrt[n]{u_n} \leq \limsup u_{n+1}/u_n$$

as claimed.

(b) We know from Proposition 1.5.7(ix), that the limit of a sequence exists if and only if the limit inferior and the limit superior are the same and equal to that limit. Since $\lim_{n\to\infty}(u_{n+1}/u_n)$ has been assumed to exist and be equal to A, we know that

$$A = \liminf \frac{u_{n+1}}{u_n} \leq \liminf \sqrt[n]{u_n} \leq \limsup \sqrt[n]{u_n} \leq \limsup \frac{u_{n+1}}{u_n} = A.$$

So we must have

$$\liminf \sqrt[n]{u_n} = \limsup \sqrt[n]{u_n} = A.$$

So $\lim_{n\to\infty} \sqrt[n]{u_n}$ exists and is equal to A by 1.5.7(ix).

(c) If the sequence u_n is defined by $u_{2n} = u_{2n+1} = 2^{-n}$ for $n = 0, 1, 2, \ldots,$ then the first few terms are

$$1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \ldots, \frac{1}{2^n}, \frac{1}{2^n}, \ldots.$$

The ratios of succeeding terms, u_{n+1}/u_n, are

$$1, \frac{1}{2}, 1, \frac{1}{2}, 1, \frac{1}{2}, \ldots.$$

So $\liminf u_{n+1}/u_n = 1/2$ while $\limsup u_{n+1}/u_n = 1$. But if $k = 2n$, then

$$\sqrt[k]{u_k} = \sqrt[2n]{2^{-n}} = 2^{-n/2n} = 2^{-1/2} = 1/\sqrt{2}.$$

While if $k = 2n+1$, then

$$\sqrt[k]{u_k} = \sqrt[2n+1]{2^{-n}} = 2^{-n/(2n+1)} \to 2^{-1/2} = 1/\sqrt{2}$$

So

$$\liminf_{k\to\infty} \sqrt[k]{u_k} = \limsup_{k\to\infty} \sqrt[k]{u_k} = \lim_{k\to\infty} \sqrt[k]{u_k} = 1/\sqrt{2}.$$

So $\liminf \sqrt[k]{u_k}$ and $\limsup \sqrt[k]{u_k}$ are the same while $\liminf(u_{k+1}/u_k)$ and $\limsup(u_{k+1}/u_k)$ are different. The converse of the result of part (b) is false.

(d) Let $a_n = \sqrt[n]{n!}\,/n$ and $u_n = n!/n^n$. Then $\sqrt[n]{u_n} = a_n$. We compute

$$\frac{u_{n+1}}{u_n} = \frac{(n+1)!}{(n+1)^{n+1}}\,\frac{n^n}{n!} = \frac{n+1}{n+1}\left(\frac{n}{n+1}\right)^n = \frac{1}{\left(1+\frac{1}{n}\right)^n} \to \frac{1}{e}.$$

(To get the last limit, take logarithms and use L'Hôpital's Rule to show that $(1+(1/n))^n \to e$.) Since this limit exists, we conclude from part (b) that $\lim \sqrt[n]{u_n}$ also exists and is $1/e$. Thus $\lim_{n\to\infty} \sqrt[n]{n!}\,/n$ exists and is equal to $1/e$. ♦

◇ **2E-52.** Test the following series for convergence.

(a) $\displaystyle\sum_{k=0}^{\infty} \frac{e^{-k}}{\sqrt{k+1}}$

(b) $\displaystyle\sum_{k=0}^{\infty} \frac{k}{k^2+1}$

(c) $\displaystyle\sum_{n=0}^{\infty} \frac{\sqrt{n+1}}{n^2-3n+1}$

(d) $\displaystyle\sum_{k=1}^{\infty} \frac{\log(k+1)-\log k}{\tan^{-1}(2/k)}$

(e) $\displaystyle\sum_{n=1}^{\infty} \sin(n^{-\alpha}), \alpha$ real, > 0

(f) $\displaystyle\sum_{n=1}^{\infty} \frac{n^3}{3^n}$

Answer. (a) Converges.

(b) Diverges.

(c) Converges.

(d) Diverges.

(e) Converges if $\alpha > 1$, diverges if $\alpha \le 1$.

(f) Converges. ◊

Solution. (a) Notice that for $k \geq 0$ we have

$$0 < \frac{e^{-k}}{\sqrt{k+1}} \leq \frac{e^{-k}}{1} = \left(\frac{1}{e}\right)^k.$$

We know that the series $\sum_{k=0}^{\infty}(1/e)^k$ converges since it is a geometric series with ratio $1/e < 1$. So $\displaystyle\sum_{k=0}^{\infty} \frac{e^{-k}}{\sqrt{k+1}}$ also converges by comparison.

(b) For $k \geq 1$, we have

$$0 < \frac{1}{2k} = \frac{k}{k^2 + k^2} \leq \frac{k}{k^2+1}.$$

We know that the harmonic series $(1/2)\sum_{k=1}^{\infty}(1/k)$ diverges, so, by comparison, $\displaystyle\sum_{k=0}^{\infty} \frac{k}{k^2+1}$ also diverges. Reintroducing the $k = 0$ term does not change the fact of divergence.

(c) Let $a_n = (\sqrt{n+1})/(n^2 - 3n + 1)$ and $b_n = (n+1)^{3/2}$. Then

$$\frac{a_n}{b_n} = \frac{(n+1)^{3/2}(n+1)^{1/2}}{n^2 - 3n + 1} = \frac{n^2 + 2n + 1}{n^2 - 3n + 1} \to 1 \quad \text{as} \quad n \to \infty.$$

We know that

$$\sum_{n=0}^{\infty} b_n = \sum_{n=0}^{\infty} \frac{1}{(n+1)^{3/2}} = \sum_{k=1}^{\infty} \frac{1}{k^{3/2}}$$

converges since it is a "p-series" with $p = 3/2 > 1$. So $\displaystyle\sum_{n=0}^{\infty} \frac{\sqrt{n+1}}{n^2 - 3n + 1}$ also converges by the ratio comparison test.

(d) Use methods of first year calculus to show that $0 < (\tan^{-1} 2)x \leq \tan^{-1} x \leq x$ for $0 < x \leq 2$ and that $0 < (\log 2)x \leq \log(1 + x) \leq x$ for $0 < x \leq 1$. So, if $k \geq 1$ we have

$$0 < \frac{\log 2}{2} = \frac{(\log 2)(1/k)}{(2/k)} \leq \frac{\log(1 + (1/k))}{\tan^{-1}(2/k)} = \frac{\log(k+1) - \log k}{\tan^{-1}(2/k)}.$$

Thus the terms of the series $\displaystyle\sum_{k=1}^{\infty} \frac{\log(k+1) - \log k}{\tan^{-1}(2/k)}$ cannot converge to 0 and the series must diverge.

(e) Use methods of first year calculus to show that $0 \le (2/\pi)x \le \sin x \le x$ for $0 \le x \le \pi/2$. So, for $n \ge 1$ and $\alpha > 0$, we have

$$0 < \frac{2}{\pi}\frac{1}{n^\alpha} \le \sin\left(n^{-\alpha}\right) \le \frac{1}{n^\alpha}.$$

The series $\sum_{n=1}^\infty (1/n^\alpha)$ converges if $\alpha > 1$ and diverges if $\alpha \le 1$. Since we have comparison in both directions (at least to a constant multiple of this series), the same is true for the series $\sum_{n=1}^\infty \sin\left(n^{-\alpha}\right)$.

(f) Let $u_n = n^3/3^n$. Then $u_n > 0$ for each $n = 1, 2, 3, \ldots$, and

$$\frac{u_{n+1}}{u_n} = \frac{(n+1)^3}{3^{n+1}}\frac{3^n}{n^3} = \frac{1}{3}\left(1 + \frac{1}{n}\right)^3 \to \frac{1}{3} < 1.$$

So the series $\displaystyle\sum_{n=1}^\infty \frac{n^3}{3^n}$ converges by the ratio test. ♦

◇ **2E-53.** Given a set A in a metric space, what is the maximum number of distinct subsets that can be produced by successively applying the operations closure, interior, and complement to A (in any order)? Give an example of a set achieving your maximum.

Answer. 14. ◊

Solution. It is convenient to set up some shorthand. Let I, C, N, and E stand for the operations of taking the interior, closure, or complement of a set S in a metric space M, or of leaving it alone.

$$E(S) = S, \quad I(S) = \mathrm{int}(S), \quad C(S) = \mathrm{cl}(S), \quad \text{and ,} \quad N(S) = M \setminus S$$

We know rather a lot about these operations. For example, if S is any subset of M, then

$$\mathrm{int}(\mathrm{int}(S)) = \mathrm{int}(S), \quad \mathrm{cl}(\mathrm{cl}(S)) = \mathrm{cl}(S), \quad \text{and } M \setminus (M \setminus S) = S.$$

In our new shorthand these become

$$II(S) = I(S), \quad CC(S) = C(S), \quad \text{and} \quad NN(S) = E(S).$$

So, as operators on the subsets of M, we have

$$II = I, \quad CC = C, \quad \text{and} \quad NN = E.$$

What we want to know is how many different operators we can get by compositions of these four, in effect, by forming words from the four letters E, N, C, and I. There are infinitely many different words, but different words may give the same operator. We have just seen, for example, that the words CC and C represent the same operator on sets. Four more such facts are in the next lemma.

Lemma. *If S is a subset of a metric space M, then*

(1) $CN(S) = NI(S)$.
(2) $NC(S) = IN(S)$.
(3) $CICI(S) = CI(S)$.
(4) $ICIC(S) = IC(A)$.

Proof: For (1)

$$\begin{aligned} x \notin CN(S) &\iff x \notin \mathrm{cl}(M \setminus S) \\ &\iff D(X, r) \cap (M \setminus S) = \emptyset \text{ for some } r > 0 \\ &\iff D(x, r) \subseteq S \text{ for some } r > 0 \\ &\iff x \in \mathrm{int}(S) \\ &\iff x \in I(S). \end{aligned}$$

So $x \in CN(S) \iff x \notin I(S) \iff x \in NI(S)$.

For (2), one can give a proof like that for (1), or apply (1) to the set $N(S)$. Since $NN(S) = S$, this gives

$$C(S) = CNN(S) = CN(N(S)) = NI(N(S)).$$

So

$$NC(S) = NNI(N(S)) = I(N(S)) = IN(S).$$

For (3) we start with the fact that any set is contained in its closure applied to the interior of S. This becomes $I(S) \subseteq CI(S)$. If $A \subseteq B$, then $\mathrm{int}(A) \subseteq \mathrm{int}(B)$, and $\mathrm{cl}(A) \subseteq \mathrm{cl}(B)$. So $I(S) = I(I(S)) \subseteq ICI(S)$, and $CI(S) \subseteq CICI(S)$. But also the interior of any set is contained in the set, so, $ICI(S) \subseteq CI(S)$, and $CICI(S) \subseteq CCI(S) = CI(S)$. We have inclusion in both directions, so $CICI(S) = CI(S)$ as claimed.

The proof of (4) is similar to that of (3). Start with the fact that the interior of a set is contained in the set. $I(C(S)) \subseteq C(S)$. Now take closures: $C(I(C(S))) \subseteq C(C(S)) = C(S)$, and then interiors: $ICIC(S) = I(C(I(C(S)))) \subseteq I(C(S)) = IC(S)$. In the other direction, start with the fact that any set is contained in its closure: $IC(S) \subseteq CIC(S)$. Then take interiors: $IC(S) = I(IC(S)) \subseteq I(CIC(S)) = ICIC(S)$. We have inclusion in both directions, so the sets are equal as claimed.

Using these facts we can look at compositions of the operators as words in the letters E, N, C, and I, noting possibly new ones and grouping ones which we know produce the same operator (that is, which are the same for all sets). If we look at the sixteen possible two letter words, we see that ten of them collapse to one letter words, while the remaining six group themselves into at most four new operators.

$$\begin{array}{cccc} EE = E & NE = E & CE = C & IE = I \\ EN = E & NN = E & CN = NI & IN = NC \\ EC = C & NC = IN & CC = C & IC \\ EI = I & NI = CN & CI & II = I \end{array}$$

In addition to the four operators with which we started,

$$o_1 = E, \quad o_2 = N, \quad o_3 = C, \quad \text{and} \quad o_4 = I,$$

we have at most four new ones,

$$o_5 = NC = IN, \quad o_6 = CN = NI, \quad o_7 = CI, \quad o_8 = IC.$$

To find three letter words, we append a letter to each of the noncollapsing two letter words. Of the twenty-four resulting words, three collapse to one letter and thirteen collapse to two letters leaving eight noncollapsing three letter words.

$$\begin{array}{cccc} ENC = NC & NNC = C & CNC = NIN & INC = IIN = IN \\ EIN = IN & NIN = NNC = NC & CIN = CNC & IIN = IN \\ ECN = CN & NCN = NNI = I & CCN = CN & ICN = INI \\ ENI = NI & NNI = I & CNI = NII = NI & INI = NCI \\ ECI = CI & NCI & CCI = CI & ICI \\ EIC = IC & NIC & CIC & IIC = IC \end{array}$$

Again we have at most four new operators:

$$\begin{array}{rcl} o_9 = NCI = ICN = INI & , & o_{10} = NIC = CIN = CNC, \\ o_{11} = CIC & , & o_{12} = ICI. \end{array}$$

To seek four letter words which might give new operators, we append a letter to one representative from each of the sets of equal three letter words. This will mean that we do not list all four letter words, but those we do not list will be equal as operators to something among those we do list. Also, we do not bother with the column appending E since it always disappears as before.

$$\begin{array}{ccc} NNCI = CI & CNCI = NICI & INCI = NCCI = NCI \\ NNIC = IC & CNIC = NIIC = NIC & INIC = NCIC \\ NCIC & CCIC = CIC & ICIC = IC \\ NICI & CICI = CI & IICI = ICI \end{array}$$

Notice where we have used parts (4) and (5) of the lemma. We find at most two new operators.

$$o_{13} = NCIC = INIC \quad \text{and} \quad o_{14} = NICI = CNCI$$

To see if we get any new operators with five letters, it is enough to try adding N, C, or I to one of the four letter words giving each of o_{13} and o_{14}.

$$\begin{array}{cc}
NNCIC = CIC & NNICI = ICI \\
CNCIC = NICIC = NIC & CNICI = NIICI = NICI \\
INCIC = NCCIC = NCIC & INICI = NCICI = NCI
\end{array}$$

Since these collapse, all five letter words can be collapsed to four or fewer letters. We get no new operators.

The argument above shows that composition of the operations of complementation, closure, and interior, together with the identity operator, generate at most fourteen different operators on the powerset of a metric space. Some of these might be equal in some metric spaces. If the metric space has only three points, then there cannot be very many different operators generated. In any set with the discrete metric all sets are both open and closed, so the closure operator and the interior operators are the same as the identity operator, and we get only two operators. But in most reasonable spaces they are different. In fact, even in $\mathbb{R}$ we can get one set S such that the fourteen sets $o_k(S)$ for $1 \le k \le 14$ are all different. The set must be fairly complicated to keep generating new sets up through the four letter words above. But it need not be completely outrageous. One example which works is

$$S = \{-\frac{1}{n} \mid n \in \mathbb{N}\} \;\cup\; [0,1[\;\cup\; \Big( ]1,2[\setminus \{1 + \frac{1}{n} \mid n \in \mathbb{N}\} \Big) \;\cup\; (]2,3[\cap \mathbb{Q}) \,.$$

Possibly the easiest way to see that this works is to sketch the fourteen sets on the line. See Figure 2-17. ♦

FIGURE 2-17. The 14 sets resulting from the example S in Exercise 2E-53

3

Compact and Connected Sets

3.1 Compactness

◇ **3.1-1.** Show that $A \subset M$ is sequentially compact iff every infinite subset of A has an accumulation point in A.

Sketch. Let $\langle x_n \rangle_1^\infty$ be a sequence in A. If $\{x_n\}$ is a finite set, some entry must be repeated infinitely often. Use it to get a convergent subsequence. If it is an infinite set, then an accumulation point must be the limit of a subsequence. (Careful, this is trickier than it looks. We need the limit of a subsequence, not just any sequence.) ◊

Solution. First suppose A is sequentially compact and that B is an infinite subset of A.

CLAIM ONE: We can pick a sequence $x_1, x_2, x_3, \ldots$ of points in B all of which are different.

This is basically what it means for the set B to be infinite. To be more precise, no finite subset can be all of B. So, pick any point $x_1 \in B$. Since B is infinite, $B \setminus \{x_1\}$ is not empty, and we can pick $x_2 \in B \setminus \{x_1\}$. Thus x_1 and x_2 are in B and x_1 is not equal to x_2. Continue in this way inductively. Having selected $x_1, x_2, \ldots, x_k$ in B all different, the set $B \setminus \{x_1, \ldots, x_k\}$ is not empty since B is infinite. So we can pick a point $x_{k+1} \in B$ different from all of $x_1, \ldots, x_k$. Continuing this process generates a sequence $x_1, x_2, x_3, \ldots$ of points in B which are all different.

From the sequence of distinct points in B obtained in CLAIM ONE, we can select a subsequence converging to a point in B since B has been assumed to be sequentially compact. That is, there are indices $n(1) < n(2) < n(3) < \dots$ and a point $x \in B$ with $\lim_{k\to\infty} x_{n(k)} = x$. The fact that x is an accumulation point of B now follows from this lemma:

Lemma. *If B is a subset of a metric space M, then a point $x \in M$ is an accumulation point of B if and only if there is a sequence of distinct points in B which converges to x.*

See Exercise 2E-4.

Now for the reverse implication: we want to assume that every infinite subset of A has an accumulation point in A and show that A must be sequentially compact. To this end, let $x_1, x_2, x_3, \dots$ be a sequence of points in$A$. Let $B = \{x_1, x_2, x_3, \dots\}$. If B is actually a finite set, then at least one point must be repeated infinitely often in the sequence. There would be a point x in B and infinitely many indices $n(1) < n(2) < \dots$ with $x_{n(k)} = x$ for all k. Then we trivially have $x_{n(k)} \to x$ as $k \to \infty$. If B is an infinite set, then it has an accumulation point x in A by hypothesis.

CLAIM TWO: The sequence $\langle x_n \rangle_1^\infty$ has a subsequence converging to x.

(Caution: This does not quite follow from the lemma given above. B is the set of points listed in the sequence, but we need not just a sequence from B converging to x. The sequence must appear as a *subsequence* of $\langle x_n \rangle_1^\infty$. That is, we need to make sure that the entries appear in the correct order.) To obtain a subsequence converging to x, first select an index $n(1)$ with $x_{n(1)} \in B$ and $0 < d(x, x_{n(1)}) < 1/2$. Then, since there are infinitely many points in B with distance to x less than $1/4$, we can select an index $n(2) > n(1)$ with $x_{n(2)} \in B$ and $0 < d(x, x_{n(2)}) < 1/4$. We continue in this way inductively. Having selected indices $n(1) < n(2) < \cdots < n(k)$ with $x_{n(j)} \in B$ and $0 < d(x, x_{n(j)}) < 1/2^j$ for $1 \le j \le k$, we note that there are infinitely many points left in B with distance to x less than $1/2^{j+1}$. So we can select an index $n(k+1) > n(k)$ with $x_{n(k+1)} \in B$ and $0 < d(x, x_{n(k+1)}) < 1/2^{k+1}$. This process generates indices $n(1) < n(2) < n(3) \dots$ with $0 < d(x, x_{n(k+1)}) < 1/2^{k+1}$. So this gives a subsequence converging to x. Each sequence in A has a subsequence converging to a point in A. So A is sequentially compact. ♦

◇ **3.1-2.** Prove that $\{(x, y) \in \mathbb{R}^2 \mid 0 \le x < 1, 0 \le y \le 1\}$ is not compact.

Sketch. The set is not closed. ◇

Solution. From the remark following Theorem 3.1.5 we know that for a set to be compact it must be closed. However, the set $A = \{(x, y) \in \mathbb{R}^2 \mid 0 \le x < 1, 0 \le y \le 1\}$ is not closed. It is a square in $\mathbb{R}^2$. The vertical edge

$\{(1, y) \mid 0 \le y \le 1\}$ is in the boundary of S but not in S. So S is not closed and cannot be compact. ♦

◇ **3.1-3.** Let M be complete and $A \subset M$ be totally bounded. Show that $\mathrm{cl}(A)$ is compact.

Sketch. The set $\mathrm{cl}(A)$ is closed. To show that it is totally bounded, let $\varepsilon > 0$ and cover A by finitely many $\varepsilon/2$-balls. Show that $\mathrm{cl}(A)$ is covered by the corresponding ε-balls. ◇

Solution. We know that a subset K of a complete metric space M is compact if and only if it is closed and totally bounded. Let $K = \mathrm{cl}(A)$. Since the closure of a set is always closed, we know that K is closed. It remains only to show that it is totally bounded. Let $\varepsilon > 0$, then $\varepsilon/2 > 0$, and the totally bounded set A can be covered with finitely many $\varepsilon/2$-balls centered in A. There is a finite set $\{a_1, a_2, \ldots, a_n\} \subseteq A$ such that $A \subseteq \bigcup_{k=1}^{n} D(a_k, \varepsilon/2)$. If $y \in \mathrm{cl}(A)$, then there is a point x in A with $d(x, y) < \varepsilon/2)$. There is at least one k with $1 \le k \le n$ and $x \in D(a_k, \varepsilon/2)$. So $d(y, a_k) \le d(y, x) + d(x, a_k) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. So $y \in D(a_k, \varepsilon)$. Since this can be done for every y in $\mathrm{cl}(A)$, we find that $K = \mathrm{cl}(A) \subseteq \bigcup_{k=1}^{n} D(a_k, \varepsilon)$. Thus K is totally bounded. Since we already know that it is a closed subset of the complete metric space M, it must be compact. ♦

◇ **3.1-4.** Let $x_k \to x$ be a convergent sequence in a metric space and let $A = \{x_1, x_2, \ldots\} \cup \{x\}$.

(a) Show that A is compact.
(b) Verify that every open cover of A has a finite subcover.

Suggestion. Parts (a) and (b) might as well be done together since (b) is the definition of compactness. ◇

Solution. (a) Since the assertion in part (b) is actually the definition of compactness, we could do both parts of the problem together by doing (b). For interest, we present an argument for part (a) using Theorem 3.1.5. When we are all done, the reader may agree that checking the abstract definition seems easier.

We know that the set A is closed from previous work. It is also totally bounded. If $\varepsilon > 0$, then There is an N such that $d(x_k, x) < \varepsilon$ whenever $k \ge N$. So $x_k \in D(x, \varepsilon)$ for such n. Since x is also in $D(x, \varepsilon)$, and $x_j \in D(xj, \varepsilon)$, we have $A \subseteq D(x, \varepsilon) \cup \bigcup_{j=1}^{N-1} D(x_j, \varepsilon)$. Thus A is totally bounded. So A is a closed, totally bounded subset of the metric space M. If we knew that M was complete, this would show that A is compact.

But unfortunately we do not know that. Thus to pursue this method we need to show that A is complete even though M might not be.
Let $\langle y_n \rangle_1^\infty$ be a Cauchy sequence in A. So each of the points y_n is either equal to x or to one of the x_k. We need to show that this sequence converges to some point in A. Let $\varepsilon > 0$. Then there are indices K and N such that $d(y_{n+p}, y_n) < \varepsilon/2$ and $d(x_k, x) < \varepsilon/2$ whenever $n \geq N$, $k \geq K$, and $p > 0$. We distinguish three cases.
CASE ONE: If y_n is eventually equal to x_{k_0} for some index k_0, then the sequence $\langle y_n \rangle_1^\infty$ certainly converges to x_{k_0}.
CASE TWO: If all but finitely many of the y_n are among the finitely many points $\{x_1, x_2, \ldots x_K\}$, then there are only finitely many different distances involved among them. If they are all 0, then we are in CASE ONE. If all but finitely many are 0, we are still in CASE ONE. If not, then one of the finitely many nonzero distances must occur infinitely often and the sequence could not have been a Cauchy sequence.
CASE THREE: If we are not in CASE ONE or CASE TWO, then there are infinitely many indices n for which y_n is not in the set $\{x_1, x_2, \ldots x_K\}$. We can pick an index $n_1 > N$ such that $y_{n_1} = x_{k_1}$ for an index $k_1 > K$. If $n \geq n_1$, we have

$$\begin{aligned} d(y_n, x) \leq d(y_n, y_{n_1}) + d(y_{n_1}, x) &= d(y_n, y_{n_1}) + d(x_{k_1}, x) \\ &\leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

So, in this case, $y_n \to x$. Every Cauchy sequence in A converges to a point in A. So A is complete. The set A is a complete, totally bounded subset of the metric space M, so it is a complete, totally bounded metric space in its own right and is compact by Theorem 3.1.5.

(b) This assertion is the definition of compactness. Suppose $\{U_\beta\}_{\beta \in B}$ is any collection of open subsets of M with $A \subseteq \bigcup_{\beta \in B} U_\beta$. (The set B is just any convenient set of indices for labeling the sets.) Then there is at least one index β_0 such that $x \in U_{\beta_0}$. Since U_{β_0} is open and $x_k \to x$, there is an index K such that $x_k \in U_{\beta_0}$ whenever $k \geq K$. For the finitely many points $x_1, x_2, \ldots, x_{K-1}$, there are indices $U_{\beta_1}, U_{\beta_2}, \ldots, U_{\beta_{K-1}}$ such that $x_k \in U_{\beta_k}$. So

$$A = \{x_1, x_2, \ldots\} \cup \{x\} \subseteq U_{\beta_0} \cup U_{\beta_1} \cup U_{\beta_2} \cup \cdots \cup U_{\beta_{K-1}}.$$

Thus every open cover of A has a finite subcover, so A is compact. ♦

◇ **3.1-5.** Let M be a set with the discrete metric. Show that any infinite subset of M is noncompact. Why does this not contradict the statement in Exercise 3.1-4?

Sketch. A sequence converges if and only if it is eventually constant. This does not contradict Exercise 3.1-4 since the entries in such a convergent sequence form a finite set. ♢

Solution. Let A be an infinite subset of M. Since we are using the discrete metric, single point sets are open. In fact, $\{x\} = D(x, 1/2)$ We certainly have $A = \bigcup_{x\in A}\{x\} = \bigcup_{x\in A} D(x, 1/2)$. This is an open cover of the set A which uses infinitely many sets, and, since they are pairwise disjoint, we cannot omit any of them. There can be no finite subcover. Thus A is not compact.

If $\langle x_k\rangle_1^\infty$ is a sequence in A converging to x, then there must be an index K such that $d(x, x_k) < 1/2$ whenever $k \geq K$. But since we are using the discrete metric which only has the values 0 and 1, this forces $d(x, x_k) = 0$ and so $x_k = x$. Thus $x_k = x$ for all $k \geq K$ and there are only finitely many points in the set $\{x_1, x_2 \ldots\} \cup \{x\}$. Such a finite set of points is compact, and the statement of Exercise 3.1-4 is not contradicted. ♦

3.2 The Heine-Borel Theorem

⋄ **3.2-1.** Which of the following sets are compact?

(a) $\{x \in \mathbb{R} \mid 0 \leq x \leq 1 \text{ and } x \text{ is irrational}\}$
(b) $\{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1\}$
(c) $\{(x, y) \in \mathbb{R}^2 \mid xy \geq 1\} \cap \{(x, y) \mid x^2 + y^2 < 5\}$

Answer. None of them. ♢

Solution. (a) No. The point 1/2 is not in the set $A = \{x \in \mathbb{R} \mid 0 \leq x \leq 1 \text{ and } x \text{ is irrational }\}$, but every interval around it contains irrational numbers which are in A. The complement of A is not open. The set A is not closed. It cannot be compact.

(b) No. The set $B = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1\}$ is not bounded. It is an infinite "vertical" strip containing the points $(1/2, y)$ for arbitrarily large y. Since B is not bounded, it cannot be compact.

(c) No. The set $C = \{(x, y) \in \mathbb{R}^2 \mid xy \geq 1\} \cap \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 5\}$ is bounded, but it is not closed. The point $(\sqrt{5/2}, \sqrt{5/2})$ is on the boundary of C but not in C. Since C is not closed it cannot be compact. ♦

⋄ **3.2-2.** Let $r_1, r_2, r_3, \ldots$ be an enumeration of the rational numbers in [0, 1]. Show that there is a convergent subsequence.

Suggestion. Use the Heine-Borel Theorem or the Bolzano-Weierstrass Property directly. ♢

Solution. The rationals in the set $[0,1]$ arranged as a sequence $\langle r_k \rangle_1^\infty$ are a sequence in the closed bounded subset $[0,1]$ of $\mathbb{R} = \mathbb{R}^1$. Since it is closed and bounded, this set is compact by the Heine-Borel Theorem. It is thus sequentially compact by the Bolzano-Weierstrass Theorem. Thus the sequence must have a convergent subsequence and the limit of that subsequence must be in $[0,1]$.

Alternatively, the sequence is a bounded sequence in $\mathbb{R}$ since it is contained in the interval $[0,1]$. It must have a convergent subsequence by the Bolzano-Weierstrass Property (Theorem 1.4.3). The limit of this subsequence must be in $[0,1]$ since that set is closed. ♦

⋄ **3.2-3.** Let $M = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ with the standard metric. Show that $A \subset M$ is compact iff A is closed.

Sketch. All subsets of M are bounded, so A compact $\iff$ (A closed and bounded) $\iff$ A closed. ♢

Solution. Suppose A is a closed subset of M. If $(x,y) \in A$, then

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

so A is bounded. Since it is a closed and bounded subset of $\mathbb{R}^2$, it is compact by the Heine-Borel Theorem.

Suppose A is compact. Then A must be closed and bounded by the Heine-Borel Theorem, so in particular it must be closed.

Thus a subset A of M is compact if and only if it is closed.

Notice that the only thing we have used here about M is that it is a bounded subset of $\mathbb{R}^n$ for some finite n.

Proposition. *If M is a bounded subset of $\mathbb{R}^n$, then a subset A of M is compact if and only if it is closed.* ♦

⋄ **3.2-4.** Let A be a bounded set in $\mathbb{R}^n$. Prove that $\mathrm{cl}(A)$ is compact.

Suggestion. Show that if $\| v \| \leq r$ for every $v \in A$, then $\| w \| \leq r + 1$ for every $w \in \mathrm{cl}(A)$. ♢

Solution. Suppose that A is a bounded set in $\mathbb{R}^n$. Since $\mathrm{cl}(A)$ is a closed set in $\mathbb{R}^n$, it will be compact by the Heine-Borel Theorem as soon as we show that it is bounded. Let r be a number large enough so that $\| v \| \leq r$

for every v in A. Suppose $w \in \mathrm{cl}(A)$. Then there is a point v in $D(w,1) \cap A$. Since $v \in A$, we have $\| v \| \le r$, and

$$\| w \| = \| w - v + v \| \le \| w - v \| + \| v \| \le \| w - v \| + 1 < r + 1.$$

Since this holds for every w in $\mathrm{cl}(A)$, we see that $\mathrm{cl}(A)$ is a bounded subset of $\mathbb{R}^n$. Since it is also closed, it is compact by the Heine-Borel Theorem. ♦

◇ **3.2-5.** Let A be an infinite set in $\mathbb{R}$ with a single accumulation point in A. Must A be compact?

Answer. No. ◊

Solution. The set might not be bounded. Consider the example

$$A = \{0, 1, 2, 1/2, 3, 1/3, 4, 1/4, 5, \ldots\}.$$

There is one accumulation point, 0, which is in A. But A is not bounded, so it is not compact.

The wording of the question is a bit ambiguous. It could be taken to allow the possibility that there was another accumulation point which was not in A. That would also prevent the set from being compact. That route to noncompactness would have been eliminated by a wording something like: "Let A be an infinite set in $\mathbb{R}^n$ with a single accumulation point which is in A. Must A be compact?" ♦

3.3 Nested Set Property

◇ **3.3-1.** Verify the nested set property for $F_k = \{x \in \mathbb{R} \mid x \ge 0,\ 2 \le x^2 \le 2 + 1/k\}$.

Sketch. $\bigcap_k F_k = \{\sqrt{2}\}$ is not empty. ◊

Solution. The sets in question are

$$F_k = \{x \in \mathbb{R} \mid x \ge 0, 2 \le x^2 \le 2 + (1/k)\} = [\sqrt{2}, \sqrt{2 + (1/k)}]$$

for $k = 1, 2, 3, \ldots$. These closed bounded intervals in $\mathbb{R}$ are all compact by the Heine-Borel Theorem, and, since $\sqrt{2 + (1/(k+1))} < \sqrt{2 + (1/k)}$, we have

$$F_1 \supseteq F_2 \supseteq F_3 \supseteq F_4 \cdots \supseteq F_k \supseteq F_{k+1} \supseteq \ldots.$$

According to the nested set property, Theorem 3.3.1, there should be at least one point in their intersection. There is: $\sqrt{2}$ is in all of them. ♦

◇ **3.3-2.** Is the nested set property true if "compact nonempty" is replaced by "open bounded nonempty"?

Answer. No. ◊

Solution. The words "compact nonempty" may not be replaced by "open bounded nonempty" in the nested set property and still retain a valid assertion. The modified statement would be:

Conjecture. *Let F_k be a sequence of nonempty open bounded subsets of a metric space M such that $F_{k+1} \subseteq F_k$ for each k. Then there is at least one point in $\bigcap_{k=1}^{\infty} F_k$.*

This assertion is false. For example, we could use the sequence of open intervals $F_k =]0, 1/k[= \{x \in \mathbb{R} \mid 0 < x < 1/k\}$ for $k = 1, 2, 3, \ldots$. Each of these is a nonempty open interval, and

$$F_1 \supseteq F_2 \supseteq F_3 \supseteq \cdots \supseteq F_k \supseteq F_{k+1} \supseteq \ldots.$$

But there is no real number which is in all of these intervals (Archimedean Property). So the intersection is empty. ♦

◇ **3.3-3.** Let $x_k \to x$ be a convergent sequence in a metric space. Verify the validity of the nested set property for $F_k = \{x_l \mid l \geq k\} \cup \{x\}$. What happens if $F_k = \{x_l \mid l \geq k\}$?

Sketch. If $F_k = \{x_l \mid l \geq k\}$, then $\bigcap_k F_k = \emptyset$. None of the sets F_k are compact. ◊

Solution. Since $x_l \to x$ as $l \to \infty$, each of the sets $F_k = \{x_l \mid l \geq k\} \cup \{x\}$ is compact by Exercise 3.1-1. (Alternatively, they are bounded since convergent sequences are bounded and closed since they contain the limit point x. So they are compact by the Heine-Borel Theorem.) Since each of the sets is obtained from the previous one by deleting the initial remaining term of the sequence, they are nested. According to the nested set property there ought to be at least one point in the intersection, and indeed there is. The limit x is in all of them.

If instead we use the sets $F_k = \{x_l \mid l \geq k\}$, then they are still nested, but they might not be compact since the limit x is not guaranteed to be in them. (It might be if it happens to be equal to x_l for infinitely many indices l, but it need not be. Consider the sequence given by $x_k = 1/k$ for $k = 1, 2, 3, \ldots$. Then the revised F_k would be $F_k = \{1/k, 1/(k+1), 1/(k+2), \ldots\}$. The intersection of these is empty. ♦

◇ **3.3-4.** Let $x_k \to x$ be a convergent sequence in a metric space. Let $\mathcal{A}$ be a family of closed sets with the property that for each $A \in \mathcal{A}$, there is an N such that $k \geq N$ implies $x_k \in A$. Prove that $x \in \cap \mathcal{A}$.

Solution. Let A be one of the sets in the collection $\mathcal{A}$. Then there is an N such that $x_N, x_{N+1}, x_{N+2}, x_{N+3}, \ldots$ are all in A. Since $x_k \to x$ as $k \to \infty$, this sequence of points in A converge to x. Since A is closed, we have $x \in A$. Since this is true for every set A in the collection $\mathcal{A}$, we have $x \in \bigcap \mathcal{A}$ as claimed. ♦

3.4 Path-Connected Sets

◇ **3.4-1.** Determine which of the following sets are path-connected:

(a) $\{x \in [0,1] \mid x \text{ is rational}\}$
(b) $\{(x,y) \in \mathbb{R}^2 \mid xy \geq 1 \text{ and } x > 1\} \cup \{(x,y) \in \mathbb{R}^2 \mid xy \leq 1 \text{ and } x \leq 1\}$
(c) $\{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 \leq z\} \cup \{(x,y,z) \mid x^2 + y^2 + z^2 > 3\}$
(d) $\{(x,y) \in \mathbb{R}^2 \mid 0 \leq x < 1\} \cup \{(x,0) \mid 1 < x < 2\}$

Sketch. (a) Not path-connected. Any path between two rationals must contain an irrational.

(b) Path-connected.

(c) Path-connected.

(d) Not path-connected. If the point $(0,1)$ were added, it would be path-connected. ◊

Solution. (a) The set $A = \{x \in [0,1] \mid x \in \mathbb{Q}\}$ is not path-connected. This is probably intuitively clear since any path from 0 to 1 would have to cross through irrational points such as $1/\sqrt{2}$ which are not in A. Actually proving this is somewhat delicate. See Exercise 3.4-2.

(b) The set $B = \{(x,y) \in \mathbb{R}^2 \mid xy \geq 1 \text{ and } x > 1\} \cup \{(x,y) \in \mathbb{R}^2 \mid xy \leq 1 \text{ and } x \leq 1\}$ is path-connected. It consists of two solid regions linked at the point $(1,1)$. Sketch it.

(c) The set $C = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 \leq z\} \cup \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 > 3\}$ is path-connected. It is the set of points inside the paraboloid of revolution $\{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 \leq z\}$ lying outside the sphere $\{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 3\}$. Since the paraboloid extends only above the xy-plane, $z \geq 0$, this solid set has only one piece and is path-connected.

(d) The set $D = \{(x, y) \in \mathbb{R}^2 \mid 0 \le x < 1\} \cup \{(x, 0) \in \mathbb{R}^2 \mid 1 < x < 2\}$ is not path-connected. The first term of the union is a vertical strip not containing any of its right boundary line, $\{(x, y) \mid x = 1\}$. The second term is the line segment along the x–axis from $(1, 0)$ to $(2, 0)$ but not including either end. The point $(1, 0)$ is not in the union, but any plausible continuous path from a point in the strip to a point in the segment would have to go through it. ♦

◇ **3.4-2.** Let $A \subset \mathbb{R}$ be path-connected. Give plausible arguments that A must be an interval (closed, open, or half-open). Are things as simple in $\mathbb{R}^2$?

Sketch. Suppose c and d are in A and that $\varphi : [a, b] \to A$ is a continuous path with $\varphi(a) = c$ and $\varphi(b) = d$. Try to give a convincing argument that $\varphi([a, b])$ must contain every point between c and d. This is basically the Intermediate Value Theorem. Our proof of this will be given with a bit more generality after more work on connected sets at the end of this chapter and continuous functions in the next. ◇

Solution. Suppose that c and d are in A and that $c < d$. Since A is path-connected, there should be a continuous path $\varphi : [a, b] \to A$ from some interval to A with $\varphi(a) = c$ and $\varphi(b) = d$. It seems plausible that such a continuous path cannot get from c to d without crossing over every point between. Since these points must be in A, we would have: $c < w < d$ and c and d in A implies w in A. This would force A to be an interval, a half line, or the whole line. The assertion we need is basically the Intermediate Value Theorem, probably familiar from a study of calculus.

Theorem. *If φ is a continuous, real valued function on the interval $[a, b]$, and w is between $\varphi(a)$ and $\varphi(b)$, then there is at least one point z in $[a, b]$ with $\varphi(z) = w$.*

Geometrically it says that a continuous curve cannot get from one side of a horizontal line to the other without crossing the line. Our proof of this will be given with somewhat greater generality in §4.5 after more study of connected sets in the next section and of continuous functions in Chapter 4. Plausible as it sounds, there is something at least moderately subtle going on. Any proof must involve the completeness of the real line somewhere. If the line were not complete, then the curve could sneak through at a "missing point" such as $\sqrt{2}$. Where does the curve $y = x^2 - 2$ cross the horizontal line $y = 0$? ♦

◇ **3.4-3.** Let $\varphi : [a, b] \to \mathbb{R}^3$ be a continuous path and $a < c < d < b$. Let $C = \{\varphi(t) \mid c \le t \le d\}$. Must $\varphi^{-1}(C)$ be path-connected?

Answer. No. ♢

Solution. No, the inverse image $\varphi^{-1}(C)$ need not be path-connected. For example, let $\varphi : [0,4] \to \mathbb{R}^3$ be a path which wraps twice around the unit circle in the xy-plane:

$$\varphi(t) = (\cos \pi t, \sin \pi t, 0) \qquad 0 \le t \le 4.$$

Then the intervals $[0,1]$ and $[2,3]$ map to the upper half circle while $[1,2]$ and $[3,4]$ map to the lower half circle. Take $c = 2$ and $d = 3$. Then C is the upper half circle. But $\varphi^{-1}(C) = [0,1] \cup [2,3]$ which is not connected. ♦

3.5 Connected Sets

◇ **3.5-1.** Is $[0,1] \cup]2,3]$ connected? Prove or disprove.

Answer. No. ♢

Solution. Let $A = [0,1] \cup]2,3]$. Put $U =]-1/2, 3/2[$ and $V =]2, 7/2[$. Then U and V are open sets, $A \subseteq U \cup V$, and $U \cap V = \emptyset$. But $1/2 \in A \cap U$, and $5/2 \in A \cap V$, so neither of these sets is empty. So this pair of open sets disconnects A. A is not connected. ♦

◇ **3.5-2.** Is $\{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1\} \cup \{(x,0) \mid 1 < x < 2\}$ connected? Prove or disprove.

Answer. Connected. ♢

Solution. Let $A = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1\}$ and $B = \{(x,0) \mid 1 < x < 2\}$. Set $C = A \cup B$. The set B is certainly path-connected, hence connected. If U and V were any pair of open sets which proposed to disconnect C, then B would have to be entirely contained within one of them or they would disconnect B. Say $B \subseteq V$. Now the set A is also path-connected. We can get from any point (x,y) in A to any other point (c,d) in A by proceeding horizontally from (x,y) to (x,d) and then vertically from (x,d) to (c,d). So A is also connected and must be completely contained within one of the sets U or V. If $A \subseteq V$, then we would have $C = A \cup B \subseteq V$, so the pair $\{U,V\}$ would not disconnect C. On the other hand, if $A \subseteq U$, then the point $(1,0)$ would be in U since it is in A. But then $(x,0)$ would be in U for x slightly larger than 1 since U is open. But these points are in B, so they are also in V. That is, for x slightly larger than 1 we would have $(x,0) \in U \cap V \cap C$. Once again, the sets U and V fail to disconnect C. We conclude that C must be connected. ♦

◇ **3.5-3.** Let $A \subset \mathbb{R}^2$ be path-connected. Regarding A as a subset of the xy-plane in $\mathbb{R}^3$, show that A is still path-connected. Can you make a similar argument for A connected?

Sketch. If $A \subseteq \mathbb{R}^2$, let $A^* = \{(x,y,0) \in \mathbb{R}^3 \mid (x,y) \in A\}$. If $\gamma(t) = (x(t), y(t))$ is a path in A, then $\gamma^*(t) = (x(t), y(t), 0)$ is a path in A^*. If U^* and V^* disconnect A^*, then $U = \{(x,y) \in \mathbb{R}^2 \mid (x,y,0) \in U^*\}$ and $V = \{(x,y) \in \mathbb{R}^2 \mid (x,y,0) \in V^*\}$ would disconnect A. ♢

Solution. If $A \subseteq \mathbb{R}^2$, let $A^* = \{(x,y,0) \in \mathbb{R}^3 \mid (x,y) \in A\}$. If $\gamma(t) = (x(t), y(t))$ is a path in A, then $\gamma^*(t) = (x(t), y(t), 0)$ is a path in A^*. Furthermore, $\gamma(t_k) = (x(t_k), y(t_k)) \to (a,b) \in A$ if and only if $\gamma^*(t_k) = (x(t_k), y(t_k), 0) \to (a,b,0) \in A^*$. So γ^* is continuous if and only if γ is continuous. If $P^* = (a,b,0)$ and $Q^* = (c,d,0)$ are in A^*, then $P = (a,b)$ and $Q = (c,d)$ are in A. Since A is path-connected, there is a continuous path γ in A from P to Q. The corresponding γ^* is a continuous path in A^* from P^* to Q^*. So A^* is path-connected.

A similar argument can be made for a connected set A. Suppose U^* and V^* were open sets in $\mathbb{R}^3$ which disconnected A^*. Put $U = \{(x,y) \in \mathbb{R}^2 \mid (x,y,0) \in U^*\}$ and $V = \{(x,y) \in \mathbb{R}^2 \mid (x,y,0) \in V^*\}$. Then U and V are open subsets of $\mathbb{R}^2$. (Why?)

If $(x,y) \in U \cap V \cap A$, then $(x,y,0)$ would be in $U^* \cap V^* \cap A^*$. But that set was to be empty. So $U \cap V \cap A = \emptyset$.

Since $A^* \cap U^*$ is not empty, there is a point $(x,y,0)$ in it. So $(x,y) \in A \cap U$, and $A \cap U$ is not empty.

Since $A^* \cap V^*$ is not empty, there is a point $(x,y,0)$ in it. So $(x,y) \in A \cap V$, and $A \cap V$ is not empty.

If $(x,y) \in A$, then $(x,y,0) \in A^*$, so $(x,y,0) \in U^* \cup V^*$, and $(x,y) \in U \cup V$. Thus $A \subseteq U \cup V$.

The last four observations taken together say that the sets U and V disconnect A. But A was assumed to be connected, so this is not possible. Thus no such sets U^* and V^* can exist and A^* must be connected. ♦

◇ **3.5-4.** Discuss the components of

(a) $[0,1] \cup [2,3] \subset \mathbb{R}$.
(b) $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\} \subset \mathbb{R}$.
(c) $\{x \in [0,1] \mid x \text{ is rational}\} \subset \mathbb{R}$.

Answer. (a) The connected components are the sets $[0,1]$ and $[2,3]$.
(b) The connected components are the single point sets $\{n\}$ such that $n \in \mathbb{Z}$.
(c) The connected components are the single point sets $\{r\}$ such that $r \in \mathbb{Q} \cap [0,1]$. ♢

Solution. (a) Let $A = [0,1] \cup [2,3]$. Each of the intervals $[0,1]$ and $[2,3]$ is connected. But any larger subset of A cannot be connected. For example, if $[0,1] \subseteq D \subseteq A$, and there is a point $x \in D \cap [2,3]$, then the open sets $U = \{x \in \mathbb{R} \mid x < 3/2\}$ and $V = \{x \in \mathbb{R} \mid x > 3/2\}$ would disconnect D. Thus $[0,1]$ is a maximal connected subset of A and is one of its connected components. Similarly, if $[2,3] \subseteq D \subseteq A$ and there is a point $x \in D \cap [0,1]$, then the same two sets would disconnect D. So $[2,3]$ is a maximal connected subset of A and is a connected component of A. Thus the connected components of A are the intervals $[0,1]$ and $[2,3]$.

(b) Let $B = \mathbb{Z} \subseteq \mathbb{R}$. The single point sets $\{n\}$ with $n \in \mathbb{Z}$ are certainly connected. But, if $D \subseteq B$ has two different integers $n < m$ in it, then the open sets $U = \{x \in \mathbb{R} \mid x < n+(1/2)\}$ and $V = \{x \in \mathbb{R} \mid x > n+(1/2)\}$ would disconnect D. So no connected subset of B can contain more than one point. The connected components are the single point sets $\{n\}$ such that $n \in \mathbb{Z}$.

(c) Let $C = \{x \in [0,1] \mid x \in \mathbb{Q}\} \subseteq \mathbb{R}$. The single point sets $\{r\}$ with $r \in \mathbb{Q} \cap [0,1]$ are certainly connected. But, if $D \subseteq C$ has two different rational numbers $r < s$ in it, then there is an irrational number z with $r < z < s$. ($z = r + (s-r)/\sqrt{2}$ will do.) The open sets $U = \{x \in \mathbb{R} \mid x < z\}$ and $V = \{x \in \mathbb{R} \mid x > z\}$ would disconnect D. So no connected subset of C can contain more than one point. The connected components are the single point sets $\{r\}$ such that $r \in \mathbb{Q}$. ♦

Exercises for Chapter 3

◇ **3E-1.** Which of the following sets are compact? Which are connected?

(a) $\{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1| \leq 1\}$
(b) $\{x \in \mathbb{R}^n \mid ||x|| \leq 10\}$
(c) $\{x \in \mathbb{R}^n \mid 1 \leq ||x|| \leq 2\}$
(d) $\mathbb{Z} = \{\text{integers in } \mathbb{R}\}$
(e) A finite set in $\mathbb{R}$
(f) $\{x \in \mathbb{R}^n \mid ||x|| = 1\}$ (distinguish between the cases $n = 1$ and $n \geq 2$)
(g) Perimeter of the unit square in $\mathbb{R}^2$
(h) The boundary of a bounded set in $\mathbb{R}$
(i) The rationals in $[0,1]$
(j) A closed set in $[0,1]$

Answer. (a) Connected, not compact.

(b) Compact and connected.

(c) $n = 1$: compact and not connected. $n \geq 2$: compact and connected.

(d) Neither compact nor connected.

(e) Compact, but not connected if it contains more than one point.

(f) $n = 1$: compact and not connected. $n \geq 2$: compact and connected.

(g) Compact and connected.

(h) Compact. Not necessarily connected.

(i) Neither compact nor connected.

(j) Compact. Not necessarily connected. $\Diamond$

Solution. (a) The set $A = \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1| \leq 1\}$ is an infinitely long vertical strip. It is closed since it contains its boundary lines, the verticals $x_1 = 1$ and $x_1 = -1$. It is not compact since it is not bounded. It contains the entire vertical x_2-axis, all points $(0, x_2)$ and the norm of such a point is $|x_2|$. This can be arbitrarily large.

(b) The set $B = \{x \in \mathbb{R}^n \mid \|x\| \leq 10\}$ is connected since it is path-connected. An easy way to establish this is to go from one point to another first by going in to the origin along a radius and then out to the second point along another radial path. It is bounded since its points have norm no larger than 10, and it is closed since it includes the boundary sphere where $\|x\| = 10$. It is a closed and bounded subset of $\mathbb{R}^n$, so it is compact by the Heine-Borel theorem.

(c) The set $C = \{x \in \mathbb{R}^n \mid 1 \leq \|x\| \leq 2\}$ is closed since it contains both boundary spheres, the points where $\|x\| = 1$ and those where $\|x\| = 2$. It is bounded since all of its points have norm no more than 2. It is a closed bounded subset of $\mathbb{R}^n$ and so is compact by the Heine-Borel Theorem. If $n \geq 2$, then it is path-connected. This is fairly obvious but not quite so easily implemented as in part (b). We cannot go all the way in to the origin. If x and y are in C, we can proceed from x in to the sphere of radius 1.5 along a radial line. Then we can go along a great circle on that sphere cut by the plane containing x, y, and the origin to the point on the same ray as y, then along that radial path to y. Thus C is path-connected and so connected if $n \geq 2$. If $n = 1$ it is not connected since it is the union of two intervals, $[-2, -1] \cup [1, 2]$.

(d) The set $D = \mathbb{Z} = \{\text{integers in } \mathbb{R}\}$ is not bounded (Archimedean Principle) so it is not compact. It is certainly not connected. The open sets $U = \{x \in \mathbb{R} \mid x < 1/2\}$ and $V = \{x \in \mathbb{R} \mid x > 1/2\}$ disconnect it.

(e) Suppose E is a finite set in $\mathbb{R}$. Say $E = \{x_1, x_2, \ldots, x_n\}$. If $\{U_\alpha\}_{\alpha \in A}$ is any open cover of E, then for each k there is an index α_k such that $x_k \in U_{\alpha_k}$. So $E \subseteq \bigcup_{k=1}^n U_{\alpha_k}$. Every open cover of E has a finite subcover. So E is compact. If E has only one point then it is certainly connected. If it has more than one, then the open sets $U = \{x \in \mathbb{R} \mid x < (x_1 + x_2)/2\}$ and $V = \{x \in \mathbb{R} \mid x > (x_1 + x_2)/2\}$ disconnect it.

(f) The set $F = \{x \in \mathbb{R}^n \mid \| x \| = 1\}$ is closed and bounded in $\mathbb{R}^n$ and so compact by the Heine-Borel theorem. If $n = 1$, then $F = \{-1, 1\}$ and is not connected. If $n \geq 2$, then we can proceed from any point x in F to any other point y in F along the great circle cut in F by the plane containing x, y and the origin. So, if $n \geq 2$, then F is path-connected and so connected.

(g) Let G be the perimeter of the unit square in $\mathbb{R}^2$. Then G is a closed and bounded subset of $\mathbb{R}^2$ and so is compact by the Heine-Borel Theorem. It is certainly path-connected. It is built as the union of four continuous straight line paths which intersect at the corners. So it is connected.

(h) Let S be a bounded set in $\mathbb{R}$ and let $H = \mathrm{bd}(S)$. Then H is a closed set since it is the intersection of the closure of S with the closure of its complement. Since S is bounded, its closure is also bounded. (There is a constant r such that $|x| \leq r$ for all x in S. If $y \in \mathrm{cl}(S)$, then there is an x in S with $|x - y| < 1$. Then $|y| \leq |y - x| + |x| < 1 + r$. So $|y| \leq r + 1$ for all y in $\mathrm{cl}(S)$.) Since $H \subseteq \mathrm{cl}(S)$, it is also bounded. Since H is a closed bounded subset of $\mathbb{R}^1$, it is compact by the Heine-Borel Theorem. The set $H = \mathrm{bd}(S)$ is most likely not connected. For example, if $S = [0, 1]$, then H is the two point set $\{0, 1\}$, which is not connected. (Can you figure out for which sets S we do have $\mathrm{bd}(S)$ connected?)

(i) Let $I = \mathbb{Q} \cap [0, 1]$. Then I is not closed. Its closure is $[0, 1]$. So it is not compact. It is not connected since the open sets $U = \{x \in \mathbb{R} \mid x < 1/\sqrt{2}\}$ and $V = \{x \in \mathbb{R} \mid x > 1/\sqrt{2}\}$ disconnect it.

(j) Let J be a closed set in $[0, 1]$. Then J is given as closed, and it is certainly bounded since $|x| \leq 1$ for every x in J. Since it is a closed bounded set in $\mathbb{R}^1$, J is compact by the Heine-Borel Theorem. It will be connected if and only if it is a closed interval. If x and y are in J and there is a point z with $x < z < y$, then the open sets $U = \{x \in \mathbb{R} \mid x < z\}$ and $V = \{x \in \mathbb{R} \mid x > z\}$ disconnect J. ♦

◇ **3E-2.** Prove that a set $A \subset \mathbb{R}^n$ is not connected iff we can write $A \subset F_1 \cup F_2$, where F_1, F_2 are closed, $A \cap F_1 \cap F_2 = \varnothing$, $F_1 \cap A \neq \varnothing$, $F_2 \cap A \neq \varnothing$.

Suggestion. Suppose U_1 and U_2 disconnect A and consider the sets $F_1 = M \setminus U_1$ and $F_2 = M \setminus U_2$. ◊

Solution. By definition, A is not connected if and only if there are open sets U_1 and U_2 such that

1. $U_1 \cap U_2 \cap A = \emptyset$

2. $U_1 \cap A$ is not empty
3. $U_2 \cap A$ is not empty
4. $A \subseteq U_1 \cup U_2$.

But U_1 and U_2 are open if and only if their complements $F_1 = M \setminus U_1$ and $F_2 = \setminus U_2$ are closed. Using DeMorgan's Laws, our four conditions translate to

1. $F_1 \cup F_2 \cup (M \setminus A) = M$
2. $F_1 \cup A$ is not all of M
3. $F_2 \cup A$ is not all of M
4. $M \setminus A \supseteq F_1 \cap F_2$.

Since none of the points in A are in $M \setminus A$, condition 1 is equivalent to $A \subseteq F_1 \cup F_2$. Condition 4 is equivalent to $A \cap F_1 \cap F_2 = \emptyset$. So, with a bit more manipulations, our conditions become

1. $A \subseteq F_1 \cup F_2$
2. $F_1 \cap A = (M \setminus U_1) \cap A = (M \setminus U_1) \cap (M \setminus (M \setminus A)) = M \setminus (U_1 \cup (M \setminus A))$. This is not empty since there are points in $U_2 \cap A$ and these cannot be in U_1.
3. $F_2 \cap A = (M \setminus U_2) \cap A = (M \setminus U_2) \cap (M \setminus (M \setminus A)) = M \setminus (U_2 \cup (M \setminus A))$. This is not empty since there are points in $U_1 \cap A$ and these cannot be in U_2.
4. $A \cap F_1 \cap F_2 = \emptyset$.

These are exactly the conditions required of the closed sets F_1 and F_2 in the problem.

In the converse direction, if we have closed sets F_1 and F_2 satisfying

1. $A \subset F_1 \cup F_2$
2. $F_1 \cap A$ not empty
3. $F_2 \cap A$ not empty
4. $A \cap F_1 \cap F_2 = \emptyset$,

we can consider the open sets $U_1 = M \setminus F_1$ and $U_2 = M \setminus F_2$ and work backwards through these manipulations to obtain

1. $U_1 \cap U_2 \cap A = \emptyset$
2. $U_1 \cap A$ is not empty
3. $U_2 \cap A$ is not empty
4. $A \subseteq U_1 \cup U_2$,

so that A is not connected. ♦

◇ **3E-3.** Prove that in $\mathbb{R}^n$, a bounded infinite set A has an accumulation point.

Sketch. cl(A) is bounded since A is (why?), and it is closed. So it is compact. Every infinite subset, such as A, must have an accumulation point. (See Exercise 3.1-1.) ◊

Solution. Let A be a bounded infinite subset of $\mathbb{R}^n$. Then cl(A) is closed since closures always are. Furthermore, it is bounded. To see this note that there is an $r > 0$ such that $\| x \| \leq r$ for every x in A. If $y \in \text{cl}(A)$, there is an $x \in A$ with $\| x - y \| < 1$. So $\| y \| \leq \| y - x \| + \| x \| < 1 + r$. So $\| y \| < r + 1$ for every $y \in \text{cl}(A)$. Thus cl(A) is a closed bounded set in $\mathbb{R}^n$ and is compact by the Heine-Borel theorem. By Exercise 3.1-1, every infinite subset of it, such as A, must have an accumulation point. The accumulation point must be in cl(A) but not necessarily in A. ♦

◇ **3E-4.** Show that a set A is bounded iff there is a constant M such that $d(x, y) \leq M$ for all $x, y \in A$. Give a plausible definition of the diameter of a set and reformulate your result.

Solution. A subset A of a metric space $\mathcal{X}$ is bounded by definition if there is a point $x_0 \in \mathcal{X}$ and a number r such that $d(x, x_0) < r$ for every x in A. That is, if $A \subseteq D(x_0, r)$ for some $x_0 \in \mathcal{X}$ and $r > 0$.

Suppose A is a bounded subset of $\mathcal{X}$. We want to show that there is a constant M such that $d(x, y) < M$ for all x and y in A. If A is empty then we can just set $M = 1$ and the condition is vacuously satisfied. If A is not empty, then there is a point $x_0 \in \mathcal{X}$ and a number $r > 0$ such that $d(z, x_0) < r$ for every $z \in A$. If x and y are in A, the triangle inequality gives

$$d(x, y) \leq d(x, x_0) + d(x_0, y) < r + r = 2r.$$

If we set $M = 2r$, we have $d(x, y) < M$ for all x and Y in A as desired.

Now suppose there is such a constant M. We want to show that there is an $x_0 \in \mathcal{X}$ and an $r > 0$ such that $d(x, x_0) < r$ for every x in A. If $A = \emptyset$, we can select any $x_0 \in \mathcal{X}$ and any $r > 0$. The condition $d(x, x_0) < r$ for every $x \in A = \emptyset$ is vacuously true. If A is not empty, select $x_0 \in A$. If $x \in A$, then $d(x, x_0) < M$. So we can take $r = M$.

The diameter of a disk is the least upper bound of the distance between points in the disk. It seems reasonable to use that as the definition of the diameter of any subset of a metric space:

$$\text{diameter}(A) = \begin{cases} \sup\{d(x, y) \mid x \in A \text{ and } y \in A\} & \text{if } A \text{ is not empty} \\ 0 & \text{if } A = \emptyset. \end{cases}$$

Then A is bounded if and only if diameter(A) is finite. ♦

◇ **3E-5.** Show that the following sets are not compact, by exhibiting an open cover with no finite subcover.

(a) $\{x \in \mathbb{R}^n \mid ||x|| < 1\}$
(b) $\mathbb{Z}$, the integers in $\mathbb{R}$

Solution. (a) Suppose $A = \{x \in \mathbb{R}^n \mid \| x \| < 1\}$. This set is not closed, so it should not be compact. For each integer $k > 0$, let $U_k = \{x \in \mathbb{R}^2 \mid \| x \| < k/(k+1)\}$. Then the sets U_k are open balls and are contained in A. Furthermore, $U_1 \subseteq U_2 \subseteq U_3 \subseteq \cdots \subseteq U_k \subseteq U_{k+1} \subseteq \ldots$. If $x \in A$, then there is an integer k such that $0 \leq \| x \| < k/(k+1) < 1$. So $x \in U_k$. Thus $A = \bigcup_{k=1}^{\infty} U_k$. However, the union of any finite subcollection of this open cover is contained in $\bigcup_{k=1}^{N} U_k = U_N$ for some N. If we put $r = (N/(N+1)) + (1 - (N/(N+1))/2$, then $N/(N+1) < r < 1$. The point $x = (r, 0, \ldots, 0) \in A \setminus U_N$. So no finite subcollection can cover A. This open cover has no finite subcover. The set A is not compact.

(b) Let $B = \mathbb{Z}$, the integers in $\mathbb{R}$. For each integer k, let U_k be the open interval $=]k - (1/3), k + (1/3)[$. Then each U_k is an open subset of $\mathbb{R}$, and $B \cap U_k = \{k\}$. So the infinite collection $\{U_k\}_{k \in \mathbb{Z}}$ is an open cover of $B = \mathbb{Z} \subseteq \mathbb{R}$. There can be no finite subcover since if U_n is deleted from the collection, then the integer n is no longer included in the union. ♦

◇ **3E-6.** Suppose that F_k is a sequence of compact nonempty sets satisfying the nested set property such that diameter$(F_k) \to 0$ as $k \to \infty$. Show that there is exactly one point in $\cap\{F_k\}$. (By definition, diameter $(F_k) = \sup\{d(x, y) \mid x, y \in F_k\}$).

Sketch. Select one point x_0 in the intersection by the nested set property. If y_0 is any other point, then it cannot be in F_k if diameter$(F_k) < d(x_0, y_0)/2$. ◊

Solution. We know from the Nested Set Property (3.3.1) that there is at least one point in the intersection. Suppose x_0 is in $\bigcap_{k=1}^{\infty} F_k$, so that $x \in F_k$ for all k. If y_0 is a point in M not equal to x_0, then $d(x_0, y_0) > 0$, and there is a K such that diameter$(F_k) < d(x_0, y_0)/2$ whenever $k \geq K$. Select any k_0 with $k_0 > K$. Then $x_0 \in F_{k_0}$, and

$$d(x_0, y_0) > \text{diameter}(F_{k_0}) = \sup\{d(x, y) \mid x \in F_{k_0} \text{ and } y \in F_{k_0}\}.$$

Since x_0 is in F_{k_0}, this means that y_0 cannot be in F_{k_0}. So it is not in the intersection. Since this is true for every y_0 not equal to x_0, we conclude that $\bigcap_{k=1}^{\infty} F_k = \{x_0\}$. It consists of exactly one point as claimed. ♦

◇ **3E-7.** Let x_k be a sequence in $\mathbb{R}^n$ that converges to x and let $A_k = \{x_k, x_{k+1}, \dots\}$. Show that $\{x\} = \cap_{k=1}^{\infty} \text{cl}(A_k)$. Is this true in *any* metric space?

Sketch. Start with $\text{cl}(A_k) = \{x\} \cup \{x_k, x_{k+1}, \dots\}$. Remember to show that if y and x are different, then $y \notin \text{cl}(A_k)$ for large k. (No such x_k can be close to y since they are close to x.) Give detail. ◊

Solution. One can proceed directly as in the sketch or one can employ the ideas of Exercise 3E-6. The basic idea is essentially the same. Let $F_k = \text{cl}(A_k)$. From Proposition 2.7.6(ii), we know that $x \in \text{cl}(A_k)$. From Exercise 2.7-2 we know that $\{x\} \cup A_k$ is closed. So $F_k = \text{cl}(A_k) = \{x\} \cup A_k$. From Exercise 3.1-4 (or directly) we know that $F_k = \{x\} \cup A_k$ is compact.

Since $F_{k+1} = \{x\} \cup \{x_{k+1}, x_{k+2}, \dots\} \subseteq \{x\} \cup \{x_k, x_{k+1}, x_{k+2}, \dots\} = F_k$, they are nested. From the nested set property, there must be at least one point in their intersection. But we already know that since x is certainly in the intersection.

Let $\varepsilon > 0$. Then there is a K such that $d(x, x_k) < \varepsilon/2$ whenever $k \geq K$. So, if z and w are in F_k for such a k, each of them must be equal either to x or do some x_j with $j \geq K$. So $d(z, w) \leq d(z, x) + d(x, w) < \varepsilon$. So $\text{diameter}(F_k) \to 0$ as $k \to \infty$. As in Exercise 3E-6, this implies that there can be no more than one point in the intersection. (The distance between two points in the intersection would have to be 0.) So $\{x\} = \bigcap_{k=1}^{\infty} \text{cl}(A_k)$ as claimed. All the steps work in any metric space. So the assertion is true in every metric space. ♦

◇ **3E-8.** Let $A \subset \mathbb{R}^n$ be compact and let x_k be a Cauchy sequence in $\mathbb{R}^n$ with $x_k \in A$. Show that x_k converges to a point in A.

Sketch. A Cauchy sequence in $\mathbb{R}^n$ must converge to some point in $\mathbb{R}^n$ since $\mathbb{R}^n$ is complete. If the sequence lies in A, then its limit must be in $\text{cl}(A)$. If A is compact, then it is closed, so $A = \text{cl}(A)$, and the limit is in A. ◊

Solution. Let A be a compact subset of $\mathbb{R}^n$ and let $\langle x_k \rangle_1^{\infty}$ be a Cauchy sequence in A. Then $\langle x_k \rangle_1^{\infty}$ is a Cauchy sequence in $\mathbb{R}^n$, so it must converge to some point $x \in \mathbb{R}^n$ by completeness (2.8.5). Since each of the points x_k is in A, their limit x must be in $\text{cl}(A)$ by Proposition 2.7.6. But A is compact, so it is closed and $\text{cl}(A) = A$. Thus the limit x is in A as claimed. ♦

◇ **3E-9.** Determine (by proof or counterexample) the truth or falsity of the following statements:

(a) (A is compact in $\mathbb{R}^n$) $\Rightarrow$ ($\mathbb{R}^n \backslash A$ is connected).
(b) (A is connected in $\mathbb{R}^n$) $\Rightarrow$ ($\mathbb{R}^n \backslash A$ is connected).

(c) (A is connected in $\mathbb{R}^n$) $\Rightarrow$ (A is open or closed).

(d) ($A = \{x \in \mathbb{R}^n \mid ||x|| \leq 1\}$) $\Rightarrow$ ($\mathbb{R}^n \backslash A$ is connected). [Hint: Check the cases $n = 1$ and $n \geq 2$.]

Answer. (a) False; $[0, 1]$ is compact, but $\mathbb{R} \setminus [0, 1]$ is not connected. In $\mathbb{R}^n$, $A = \{x \in \mathbb{R}^n \mid 1 \leq \| x \| \leq 2\}$ is compact, but $\mathbb{R}^n \setminus A$ is not connected.

(b) False; same examples as in (a).

(c) False; $]a, b]$ is connected but is neither open nor closed.

(d) False for $n = 1$, true for $n \geq 2$. ($\mathbb{R}^n \setminus A$ is path-connected if $n \geq 2$.) ♢

Solution. (a) Given above.

(b) Given above.

(c) Given above.

(d) If $n = 1$, then A is the closed interval $[-1, 1]$, and its complement $\mathbb{R} \setminus A$, is the union of two disjoint open sets, $]-\infty, -1[\cup]1, \infty[$ and is not connected. In more than one dimension, A is the exterior of the unit ball and is path-connected. Connect each point to the point on the same ray from the origin on the sphere of radius 1. Then connect any two points on this sphere by following a great circle route (a circle containing those two points and the origin). ♦

♢ **3E-10.** A metric space M is said to be ***locally path-connected*** if each point in M has a neighborhood U such that U is path-connected. (This terminology differs somewhat from that of some topology books.) Show that (M is connected and locally path-connected) $\Leftrightarrow$ (M is path-connected).

Suggestion. Look at the proof in Worked Example 3WE-4 that an open connected subset of $\mathbb{R}^n$ is path-connected. ♢

Solution. If M is path-connected and $x_0 \in M$, then M is connected and M itself is a neighborhood of x_0 which is path-connected. So (by this definition) M is locally path-connected.

For the converse, suppose M is connected and locally path-connected. The proof from here is the same as the proof that an open connected set in $\mathbb{R}^n$ is path-connected. The hypothesized path-connected neighborhoods of points take the place of balls centered at the points. Let $x_0 \in A$. Define sets A and B by

$$A = \{x \in M \mid \text{ there is a path in } M \text{ from } x_0 \text{ to } x\}$$
$$B = \{x \in M \mid \text{ there is no path in } M \text{ from } x_0 \text{ to } x\}.$$

The sets A and B are certainly disjoint and their union is M. They are both open.

A is open: Suppose $x \in A$, then there is path-connected neighborhood U of x. If $y \in U$, then we get a path from x_0 to y by proceeding first to x and then from x to y through U. Thus $y \in A$. This is true for every $y \in U$, so $U \subseteq A$. For each point x in A there is a neighborhood of x contained in A. So A is open.

B is open: Suppose $x \in B$, then there is path-connected neighborhood U of x. Let $y \in U$. If there were a path from x_0 to y, then we could continue it by a path through U to x, but that is impossible if $x \in B$. So y is not in A and it must be in B. This is true for every $y \in B$, so $U \subseteq B$. For each point in B there is a neighborhood of x contained in B. So B is open.

The sets A and B are both open and their disjoint union is M. Since M is connected, one of them must be empty. Since $x_0 \in A$, B must be empty and $A = M$. Thus every point in M can be connected by a path to x_0. If x_1 and x_2 are in M then we can connect x_1 to x_2 by going through x_0. So M is path-connected. ♦

◇ **3E-11.** (a) Prove that if A is connected in a metric space M and $A \subset B \subset \mathrm{cl}(A)$, then B is connected.

(b) Deduce from (a) that the components of a set A are relatively closed. Give an example in which they are not relatively open. ($C \subset A$ is called ***relatively closed*** in A if C is the intersection of some closed set in M with A, *i.e.,* if C is closed in the metric space A.)

(c) Show that if a family $\{B_i\}$ of connected sets is such that $B_i \cap B_j \neq \varnothing$ for all i, j, then $\cup_i B_i$ is connected.

(d) Deduce from (c) that every point of a set lies in a unique component.

(e) Use (c) to show that $\mathbb{R}^n$ is connected, starting with the fact that lines in $\mathbb{R}^n$ are connected.

Sketch. (a) If U and V disconnect B and $x \in B \cap U$, then $A \cap U \neq \emptyset$. (Why?) Similarly $A \cap V \neq \emptyset$. So A would be disconnected.

(c) Establish a lemma: If B is connected, $B \subseteq C$, and C is disconnected by U and V, then either $B \subseteq U$ or $B \subseteq V$. ◇

Solution. (a) Suppose U and V were open sets which disconnect B. Then $B \cap U$ would not be empty. Select $x \in B \cap U$, then $x \in \mathrm{cl}(A)$ since $B \subseteq \mathrm{cl}(A)$. Since U is an open set containing x, there would be a point y in $U \cap A$. So $A \cap U$ would not be empty. Similarly, there would be a point x in $B \cap V$ and hence a corresponding y in $A \cap V$. With $A \cap U$ and $A \cap V$ both nonempty, the open sets U and V would disconnect A. Since A was assumed to be connected, this is not possible. So B must be connected as claimed.

(b) Recall that a connected component of a set A is a maximal connected subset of A. That is, a subset A_0 of A such that $A_0 \subseteq B \subseteq A$ and B connected imply that $A_0 = B$. To show that such a subset is relatively closed in A, we need to show that there is a closed set F such that $A_0 = F \cap A$. The obvious candidate is $\mathrm{cl}(A_0)$.

Proposition. *If A_0 is a connected component of a set A, then $A_0 = A \cap \mathrm{cl}(A_0)$.*

Proof: We always have $A_0 \subseteq \mathrm{cl}(A_0)$. Since $A_0 \subseteq A$, we have $A_0 \subseteq A \cap \mathrm{cl}(A_0) \subseteq \mathrm{cl}(A_0)$. Since A_0 is a connected component of A, it is connected, and part (a) implies that $A \cap \mathrm{cl}(A_0)$ is connected. Since $A_0 \subseteq A \cap \mathrm{cl}(A_0) \subseteq A$, and $A \cap \mathrm{cl}(A_0)$ is connected, and A_0 is a maximal connected subset of A, we must have $A_0 = A \cap \mathrm{cl}(A_0)$. Since $\mathrm{cl}(A_0)$ is a closed set, we conclude that A_0 is closed relative to A as claimed.

(c) Suppose $\{B_j\}_{j \in J}$ is a family of connected subsets of a metric space M, and let $C = \bigcup_{j \in J} B_j$. We want to show that C is connected. To this end, we may as well assume that all of the sets B_j are nonempty. We establish a lemma:

Lemma. *If B is connected, $B \subseteq C$, and C is disconnected by U and V, then either $B \subseteq U$, or $B \subseteq V$.*

Proof: The sets U and V are open sets such that

(1) $B \subseteq C \subseteq U \cup V$.

(2) $(B \cap U) \subseteq (C \cap U)$ which is not empty

(3) $(B \cap V) \subseteq (C \cap V)$ which is not empty

(4) $(B \cap U \cap V) \subseteq (C \cap U \cap V) = \emptyset$.

If $(B \cap U)$ and $(B \cap V)$ were both nonempty, then U and V would disconnect B. Since B is connected, this is not possible. One of the sets $(B \cap U)$ and $(B \cap V)$ must be empty. From (1) we have $B \subseteq (B \cap U) \cup (B \cap V)$ and the other must be all of B. This establishes the lemma.

Suppose U and V were open sets which disconnected C, and let i and j be in J. Applying the lemma to B_i, we find that $B_i \subseteq W_i$ where W_i is either U or V and the intersection of B_i with the other is empty. Similarly $B_j \subseteq W_j$. By hypothesis $B_i \cap B_j$ is not empty, so $W_i = W_j$. This works for every pair i, j. So either all of the B_j are contained in U or all are contained in V and the intersections with the other are all empty. If all are contained in U, then $C \subset U$ and $C \cap V = \emptyset$. So U and V did not really disconnect C after all. The set C must be connected.

(d) Suppose A is a subset of a metric space M and let $x_0 \in A$. Let

$$A_0 = \bigcup \{B \subseteq A \mid x_0 \in B \text{ and } B \text{ is connected }\}.$$

From part (c) we see that A_0 is a connected subset of A. It certainly contains x_0. If B is any connected subset of A which contains x_0, then we have $x_0 \in B \subseteq A_0 \subseteq A$. In particular, if B is a connected component of A containing x_0, we would have $B \subseteq A_0$. We would also have $A_0 \subseteq B$ since B is a maximal connected subset. Thus $B = A_0$. Thus A_0 is a connected component of A containing the point x_0 and it is the only such component.

(e) Every line through the origin in $\mathbb{R}^n$ is parameterized as $\gamma(t) = tv$ for some fixed vector v in $\mathbb{R}^n$. This parameterization supplies a continuous path between any two points on that line. So each such line is path-connected and hence connected. They all contain the origin, so the lemma developed in part (c) shows that their union is connected. But that union is all of $\mathbb{R}^n$. Thus the space $\mathbb{R}^n$ is connected. ♦

◇ **3E-12.** Let S be a set of real numbers that is nonempty and bounded above. Let $-S = \{x \in \mathbb{R} \mid -x \in S\}$. Prove that

(a) $-S$ is bounded below.
(b) $\sup S = -\inf(-S)$.

Solution. (a) The set S is bounded above, so there is a constant B such that $y \leq B$ for every y in S. If $x \in -S$, then $-x \in S$, so $-x \leq B$ and $x \geq -B$. This is true for every x in $-S$, so $-B$ is a lower bound for the set $-S$.

(b) Since $\sup(S)$ is an upper bound for the set S, the computation of part (a) shows that $-\sup S$ is a lower bound for $-S$. So $-\sup S \leq \inf(-S)$ and $\sup S \geq -\inf(-S)$. On the other hand, there is a sequence $x_1, x_2, x_3, \ldots$ of points in S with $x_n \to \sup S$. So $-x_1, -x_2, -x_3, \ldots$ are in $-S$, and $-x_n \to -\sup S$. Thus $\inf(-S) \leq -\sup S$, and $\sup S \leq -\inf(-S)$. We have inequality in both directions and so equality as claimed. ♦

◇ **3E-13.** Let M be a complete metric space and F_n a collection of closed nonempty subsets (not necessarily compact) of M such that $F_{n+1} \subset F_n$ and diameter $(F_n) \to 0$. Prove that $\cap_{n=1}^{\infty} F_n$ consists of a single point; compare to Exercise 3E-6.

Sketch. Pick $x_n \in F_n$. Show $\langle x_n \rangle_1^{\infty}$ is a Cauchy sequence. Its limit must be in $\mathrm{cl}(F_n)$ for all n. (Why?) There cannot be two such points since $\mathrm{diam}(F_n) \to 0$. ◇

Solution. In Exercise 3E-6 we were given that the sets were compact so we could use the Nested Set Property to obtain a point in the intersection. Here the sets are not known to be compact. Since the diameters tend to 0, they are certainly finite for large n, so those sets are closed and bounded. If we were in $\mathbb{R}^n$, we would know that they were compact and could proceed as in Exercise 3E-6. In other metric spaces closed bounded sets might not be compact, and must work directly with the hypotheses we have.

The sets F_n are all nonempty, so for each n we can select a point $x_n \in F_n$. Since $F_1 \supseteq F_2 \supseteq F_3 \supseteq \dots$, we have $x_{n+p} \in F_n$ for all n and all $p \geq 0$. Let $\varepsilon > 0$. There is an N such that $\operatorname{diam}(F_n) < \varepsilon$ whenever $n \geq N$. So, if $n \geq N$ and $p \geq 0$, we have $d(x_{n+p}, x_n) \leq \operatorname{diam}(F_n) < \varepsilon$. The sequence $\langle x_n \rangle_1^\infty$ is a Cauchy sequence in M. Since M is complete, it must converge to some point $x \in M$. Since $x_k \in F_n$ for all $k \geq n$, and F_n is closed, $x \in F_n$. Since this is true for every n, $x \in \bigcap_{n=1}^\infty F_n$. There can be no other points in the intersection. If y and x are different, we need only pick an n large enough so that $\operatorname{diam}(F_n) < (1/2)d(x, y)$. Then $x \in F_n$, so y cannot be in F_n, and y is not in the intersection. ♦

◇ **3E-14.** (a) A point $x \in A \subset M$ is said to be ***isolated*** in the set A if there is a neighborhood U of x such that $U \cap A = \{x\}$. Show that this is equivalent to saying that there is an $\varepsilon > 0$ such that for all $y \in A$, $y \neq x$, we have $d(x, y) > \varepsilon$.

(b) A set is called ***discrete*** if all its points are isolated. Give some examples. Show that a discrete set is compact iff it is finite.

Solution. (a) If x is an isolated point of A, then there is a neighborhood U of x such that $U \cap A = \{x\}$. Since U is a neighborhood of x, there is an $\varepsilon > 0$ such that $D(x, \varepsilon) \subseteq U$. If $y \in A$ and y is not x, then y is not in U, so it is not in $D(x, \varepsilon)$. Thus $d(x, y) > \varepsilon$.

For the converse, suppose $x \in A$ and that there is an $\varepsilon > 0$ such that for all $y \in A$ different from x, we have $d(x, y) > \varepsilon$. Let $U = D(x, \varepsilon)$. Then U is a neighborhood of x and $U \cap A = \{x\}$, so x is an isolated point of A.

(b) Suppose A is a discrete set. For each $x \in A$ there is a $\delta_x > 0$ such that $D(x, \delta_x) \cap A = \{x\}$. Since $x \in D(x, \delta_x)$, we have $A \subseteq \bigcup_{x \in A} D(x, \delta_x)$. If any of the disks in this open cover of A are deleted, we would no longer have a cover of A since each intersects A in exactly one point. So this open cover cannot have a finite subcover unless A is a finite set. If $A = \{x_1, \dots x_n\}$ is a finite set, then any open cover $\{U_\alpha\}$ has a finite subcover since we need only pick $\alpha_1, \dots \alpha_n$ with $x_k \in U_{\alpha_k}$. ♦

◇ **3E-15.** Let $K_1 \subset M_1$ and $K_2 \subset M_2$ be path-connected (respectively, connected, compact). Show that $K_1 \times K_2$ is path-connected (respectively, connected, compact) in $M_1 \times M_2$.

Sketch. If $\gamma(t)$ and $\mu(t)$ are paths from x to a in K_1 and from y to b in K_2, then $\varphi(t) = (\gamma(t), \mu(t))$ is a path from (x, y) to (a, b) in $K_1 \times K_2$. $\Diamond$

Solution. This is actually a collection of three exercises:

(a) If K_1 and K_2 are path-connected, then $K_1 \times K_2$ is path-connected.
(b) If K_1 and K_2 are connected, then $K_1 \times K_2$ is connected.
(c) If K_1 and K_2 are compact, then $K_1 \times K_2$ is compact.

Let M be the cross product space $M_1 \times M_2$. Before the exercise can even make sense, we need to specify what convergence and open sets mean in M by specifying a metric on M. There are several equivalent ways to do this. One is to consider a set in M to be open if it contains a "taxicab" disk around each of its points.

Definition. *A set $S \subseteq M = M_1 \times M_2$ is open if for each $v = (v_1, v_2)$ in S, there is an $r > 0$ such that $d_1(w_1, v_1) + d_2(w_2, v_2) < r$ implies $w = (w_1, w_2) \in S$.*

This is the same as putting the "taxicab metric" on M.

$$d(v, w) = d((v_1, v_2), (w_1, w_2)) = d_1(v_1, w_1) + d_2(v_2, w_2).$$

This is a metric on the cross product. (Proof ?) We could also use a formula like that for the Euclidean metric in $\mathbb{R}^2$:

$$\rho(v, w) = \rho((v_1, v_2), (w_1, w_2)) = \sqrt{d_1(v_1, w_1)^2 + d_2(v_2, w_2)^2}.$$

The same sorts of inequalities that we used in $\mathbb{R}^2$ show that these two metrics produce the same open sets in the cross product and that if $v^{(j)} = (v_1^{(j)}, v_2^{(j)})$ for $j = 1, 2, 3, \ldots$ is a sequence in $M = M_1 \times M_2$, and $w = (w_1, w_2) \in M$, then

$$\begin{aligned} d(v^{(j)}, w) \to 0 &\iff \rho(v^{(j)}, w) \to 0 \\ &\iff d_1(v_1^{(j)}, w_1) \to 0 \text{ and } d_2(v_2^{(j)}, w_2) \to 0. \end{aligned}$$

(a) Suppose K_1 and K_2 are path-connected and that $v = (v_1, v_2)$ and $w = (w_1, w_2)$ are in $K_1 \times K_2$. Then v_1 and v_2 are in the path-connected set K_1 and w_1 and w_2 are in the path-connected set K_2. If $\gamma(t)$ and $\mu(t)$ are paths from v_1 to v_2 in K_1 and from w_1 to w_2 in K_2, then $\varphi(t) = (\gamma(t), \mu(t))$ is a path from v to w in $K_1 \times K_2$. The path $\varphi(t)$ is continuous since each of the coordinate paths are continuous. (Why?)
(b) Suppose K_1 and K_2 are connected and let $C = K_1 \times K_2 \subseteq M = M_1 \times M_2$. We want to show that C is connected. The idea is that each "cross section" $\{x\} \times K_2$ of C is basically a copy of K_2 and is connected. C

is the union of these over all x in K_1. We can see that this union is connected by tying them together with a cross section $K_1 \times \{y_0\}$ taken in the other direction.

We will use three lemmas.

Lemma. *If $\{B_\alpha\}_{\alpha \in A}$ is a family of connected sets such that none of the intersections $B_\alpha \cap B_\beta$ are empty, then $\bigcup_{\alpha \in A} B_\alpha$ is connected.*

This is Exercise 3E-11(c). See that solution.

Lemma. *If S is an open subset of $M = M_1 \times M_2$, then the projections*

$$\pi_1(S) = \{v \in M_1 \mid (v, w) \in S \textit{ for some } w \in M_2\}$$
$$\pi_2(S) = \{w \in M_2 \mid (v, w) \in S \textit{ for some } v \in M_1\}$$

are open subsets of M_1 and M_2 respectively.

Proof: If $v_0 \in \pi_1(S)$, then there is a $w_0 \in M_2$ with $(v_0, w_0) \in S$. Since S is open, there is an $r > 0$ such that $(v, w) \in S$ whenever $d((v, w), (v_0, w_0)) < r$. If $v \in M_1$ and $d_1(v, v_0) < r$, then

$$d((v, w_0), (v_0, w_0)) = d_1(v, v_0) + d_2(w_0, w_0) = d_1(v, v_0) + 0 < r.$$

So $(v, w_0) \in S$ and $v \in \pi_1(S)$. Thus $\pi_1(S)$ is an open subset of M_1. The proof that $\pi_2(S)$ is an open subset of M_2 is similar.

The third is a technical addition to the second.

Lemma. *If S is an open subset of $M = M_1 \times M_2$, then for each fixed y in M_2 and each fixed x in M_1, the sets*

$$\pi_1(S \cap (M_1 \times \{y\})) \textit{ and } \pi_2(S \cap (\{x\} \times M_2))$$

are open subsets of M_1 and M_2 respectively.

The proof is much like the last. If $x_0 \in \pi_1(S \cap (M_1 \times \{y\}))$, then $(x_0, y) \in S$. Since S is open, there is an $r > 0$ such that $(v, w) \in S$ whenever $d((v, w), (x_0, y)) < r$. If $x \in M_1$ and $d_1(x, x_0) < r$, then

$$d((x, y), (x_0, y)) = d_1(x, x_0) + d_2(y, y) = d_1(x, x_0) + 0 < r.$$

So $(x, y) \in S \cap (M_1 \times \{y\})$, and $x \in \pi_1(S \cap (M_1 \times \{y\}))$. Thus $\pi_1(S \cap (M_1 \times \{y\}))$ is an open subset of M_1. The proof for the other assertion is similar.

We are now prepared for the main argument.

STEP ONE: *For each fixed y in M_2, the set $C_y = K_1 \times \{y\}$ is a connected subset of $M = M_1 \times M_2$.*

Suppose U and V were open subsets of M which disconnected C_y. Let $U_y = U \cap C_y = U \cap (K_1 \times \{y\})$ and $V_y = V \cap C_y = V \cap (K_1 \times \{y\})$. Then to say that U and V disconnect Cy is to say that

(1) $C_y = U_y \cup V_y$.

(2) Neither U_y nor V_y is empty.

(3) $U_y \cap V_y = \emptyset$.

Now let $U_1 = \pi_1(U \cap (M_1 \times \{y\}))$ and $V_1 = \pi_1(V \cap (M_1 \times \{y\}))$. By the third lemma, these are open subsets of M_1. If $x \in K_1$, then $(x, y) \in K_1 \times \{y\} = C_y$. So it must be in either U_y or V_y. This implies that x is in either U_1 or V_1. So $K_1 \subseteq U_1 \cup V_1$. On the other hand, if x were in both U_1 and V_1, then (x, y) would be in $U_y \cap V_y$ which is not possible. So $K_1 \cap U_1 \cap V_1 = \emptyset$. Finally, neither $U_1 \cap K_1$ nor $V_1 \cap K_1$ is empty since neither U_y nor V_y is empty. Thus U_1 and V_1 would disconnect K_1. Since K_1 is connected, this is not possible. We conclude that C_y is connected as claimed.

STEP TWO: *For each fixed x in M_1, the set $C_x = \{x\} \times K_2$ is a connected subset of $M = M_1 \times M_2$.*

This is essentially the same as the first step with all appropriate roles reversed. (Do it.)

STEP THREE: *If y_0 in K_2 and $x \in K_1$, the set $B_x = (\{x\} \times K_2) \cup (K_1 \times \{y_0\})$ is a connected subset of $M = M_1 \times M_2$.*

The sets $\{x\} \times K_2$ and $K_1 \times \{y_0\}$ are connected by steps one and two, and (x, y_0) is in their intersection. So their union, B_x, is connected by the first lemma.

STEP FOUR: $C = \bigcup_{x \in K_1} B_x$.

If $(x, y) \in C$, then $(x, y) \in B_x$, so $C \subseteq \bigcup_{x \in K_1} B_x$. Conversely, each element of B_x is either (x, y) for some y in K_2 or (t, y_0) for some $t \in K_1$. So $\bigcup_{x \in K_1} B_x \subseteq C$.

FINALLY: *C is connected.*

Each of the sets B_x for $x \in K_1$ is connected by step three. Their intersection is not empty since it contains $K_1 \times \{y_0\}$. So their union is connected by the first lemma. But that union is C by step four.

(c) Suppose K_1 and K_2 are compact and let $C = K_1 \times K_2 \subseteq M = M_1 \times M_2$. Since M is a metric space, we can show that C is compact by showing that it is sequentially compact. To do this we repeat the pattern of the proof of the Bolzano-Weierstrass property in $\mathbb{R}^2$. Suppose $v_n = (x_n, y_n)$ for $n = 1, 2, 3, \ldots$ is a sequence in C with $x_n \in K_1$ and $y_n \in K_2$ for each n. Then $\langle x_n \rangle_1^\infty$ is a sequence in the compact set K_1 and $\langle y_n \rangle_1^\infty$ is a sequence in the compact set K_2. Each must have a convergent subsequence. It is tempting merely to take these convergent subsequences and say that they form a convergent subsequence of the vectors $\langle v_n \rangle_1^\infty$, but this is not correct. Nothing has been done to guarantee that the same indices have been used on the two coordinate sequences. We might have selected $x_1, x_3, x_5, \ldots$, and $y_2, y_4, y_6, \ldots$. The points $(x_1, y_2), (x_3, y_4), (x_5, y_6), \ldots$ are not a subsequence of $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots$. To overcome this problem we perform the selection in two steps.

FIRST: The first coordinate sequence $\langle x_n \rangle_1^\infty$ is a sequence in the compact set K_1, so there are indices

$$n(1) < n(2) < n(3) < \dots$$

and a point $a \in K_1$ such that the sequence

$$x_{n(1)}, x_{n(2)}, x_{n(3)}, x_{n(4)}, \dots$$

converges to a.
SECOND: In considering the second coordinates we restrict attention to the indices appearing in the subsequence found in the first step. The sequence

$$y_{n(1)}, y_{n(2)}, y_{n(3)}, y_{n(4)}, \dots$$

is a sequence in the compact set K_2. So it has a convergent subsequence. There are integers $j(1) < j(2) < j(3) < \dots$ and a point $b \in K_2$ such that the sequence

$$y_{n(j(1))}, y_{n(j(2))}, y_{n(j(3))}, y_{n(j(4))}, \dots$$

converges to b.
The corresponding x's

$$x_{n(j(1))}, x_{n(j(2))}, x_{n(j(3))}, x_{n(j(4))}, \dots$$

are a sub-subsequence of the convergent subsequence found in the first step. So they converge to a. Since we now have both coordinates converging with the *same selection of indices*, we conclude that the sequence of vectors

$$(x_{n(j(1))}, y_{n(j(1))}), (x_{n(j(2))}, y_{n(j(2))}), (x_{n(j(3))}, y_{n(j(3))}), \dots$$

converges to the point $w = (a, b) \in C = K_1 \times K_2$. Here we assume that we have established the fact mentioned at the beginning that a sequence in the cross product converges to w if and only if both coordinate sequences converge to the corresponding coordinates of w. Each sequence in C has a subsequence converging to a point in C. The subset C of the metric space M is sequentially compact, so it is compact.

CHALLENGE: Give a proof of the compactness of $K_1 \times K_2$ directly from the definition of compactness without using the equivalence to sequential compactness. This is a bit trickier than it looks at first glance. Here is a very tempting wrong proof somewhat analogous to the wrong proof mentioned above for sequential compactness.

Let $\{S_\alpha\}_{\alpha \in A}$ be any open cover of C in M. That is, $C \subseteq \bigcup_{\alpha \in A} S_\alpha$ and each of the sets S_α is an open subset of M. From the lemma, the sets

$U_\alpha = \pi_1(S_\alpha)$ and $V_\alpha = \pi_1(S_\alpha)$ are open in M_1 and M_2 respectively. We must also have

$$K_1 \subseteq \bigcup_{\alpha \in A} U_\alpha \quad \text{and} \quad K_2 \subseteq \bigcup_{\alpha \in A} V_\alpha$$

since if $v \in K_1$ and $w \in K_2$, then $(v, w) \in K_1 \times K_2 = C$. There is an index α such that $(v, w) \in S_\alpha$. So $v_1 \in \pi_1(S_\alpha) = U_\alpha$, and $w \in \pi_2(S_\alpha) = V_\alpha$. Since K_1 and K_2 are compact, there are finite subcovers of these open covers. There are finite subsets B_1 and B_2 of the index set A such that

$$K_1 \subseteq \bigcup_{\alpha \in B_1} U_\alpha \quad \text{and} \quad K_2 \subseteq \bigcup_{\beta \in B_2} V_\beta.$$

So far everything is all right. It is now tempting to conclude that

$$C = K_1 \times K_2 \subseteq \bigcup_{\gamma \in B_1 \cup B_2} S_\gamma,$$

but this is not correct. The problem can be seen by considering the example sketched in the figure. ♦

FIGURE 3-1. $\pi_1(S_3)$ covers K_1 and $\pi_2(S_3)$ covers K_2, but S_3 does not cover $K_1 \times K_2$

◇ **3E-16.** If $x_k \to x$ in a normed space, prove that $||x_k|| \to ||x||$. Is the converse true? Use this to prove that $\{x \in \mathbb{R}^n \mid ||x|| \leq 1\}$ is closed, using sequences.

Sketch. Use the alternative form of the triangle inequality:

$$\Big| \, \| v \| - \| w \| \, \Big| \le \| v - w \| \, . \qquad \Diamond$$

Solution. Recall that we have an alternative form of the triangle inequality.

Lemma. *If v and w are vectors in a normed space $\mathcal{V}$, then $\big| \| v \| - \| w \| \big| \le \| v - w \|$.*

Proof: First we compute $\| v \| = \| v - w + w \| \le \| v - w \| + \| w \|$. So

$$\| v \| - \| w \| \le \| v - w \| \, . \tag{1}$$

Reversing the roles of v and w, we find that $\| w \| = \| w - v + v \| \le \| w - v \| + \| v \| = \| v - w \| + \| v \|$. So that

$$\| w \| - \| v \| \le \| v - w \| \, . \tag{2}$$

Combining (1) and (2), we find that $\big| \, \| v \| - \| w \| \, \big| \le \| v - w \|$ as claimed.

With this version of the triangle inequality, the first assertion of the exercise is almost immediate. Let $\varepsilon > 0$. Since $x_k \to x$, there is an N such that $\| x_k - x \| < \varepsilon$ whenever $k \ge N$. For such k we have $\big| \, \| x_k \| - \| x \| \, \big| \le \| x_k - x \| < \varepsilon$. So $\| x_k \| \to \| x \|$ as claimed.

The converse is not true, even in one dimension. Let $x_k = (-1)^k$ and $x = 1$ then $\| x_k \| = 1 = \| x \|$ for each index k. We certainly have $\| x_k \| \to \| x \|$, but the points x_k do not tend to x. Now suppose $\mathcal{V}$ is a normed space and let $F = \{x \in \mathcal{V} \mid \| x \| \le 1\}$. If $x \in \mathrm{cl}(F)$, then there is a sequence $\langle x_k \rangle_1^\infty$ in F with $x_k \to x$. We now know that this implies that $\| x_k \| \to \| x \|$. But $\| x_k \| \le 1$ for each k, so we must have $\| x \| \le 1$ also. Thus $x \in F$. This shows that $\mathrm{cl}(F) \subseteq F$ so that F is closed as claimed. ♦

◇ **3E-17.** Let K be a nonempty closed set in $\mathbb{R}^n$ and $x \in \mathbb{R}^n \backslash K$. Prove that there is a $y \in K$ such that $d(x, y) = \inf\{d(x, z) \mid z \in K\}$. Is this true for open sets? Is it true in general metric spaces?

Sketch. As in Worked Example 1WE-2, get $z_k \in K$ with $d(x, z_k) \to d(z, K) = \inf\{d(x, z) \mid z \in K\}$. For large k they are all in the closed ball of radius $1 + d(z, K)$ around x. Use compactness to get a subsequence converging to some z. Then $z \in K$ (why?) and $d(z_{n(j)}, x) \to d(z, x)$. (Why?) So $d(x, z) = d(x, K)$. (Why?) This does not work for open sets. The proof does not work unless closed balls are compact. ◇

Solution. Let $x \in \mathbb{R}^n \setminus K$ and $S = \{d(x,z) \in \mathbb{R} \mid z \in K\}$. Since K is not empty, neither is S, and S is certainly bounded below by 0. So $a = \inf S$ exists as a nonnegative real number. There must be a sequence $\langle t_k \rangle_1^\infty$ of points in S with $t_k \to a$, and thus a sequence $\langle z_k \rangle_1^\infty$ of points in K with $d(x, z_k) = t_k \to a$. There is an N such that $d(x, z_k) < a + 1$ whenever $k \geq N$. So for these k we have $z_k \in F = \{z \mid d(x,z) \leq a+1\}$. The set F is a closed bounded set in $\mathbb{R}^n$ and so is compact. The sequence $z_K, z_{K+1}, z_{K+2}, \ldots$ in F must have a subsequence converging to a point in F. Thus there are indices $k(1) < k(2) < k(3) < \ldots$ and a point $z \in F$ with $z_{k(j)} \to z$ as $j \to \infty$. Since each z_k is in K and K is closed, we have $z \in K$.

From the triangle inequality we know that $d(z_{k(j)}, x) \leq d(z_{k(j)}, z) + d(z,x)$ and that $d(z,x) \leq d(z, z_{k(j)}) + d(z_{k(j)}, x)$. So

$$d(z,x) - d(z, z_{k(j)}) \leq d(z_{k(j)}, x) \leq d(z_{k(j)}, z) + d(z,x).$$

Since $d(z, z_{k(j)}) \to 0$, we conclude (Sandwich Lemma) that $d(z_{k(j)}, x) \to d(z,x)$. But, we know that $d(z_{k(j)}, x) \to a$. Since limits are unique, we must have $d(x,z) = a = d(x,K)$ as desired.

This certainly does not work for open sets. If $K = \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 < 1\}$ is the open unit disk in $\mathbb{R}^2$ and $v = (1,0)$, then $\inf\{d(v,w) \mid w \in K\} = 0$, but there certainly is no point w in K with $d(v,w) = 0$ since v is not in K.

The proof just given used the fact that closed bounded sets in $\mathbb{R}^n$ are sequentially compact. This is not true in every metric space, and, in fact, there are complete metric spaces in which the assertion is false.

We will see in Chapter 5 that the space $\mathcal{V} = \mathcal{C}([0,1], \mathbb{R})$ of all continuous real valued functions on the closed unit interval with the norm $\| f \| = \sup\{|f(x)| \mid x \in [0,1]\}$ is a complete metric space. For $k = 1, 2, 3, \ldots$, let f_k be the function whose graph is sketched in Figure 3-2. Then $\| f_k \| = 1 + (1/k)$, and, if n and k are different, then each is 0 wherever the other is nonzero. So $\| f_n - f_k \| > 1$. So the f_k are all smaller than 2 in norm and form a bounded set. Furthermore the set $K = \{f_1, f_2, f_3 \ldots\}$ can have no accumulation points and so is a closed bounded set. $\| f_k - 0 \| \to 1$ as $k \to \infty$, but the distance is not equal to 1 for any function in the set.

♦

◇ **3E-18.** Let $F_n \subset \mathbb{R}$ be defined by $F_n = \{x \mid x \geq 0 \text{ and } 2 - 1/n \leq x^2 \leq 2 + 1/n\}$. Show that $\cap_{n=1}^\infty F_n \neq \varnothing$. Use this to show the existence of $\sqrt{2}$.

Suggestion. Use the nested set property. ◇

Solution. Suppose $x \in F_{n+1}$. Then

$$2 - \frac{1}{n} < 2 - \frac{1}{n+1} \leq x^2 \leq 2 + \frac{1}{n+1} < 2 + \frac{1}{n}.$$

FIGURE 3-2. Graph of the function f_k.

So $F_{n+1} \subseteq F_n$ for each $n = 1, 2, 3, \ldots$. If $x \in F_n$, then $x > 0$, and $x < 3$, since if $x \geq 3$ we would have $x^2 \geq 9 > 2 + (1/n)$ and x would not be in F_n. So the sets F_n are all bounded. One way to see that they are closed is to let $\langle x_k \rangle_1^\infty$ be a sequence of points in F_n which converge to x. Then $x_k^2 \to x^2$. We have $2-(1/n) \leq x_k^2 \leq 2+(1/n)$ for each index k, so the same is true of the limit x^2. Thus $x \in F_n$. This shows that the set F_n is closed. The sets F_n thus form a nested sequence of compact subsets of $\mathbb{R}$. By the nested set property, 3.3.1, there is at least one point x_0 in the intersection $\bigcap_{n=1}^\infty F_n$. For this point we must have $x_0 \geq 0$, and

$$2 - \frac{1}{n} \leq x_0^2 \leq 2 + \frac{1}{n}$$

for every positive integer n. So $\left|x_0^2 - 2\right| \leq 1/n$ for all $n = 1, 2, 3, \ldots$. The only way this can happen is for $x_0^2 = 2$. Thus x_0 is a positive square root for 2. ♦

◇ **3E-19.** Let $V_n \subset M$ be open sets such that $\mathrm{cl}(V_n)$ is compact, $V_n \neq \varnothing$, and $\mathrm{cl}(V_n) \subset V_{n-1}$. Prove $\cap_{n=1}^\infty V_n \neq \varnothing$.

Sketch. $\mathrm{cl}(V_{n+1}) \subseteq V_n \subseteq \mathrm{cl}(V_n)$. Use the nested set property. ◊

Solution. Let $K_n = \mathrm{cl}(V_n)$, we have assumed that the sets K_n are compact, not empty, and that $\mathrm{cl}(V_k) \subseteq V_{k-1}$ for each k. Applying this with $k = n+1$ gives

$$K_{n+1} = \mathrm{cl}(V_{n+1}) \subseteq V_n \subseteq \mathrm{cl}(V_n) = K_n.$$

So we have a nested sequence of nonempty compact sets. By the nested set property, there must be at least one point x_0 in the intersection $\bigcap_1^\infty K_n$. For each n we have $x_0 \in K_n - 1 \subseteq V_n$. Thus $x_0 \in \bigcap_1^\infty V_n$, and this intersection is not empty. ♦

◇ **3E-20.** Prove that a compact subset of a metric space must be closed as follows: Let x be in the complement of A. For each $y \in A$, choose disjoint neighborhoods U_y of y and V_y of x. Consider the open cover $\{U_y\}_{y \in A}$ of A to show the complement of A is open.

Solution. Suppose A is a compact subset of a metric space M, and suppose $x \in M \setminus A$. If $y \in A$, then let $r = d(x, y)/2 > 0$, and set $U_y = D(y, r)$ and $V_y = D(x, r)$. Then U_y and V_y are disjoint open sets with $y \in U_y$ and $x \in V_y$. We certainly have $A \subseteq \bigcup_{y \in A} U_y$. Since A is compact, this open cover must have a finite subcover:

$$A \subseteq U_{y_1} \cup U_{y_2} \cup \cdots \cup U_{y_N}.$$

If $\rho = \min\{d(x, y_k) \mid 1 \leq k \leq N\}$, then

$$D(x, \rho) \cap A \subseteq D(x, \rho) \cap (U_{y_1} \cup U_{y_2} \cup \cdots \cup U_{y_N}) = \bigcup_{k=1}^{N} (D(x, \rho) \cap U_{y_k}) = \emptyset.$$

Thus $D(x, \rho) \subseteq M \setminus A$. This shows that $M \setminus A$ is open. so A is closed as claimed. ♦

◇ **3E-21.** (a) Prove: a set $A \subset M$ is connected iff $\varnothing$ and A are the only subsets of A that are open and closed relative to A. (A set $U \subset A$ is called ***open relative to*** A if $U = V \cap A$ for some open set $V \subset M$; "closed relative to A" is defined similarly.)

(b) Prove that $\varnothing$ and $\mathbb{R}^n$ are the only subsets of $\mathbb{R}^n$ that are both open and closed.

Sketch. Look closely at the definitions for the general statement. We know $\mathbb{R}^n$ is path-connected, so it is connected. ◊

Solution. (a) If A is not connected, then there are open sets U and V such that

(1) $U \cap V \cap A = \emptyset$.
(2) $A \cap U$ is not empty.
(3) $A \cap V$ is not empty.
(4) $A \subseteq U \cup V$.

Let $B = A \cap U$ and $C = A \cap V$. Since U and V are open, each of B and C are open relative to A. From (1) and (4), we obtain $B = A \cap (M \setminus V)$ and $C = A \cap (M \setminus U)$. (Details?) So B and C are each closed relative to

A. From (2) and (3), neither of the sets B and C are empty. So if A is not connected, then there are proper subsets of A which are both open and closed relative to A.

For the converse, suppose there is a nonempty proper subset B of A which is both open and closed relative to A. Then there are an open set U and a closed set F such that $B = A \cap U$ and $B = A \cap F$. Let $V = M \setminus F$ and $C = A \cap V$. Then U and V are open and we have

(1) $U \cap V \cap A = \emptyset$: To see this compute

$$\begin{aligned} U \cap V \cap A &= (A \cap U) \cap (A \cap V) \\ &= B \cap [A \cap (M \setminus F)] \\ &= (B \cap A) \cap (M \setminus F) \\ &= B \cap (M \setminus F) \\ &= (A \cap F) \cap (M \setminus F) \\ &= \emptyset. \end{aligned}$$

(2) $A \cap U$ is not empty: This is true since $A \cap U = B$ which is nonempty by hypothesis.

(3) $A \cap V$ is not empty: To see this, compute

$$A \cap V = A \cap (M \setminus F) = A \setminus (F \cap A) = A \setminus B.$$

This is not empty since B is a proper subset of A.

(4) $A \subseteq U \cup V$: From (3) we know that $A \cap V = A \setminus B$, so

$$A \cap (U \cup V) = (A \cap U) \cup (A \cap V) = B \cup (A \setminus B) = A.$$

Thus $A \subseteq U \cap V$ as claimed.

Properties (1) - (4) show that the open sets U and V disconnect A. Thus $A \subseteq M$ is connected if and only if $\emptyset$ and A are the only subsets of A that are both open and closed relative to A as claimed.

(b) $\mathbb{R}^n$ is path-connected. If v and w are in $\mathbb{R}^n$, then the path $\gamma(t) = tw + (1-t)v$ is a continuous path from v to w. Since it is path-connected, $\mathbb{R}^n$ is connected. Since $\mathbb{R}^n$ is the whole space, "open relative to $\mathbb{R}^n$" and "closed relative to $\mathbb{R}^n$" mean the same as "open" and "closed". It follows immediately from part (a) that $\emptyset$ and $\mathbb{R}^n$ are the only subsets of $\mathbb{R}^n$ that are both open and closed. ♦

◇ **3E-22.** Find two subsets $A, B \subset \mathbb{R}^2$ and a point $x_0 \in \mathbb{R}^2$ such that $A \cup B$ is not connected but $A \cup B \cup \{x_0\}$ is connected.

Suggestion. Pick two disjoint open sets whose closures have a point in common. ◇

Solution. If A and B are nonempty open subsets of $\mathbb{R}^2$ with $A \cap B = \emptyset$ and $C = A \cup B$, then C is not connected. If we arrange it so that each of A and B are connected and their closures have a point v_0 in common, then $A \cup B \cup \{v_0\}$ will be connected. For example, let $A = \{(x, y) \in \mathbb{R}^2 \mid x > 0\}$ and $B = \{(x, y) \in \mathbb{R}^2 \mid x < 0\}$ and $v_0 = (0, 0)$. The half-planes A and B are both open and path-connected, hence connected, but their union is not connected. However, with the inclusion of $(0, 0)$, the union becomes path-connected. Any point v_0 in A can be connected to any point w_0 in B by going along a straight line $\gamma(t) = tv_0$ to $(0, 0)$ and then out along the straight line $\mu(t) = tw_0$ to w_0. Thus $A \cup B \cup \{(0, 0)\}$ is path-connected and so connected. ♦

◇ **3E-23.** Let $\mathbb{Q}$ denote the rationals in $\mathbb{R}$. Show that both $\mathbb{Q}$ and the irrationals $\mathbb{R}\backslash\mathbb{Q}$ are not connected.

Sketch. $\mathbb{Q} \subset]-\infty, \sqrt{2}\,[\, \cup \,[\sqrt{2}\,, \infty[$; both intervals are open, they are disjoint. They disconnect $\mathbb{Q}$. Similarly $\mathbb{R}\backslash\mathbb{Q} \subset]-\infty, 0[\, \cup\,]0, \infty[$ disconnects $\mathbb{R} \setminus \mathbb{Q}$. ◊

Solution. To show that $\mathbb{Q} \subseteq \mathbb{R}$ is not connected, recall that $\sqrt{2}$ is not rational. The two open half lines $U = \{x \in \mathbb{R} \mid x < \sqrt{2}\}$ and $V = \{x \in \mathbb{R} \mid x > \sqrt{2}\}$ are disjoint. Each intersects $\mathbb{Q}$ since $0 \in U$ and $3 \in V$. Their union is $\mathbb{R} \setminus \{\sqrt{2}\}$ which contains $\mathbb{Q}$. Thus U and V disconnect $\mathbb{Q}$.

To show that $\mathbb{R} \setminus \mathbb{Q}$ is not connected, we do essentially the same thing but use a rational point such as 0 as the separation point. Let $U = \{x \in \mathbb{R} \mid x < 0\}$ and $V = \{x \in \mathbb{R} \mid x > 0\}$. Then U and V are disjoint open half lines. Each intersects $\mathbb{R} \setminus \mathbb{Q}$ since $-\sqrt{2} \in U$ and $\sqrt{2} \in V$. Their union is $\mathbb{R} \setminus \{0\}$ which contains $\mathbb{R} \setminus \mathbb{Q}$. Thus U and V disconnect $\mathbb{R} \setminus \mathbb{Q}$. ♦

◇ **3E-24.** Prove that a set $A \subset M$ is not connected if we can write A as the disjoint union of two sets B and C such that $B \cap A \neq \varnothing$, $C \cap A \neq \varnothing$, and neither of the sets B or C has a point of accumulation belonging to the other set.

Suggestion. Show $\mathrm{cl}(B) \cap C$ and $\mathrm{cl}(C) \cap B$ are empty. Show that the complements of these closures disconnect A. ◊

Solution. Suppose A is a subset of a metric space M with nonempty disjoint subsets B and C such that neither B nor C has an accumulation point in the other. Then $\mathrm{cl}(B) \cap C = \emptyset$ since a point is in the closure of B if and only if it is either in B or is an accumulation point of B. None of these are in C. Similarly, $\mathrm{cl}(C) \cap B = \emptyset$. Let $U = M \setminus \mathrm{cl}(B)$ and $V = M \setminus \mathrm{cl}(C)$. Then U and V are open since they are the complements of the closed sets $\mathrm{cl}(B)$ and $\mathrm{cl}(C)$. Since $B \cap \mathrm{cl}(C) = \emptyset$, we have $B \subseteq V$, and since $C \cap \mathrm{cl}(B) = \emptyset$, we have $C \subseteq U$. In fact:

(1) $U \cap V \cap A = \emptyset$: To see this compute

$$\begin{aligned} U \cap V &= (M \setminus \mathrm{cl}(B)) \cap (M \setminus \mathrm{cl}(C)) = M \setminus (\mathrm{cl}(B) \cup \mathrm{cl}(C)) \\ &\subseteq M \setminus (B \cup C) = M \setminus A. \end{aligned}$$

(2) $B \subseteq V$, so $A \cap V$ is not empty.

(3) $C \subseteq U$, so $A \cap U$ is not empty.

(4) $A = B \cup C \subseteq U \cup V$.

So the open sets U and V disconnect A.

For a related result, see Exercise 3E-21. ♦

◇ **3E-25.** Prove that there is a sequence of distinct integers $n_1, n_2, \ldots \to \infty$ such that $\lim_{k\to\infty} \sin n_k$ exists.

Sketch. The sequence $\sin n$ for $n = 1, 2, 3, \ldots$ is contained in the compact set $[-1, 1]$ and hence has a convergent subsequence $\sin(n_k)$. ◊

Solution. We know that $|\sin x| \leq 1$ for every real number x. In particular, $|\sin n| \leq 1$ for every integer n. The sequence $\langle a_n \rangle_1^\infty$ defined by $a_n = \sin n$ for $n = 1, 2, 3, 4, \ldots$ is a sequence in the closed bounded set $[-1, 1] \subseteq \mathbb{R}$. Since this set is sequentially compact, there must be a convergent subsequence. There are integers $n_1 < n_2 < n_3 < \ldots$ and a number $\lambda \in [-1, 1]$ such that $\lim_{k\to\infty} \sin n_k = \lambda$. ♦

◇ **3E-26.** Show that the completeness property of $\mathbb{R}$ may be replaced by the *Nested Interval Property*: If $\{F_n\}_1^\infty$ is a sequence of closed bounded intervals in $\mathbb{R}$ such that $F_{n+1} \subset F_n$ for all $n = 1, 2, 3, \ldots$, then there is at least one point in $\bigcap_{n=1}^\infty F_n$.

Solution. We want to show that the following two assertions are equivalent:

Completeness Property: *If $\langle a_n \rangle_1^\infty$ is a monotonically increasing sequence bounded above in $\mathbb{R}$, then it converges to some point in $\mathbb{R}$.*

Nested Interval Property: *If $\langle F_n \rangle_1^\infty$ is a sequence of closed bounded intervals in $\mathbb{R}$ such that $F_{n+1} \subseteq F_n$ for all $n = 1, 2, 3, \ldots$, then there is at least one point in $\bigcap_{n=1}^\infty F_n$.*

The development of the text starts with the assumption of the completeness property and proves that closed bounded intervals are compact. The nested interval property as stated then follows from the Nested Set Property (Theorem 3.3.1) which says that the intersection of a nested sequence of nonempty compact sets is not empty.

What is needed now is to start with the assumption of the nested interval property and to show that the completeness property can be proved from it. So, suppose the nested interval property is true in $\mathbb{R}$ and $\langle a_n \rangle_1^\infty$ is an increasing sequence in $\mathbb{R}$ with $a_n \le B \in \mathbb{R}$ for all n. We want to show that the a_n must converge to a limit in $\mathbb{R}$. The idea is to use the points a_n as the left ends of a nested family of intervals. The right ends, b_n, are to be selected so that the lengths of the intervals $[a_n, b_n]$ tend to 0. If we can manage this we will apply the following variation on the nested interval property.

Lemma. *If $F_n = [a_n, b_n]$, $n = 1, 2, 3, \ldots$ are closed bounded intervals in $\mathbb{R}$ such that $F_{n+1} \subseteq F_n$ for each n and $b_n = a_n \to 0$ as $n \to \infty$, then there is exactly one point λ in $\bigcap_{n=1}^\infty F_n$ and $a_n \to \lambda$ and $b_n \to \lambda$.*

Proof: That there is at least one point λ in the intersection follows from the nested interval property. Any other point μ would be excluded from any of the intervals with $b_n - a_n < |\lambda - \mu|$, so there is exactly one point in the intersection. If $\varepsilon > 0$, then there is an index N with $b_N - a_N < \varepsilon$. If $n \ge N$, we have $a_N \le a_n \le \lambda \le b_n \le b_N$. So $|\lambda - a_n| < \varepsilon$, and $|b_n - \lambda| < \varepsilon$. Thus $a_n \to \lambda$ and $b_n \to \lambda$ as claimed.

Returning to our original problem, we let $b_0 = B$ so that $b_0 \ge a_m$ for all m. The right hand endpoints b_n are produced inductively. Having selected $b_0 \ge b_1 \ge b_2 \ge \cdots \ge b_n$ with $b_n \ge a_m$ for all m, we need to specify b_{n+1}. If there is a positive integer k with $b_n - (1/k) \ge x_m$ for all m, then we let k_n be the smallest such integer and put $b_{n+1} = b_n - (1/k_n)$. If there is no such integer, put $b_{n+1} = b_n$. This produces a sequence $\langle b_n \rangle_1^\infty$ such that $b_{n+1} \le b_n$ for each n and $b_n \ge a_m$ for all indices n and m. We will be able to use the lemma to conclude that there is a real number λ to which the points a_n converge as soon as we know that $b_n - a_n \to 0$.

If $b_n - a_n$ did not tend to 0, there would be an $\varepsilon > 0$ such that $b_m - a_m > \varepsilon$ for all indices m. Focus attention on one of the b_n. If $m \ge n$ we would have $a_n \le a_m < b_m \le b_n$ so $b_n - a_m \ge b_m - a_m > \varepsilon$. If $m < n$ we would have $a_m \le a_n < b_n \le b_m$, so again $b_n - a_m \ge b_n - a_n > \varepsilon$. So $b_n - \varepsilon > a_m$ for all indices m and n. Fix a positive integer N with $0 < 1/N < \varepsilon$. Then $b_n - (1/N) > x_m$ for all m, so $1 \le k_n \le N$. This would imply that

$$b_{n+1} = b_n - \frac{1}{k_n} \le b_n - \frac{1}{N} \qquad \text{for every index } n.$$

This would mean that the points b_n would tend to $-\infty$ and in a finite number of steps would be smaller than a_1. Since this is not the case, we conclude that $b_n - a_n \to 0$. The length of our closed intervals tends to 0, we can apply the lemma to conclude that there is exactly one point in their intersection and that the right ends, a_n converge to that point.

It is worth noting that the Nested *Set* Property as phrased in Theorem 3.3.1 does not imply the completeness property. It is a perfectly good theorem about compact sets which follows directly from the abstract definition of compactness. It would apply perfectly well to nested sequences of compact subsets of the rational line. Of course it does not imply that the rational line is complete. The difficulty is that to make the argument work we would need to know that closed bounded intervals are compact, and they are not in the rational line. ♦

◇ **3E-27.** Let $A \subset \mathbb{R}$ be a bounded set. Show that A is closed iff for every sequence $x_n \in A$, $\limsup x_n \in A$ and $\liminf x_n \in A$.

Sketch. If $x \in \mathrm{cl}(A)$, there is a sequence $\langle x_n \rangle_1^\infty$ in A converging to x. If the condition holds, then $x \in A$. (Why?) So $\mathrm{cl}(A) \subseteq A$, and A is closed. For the converse, if A is closed and bounded, the lim inf and lim sup of a sequence in A are the limits of subsequences, so they are in A. ◇

Solution. First suppose A is a bounded set in $\mathbb{R}$ and that $\limsup x_n$ and $\liminf x_n$ are in A for every sequence $\langle x_n \rangle_1^\infty$ of points in A. If $x \in \mathrm{cl}(A)$, then there is a sequence $\langle x_n \rangle_1^\infty$ of points in A converging to x. Since the limit exists, we have

$$x = \lim_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n.$$

By hypothesis, this is in A. This is true for every x in $\mathrm{cl}(A)$. So $\mathrm{cl}(A) \subseteq A$, and A is closed.

For the converse, suppose A is a closed, bounded set in $\mathbb{R}$ and that $\langle x_n \rangle_1^\infty$ is a sequence in A. Since A is a bounded set, the sequence is bounded and $a = \liminf x_n$ and $b = \limsup x_n$ exist as finite real numbers. Furthermore, there are subsequences $x_{n(j)}$ and $x_{k(j)}$ converging to a and b respectively. Since A is closed, the limits of these subsequences must be in A. ♦

◇ **3E-28.** Let $A \subset M$ be connected and contain more than one point. Show that every point of A is an accumulation point of A.

Solution. Suppose a and b are different points in A and that a is not an accumulation point of A. Then there is a radius $r > 0$ such that $D(a,r) \cap A = \{a\}$. Let $U = \{x \in M \mid d(x,a) < r/3\}$ and $V = \{x \in M \mid d(x,a) > 2r/3\}$. Then U and V are open (why?) and disjoint. Neither $A \cap U$ nor $A \cap V$ is empty since $a \in A \cap U$ and $b \in A \cap V$. But $A \subseteq U \cup V$ since there are no points x in A with $r/3 \le d(x,a) \le 2r/3$. So the sets U and V would disconnect A. Since A is connected, this is not possible. Thus a must be an accumulation point of A. Since a was arbitrary in A assuming that there was at least one other point in A, this establishes our assertion. ♦

◇ **3E-29.** Let $A = \{(x, y) \in \mathbb{R}^2 \mid x^4 + y^4 = 1\}$. Show that A is compact. Is it connected?

Sketch. A is both compact and connected. ♢

Solution. If x and y are real numbers with $x^4 + y^4 = 1$, then $0 \leq x^4 \leq 1$ and $0 \leq y^4 \leq 1$, so $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$ and $\| (x, y) \| = \sqrt{x^2 + y^2} \leq \sqrt{2}$. This shows that the set A is bounded. Suppose $v_n = (x_n, y_n)$ is a sequence of points in A which converge to $w = (x, y)$. Then $x_n \to x$ and $y_n \to y$. So $x_n^4 \to x^4$ and $y_n \to y^4$. Thus $x_n^4 + y_n^4 \to x^4 + y^4$. But $x_n^4 + y_n^4 = 1$ for each n, so $x^4 + y^4 = 1$, and $w = (x, y) \in A$. This shows that the set A is closed. Since it is a closed bounded subset of $\mathbb{R}^2$, it is a compact set.

One way to see that A is connected is to observe that it is path-connected. This can be established by writing A as the union of the graphs of two continuous functions. $A = A^+ \cup A^-$ where

$$A^+ = \{(x, y) \in \mathbb{R}^2 \mid -1 \leq x \leq 1 \text{ and } y = \sqrt[4]{1 - x^4}\,\}$$

and

$$A^- = \{(x, y) \in \mathbb{R}^2 \mid -1 \leq x \leq 1 \text{ and } y = -\sqrt[4]{1 - x^4}\,\}.$$

Two points in the same part of A can be connected by following along the graph of the appropriate continuous function. Points in different parts of A can be connected by connecting each to $(1, 0)$. ♦

◇ **3E-30.** Let U_k be a sequence of open bounded sets in $\mathbb{R}^n$. Prove or disprove:

(a) $\bigcup_{k=1}^{\infty} U_k$ is open.
(b) $\bigcap_{k=1}^{\infty} U_k$ is open.
(c) $\bigcap_{k=1}^{\infty} (\mathbb{R}^n \backslash U_k)$ is closed.
(d) $\bigcap_{k=1}^{\infty} (\mathbb{R}^n \backslash U_k)$ is compact.

Answer. (a) Yes.
(b) Not necessarily.
(c) Yes.
(d) Not necessarily. ♢

Solution. (a) The union of any collection of open subsets of $\mathbb{R}^n$ is open, so in particular $\bigcup_{k=1}^{\infty} U_k$ is open.

(b) The intersection of an infinite collection of open sets need not be open. Boundedness does not help. Let $U_k = \{v \in \mathbb{R}^n \mid \| v \| < 1/k\}$. Each of the sets U_k is open and bounded. But $\bigcap_{k=1}^{\infty} U_k = \{0\}$ which is not open.

(c) The sets $\mathbb{R}^n \setminus U_k$ are closed since the U_k are open. The intersection of any family of closed subsets of $\mathbb{R}^n$ is closed, so in particular $\bigcap_{k=1}^{\infty} (\mathbb{R}^n \setminus U_k)$ is closed.

(d) As in part (c) the intersection is closed, but it need not be bounded. Let U_k be the sets defined in part (b). Then $U_1 \supseteq U_2 \supseteq U_3 \supseteq U_4 \supseteq \dots$. So

$$(\mathbb{R}^n \setminus U_1) \subseteq (\mathbb{R}^n \setminus U_2) \subseteq (\mathbb{R}^n \setminus U_3) \subseteq \dots.$$

Thus

$$\bigcap_{k=1}^{\infty} (\mathbb{R}^n \setminus U_k) = \mathbb{R}^n \setminus U_1 = \{v \in \mathbb{R}^2 \mid \| v \| \geq 1\}.$$

This set is not bounded and so is not compact. ♦

◊ **3E-31.** Suppose $A \subset \mathbb{R}^n$ is not compact. Show that there exists a sequence $F_1 \supset F_2 \supset F_3 \cdots$ of closed sets such that $F_k \cap A \neq \varnothing$ for all k and

$$\left(\bigcap_{k=1}^{\infty} F_k \right) \bigcap A = \varnothing.$$

Suggestion. The set A must be either not closed or not bounded or both. Treat these cases separately. If A is not closed, there must be an accumulation point of A which is not in A. ◊

Solution. If A is not bounded, let $F_k = \{v \in \mathbb{R}^n \mid \| v \| \geq k\}$. Since A is not bounded, $F_k \cap A$ is not empty. However, $\bigcap_{k=1}^{\infty} F_k = \emptyset$, so $(\bigcap_{k=1}^{\infty} F_k) \cap A$ is certainly empty.

If A is not closed, then there is a point $v_0 \in \mathrm{cl}(A) \setminus A$. Such a point must be an accumulation point of A. For $k = 1, 2, 3, 4, \dots$, let $F_k = \{v \in \mathbb{R}^n \mid \| v - v_0 \| \leq 1/k\}$. Then each F_k is closed and each F_k intersects A since v_0 is an accumulation point of A. We certainly have $F_1 \supseteq F_2 \supseteq F_3 \supseteq \dots$. But $\bigcap_{k=1}^{\infty} F_k = \{v_0\}$ and v_0 is not in A. So $(\bigcap_{k=1}^{\infty} F_k) \cap A = \emptyset$.

If A is not compact then it is either not bounded or not closed or both, so at least one of the previous paragraphs applies. ♦

◊ **3E-32.** Let x_n be a sequence in $\mathbb{R}^3$ such that $||x_{n+1} - x_n|| \leq 1/(n^2 + n)$, $n \geq 1$. Show that x_n converges.

Suggestion. Use the triangle inequality and the hypothesis to show that the sequence is a Cauchy sequence. ◊

Solution. For each index k we have

$$\| x_{k+1} - x_k \| \le \frac{1}{k^2 + k} = \frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}.$$

So we can compute

$$\begin{aligned}
\| x_{n+p} - x_n \| &= \| x_{n+p} - x_{n+p-1} + x_{n+p-1} + \cdots - x_{n+1} + x_{n+1} - x_n \| \\
&\le \| x_{n+p} - x_{n+p-1} \| + \| x_{n+p-1} - x_{n+p-2} \| + \cdots + \| x_{n+1} - x_n \| \\
&\le \left(\frac{1}{n+p-1} - \frac{1}{n+p} \right) + \left(\frac{1}{n+p-2} - \frac{1}{n+p-1} \right) + \ldots \\
&\qquad + \left(\frac{1}{n} - \frac{1}{n+1} \right) \\
&\le \frac{1}{n} - \frac{1}{n+p} \le \frac{1}{n}.
\end{aligned}$$

If $\varepsilon > 0$, and N is large enough so that $1/N < \varepsilon$, then $\| x_{n+p} - x_n \| < 1/n \le 1/N < \varepsilon$ whenever $n \ge N$ and $p > 0$. Our sequence is thus a Cauchy sequence and must converge since $\mathbb{R}^3$ is complete.

Here is a solution which uses a little more knowledge and less of a clever trick. We know from the integral test that the infinite series $\sum_{k=1}^{\infty}(1/k^2)$ converges. So its partial sums must be a Cauchy sequence in $\mathbb{R}$. Given $\epsilon > 0$, there is an $N(\varepsilon)$ such that $n \ge N(\varepsilon)$ and $p > 0$ imply that $\sum_{k=n+1}^{n+p}(1/k^2) < \varepsilon$.

Now suppose $n \ge N(\varepsilon)$ and $p > 0$. Then

$$\begin{aligned}
\| x_{n+p} - x_n \| &= \| x_{n+p} - x_{n+p-1} + x_{n+p-1} + \cdots - x_{n+1} + x_{n+1} - x_n \| \\
&\le \| x_{n+p} - x_{n+p-1} \| + \| x_{n+p-1} - x_{n+p-2} \| + \cdots + \| x_{n+1} - x_n \| \\
&\le \frac{1}{(n+p-1)(n+p)} + \frac{1}{(n+p-2)(n+p-1)} + \cdots + \frac{1}{n(n+1)} \\
&\le \frac{1}{(n+p-1)^2} + \frac{1}{(n+p-2)^2)} + \cdots + \frac{1}{n^2} < \varepsilon.
\end{aligned}$$

Again the sequence is a Cauchy sequence in $\mathbb{R}^3$ and must converge. ♦

◇ **3E-33.** ***Baire category theorem.*** A set S in a metric space is called ***nowhere dense*** if for each nonempty open set U, we have $\mathrm{cl}(S) \cap U \ne U$, or equivalently, $\mathrm{int}(\mathrm{cl}(S)) = \varnothing$. Show that $\mathbb{R}^n$ cannot be written as the countable union of nowhere dense sets.

Sketch. If $A_1, A_2, A_3, \ldots$ are closed nowhere dense sets, put $B_k = \mathbb{R}^n \backslash A_k$ to see that the assertion is implied by the following theorem.

Theorem. *If $B_1, B_2, \ldots$ are open dense subsets of $\mathbb{R}^n$, then $B = \bigcap_k B_k$ is dense in $\mathbb{R}^n$.*

Let $x \in \mathbb{R}^n$ and $\epsilon > 0$. Inductively define a nested sequence of closed disks $B(x_k, r_k) = \{y \mid \| x_k - y \| \leq r_k\}$ by

$$\begin{aligned}
D_1 = B(x_1, r_1) &\text{ where } x_1 \in B_1, \quad \| x - x_1 \| < \varepsilon/3, \quad r_1 < \varepsilon/3, \\
&\text{ and } B(x_1, r_1) \subset B_1 \\
D_2 = B(x_2, r_2) &\text{ where } x_2 \in B_2, \quad \| x_1 - x_2 \| < r_1/3, \quad r_2 < r_1/3, \\
&\text{ and } B(x_2, r_2) \subset B_2 \\
\text{etc.}
\end{aligned}$$

Check that $D_k \subseteq B_k$ for each k and $D(x, \varepsilon) \supseteq \mathrm{cl}(D_1) \supseteq D_1 \supseteq \mathrm{cl}(D_2) \supseteq D_2 \supseteq \cdots$. Existence of $y \in \bigcap_k \mathrm{cl}(D_k)$ shows that B is dense in $\mathbb{R}^n$. (Why?) $\diamondsuit$

Solution. To say that a set is "nowhere dense" is to say that its closure has empty interior.

$$\begin{aligned}
A \text{ is nowhere dense} &\iff (U \text{ open} \implies U \cap \mathrm{cl}(A) \neq U) \\
&\iff \mathrm{int}(\mathrm{cl}(A)) = \emptyset.
\end{aligned}$$

Suppose A is such a set and B is the complement of A. If $v_0 \in A$, then every neighborhood of v_0 must intersect B since otherwise v_0 would be an interior point of A, and there are none. So $v_0 \in \mathrm{cl}(B)$. Thus $A \subseteq \mathrm{cl}(B)$. Since $B \subseteq \mathrm{cl}(B)$ also, we see that the whole space is contained in the closure of B. That is, B is "dense". Conversely, if B is dense, then its complement can have no interior. If A is closed so that $A = \mathrm{cl}(A)$, this says that A is nowhere dense and that B is open. We summarize these observations.

Lemma. *The set A is closed and nowhere dense if and only if its complement is an open dense set.*

The Baire category theorem asserts that $\mathbb{R}^n$, or more generally, any complete metric space, is not the union of countably many closed, nowhere dense sets. Suppose $A_1, A_2, A_3, \ldots$, are closed nowhere dense sets and $B_1, B_2, B_3, \ldots$, are their complements in a metric space M. Then

$$\bigcup_{k=1}^{\infty} A_k = M \iff \emptyset = M \setminus \bigcup_{k=1}^{\infty} A_k = \bigcap_{k=1}^{\infty} (M \setminus A_k) = \bigcap_{k=1}^{\infty} B_k.$$

So the assertion that M is not such a countable union is equivalent to saying that the last intersection is not empty. In fact, with a bit more care, we see that it must be dense. Here are several forms of the theorem:

Baire Theorem (version (i)) *Suppose M is a nonempty complete metric space and that $A_1, A_2, A_3, \ldots$, are closed nowhere dense subsets of M. Then $\bigcup_{k=1}^{\infty} A_k \neq M$.*

Baire Theorem (version (ii)) *Suppose M is a complete metric space and that $A_1, A_2, A_3, \ldots$, are closed subsets of M with $\bigcup_{k=1}^{\infty} A_k = M$. Then at least one of the sets A_k must have nonempty interior.*

Baire Theorem (version (iii)) *Suppose M is a complete metric space and that $B_1, B_2, B_3, \ldots$, are open dense subsets of M. Then $\bigcap_{k=1}^{\infty} B_k$ is not empty.*

Baire Theorem (version (iv)) *Suppose M is a complete metric space and that $B_1, B_2, B_3, \ldots$, are open dense subsets of M. Then $\bigcap_{k=1}^{\infty} B_k$ is dense in M.*

Version (iv) certainly implies version (iii) since an empty set cannot be dense in the nonempty space M. If version (iii) holds and $A_1, A_2, A_3, \ldots$, are closed nowhere dense subsets of M, then their complements B_k are open dense sets, and their intersection would not be empty. The computation above shows that the union of the A_k cannot be all of M. Thus version (iii) implies version (i). Versions (i) and (ii) are essentially contrapositives of each other and so are equivalent.

We prove version (iv):

Let $B_1, B_2, B_3, \ldots$ be open dense subsets of M, and let $B = \bigcap_{k=1}^{\infty} B_k$. To show that B is dense in M, let $x \in M$ and $\varepsilon > 0$. We need to show that there is a point y in B with $d(x, y) < \varepsilon$. To do this we will construct a Cauchy sequence $\langle x_k \rangle_1^{\infty}$ with $x_k \in B_k \cap \mathrm{cl}(D(x, \varepsilon/2))$ for each k and obtain y as the limit of that sequence. The construction proceeds inductively.

STEP ONE: Since B_1 is dense, there is a point $x_1 \in B_1$ with $d(x, x_1) < \varepsilon/3$. Since B_1 is open, there is an $r_1 < \varepsilon/3$ with $D_1 = D(x_1, r_1) \subseteq B_1$. If $x \in D(x_1, r_1)$, then $d(x, z) \le d(x, x_1) + d(x_1, z) < 2\varepsilon/3$. So $\mathrm{cl}(D_1) \subseteq D(x, \varepsilon)$.

STEP TWO: Since B_2 is dense, there is a point $x_2 \in B_2$ with $d(x_2, x_1) < r_1/3$. Since B_2 is open, there is a positive r_2 with $r_2 < r_1/3 < \varepsilon/9$ and $D_2 = D(x_2, r_2) \subseteq B_2$. If $z \in D_2$, then $d(z, x_1) \le d(z, x_2) + d(x_2, x_1) < 2r_1/3$, so $D_2 \subseteq \mathrm{cl}(D_2) \subseteq D_1$.

$$\vdots$$

STEP $k+1$: We have found points $x_1, x_2, x_3, \ldots, x_k$ and positive radii $r_1, r_2,, r_3, \ldots\ r_k$ with

1. $D_j = D(x_j, r_j) \subseteq B_j$ for each $j = 1, 2, 3, \ldots k$.
2. $D(x, \varepsilon) \supseteq \mathrm{cl}(D_1) \supseteq D_1 \supseteq \mathrm{cl}(D_2) \supseteq D_2 \supseteq \cdots \supseteq D_{k-1} \supseteq \mathrm{cl}(D_k) \supseteq D_k$.
3. $r_j < \varepsilon/3^j$ for $j = 1, 2, 3, \ldots, k$.

Use the fact that B_{k+1} is dense to pick $x_{k+1} \in B_{k+1}$ with $d(x_{k+1}, x_k) < r_k/3$ and the fact that B_{k+1} is open to select an r_{k+1} with $0 < r_{k+1} < r_k/3 < \varepsilon/3^{k+1}$ such that $D_{k+1} = D(x_{k+1}, r_{k+1}) \subseteq B_{k+1}$. If $z \in D_{k+1}$, then $d(z, x_k) \le d(z, x_{k+1}) + d(x_{k+1}, x_k) < 2r_k/3$, so $D_{k+1} \subseteq \mathrm{cl}(D_{k+1}) \subseteq D_k$.

$$\vdots$$

Continued inductively, this process produces a sequence of points $x_1, x_2, x_3, \ldots$ and positive radii $r_1, r_2,, r_3, \ldots$ with

1. $D_k = D(x_k, r_k) \subseteq B_k$ for each $k = 1, 2, 3, \ldots$.
2. $D(x, \varepsilon) \supseteq \mathrm{cl}(D_1) \supseteq D_1 \supseteq \mathrm{cl}(D_2) \supseteq \cdots \supseteq D_{k-1} \supseteq \mathrm{cl}(D_k) \supseteq D_k \supseteq \ldots$.
3. $r_k < \varepsilon/3^k$ for $k = 1, 2, 3, \ldots$.

See the figure.

Figure 3-3. The construction for the Baire Category Theorem.

In $\mathbb{R}^n$ we can finish the proof very quickly. The sets $F_k = \mathrm{cl}(D_k)$ are a nested sequence of nonempty compact sets. From the Nested Set Property (3.3.1), we know that there must be at least one point y in their intersection. For such a y we have

$$y \in \bigcap_{k=2}^{\infty} \mathrm{cl}(D_k) \subseteq \bigcap_{k=1}^{\infty} D_k \subseteq D(x, \varepsilon) \cap B$$

and we have the point we need.

In some other complete metric space, the closed balls $\mathrm{cl}(D_k)$ might not be compact, and we need to do a little more work. These closed balls are nested, and their radii, r_k tend to 0. So, if $\varepsilon' > 0$, there is a K such that $r_k < \varepsilon'$ whenever $k \geq K$. For such k and $p > 0$, we know that x_k and x_{k+p} are both in $D_k = D(x_k, r_k)$. So $d(x_{k+p}, x_k) < \varepsilon'$. Thus the sequence $\langle x_k \rangle_1^\infty$ is a Cauchy sequence in the complete metric space M and must converge to some point $y \in M$. Since $x_k \in D_m$ whenever $k \geq m$ and $x_k \to y$, we must have $y \in \mathrm{cl}(D_m)$. Since this holds for each m we have

$$y \in \bigcap_{m=2}^{\infty} \mathrm{cl}(D_m) \subseteq \bigcap_{k=m}^{\infty} D_m \subseteq D(x, \varepsilon) \cap B$$

just as before, and y is the point we need.

Some care is needed in the interpretation of this theorem. We know that a closed subset of a complete metric space is complete. For example, the set $\mathbb{Z}$ of integers is complete since they are all separated from each other by a distance of at least one. The only way that a sequence of integers can be a Cauchy sequence is for it to be eventually constant. So every Cauchy sequence of integers certainly converges to an integer. But $\mathbb{Z}$ is a countable union of single points. Why does this not contradict the Baire Category Theorem. The trick is in being very careful about what we mean by open sets. If we are considering the integers as a metric *space* in its own right and not as a subset of a larger space, then those single point sets are open $\{n\} = D(n, 1/2)$. So $\{n\}$ is both closed and open. It is not a closed nowhere dense set. It has interior $\mathrm{int}(\{n\}) = \{n\}$. ♦

◇ **3E-34.** Prove that each closed set $A \subset M$ is an intersection of a countable family of open sets.

Suggestion. Consider $U_k = \bigcup_{x \in A} D(x, 1/k)$. ◇

Solution. For each positive integer k, let $U_k = \bigcup_{x \in A} D(x, 1/k)$. Since each of the disks $D(x, 1/k)$ is open, their union U_k is also open. Since $x \in D(x, 1/k)$, we certainly have $A \subseteq \bigcup_{x \in A} D(x, 1/k) = U_k$. Since this is true for each k, we have $A \subseteq \bigcap_{k+1}^{\infty} U_k$. On the other hand, if y is in this intersection and $\varepsilon > 0$, then we can find an integer k with $0 < 1/k < \varepsilon$. Since $y \in U_k$, there is an $x \in A$ with $y \in D(x, 1/k) \subseteq D(x, \varepsilon)$. Since this can be done for each $\varepsilon > 0$, we have $y \in \mathrm{cl}(A)$. But since A is closed, we have $\mathrm{cl}(A) = A$, and $y \in A$. Thus $\bigcap_{k=1}^{\infty} U_k \subseteq A$. We have inclusion in both directions, so $A = \bigcap_{k=1}^{\infty} U_k$ which is the intersection of a countable collection of open sets. ♦

◇ **3E-35.** Let $a \in \mathbb{R}$ and define the sequence $a_1, a_2, \ldots$ in $\mathbb{R}$ by $a_1 = a$, and $a_n = a_{n-1}^2 - a_{n-1} + 1$ if $n > 1$. For what $a \in \mathbb{R}$ is the sequence

(a) Monotone?
(b) Bounded?
(c) Convergent?

Compute the limit in the cases of convergence.

Answer. (a) All a. If $a = 0$ or 1, the sequence is constant.

(b) $0 \leq a \leq 1$.

(c) $0 \leq a \leq 1$. ◊

Solution. Let $f(x) = x^2 - x + 1$. Our sequence is defined by $a_1 = a$ and $a_{n+1} = f(a_n)$ for $n = 1, 2, 3, \ldots$.. The graph of $y = f(x)$ is a parabola opening upward. Its vertex is at the point $x = 1/2$, $y = 3/4$.

(a) For each $n \geq 1$ we have

$$a_{n+1} - a_n = a_n^2 - 2a_n + 1 = (a_n - 1)^2 \geq 0.$$

So the sequence is monotonically increasing (or at least nondecreasing), no matter what the starting point is. If $a = 1$, then the differences are always 0 and the sequence is constant. If $a = 0$, then $a_1 = 1$, and the sequence is constant beyond that point.

$a_{n+1} = a_n$ if and only if $a_n = 1$, so the sequence is strictly increasing unless there is some n with $a_n = 1$. But $a_n = 1$ if and only if a_{n-1} is either 1 or 0, and 0 has no possible predecessor since the equation $x^2 - x + 1 = 0$ has no real root. Thus the sequence is increasing for all a and strictly increasing unless $a = 0$ or $a = 1$.

(b) The function $f(x) = x^2 - x + 1$ defining our sequence has an absolute minimum value of $3/4$ occurring at $x = 1/2$. Also $f(0) = f(1) = 1$, and if $0 \leq x \leq 1$, then $3/4 \leq f(x) \leq 1$. If $0 \leq a_k \leq 1$, this shows that $0 \leq f(a_k) = a_{k+1} \leq 1$. If $0 \leq a = a_1 \leq 1$, then it follows by induction that $0 \leq a_n \leq 1$ for all n, and the sequence is bounded.

If $a_k > 1$, then $a_m > a_k > 1$ for all $m > k$ by part (a). Furthermore,

$$\begin{aligned}
(a_{m+2} - a_{m+1}) - (a_{m+1} - a_m) &= (a_{m+1} - 1)^2 - (a_m - 1)^2 \\
&= (a_{m+1}^2 - 2a_{m+1} + 1) - (a_m - 2a_m + 1) \\
&= a_{m+1}^2 - a_m^2 - 2(a_{m+1} - a_m) \\
&= (a_{m+1} + a_m - 2)(a_{m+1} - a_m) \\
&\geq 0.
\end{aligned}$$

So the differences between successive terms is increasing. The terms must diverge to $+\infty$.

If $a = a_1 > 1$, then this analysis applies directly and the sequence must diverge to $+\infty$.

If $a = a_1 < 0$, then $a_2 = f(a_1) = f(a) = a^2 - a + 1 = a^2 + |a| + 1 > 1$. So the analysis still applies and the sequence must diverge to $+\infty$.

Combining these observations, we see that the sequence is bounded if and only if $0 \leq a \leq 1$.

(c) From parts (a) and (b), we know that if $0 \le a \le 1$, then the sequence is a bounded monotone sequence in $\mathbb{R}$ and must converge to some $\lambda \in \mathbb{R}$. From part (b), we know that if a is not in this interval, then the sequence diverges to $+\infty$. Thus the sequence converges if and only if $0 \le a \le 1$. If the limit λ exists, then $a_{n+1} \to \lambda$ and $a_n^2 - a_n + 1 \to \lambda^2 - \lambda + 1$. Since $a_{n+1} = a_n^2 - a_n + 1$ and limits are unique, we must have $\lambda = \lambda^2 - \lambda - 1$. So $0 = \lambda^2 - 2\lambda + 1 = (\lambda - 1)^2$. Thus we must have $\lim_{n\to\infty} a_n = \lambda = 1$.

The action of f on a to produce the sequence $\langle a_n \rangle_1^\infty$ can be very effectively illustrated in terms of the graph of the function $y = f(x)$. Starting with a_k on the x-axis, the point $(a_k, f(a_k))$ is located on the graph of f. The point a_{k+1} is then located on the x-axis by moving horizontally to the line $y = x$, and then vertically to the x-axis. The fixed point 1 occurs when the graph of f touches the line $y = x$. Repetition produces a visual representation of the behavior of the sequence. See the figure.

FIGURE 3-4. Iteration of a function to produce a sequence.

The iteration of the function f to produce the sequence $\langle a_n\rangle_1^\infty$ from the starting point a is an example of a *discrete dynamical system*. The sequence obtained is called the *orbit* of the point. The study of the behavior of sequences obtained from the iteration of functions is a rich field with many applications in mathematics and other fields. It can lead to nice clean behavior such as we have seen here or to much more complicated behavior now characterized by the term "chaos". A couple of nice references are:

An Introduction to Chaotic Dynamical Systems, Robert L. Devaney, Addison-Wesley Publishing Company.

Encounters with Chaos, Denny Gulick, McGraw-Hill, Inc. ♦

◇ **3E-36.** Let $A \subset \mathbb{R}^n$ be uncountable. Prove that A has an accumulation point.

Suggestion. Consider the sets $A_k = A \cap \{v \in \mathbb{R}^n \mid \|v\| \le k\}$. ◊

Solution. For $k = 1, 2, 3, \ldots$, let $A_k = A \cap \{v \in \mathbb{R}^n \mid \|v\| \le k\}$. Then $A = \bigcup_{k=1}^\infty A_k$. If each of the countably many sets A_k were finite, then their union A would be countable. Since it is not, at least one of the sets A_k must have infinitely many points. We can choose a sequence of points $a_1, a_2, a_3, \ldots$ in A_k all different. This is a bounded sequence in $\mathbb{R}^n$, so by the Bolzano-Weierstrass property of $\mathbb{R}^n$, it must have a subsequence converging to some point $w \in \mathbb{R}^n$. (The closed bounded set $\mathrm{cl}(A_k) \subseteq \mathbb{R}^n$ is sequentially compact.) Points in A_k are in A, so there is a sequence of distinct points of A converging to w and w must be an accumulation point of A. ♦

◇ **3E-37.** Let $A, B \subset M$ with A compact, B closed, and $A \cap B = \varnothing$.

(a) Show that there is an $\varepsilon > 0$ such that $d(x, y) > \varepsilon$ for all $x \in A$ and $y \in B$.

(b) Is (a) true if A, B are merely closed?

Sketch. (a) For each $x \in A$ there is a δ_x such that $D(x, \delta_x) \subseteq M \setminus B$; apply compactness to the covering $\{D(x, \delta_x/2) \mid x \in A\}$.

(b) No; let A be the y-axis and B be the graph of $y = 1/x$. ◊

Solution. (a) If either A or B is empty, then the assertion is vacuously true for any $\varepsilon > 0$. Otherwise for $x \in A$, let $\Delta_x = \inf\{d(x, y) \mid y \in B\}$. Then there is a sequence of points $y_1, y_2, y_3, \ldots$ in B with $d(x, y_k) \to \Delta_x$. If $\Delta_x = 0$, this would say that $y_k \to x$. Since B is closed, this would say that $x \in B$. But it is not since $A \cap B = \emptyset$. Thus $\Delta_x > 0$. We have

$d(y, x) \geq \Delta_x$ for every $y \in B$. If we pick any δ_x with $0 < \delta_x < \Delta_x$, then we have $d(y, x) > \delta_x$ for every $y \in B$ and $x \in D(x, \delta_x) \subseteq M \setminus B$. So

$$A \subseteq \bigcup_{x \in A} D\left(x, \frac{\delta_x}{2}\right) \subseteq M \setminus B.$$

This open cover of the compact set A must have a finite subcover, so there are points $x_1, x_2, x_3, \ldots\ a_N$ in A such that

$$A \subseteq D\left(x_1, \frac{\delta_{x_1}}{2}\right) \cup D\left(x_1, \frac{\delta_{x_1}}{2}\right) \cup \cdots \cup D\left(x_N, \frac{\delta_{x_N}}{2}\right) \subseteq M \setminus B.$$

Let

$$\varepsilon = \min\left\{\frac{\delta_1}{2}, \frac{\delta_2}{2}, \ldots, \frac{\delta_N}{2}\right\}$$

and suppose $x \in A$ and $y \in B$. Then there is an index j with $1 \leq j \leq N$ and $x \in D(x_j, \delta_{x_j}/2) \subseteq M \setminus B$. So $d(x_j, y) \leq d(x_j, x) + d(x, y)$. This gives

$$d(x, y) \geq d(x_j, y) - d(x_j, x) > \delta_j - d(x_j, x) > \delta_j - \frac{\delta_j}{2} = \frac{\delta_j}{2} > \varepsilon.$$

So $d(x, y) > \varepsilon$ for all x in A and y in B as required.

(b) This is not true if we only assume that both A and B are closed. Let $A = \{(0, y) \in \mathbb{R}^2 \mid y \in \mathbb{R}\}$ and $B = \{(x, 1/x) \in \mathbb{R}^2 \mid x > 0\}$. Then A is a straight line (the y-axis), and B is one branch of a hyperbola (the graph of $y = 1/x$ for $x > 0$). Each of these sets is closed and their intersection is empty. But $d((0, 1/x), (x, 1/x)) = x \to 0$ as $x \to 0$. So no such ε can exist for these sets. ♦

◇ **3E-38.** Show that $A \subset M$ is not connected iff there exist two *disjoint* open sets U, V such that $U \cap A \neq \varnothing$, $V \cap A \neq \varnothing$, and $A \subset U \cup V$.

Suggestion. Suppose G and H disconnect A. Let $U = G \cap (M \setminus \mathrm{cl}(H))$ and $V = H \cap (M \setminus \mathrm{cl}(G))$. ◇

Solution. If there are such open sets, then they satisfy the conditions of Definition 3.5.1 and disconnect A. For the converse, suppose A is disconnected and that G and H are open sets which disconnect A. That is

1. $G \cap H \cap A = \emptyset$.
2. $A \cap G$ is not empty.
3. $A \cap H$ is not empty.
4. $A \subseteq G \cup H$.

Let $U = G \cap (M \setminus \mathrm{cl}(H))$ and $V = H \cap (M \setminus \mathrm{cl}(G))$. Then U and V are open since each is the intersection of a pair of open sets. We certainly have $A \cap U = A \cap G \cap (M \setminus \mathrm{cl}(H)) \subseteq A \cap G$. On the other hand, if $x \in A \cap G$, then x is not in H since $A \cap G \cap H = \emptyset$. So x is certainly not in $\mathrm{cl}(H)$, and $x \in A \cap G \cap (M \setminus \mathrm{cl}(H)) = A \cap U$. Thus $A \cap U = A \cap G$. Similarly, $A \cap V = A \cap H$. (Do it.). So neither $A \cap U$ nor $A \cap V$ is empty, and

$$(A \cap U) \cup (A \cap V) = (A \cap G) \cup (A \cap H) = A.$$

So $A \subseteq U \cup V$. Finally,

$$U \cap V = (A \cap G) \cap (A cap H) = A \cap G \cap H = \emptyset$$

as required. ♦

◇ **3E-39.** Let $F_1 = [0, 1/3] \cup [2/3, 1]$ be obtained from $[0, 1]$ by removing the middle third. Repeat, obtaining

$$F_2 = [0, 1/9] \ \cup \ [2/9, 1/3] \ \cup \ [2/3, 7/9] \ \cup \ [8/9, 1].$$

In general, F_n is a union of intervals and F_{n+1} is obtained by removing the middle third of these intervals. Let $C = \cap_{n=1}^{\infty} F_n$, the ***Cantor set.*** Prove:

(a) C is compact.
(b) C has infinitely many points. [Hint: Look at the endpoints of F_n.]
(c) $\mathrm{int}(C) = \varnothing$.
(d) C is ***perfect;*** that is, it is closed with no isolated points.
(e) Show that C is ***totally disconnected;*** that is, if $x, y \in C$ and $x \neq y$ then $x \in U$ and $y \in V$ where U and V are open sets that disconnect C.

Solution. (a) Each F_n is a union of finitely many closed intervals and so is closed. The set $C = \bigcap_{n=1}^{\infty} F_n$ is thus an intersection of closed sets and so is closed. It is contained in the unit interval and so is bounded. It is a closed bounded subset of $\mathbb{R}$ and so is compact.

(b) At the nth stage, none of the endpoints of the subintervals making up F_n are removed, and two new endpoints are created within each of these subintervals by the removal of the open middle third of it. Since all these endpoints remain in the sets from step to step and each F_n is a subset of the previous one, they are all in the intersection. Since their number increases by more than one at each step and there are infinitely many steps, there are infinitely many of these endpoints in C.

(c) The length of the subintervals making up the set F_n is 3^{-n}, so the intersection can contain no interval longer than this. Since this is true for every positive integer n and $3^{-n} \to 0$ as $n \to \infty$, the intersection can contain no interval of positive length. Any interior point of C would have to have an interval around it of positive length contained in C, and there are no such intervals. Thus there can be no interior points and $\mathrm{int}(C) = \emptyset$.

(d) As noted in part (b), the endpoints of the subintervals making up the set F_n are never removed in succeeding steps. Thus they are all in the intersection. The distance between the endpoints of one of these subintervals is 3^{-n}. If x is in the intersection, then at the nth stage it must be in one of the subintervals of F_n. If we let this subinterval be $[a_n(x), b_n(x)]$, then we have

$$a_n(x) \le x \le b_n(x) = a_n(x) + \frac{1}{3^n}.$$

If $x = a_n(x)$ for some n, let $c_k(x) = b_k(x)$ for all k.
If $x = b_n(x)$ for some n, let $c_k(x) = a_k(x)$ for all k.
Otherwise, let $c_k(x) = b_k(x)$ for all k.
Then $|c_k(x) - x| < 3^{-n}$, so $c_k(x) \to x$ as $k \to \infty$, and the $c_k(x)$ are all different from x and from each other. As noted in part (b), each of the points $c_k(x)$ is in C, so x is an accumulation point of C and is not isolated. C is thus a closed set in which every point is an accumulation point. That is, it is a *perfect set*.

(e) Let x and y be distinct points of C. Then they are both in F_n for every n. Since the subintervals making up F_n have length 3^{-n}, the points x and y must lie in different subintervals if n is large enough so that $3^{-n} < |x - y|$. In the language of part (d), if $x < y$ we must have

$$a_n(x) \le x \le b_n(x) < a_n(y) \le y \le b_n(y).$$

At least one of the subintervals removed from F_{n-1} to create F_n lies between $b_n(x)$ and $a_n(y)$. Pick a point z in such a removed subinterval, then z is not in F_n, so z is not in C. Furthermore $x < y < z$. Let $U = \{s \in \mathbb{R} \mid s < z\}$ and $V = \{t \in \mathbb{R} \mid z < t\}$. Then U and V are disjoint open sets, $x \in U$, $y \in V$, and $C \subseteq \mathbb{R} \setminus \{z\} = U \cup V$. Thus U and V disconnect C with x in one component and y in another. So C is *totally disconnected* as claimed.

The set C is called the *Cantor set*. It was used by Georg Cantor as an example of several interesting things in the theory of sets which he was developing. It is a very useful device for creating examples and counterexamples for many things involving functions on the unit interval. For example, by arranging for the function to be constant on the removed subintervals one can create a continuous monotonically increasing function with $f(0) = 0$ and $f(1) = 1$, which is, in a fairly reasonable sense, *constant almost everywhere*. This is because the total length of the segments removed is one.

Another aspect of the Cantor set of interest is that it is an example of a fractal set. If we "zoom in" and look at a piece of it magnified, we find that it looks essentially the same at every scale.

One way of thinking about the Cantor set is to expand numbers between 0 and 1 in a base 3 expansion instead of the customary base 10. In such an expansion, we need only three symbols – 0, 1, and 2 – instead of 10. The Cantor set consists of 0 and 1 together with those numbers between 0 and 1 which have no 1's in their ternary expansion. ♦

◇ **3E-40.** Let F_k be a nest of compact sets (that is, $F_{k+1} \subset F_k$). Furthermore, suppose each F_k is connected. Prove that $\cap_{k=1}^{\infty}\{F_k\}$ is connected. Give an example to show that compactness is an essential condition and we cannot just assume that "F_k is a nest of closed connected sets."

Suggestion. Use the Nested Set Property. ◇

Solution. From the nested set property, there is at least one point x_0 in the intersection $F = \bigcap_{k=1}^{\infty} F_k$. Suppose U and V are open sets with $F \cap U \cap V = \emptyset$ and $F \subseteq U \cup V$.

Claim. *There is an integer k such that $F_k \subset U \cap V$.*

Proof of claim: For each k let $C_k = F_k \cap (M \setminus (U \cup V))$. Then C_k is a closed subset of the compact set F_k, so C_k is compact. (Why?) Furthermore $C_{k+1} \subseteq C_k$ for each k. (Why?) If each of the sets C_k were nonempty, then by the nested set property there would be at least one point y in their intersection. But such a point would be in all of the F_k and so in F. We would have $y \in F \cap U \cap V$ which was supposed to be empty. Thus at least one of the sets C_k must be empty and for that k we have $F_k \subseteq U \cap V$.

Let k be an index with $F_k \subseteq U \cap V$. The point x_0 must be in one of the sets U or V. Say $x_0 \in U$. Then $x_0 \in U \cap F_k$, so $V \cap F_k$ must be empty or the sets U and V would disconnect the connected set F_k. Thus $F \subseteq F_k \subseteq U$. So $F \subseteq U$ and $F \cap V = \emptyset$. The sets U and V cannot disconnect F. Thus there can be no pair of open sets which disconnect F and F must be connected.

For a counterexample in which the sets are closed and connected but not compact, let $F_k = \{(x, y) \in \mathbb{R}^2 \mid |x| \geq 1 \text{ or } |y| \geq k\}$ for $k = 1, 2, 3, ...$ Each of these is closed and connected. But their intersection is $\{(x, y) \in \mathbb{R}^2 \mid |x| \geq 1\}$ which is not connected. ♦

4
Continuous Mappings

4.1 Continuity

◇ **4.1-1.** (a) Let $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$. Prove that f is continuous.
(b) Let $f : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x$. Prove that f is continuous.

Suggestion. For $x_0 \in \mathbb{R}$ and $\varepsilon > 0$, try $\delta = \min(1, \varepsilon/(1 + 2|x_0|))$. ◇

Solution. (a) To find the form of the solution, compute the quantity which is to be made small. For $x_0 \in \mathbb{R}$ we have

$$|f(x) - f(x_0)| = \left|x^2 - x_0^2\right| = |x + x_0|\,|x - x_0|\,.$$

If $|x - x_0| < \delta$, then $|x + x_0| \le |x - x_0| + |2x_0| \le \delta + 2|x_0|$. If $\delta < 1$, this leaves us with $|f(x) - f(x_0)| < (1 + 2|x_0|)\,|x - x_0|$. So, if we take δ to be the smaller of the two numbers 1 and $\varepsilon/((1 + 2|x_0|)$, we have

$$\begin{aligned} |f(x) - f(x_0)| &\le (\delta + 2|x_0|)\,|x - x_0| \le (1 + 2|x_0|)\,|x - x_0| \\ &\le (1 + 2|x_0|)\frac{\varepsilon}{1 + 2|x_0|} = \varepsilon. \end{aligned}$$

Thus f is continuous at x_0. Since x_0 was an arbitrary point in $\mathbb{R}$, we conclude that f is continuous on all of $\mathbb{R}$.

(b) Solution One: If $v = (a, b) \in \mathbb{R}$, then $f(v) = a$. This is called *projection onto the first coordinate*. If (x, y) is another point, then

$$|f(x, y) - f(a, b)| = |x - a| \le \sqrt{(x - a)^2 + (y - b)^2} = \|\,(x, y) - (a, b)\,\|\,.$$

Let $\varepsilon > 0$. If we take $\delta = \varepsilon$, we find that if $\| (x, y) - (a, b) \| < \delta$, then

$$|f(x, y) - f(a, b)| \leq \| (x, y) - (a, b) \| < \delta = \varepsilon.$$

(See Figure 4-1.)

FIGURE 4-1. The projections onto the coordinate axes are continuous.

Thus f is continuous at (a, b). Since (a, b) was an arbitrary point in $\mathbb{R}^2$, we conclude that f is continuous on $\mathbb{R}^2$.

Solution Two: Another way to handle this exercise is to use the characterization of continuity in terms of convergent sequences given in Theorem 4.1.4(ii) together with facts we already know about the convergence of sequences in $\mathbb{R}^2$. Let $(x_k, y_k) \to (a, b))$ in $\mathbb{R}^2$. We know from Chapter 2 that this happens (with respect to the usual Euclidean distance in $\mathbb{R}^2$) if and only if $x_k \to a$ and $y_k \to b$ in $\mathbb{R}$. In particular, $f((x_k, y_k)) = x_k \to a$. Since this happens for every sequence in $\mathbb{R}^2$ converging to (a, b), Theorem 4.1.4(ii) says that f is continuous at (a, b). Since (a, b) was arbitrary in $\mathbb{R}^2$, we conclude that f is continuous on $\mathbb{R}^2$.

The function studied in part (b) is called the projection of $\mathbb{R}^2$ onto the first coordinate. It is sometimes denoted by π_1. The projection onto the second coordinate is defined similarly: $\pi_2((x, y)) = y$. This notation might or might not be a good idea, but it is fairly common. It is one of the very few times when it is permissible to use the symbol "π" for something other than the ratio of the circumference of a circle to its diameter.

For related material, see Exercise 3E-15. ♦

◇ **4.1-2.** Use (b) in Exercise 4.1-1 to show that if $U \subset \mathbb{R}$ is open, then $A = \{(x, y) \in \mathbb{R}^2 \mid x \in U\}$ is open.

Solution. For $(x, y) \in \mathbb{R}^2$, let $\pi_1(x, y) = x$ (the projection onto the first coordinate). From part (b) of Exercise 4.1-1, we know that the function $\pi_1 : \mathbb{R}^2 \to \mathbb{R}$ is continuous. Since $A = \pi_1^{-1}(U)$ and U is open in $\mathbb{R}$, we conclude from Theorem 4.1.4(iii) that A is open in $\mathbb{R}^2$. ♦

◇ **4.1-3.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be continuous. Show $A = \{(x, y) \in \mathbb{R}^2 \mid 0 \le f(x, y) \le 1\}$ is closed.

Solution. $A = f^{-1}([0, 1])$. By Theorem 4.1.4(iv), this is closed in $\mathbb{R}^2$ since f is continuous and $[0, 1]$ is closed in $\mathbb{R}$. ♦

◇ **4.1-4.** Prove directly that condition (iii) implies condition (iv) in Theorem 4.1.4.

Solution. The conditions in question are:

4.1.4 Theorem *Let $f : A \subseteq M \to N$ be a mapping from a subset A of a metric space M into a metric space N. Then the following are equivalent:*

iii. *For each open set U in N, the inverse image, $f^{-1}(U) \subseteq A$ is open relative to A. That is, $f^{-1}(U) = V \cap A$ for some open set V.*

iv. *For each open set F in N, the inverse image, $f^{-1}(F) \subseteq A$ is closed relative to A. That is, $f^{-1}(F) = G \cap A$ for some open set V.*

We are asked to prove $(iii) \implies (iv)$.

Suppose F is a closed set in N. Let $U = N \setminus F$. Then U is open in N, and $f^{-1}(U) = f^{-1}(N \setminus F) = A \cap (M \setminus f^{-1}(F)) = A \setminus f^{-1}(F)$. Since for $x \in A$ we have

$$\begin{aligned} x \in f^{-1}(U) &\iff x \in f^{-1}(N \setminus F) \\ &\iff f(x) \in N \setminus F \\ &\iff f(x) \notin F \\ &\iff x \notin f^{-1}(F) \\ &\iff x \in A \setminus f^{-1}(F) \end{aligned}$$

With the assumption of (iii) we have $f^{-1}(U)$ open relative to A. So there is an open set V such that $f^{-1}(U) = V \cap A$. So

$$\begin{aligned} f^{-1}(F) &= A \setminus (A \setminus f^{-1}(F)) = A \setminus f^{-1}(U) \\ &= A \setminus (A \cap V) = A \cap (M \setminus V). \end{aligned}$$

Since V is open, $G = M \setminus V$ is closed, and $f^{-1}(F) = A \cap G$. So $f^{-1}(F)$ is closed relative to A as claimed. ♦

◇ **4.1-5.** Give an example of a continuous function $f : \mathbb{R} \to \mathbb{R}$ and an open set $U \subset \mathbb{R}$ such that $f(U)$ is *not* open.

Sketch. $f(x) = 1$, $U =$ any open set;

$$f(x) = \begin{cases} 0, & \text{if } x \le 0 \\ x, & \text{if } 0 < x < 1\,, \quad U =]-1, 2[\,;\ f(U) = [0,1] \text{ is closed.} \\ 1, & \text{if } x \ge 1. \end{cases}$$ ♢

Solution. For a very easy example, take the constant function $f(x) = 1$ for all $x \in \mathbb{R}$. If $a \in \mathbb{R}$ and $x_k \to a$, then $f(x_k) = f(a) = 0$ for every k. So we certainly have $f(x_k) \to f(a)$. Thus f is continuous at a by 4.1.4(ii), and, since a was an arbitrary point in $\mathbb{R}$, f is continuous on $\mathbb{R}$. If U is any nonempty open subset of $\mathbb{R}$ then $f(U) = \{1\}$. This one point set is not an open subset of $\mathbb{R}$.

For a slightly more imaginative example, we could take

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

See Figure 4-2.

Figure 4-2. A continuous function.

Then if $U =]-1, 2[\,$, we have $f(U) = [0, 1]$, which is closed. (Show f is continuous.) ♦

4.2 Images of Compact and Connected Sets

♢ **4.2-1.** Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Which of the following sets are necessarily closed, open, compact, or connected?

(a) $\{x \in \mathbb{R} \mid f(x) = 0\}$

(b) $\{x \in \mathbb{R} \mid f(x) > 1\}$
(c) $\{f(x) \in \mathbb{R} \mid x \geq 0\}$
(d) $\{f(x) \in \mathbb{R} \mid 0 \leq x \leq 1\}$

Answer. (a) Closed, not necessarily compact or connected.
(b) Open, not necessarily compact or connected.
(c) Connected, not necessarily compact, open, or closed.
(d) Compact, closed, and connected; not necessarily open. ◊

Solution. (a) If $A = \{x \in \mathbb{R} \mid f(x) = 0\}$, then $A = f^{-1}(\{0\})$. Since the one point set $\{0\}$ is closed in $\mathbb{R}$, and f is continuous, A must be closed by Theorem 4.1.4(iv).

If we use the function $f(x) = 0$ for all x, then f is continuous on all of $\mathbb{R}$. One way to see this is to let $x_0 \in \mathbb{R}$ and $\varepsilon > 0$. If $x \in \mathbb{R}$, we have $|f(x) - f(x_0)| = |0 - 0| = 0 < \varepsilon$. So we can let δ be any positive number to get $|x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon$. Thus f is continuous at x_0, and, since x_0 was arbitrary in $\mathbb{R}$, f is continuous on $\mathbb{R}$. For this function we have $f^{-1}(\{0\}) = \mathbb{R}$ which is not bounded and not compact. So A need not be compact.

Now consider $f(x) = x^2 - 1$. For this function we have $A = f^{-1}(\{0\}) = \{\pm 1\}$, a two point set which is not connected. To see that f is continuous, note that $|f(x) - f(x_0)| = \left|x^2 - x_0^2\right|$. Given $\varepsilon > 0$, select $\delta > 0$ as in Exercise 4.1-1(a). Thus A need not be connected.

(b) If $B = \{x \in \mathbb{R} \mid f(x) > 1\}$, then $B = f^{-1}(U)$ where $U = \{y \in \mathbb{R} \mid y > 1\}$. Since U is open in $\mathbb{R}$ and f is continuous, B must be open by Theorem 4.1.4(iii).

To see that B need not be compact, consider the constant function $f(x) = 2$ for all $x \in \mathbb{R}$. Modify the argument in part (a) slightly to show that f is continuous on $\mathbb{R}$. Since $2 > 1$, we have $B = f^{-1}(U) = \mathbb{R}$ which is not compact.

To see that B need not be connected, let $f(x) = x^2 - 1$. Then f is continuous as in part (a). $f(x) > 1 \iff |x| > \sqrt{2}$, so $B = \{x \in \mathbb{R} \mid x < -\sqrt{2}\} \cup \{x \in \mathbb{R} \mid x > \sqrt{2}\}$. Since B is the union of two disjoint, nonempty open sets, it is not connected.

(c) If $C = \{f(x) \in \mathbb{R} \mid x \geq 0\}$, then $C = f(J)$ where J is the closed half line $\{x \in \mathbb{R} \mid x \geq 0\}$. Since J is path-connected it is connected. Since f is continuous on J, $C = f(J)$ must be connected by Theorem 4.2.1.

To see that $f(J)$ need not be open, consider the function $f(x) = 0$ for all x used in part (a). We know that f is continuous, and $C = f(J) = \{0\}$. This one point set is not open in $\mathbb{R}$.

To see that C need not be closed, consider the function $f(x) = 1/(x^2+1)$ for all $x \in \mathbb{R}$. The arithmetic of limits allows us to conclude from 4.1.4(ii) that f is continuous. If $x_k \to a$, then $x_k^2 \to a^2$, and $x_k^2+1 \to a^2+1 \geq 1$. So $f(x_k) = 1/(x_k^2+1) \to 1/(a^2+1) = f(a)$. This shows that f is continuous at a, and, since a was an arbitrary point in $\mathbb{R}$, that f is continuous on $\mathbb{R}$. If $x \in \mathbb{R}$, then $1 \leq x^2+1 < \infty$, so $0 < f(x) \leq 1$. On the other hand, if $0 < y \leq 1$, then we can put $x = \sqrt{(1-y)/y}$ and compute that $f(x) = y$. So $f(J)$ is the half-open interval $]0,1]$. This is not closed since $0 \in \mathrm{cl}(]0,1]) \setminus]0,1]$. This supplies the required counterexample.

Since the half-open interval $]0,1]$ is not closed, it is not compact, and the same example as in the last paragraph shows that C need not be compact.

(d) If $D = \{f(x) \in \mathbb{R} \mid 0 \leq x \leq 1\}$, then $D = f(K)$ where K is the closed unit interval $[0,1]$. Since K is compact, the image $D = f(K)$ must be compact by Theorem 4.2.2. Since it is a compact subset of $\mathbb{R}$, the set D must be closed. Since the interval K is connected, the image $D = f(K)$ must be connected by Theorem 4.2.1.

To see that D need not be open, use the function $f(x) = x$ for all x. Then f is continuous by Example 4.1.5, and $D = f(K) = K = [0,1]$. The image D is not open since $0 \in D$ but no small interval around 0 is contained in D. ♦

◇ **4.2-2.** Verify Theorems 4.2.1 and 4.2.2 for $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2+y$, $K = B = \{(x,y) \mid x^2+y^2 \leq 1\}$.

Suggestion. Show $f(K) = [-1,2]$. ◇

Solution. If $K = B = \{(x,y) \mid x^2+y^2 \leq 1\}$, then the closed disk K is both compact and connected. If we believe that the function $f(x,y) = x^2+y$ is continuous, then Theorem 4.2.1 says that the image $f(K)$ should be connected, and Theorem 4.2.2 says that it should be compact. To check this out, we show that that image is a closed bounded interval in $\mathbb{R}$. To be precise, $f(K) = [-1,2]$.

For (x,y) in K we have $-1 \leq x \leq 1$, and $-1 \leq y \leq 1$. So $0 \leq x^2 \leq 1$. Thus $-1 \leq f(x,y) = x^2+y \leq 2$. On the other hand, we have:

If $-1 \leq t \leq 0$, then $t = (\sqrt{t+1}\,)^2 - 1 = f(\sqrt{t+1}, -1)$. So $t \in f(K)$.

If $0 \leq t \leq 1$, then $t = (\sqrt{t}\,)^2 + 0 = f(\sqrt{t}, 0)$. So $t \in f(K)$.

If $1 \leq t \leq 2$, then $t = (\sqrt{t-1}\,)^2 + 1 = f(\sqrt{t-1}, +1)$. So $t \in f(K)$.

Thus $f(K) = [-1,2]$. This closed bounded interval is both compact and connected just as Theorems 4.2.1 and 4.2.2 say it should be. ♦

◇ **4.2-3.** Give an example of a continuous map $f : \mathbb{R} \to \mathbb{R}$ and a closed subset $B \subset \mathbb{R}$ such that $f(B)$ is not closed. Is this possible if B is bounded as well?

Answer. It is not possible if B is closed and bounded since then it is compact. ♢

Solution. An example is given in part (c) of Exercise 4.2-1. $B = \{x \in \mathbb{R} \mid x \geq 0\}$, and $f(x) = 1/(x^2 + 1)$. Then $f(B) =]0, 1]$ which is not closed.

If B is a subset of $\mathbb{R}$, which is both closed and bounded, then it is compact. Theorem 4.2.2 says that $f(B)$ would also be compact. It would thus be a closed subset of $\mathbb{R}$. ♦

⋄ **4.2-4.** Let $A, B \subset \mathbb{R}$, and suppose $A \times B \subset \mathbb{R}^2$ is connected.

(a) Prove that A is connected.
(b) Generalize to metric spaces.

Suggestion. Use Exercise 4.1-1(b). ♢

Solution. (a) Let $\pi_1 : \mathbb{R}^2 \to \mathbb{R}$ be the projection onto the first coordinate. That is, $\pi_1((x, y)) = x$. Then π_1 is continuous. (See Exercise 4.1-1(b).) If B is not empty, then $\pi_1(A \times B) = A$. So A is connected by Theorem 4.2.1. If $B = \emptyset$, then $A \times B = \emptyset$, and $\pi_1(A \times B) = \emptyset$. We are still all right provided we realize that the empty set is connected. (Could there be sets U and V which disconnect it?)

(b) To generalize to metric spaces, let M_1 and M_2 be metric spaces with metrics d_1 and d_2. One way to put a metric on the cross product $M = M_1 \times M_2$ is to imitate the "taxicab" metric on $\mathbb{R}^2$

$$d((x, y), (a, b)) = d_1(x, a) + d_2(y, b)$$

so that a set $S \subseteq M$ would be open if and only if for each $(a, b) \in S$ there is an $r > 0$ such that $(x, y) \in S$ whenever $d_1(x, a) + d_2(y, b) < r$. We could also use a formula analogous to the Euclidean metric on $\mathbb{R}^2$:

$$\rho((x, y), (a, b)) = \sqrt{d_1(x, a)^2 + d_2(y, b)^2}.$$

Then

$$\begin{aligned} d((x, y), (a, b))^2 &= (d_1(x, a) + d_2(y, b))^2 \\ &= d_1(x, a)^2 + 2d_1(x, a)d_2(x, a) + d_2(x, a)^2 \\ &\geq d_1(x, a)^2 + d_2(x, a)^2 \\ &= (\rho((x, y), (a, b)))^2 . \end{aligned}$$

On the other hand,

$$\begin{aligned} d_1(x, a) + d_2(y, b) &\leq \sqrt{d_1(x, a)^2 + d_2(y, b)^2} + \sqrt{d_1(x, a)^2 + d_2(y, b)^2} \\ &\leq 2\rho((x, y), (a, b)). \end{aligned}$$

So
$$\rho((x,y),(a,b)) \le d((x,y),(a,b)) \le 2\rho((x,y),(a,b)).$$

These inequalities show that d and ρ produce the same open sets, the same closed sets, and the same convergent sequences with the same limits in M. Also, since

$$d_1(x,a) \le d((x,y),(a,b)) \text{ and } d_2(y,b) \le d((x,y),(a,b)),$$
$$\text{and } d((x,y),(a,b)) \le \max(d_1(x,a), d_2(y,b)),$$

we see that

$$\begin{aligned}
(x_k, y_k) \to (a,b) \text{ in } M &\iff \rho((x_k,y_k),(a,b)) \to 0\\
&\iff d((x_k,y_k),(a,b)) \to 0\\
&\iff d_1(x_k,a) \to 0 \text{ and } d_2(y_k,b) \to 0\\
&\iff x_k \to a \text{ in } M_1 \text{ and } y_k \to b \text{ in } M_2.
\end{aligned}$$

This last observation shows that if we define the projections $\pi_1 : M \to M_1$ and $\pi_2 : M \to M_2$ by

$$\pi_1((x,y)) = x \in M_1 \text{ and } \pi_2((x,y)) = y \in M_2,$$

then π_1 and π_2 are continuous by Theorem 4.1.4(ii).
If $A \subseteq M_1$ and $B \subseteq M_2$, then $A = \pi_1(A \times B)$ and $B = \pi_2(A \times B)$. So Theorem 4.2.1 says that if $A \times B$ is connected, then so are A and B. If $A \times B$ is compact, then so are A and B.

For implications in the opposite direction, see Exercise 3E-15. ♦

◇ **4.2-5.** Let A and B be subsets of $\mathbb{R}$ with B not empty. If $A \times B \subseteq \mathbb{R}^2$ is open, must A be open?

Answer. Yes. ◊

Solution. Since B is not empty, there is a point $b \in B$. If $a \in A$, then $(a,b) \in A \times B$. Since $A \times B$ is open, there is an $r > 0$ such that $\sqrt{(x-a)^2 + (y-b)^2} < r$ implies $(x,y) \in A \times B$. If $|x-a| < r$, then $\sqrt{(x-a)^2 + (b-b)^2} = |x-a| < r$, so $(x,b) \in A \times B$. Thus $x \in A$. For each $a \in A$ there is an $r > 0$ such that $x \in A$ whenever $|x-a| < r$. So A is open. ♦

4.3 Operations on Continuous Mappings

◇ **4.3-1.** Where are the following functions continuous?

(a) $f(x) = x\sin(x^2)$.
(b) $f(x) = (x + x^2)/(x^2 - 1)$, $x^2 \neq 1$, $f(\pm 1) = 0$.
(c) $f(x) = (\sin x)/x$, $x \neq 0$, $f(0) = 1$.

Answer. (a) Everywhere.
(b) f is continuous on $\mathbb{R} \setminus \{-1, 1\}$.
(c) Everywhere. ♢

Solution. (a) If $f(x) = x\sin(x^2)$, then f is the product of the continuous function $x \mapsto x$ (Example 4.1.5) with the composition $x \mapsto \sin(x^2)$. The function $x \mapsto x^2$ is continuous. This was seen directly in Exercise 4.1-1(a). We now have an easier indirect proof since it is the product of $x \mapsto x$ with itself. The final fact we need is that $\vartheta \mapsto \sin\vartheta$ is continuous. How this is proved depends on just how the sine function is defined. We will just assume it here. Since products and compositions of continuous functions are continuous, this function is continuous everywhere.

(b) The numerator and denominator of $f(x) = (x + x^2)/(x^2 - 1)$ are continuous everywhere since products and sums of continuous functions are continuous. Since quotients are continuous except where the denominator is 0, the only possible discontinuities are at 1 and -1. The limit at 1 does not exist, and the limit at -1 is $1/2$ since

$$\lim_{x\to-1} \frac{x + x^2}{x^2 - 1} = \lim_{x\to-1} \frac{x(x+1)}{(x+1)(x-1)} = \lim_{x\to-1} \frac{x}{x-1} = \frac{-1}{-1-1} = \frac{1}{2}.$$

Since $1/2 \neq 0 = f(-1)$, f is not continuous at -1. Thus f is continuous on $\mathbb{R} \setminus \{\pm 1\}$.

(c) Again we assume the the sine function is continuous everywhere. So the numerator and denominator of $f(x) = (\sin x)/x$ are continuous everywhere. The only possible discontinuity is at $x = 0$ where the denominator is 0. But the numerator is also 0 there. We know from L'Hôpital's Rule that

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{\cos x}{1} = \frac{\cos 0}{1} = 1 = f(0).$$

So f is continuous at 0. It is thus continuous everywhere. ♦

♢ **4.3-2.** Let $a_k \to a$ and $b_k \to b$ be convergent sequences in $\mathbb{R}$. Prove that $a_k b_k \to ab$, directly and as a consequence of Corollary 4.3.3(ii).

Solution. First compute that

$$\begin{aligned} |a_k b_k - ab| &= |a_k b_k - a_k b + a_k b - ab| \le |a_k b_k - a_k b| + |a_k b - ab| \\ &\le |a_k|\,|b_k - b| + |b|\,|a_k - a|\,. \end{aligned}$$

Let $\varepsilon > 0$. Since $a_k \to a$ and $b_k \to b$, there are N_1, N_2, and N_3 such that

$$\begin{aligned} k \ge N_1 &\implies |a_k - a| < 1 \implies |a_k| < |a| + 1 \\ k \ge N_2 &\implies |b_k - b| < \frac{\varepsilon}{2(|a|+1)} \\ k \ge N_3 &\implies |a_k - a| < \frac{\varepsilon}{2(|b|+1)}. \end{aligned}$$

So, if $k \ge \max(N_1, N_2, N_3)$, then

$$\begin{aligned} |a_k b_k - ab| &\le |a_k|\,|b_k - b| + |b|\,|a_k - a| \\ &\le (|a|+1)\frac{\varepsilon}{2(|a|+1)} + |b|\,\frac{\varepsilon}{2(|b|+1)} \\ &\le \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Thus $a_k b_k \to ab$ as claimed.

According to 4.3.3(ii),

Proposition. *Let $f : A \to \mathbb{R}$ and $g : A \to \mathcal{V}$ be continuous at x_0. Then the product $f \cdot g : A \to \mathcal{V}$ is continuous at x_0.*

We can apply this to our sequence problem by letting $A = \mathbb{R}^2$, $f(x,y) = \pi_1(x,y) = x$, and $g(x,y) = \pi_2(x,y) = y$ for $(x,y) \in \mathbb{R}^2$. Then f and g are continuous as in Exercise 4.1-1(b). Then 4.3.3(ii) implies that $f \cdot g : \mathbb{R}^2 \to \mathbb{R}$ is continuous. Using the fact that a sequence in $\mathbb{R}^2$ converges to w if and only if the coordinate sequences converge to the corresponding coordinates of w, we have

$$\begin{aligned} a_k \to a \text{ and } b_k \to b &\implies (a_k, b_k) \to (a, b) \\ &\implies a_k b_k = (f \cdot g)((a_k, b_k)) \to (f \cdot g)((a, b)) = ab \end{aligned}$$

as desired. ♦

◇ **4.3-3.** Let $A = \{x \in \mathbb{R} \mid \sin x = 0.56\}$. Show that A is a closed set. Is it compact?

Suggestion. Use the fact that $\{0.56\}$ is closed and $\sin x$ is continuous. A is not compact. ◇

Solution. Again we assume that the function $f(x) = \sin x$ is continuous. Then $A = \{x \in \mathbb{R} \mid \sin x = 0.56\} = f^{-1}(\{0.56\})$. Since the one point set

$\{0.56\}$ is closed in $\mathbb{R}$ and f is continuous, A must also be closed. It is not compact since if x_0 is any one point in A, then $x_0 + 2\pi k$ is also in A for every integer k. Thus A is not bounded and cannot be compact. ♦

◇ **4.3-4.** Show that $f(x) = \sqrt{|x|}$ is continuous.

Solution. The function $f(x) = \sqrt{|x|}$ is the composition of the two functions $h : \mathbb{R} \to \mathbb{R}$ and $g : A = \{y \in \mathbb{R} \mid y \geq 0\} \to \mathbb{R}$ defined by $h(x) = |x|$ for $x \in \mathbb{R}$ and $g(y) = \sqrt{y}$ for $y \geq 0$. Since $h : \mathbb{R} \to A$, continuity of f on $\mathbb{R}$ will follow from Theorem 4.3.1 as soon as we know that g and h are continuous.

Proposition. *The function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(x) = |x|$ for $x \in \mathbb{R}$ is continuous on $\mathbb{R}$.*

Proof: Let $x_0 \in \mathbb{R}$ and $\varepsilon > 0$. If we take $\delta = \varepsilon$, we can use the alternate form of the triangle inequality to compute

$$|x - x_0| < \delta \implies |h(x) - h(x_0)| = ||x| - |x_0|| \leq |x - x_0| < \delta = \varepsilon.$$

This shows that h is continuous at x_0, and, since x_0 was an arbitrary point in $\mathbb{R}$, that h is continuous on $\mathbb{R}$.

Proposition. *The function $g : A \to \mathbb{R}$ defined on the set $A = \{y \in \mathbb{R} \mid y \geq 0\}$ by $g(y) = \sqrt{y}$ is continuous on A.*

Proof: To show that g is continuous at 0, let $\varepsilon > 0$. Set $\delta = \varepsilon^2$. If $0 \leq y \leq \delta$, we have

$$|g(y) - g(0)| = \sqrt{y} < \sqrt{\delta} = \sqrt{\varepsilon^2} = \varepsilon.$$

This shows that g is continuous at 0.

Now suppose $y_0 > 0$, and let $\varepsilon > 0$. Set $\delta = \varepsilon\sqrt{y_0}$. If $|y - y_0| < \delta$ we have

$$|g(y) - g(y_0)| = |\sqrt{y} - \sqrt{y_0}| = \left| \frac{y - y_0}{\sqrt{y} + \sqrt{y_0}} \right| \leq \frac{|y - y_0|}{\sqrt{y_0}} \leq \frac{\varepsilon\sqrt{y_0}}{\sqrt{y_0}} = \varepsilon.$$

This shows that g is continuous at y_0. Since y_0 was an arbitrary positive number and we have already taken care of the case $y = 0$, we see that g is continuous on A as claimed.

Since h is continuous on $\mathbb{R}$, g is continuous on A, $h : \mathbb{R} \to A$ and $f = g \circ h$, Theorem 4.3.1 shows that f is continuous. ♦

◇ **4.3-5.** Show that $f(x) = \sqrt{x^2 + 1}$ is continuous.

Sketch. $f = g \circ h$ where $g(x) = \sqrt{x}$ and $h(x) = x^2 + 1$. f is continuous since g and h are continuous. ◊

Solution. We know that $x \mapsto x$ is continuous from Example 4.1.5 and that constant functions are continuous. So $h(x) = x^2 + 1$ is continuous on $\mathbb{R}$ since products and sums of continuous functions are continuous. The function $g(y) = \sqrt{y}$ is continuous on the set $\{y \in \mathbb{R} \mid y \geq 0\}$. (See Exercise 4.3-4.) Since $x^2 + 1 \geq 1 \geq 0$ for every real x, Theorem 4.3.1 says that the composition $f(x) = \sqrt{x^2 + 1} = g(h(x)) = (g \circ h)(x)$ is continuous on $\mathbb{R}$. ♦

4.4 The Boundedness of Continuous Functions on Compact Sets

◇ **4.4-1.** Give an example of a continuous and bounded function on all of $\mathbb{R}$ that does not attain its maximum or minimum.

Answer. One possibility is $f(x) = x/(1 + |x|)$. ◇

Solution. If we let $f(x) = x/(1 + |x|)$ for all $x \in \mathbb{R}$, then the numerator and denominator are continuous everywhere and the denominator is never 0, so f is continuous everywhere on $\mathbb{R}$. Furthermore, $|f(x)| = |x| / (1+|x|) < 1$. So $-1 < f(x) < 1$ for all $x \in \mathbb{R}$. Finally

$$\lim_{x \to +\infty} f(x) = \lim_{x \to +\infty} \frac{x}{1+x} = \lim_{x \to +\infty} \frac{1}{1 + (1/x)} = \frac{1}{1+0} = 1$$
$$\lim_{x \to -\infty} f(x) = \lim_{x \to -\infty} \frac{x}{1-x} = \lim_{x \to +\infty} \frac{1}{-1 + (1/x)} = \frac{1}{-1+0} = -1$$

So $\sup\{f(x) \mid x \in \mathbb{R}\} = 1$ and $\inf\{f(x) \mid x \in \mathbb{R}\} = -1$ and neither of these is attained anywhere in $\mathbb{R}$. See Figure 4-3. ♦

Figure 4-3. The function $f(x) = x/(1 + |x|)$.

◇ **4.4-2.** Verify the maximum-minimum theorem for $f(x) = x^3 - x$ on $[-1, 1]$.

Solution. The function $f(x) = x^3 - x$ is continuous and differentiable everywhere. $f'(x) = 3x^2 - 1$. The derivative is 0 at $x = \pm 1/\sqrt{3}$. Values at these points and at the ends are:

$$f(-1) = 0, \qquad f(-1/\sqrt{3}) = \frac{2}{3\sqrt{3}}, \qquad f(1/\sqrt{3}) = -\frac{2}{3\sqrt{3}}, \qquad f(1) = 0.$$

The maximum value is $2/3\sqrt{3}$ and is attained at $x = -1/\sqrt{3}$. The minimum value is $-2/3\sqrt{3}$ and is attained at $x = 1/\sqrt{3}$. See Figure 4-4. ♦

FIGURE 4-4. The function $f(x) = x^3 - x$.

◇ **4.4-3.** Let $f : K \subset \mathbb{R}^n \to \mathbb{R}$ be continuous on a compact set K and let $M = \{x \in K \mid f(x) \text{ is the maximum of } f \text{ on } K\}$. Show that M is a compact set.

Suggestion. $M = f^{-1}(\sup(f(x)))$. Why does this exist; why compact? ♢

Solution. The function f is a continuous function from the compact set K into $\mathbb{R}$. We know from the Maximum-Minimum Theorem that $b = \sup\{f(x) \in \mathbb{R} \mid x \in K\}$ exists as a finite real number and that there is at least one point x_0 in K such that $f(x_0) = b$. So $M = \{x \in K \mid f(x) \text{ is the maximum of } f \text{ on } K\}$ is not empty. In particular, $x_0 \in M$. The single point set $\{b\}$ is a closed set in $\mathbb{R}$. So $M = f^{-1}(\{b\})$ is a closed set in K. Since it is a closed subset of the compact set K, it is also a compact set by Lemma 2 to Theorem 3.1.3 (p. 165 of the text). ♦

◇ **4.4-4.** Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be continuous, c be a continuous function, $x, y \in A$, and $c : [0,1] \to A \subset \mathbb{R}^n$ be a curve joining x and y. Show that along this curve, f assumes its maximum and minimum values (among all values along the curve).

Sketch. The composition $f \circ c$ is a continuous function from the compact set $[0,1]$ into $\mathbb{R}$. ◇

Solution. Let $C \subseteq A$ be the image curve of the function $c : [0,1] \to A$. Let $b = \sup\{f(v) \in \mathbb{R} \mid v \in C\}$. If $v \in C$, then $v = c(t)$ for some $t \in [0,1]$. Conversely, if $t \in [0,1]$, then $c(t) \in C$ and $f(c(t)) \in \{f(v) \in \mathbb{R} \mid v \in C\}$. So $b = \sup\{f(c(t)) \in \mathbb{R} \mid t \in [0,1]\}$.

Since c and f are both continuous and c maps the interval $[0,1]$ into the domain A of f, the composition $f \circ c$ is a continuous function from the compact set $[0,1]$ into $\mathbb{R}$. From the maximum-minimum theorem we conclude that b exists as a finite real number and that there is at least one point $t_0 \in [0,1]$ at which $f(c(t_0)) = b$. Thus $v_0 = c(t_0)$ is a point on the curve C at which $f(v_0) = \sup\{f(v) \in \mathbb{R} \mid v \in C\}$. This is exactly the conclusion we wanted. ♦

◇ **4.4-5.** Is a version of the maximum-minimum theorem valid for the function $f(x) = (\sin x)/x$ on $]0,\infty[$? On $[0,\infty[$?

Sketch. $\sup(f(\,]0,\infty[\,)) = 1$ is not attained on $]0,\infty[$. Extend by $f(0) = 1$ (continuous?) to get it on $[0,\infty[$. ◇

Solution. If $f(x) = (\sin x)/x$ for x not 0, then we know from calculus or elsewhere that $\lim_{x\to 0} f(x) = 1$. So if we define $f(0)$ to be 0, we obtain a continuous function on $\mathbb{R}$. $f'(x) = (x\cos x - \sin x)/x^2$. This is 0 only when $x = \tan x$. This occurs at $x = 0$ in the limit and not again until $|x| > \pi$. See Figure 4-5.

We have $f(\pm\pi) = 0$, $f'(x) > 0$ for $-\pi < x < 0$, and $f'(x) < 0$ for $0 < x < \pi$. Consequently, $0 < f(x) < \lim_{x\to 0} f(x) = 1 = f(0)$ for $0 < x < \pi$. For $|x| > \pi$, we have $|f(x)| = |\sin x| \,/\, |x| < 1/\,|x| < 1/\pi$. Thus $\sup\{f(x) \mid x \in \mathbb{R}\} = 1 = f(1)$ and $\inf\{f(x) \mid x \in \mathbb{R}\}$ occurs at the two points with $\pi/2 < |x| < \pi$ at which the derivative is 0. The supremum is attained on $[0,\infty[$, but not on $]0,\infty[$. See Figure 4-6.

What is going on might be summarized by something like the following.

FIGURE 4-5. $x = \tan x$?

FIGURE 4-6. The function $f(x) = (\sin x)/x$.

Proposition. *If f is a continuous real valued function on the closed half line $[a, \infty[$ and*

$$\limsup_{x\to\infty} f(x) < \sup\{f(x) \mid x \in [a, \infty[\,\} < \infty,$$

then there is at least one point $x_1 \in [a, \infty[$ at which $f(x_1) = \sup\{f(x) \mid x \in [a, \infty[\,\}$. If

$$\liminf_{x\to\infty} f(x) > \inf\{f(x) \mid x \in [a, \infty[\,\} > -\infty,$$

then there is at least one point $x_2 \in [a, \infty[$ at which $f(x_2) = \inf\{f(x) \mid x \in [a, \infty[\,\}$.

The hypotheses say that there is a number $B > a$ such that $f(x)$ is smaller than $\sup\{f(x) \mid x \in [a, \infty[\,\}$ and larger than $\inf\{f(x) \mid x \in [a, \infty[\,\}$ for $x > B$. The supremum and infimum over $[a, \infty[$ are thus the same as those over the compact interval $[a, B]$. We can use the maximum-minimum theorem to conclude that they are attained at points in that interval. ♦

4.5 The Intermediate Value Theorem

◇ **4.5-1.** What happens when you apply the method used in Example 4.5.4 to quadratic polynomials? To quintic polynomials?

Sketch. Quadratic polynomials need not be negative anywhere, so the method fails. The method works for quintic polynomials and, in general, for all odd-degree polynomials. ◇

Solution. The method used to show that every cubic polynomial with real coefficients must have at least one real root depends first on the fact that they define continuous real valued functions on all of $\mathbb{R}$. This is true of all polynomials with real coefficients. The other fact needed was that the value of the function defined by such a polynomial must be positive for some values of x and negative for others. We were then able to apply the intermediate value theorem to conclude that the value must be 0 somewhere between these. This fails for quadratic polynomials. For example, the polynomial $p(x) = x^2 + 1$ for real x takes on values only at least as large as 1. It is never negative. Similar examples can be found for any even degree. However, if the degree is odd, such as the degree 5 of a quintic polynomial, then the same basic idea used for degree 3 will work. Suppose p is an odd degree polynomial

$$\begin{aligned} p(x) &= a_{2n+1}x^{2n+1} + a_{2n}x^{2n} + \cdots + a_1 x + a_0 \\ &= a_{2n+1}x^{2n+1}\left(1 + \frac{a_{2n}}{x} + \frac{a_{2n-1}}{x_2} + \cdots + \frac{a_1}{x^{2n}} + \frac{a_0}{x^{2n+1}}\right) \end{aligned}$$

with a_{2n+1} not 0, and all of the coefficients a_k real. If $|x|$ is very large, then the expression in parentheses is close to 1. In particular, it is positive. So, if x is large positive, the sign of $p(x)$ is the same as that of a_{2n+1}, while if it is large negative, then $p(x)$ and a_{2n+1} have opposite sign. Thus there are a and b such that $p(a) < 0 < p(b)$. Since $p(x)$ is continuous for $x \in [a, b]$, the intermediate value theorem implies that there is at least one point c in $[a, b]$ at which $p(c) = 0$. ♦

◇ **4.5-2.** Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be continuous. Let $\Gamma = \{(x, f(x)) \mid x \in \mathbb{R}^n\}$ be the graph of f in $\mathbb{R}^n \times \mathbb{R}^m$. Prove that Γ is closed and connected. Generalize your result to metric spaces.

Solution. $g : \mathbb{R}^n : \mathbb{R}^n \times \mathbb{R}^m$ by $g(x) = (x, f(x))$. If $x_k \to x$ in $\mathbb{R}^n$, then $f(x_k) \to f(x)$ in $\mathbb{R}^m$ since f is continuous. So $g(x_k) = (x_k, f(x_k)) \to (x, f(x)) \in \Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^m$ since each of the $n + m$ coordinates converges correctly. Since $\mathbb{R}^n$ is connected and g is continuous, the image $\Gamma = g(\mathbb{R}^n)$ is also connected. Furthermore, if $(x, y) \in \mathrm{cl}(\Gamma)$, then there is a sequence

$(x_k, y_k) \in \Gamma$ with $(x_k, y_k) \to (x, y)$. This means that $x_k \to x$ and $y_k \to y$. Since $(x_k, y_k) \in \Gamma$, we have $y_k = f(x_k)$, and since f is continuous, we have $y_k = f(x_k) \to f(x)$. Since limits are unique, we must have $y = f(x)$. So $(x, y) \in \Gamma$. This shows that $\mathrm{cl}(\Gamma) \subseteq \Gamma$. So Γ is closed as claimed.

To generalize this, suppose M_1 and M_2 are metric spaces with metrics d_1 and d_2. We can establish a metric on the cross product $M = M_1 \times M_2$ by either of

$$d((x,y),(a,b)) = d_1(x,a) + d_2(y,b) \quad \text{or}$$
$$\rho((x,y),(a,b)) = \sqrt{d_1(x,a)^2 + d_2(y,b)^2}.$$

As we have seen earlier, these metrics produce the same open sets, the same closed sets, and the same convergent sequences converging to the same limits in M. Furthermore $(x_k, y_k) \to (x, y)$ in M if and only if $x_k \to x$ in M_1 and $y_k \to y$ in M_2. If $f : M_1 \to M_2$, its graph is defined by $\Gamma = \{(x, f(x)) \in M \mid x \in M_1\}$. Essentially the same argument as given above shows that if f is continuous, then Γ is closed and that if f is continuous and M_1 is connected, then Γ is connected. ♦

◇ **4.5-3.** Let $f : [0,1] \to [0,1]$ be continuous. Prove that f has a fixed point.

Suggestion. Apply the intermediate value theorem to $g(x) = f(x) - x$. ◇

Solution. For $x \in [0,1]$, let $g(x) = f(x) - x$. Since f is continuous and $x \mapsto -x$ is continuous, g is continuous. We have

$$g(0) = f(0) - 0 = f(0) \geq 0 \quad \text{and} \quad g(1) = f(1) - 1 \leq 1 - 1 = 0$$

By the intermediate value theorem, there must be at least one point c in $[0,1]$ with $g(c) = 0$. For such a point we have $f(c) - c = 0$. So $f(c) = c$. Thus c is a fixed point for the mapping f as required.

See Figure 4-7. ♦

FIGURE 4-7. A fixed point theorem.

◇ **4.5-4.** Let $f : [a,b] \to \mathbb{R}$ be continuous. Show that the range of f is a bounded closed interval.

Sketch. The interval $[a,b]$ is compact and connected and f is continuous. ♢

Solution. Suppose $f : [a,b] \to \mathbb{R}$ is continuous. The interval $[a,b]$ is a connected set and f is continuous. So the image $f([0,1])$ must be a connected subset of $\mathbb{R}$. So it must be an interval, a half-line, or the whole line. But $[a,b]$ is a closed, bounded subset of $\mathbb{R}$, so it is compact. Since f is continuous, the image $f([0,1])$ must be compact in $\mathbb{R}$. It must thus be closed and bounded. Of the possibilities listed above, this leaves a closed, bounded interval. ♦

◇ **4.5-5.** Prove that there is no continuous map taking $[0,1]$ onto $]0,1[$.

Sketch. $f([0,1])$ would be compact, and $]0,1[$ is not compact. ♢

Solution. The interval $[0,1]$ is a closed bounded subset of $\mathbb{R}$ and so it is compact. If f were a continuous function on it, then the image would have to be compact. The open interval $]0,1[$ is not compact. So it cannot be the image of such a map. ♦

4.6 Uniform Continuity

◇ **4.6-1.** Demonstrate the conclusion in Example 4.6.3 directly from the definition.

Sketch. $|(1/x) - (1/y)| = |(x-y)/xy| \le |x-y|/a^2$. Take $\delta = a^2\varepsilon$. ♢

Solution. We are asked to show that the function $f(x) = 1/x$ is uniformly continuous on the interval $[a,1]$ if $0 < a < 1$. To do this we compute

$$\left|\frac{1}{x} - \frac{1}{y}\right| = \frac{|x-y|}{|xy|} \le \frac{|x-y|}{a^2}.$$

If $\varepsilon > 0$, let $\delta = a^2\varepsilon$. If x and y are in $[a,1]$ and $|x-y| < \delta$, then the last computation shows that

$$|f(x) - f(y)| = \left|\frac{1}{x} - \frac{1}{y}\right| =\le \frac{|x-y|}{a^2} < \frac{a^2\varepsilon}{a^2} = \varepsilon.$$

The same δ works everywhere on $[a,1]$, so f is uniformly continuous on that domain. ♦

◇ **4.6-2.** Prove that $f(x) = 1/x$ is uniformly continuous on $[a, \infty[$ for $a > 0$.

Suggestion. Compare to Exercise 4.6-1. ◇

Solution. We are asked to show that the function $f(x) = 1/x$ is uniformly continuous on the half line $[a, \infty[$ if $0 < a$. To do this we compute

$$\left|\frac{1}{x} - \frac{1}{y}\right| = \frac{|x-y|}{|xy|} \le \frac{|x-y|}{a^2}.$$

If $\varepsilon > 0$, let $\delta = a^2\varepsilon$. If x and y are in $[a, \infty[$ and $|x - y| < \delta$, then the last computation shows that

$$|f(x) - f(y)| = \left|\frac{1}{x} - \frac{1}{y}\right| = \le \frac{|x-y|}{a^2} < \frac{a^2\varepsilon}{a^2} = \varepsilon.$$

The same δ works everywhere on $[a, \infty[$, so f is uniformly continuous on that domain. ♦

◇ **4.6-3.** Must a bounded continuous function on $\mathbb{R}$ be uniformly continuous?

Sketch. No. Consider $f(x) = \sin(x^2)$. ◇

Solution. To get a bounded continuous function which is not uniformly continuous, we want a function whose derivative stays large over nontrivial regions. If we put $f(x) = \sin(x^2)$, then $f'(x) = 2x\cos(x^2)$. So f has a large derivative near $x = \sqrt{2\pi n}$ for integer n. See Figure 4-8.

Figure 4-8. $f(x) = \sin(x^2)$.

The function f is continuous. To show that it is not uniformly continuous, we take $\varepsilon = 1/2$, and show that for any $\delta > 0$ there are points x and y

with $|x-y|<\delta$ and $|f(x)-f(y)|>1/2$. To this end let $x=\sqrt{2\pi n}$ and $y=\sqrt{2\pi n+(\pi/2)}$. Then

$$y-x=\sqrt{2\pi n+\frac{\pi}{2}}-\sqrt{2\pi n}=\frac{(2\pi n+\frac{\pi}{2})-2\pi n}{\sqrt{2\pi n+\frac{\pi}{2}}+\sqrt{2\pi n}}<\frac{\pi/2}{2\sqrt{2\pi n}}.$$

Since this tends to 0 as n increases, we can pick n large enough so that $|y-x|<\delta$. But $|f(y)-f(x)|=|\sin(2\pi n)-\sin(2\pi n+(\pi/2))|=|0-1|=1$. So no choice of $\delta>0$ can work everywhere in $\mathbb{R}$. So f is not uniformly continuous on $\mathbb{R}$. ♦

◇ **4.6-4.** If f and g are uniformly continuous maps of $\mathbb{R}$ to $\mathbb{R}$, must the product $f\cdot g$ be uniformly continuous? What if f and g are bounded?

Answer. No; yes. ◊

Solution. The function $f(x)=x$ is uniformly continuous on $\mathbb{R}$ since if $\varepsilon>0$, we can simply take $\delta=\varepsilon$. If $|x-y|<\delta$, then $|f(x)-f(y)|=|x-y|<\delta=\varepsilon$. The same δ works everywhere on $\mathbb{R}$, so f is uniformly continuous on $\mathbb{R}$. If we take $g(x)=x$ also, then f and g are both uniformly continuous on $\mathbb{R}$. The product is $(f\cdot g)(x)=x^2$. This is not uniformly continuous on $\mathbb{R}$.

Claim. *The function $h(x)=x^2$ is not uniformly continuous on $\mathbb{R}$.*

If we take $\varepsilon=1$, then no matter how small we take $\delta>0$, there are always points x and y with $|x-y|<\delta$ and $\left|x^2-y^2\right|>1$. For positive integers n, take $x=n+(1/n)$ and $y=n$. Then $|x-y|=1/n$, and we can certainly arrange for this to be smaller than δ. However,

$$x^2-y^2=\left(n+\frac{1}{n}\right)^2-n^2=n^2+2+\frac{1}{n^2}-n^2=2+\frac{1}{n^2}>2.$$

Thus $h(x)=x^2$ is not uniformly continuous on $\mathbb{R}$.

If both functions are bounded and continuous, then the product is uniformly continuous.

Proposition. *If f and g are uniformly continuous, bounded real valued functions on a metric space M, then the product $h(x)=f(x)g(x)$ is uniformly continuous on M.*

Proof: Suppose f and g are continuous on M, that $|f(x)|\le A$ and $|g(x)|\le B$ for all x in M. We compute

$$\begin{aligned}|f(x)g(x)-f(y)g(y)|&=|f(x)g(x)-f(x)g(y)+f(x)g(y)-f(y)g(y)|\\&\le|f(x)g(x)-f(x)g(y)|+|f(x)g(y)-f(y)g(y)|\\&\le|f(x)|\,|g(x)-g(y)|+|g(y)|\,|f(x)-f(y)|\\&\le A\,|g(x)-g(y)|+B\,|f(x)-f(y)|\,.\end{aligned}$$

Let $\varepsilon > 0$. Since f is uniformly continuous on M, we can find a $\delta_1 > 0$ small enough so that $|f(x) - f(y)| < \varepsilon/2(B+1)$. Since g is uniformly continuous on M, we can find a $\delta_2 > 0$ small enough so that $|g(x) - g(y)| < \varepsilon/2(A+1)$. If $|x - y| < \delta = \min(\delta_1, \delta_2)$, then the computation above shows that

$$\begin{aligned}|f(x)g(x) - f(y)g(y)| &\leq A\,|g(x) - g(y)| + B\,|f(x) - f(y)| \\ &\leq A\frac{\varepsilon}{2(A+1)} + B\frac{\varepsilon}{2(B+1)} < \varepsilon.\end{aligned}$$

The same δ works everywhere in M. So $h = f \cdot g$ is uniformly continuous on M.

It is not enough to assume that only one of f and g is bounded.

Challenge: Find a bounded function f and an unbounded function g, both uniformly continuous on $\mathbb{R}$ such that the product $f \cdot g$ is not uniformly continuous on $\mathbb{R}$. ♦

◇ **4.6-5.** Let $f(x) = |x|$. Show that $f : \mathbb{R} \to \mathbb{R}$ is uniformly continuous.

Sketch. $\delta = \varepsilon$ works everywhere. ◇

Solution. The alternative form of the triangle inequality gives us our result almost immediately. If $\varepsilon > 0$, let $\delta = \varepsilon$. If x and y are in $\mathbb{R}$ and $|x - y| < \delta$, then

$$|f(x) - f(y)| = \big|\, |x| - |y| \,\big| \leq |x - y| < \delta = \varepsilon.$$

The same δ works everywhere, so f is uniformly continuous on $\mathbb{R}$. ♦

◇ **4.6-6.** (a) Show that $f : \mathbb{R} \to \mathbb{R}$ is *not* uniformly continuous iff there exist an $\varepsilon > 0$ and sequences x_n and y_n such that $|x_n - y_n| < 1/n$ and $|f(x_n) - f(y_n)| \geq \varepsilon$. Generalize this statement to metric spaces.

(b) Use (a) on $\mathbb{R}$ to prove that $f(x) = x^2$ is not uniformly continuous.

Solution. (a) First suppose that $f : \mathbb{R} \to \mathbb{R}$ is not uniformly continuous. Then there is an $\varepsilon > 0$ for which no $\delta > 0$ will work in the definition of uniform continuity. In particular, $\delta = 1/n$ will not work. So there must be a pair of numbers x_n and y_n such that $|x_n - y_n| < 1/n$ but $|f(x_n) - f(y_n)| > \varepsilon$. These x_n and y_n form the required sequences.

Conversely, if there are such sequences, then for that ε no $\delta > 0$ can work in the definition of uniform continuity. No matter what $\delta > 0$ is proposed, we can, by the Archimedean Principle, select an integer n with $0 < 1/n < \delta$. For the corresponding x_n and y_n we have $|f(x_n) - f(y_n)| > \varepsilon$ even though $|x_n - y_n| < 1/n < \delta$. Since this can be done for every $\delta > 0$, the function f cannot be uniformly continuous on $\mathbb{R}$.

This can be reformulated with very little change for a map between metric spaces.

Proposition. *Suppose M_1 and M_2 are metric spaces with metrics d_1 and d_2 respectively. If $A \subseteq M_1$ and $f : A \to M_2$, then f fails to be uniformly continuous on A if and only if there are a number varepsilon > 0 and sequences $\langle x_n \rangle_1^\infty$ and $\langle y_n \rangle_1^\infty$ in A such that $d_1(x_n, y_n) < 1/n$ and $d_2(f(x_n), f(y_n)) > \varepsilon$ for each $n = 1, 2, 3, \ldots$.*

The proof is essentially the same as that given above. First suppose that $f : A \to M_2$ is not uniformly continuous. Then there is an $\varepsilon > 0$ for which no $\delta > 0$ will work in the definition of uniform continuity. In particular, $\delta = 1/n$ will not work. So there must be a pair of points x_n and y_n in A such that $d_1(x_n, y_n) < 1/n$ but $d_2(f(x_n), f(y_n)) > \varepsilon$. These x_n and y_n form the required sequences.

Conversely, if there are such sequences, then for that ε no $\delta > 0$ can work in the definition of uniform continuity. No matter what $\delta > 0$ is proposed, we can, by the Archimedean Principle, select an integer n with $0 < 1/n < \delta$. For the corresponding x_n and y_n we have $d_2(f(x_n), f(y_n)) > \varepsilon$ even though $d_1(x_n, y_n) < 1/n < \delta$. Since this can be done for every $\delta > 0$, the function f cannot be uniformly continuous on A.

(b) To apply the result of part (a) to the function $f(x) = x^2$ on $\mathbb{R}$, let $\varepsilon = 1/2$, $x_n = n + (1/2n)$ and $y_n = n$. Then $|x_n - y_n| = 1/2n < 1/n$ and $|f(x_n) - f(y_n)| = n^2 + 1 + 1/(4n^2) - n^2 = 1 + 1/(4n^2) > 1 > \varepsilon$. According to the result of part (a), the function f cannot be uniformly continuous on $\mathbb{R}$. ♦

◇ **4.6-7.** Let $f(x) = \sqrt{x}$.

(a) Show that f is uniformly continuous on the interval $[0, 1]$.

(b) Discuss the relationship between uniform continuity and bounded slopes in light of this example.

Sketch. (a) f is continuous on a compact domain. ◇

Solution. (a) We know that the function $f(x) = \sqrt{x}$ is continuous on $\{x \in \mathbb{R} \mid x \geq 0\}$. (See Exercise 4.3-4.) So it is certainly continuous on the smaller domain $[0, 1]$. Since the closed bounded interval $[0, 1]$ is a compact subset of $\mathbb{R}$ and f is continuous on this domain, it is uniformly continuous there by Theorem 4.6.2.

(b) The function $f(x) = \sqrt{x}$ studied in part (a) is differentiable except at $x = 0$ with derivative $f'(x) = 1/(2\sqrt{x})$. This becomes arbitrary large as x approaches 0. Thus the slope is not bounded on the domain $[0, 1]$. The function is nonetheless uniformly continuous on that domain. Thus unbounded slope does not necessarily imply lack of uniform continuity

even though it is a good clue. To destroy uniform continuity the slope must remain large over a significant interval as it does in the example of $\sin(x^2)$ considered in the solution to Exercise 4.6-3.
Bounded slope should imply uniform continuity although, strictly speaking, this makes sense only for differentiable functions. For continuous, but not differentiable functions, bounded slope might mean something like a **Lipschitz condition**.

Definition. *A function f is said to satisfy a Lipschitz condition on the domain A if there is a constant M such that $|f(x) - f(y)| \leq M\,|x - y|$ whenever x and y are in A.*

Proposition. *If f satisfies a Lipschitz condition on the domain A, then f is uniformly continuous on A.*

Suppose $f : A \to \mathbb{R}$ satisfies a Lipschitz condition on A with constant M. If $M = 0$, then f is a constant function and is surely uniformly continuous. If $M > 0$ and $\varepsilon > 0$, put $\delta = \varepsilon/M$. If x and y are in A and $|x - y| < \delta$, then

$$|f(x) - f(y)| \leq M\,|x - y| < M\delta = M\frac{\varepsilon}{M} = \varepsilon.$$

The same δ works everywhere in A. So f is uniformly continuous on A.
Challenge: Rewrite the definition of Lipschitz condition for functions between a pair of metric spaces and prove the corresponding proposition.

♦

4.7 Differentiation of Functions of One Variable

◇ **4.7-1.** Give an example of a function defined on $\mathbb{R}$, which is continuous everywhere and which fails to be differentiable at exactly n given points $x_1, \ldots, x_n$.

Answer. One possibility is $f(x) = \sum_{i=1}^{n} a_i\,|x - x_i|$, $a_i \neq 0$. ◇

Solution. The function $x \mapsto |x|$ is continuous (in fact uniformly so) on all of $\mathbb{R}$ as are the functions $x \mapsto x - x_i$ for each of the constants x_i. Thus the compositions $x \mapsto |x - x_i|$ are also. So the sum, $f(x) = \sum_{i=1}^{n} |x - x_i|$ is continuous on $\mathbb{R}$.

If x is slightly greater than x_k, then $\dfrac{f(x) - f(x_k)}{x - x_k} =$

$$\frac{\Big((x - x_k) + \sum_{i \neq k} |x - x_i|\Big) - \Big((x_k - x_k) + \sum_{i \neq k} |x - x_i|\Big)}{x - x_k} = 1.$$

While if x is slightly less than x_k, then $\dfrac{f(x) - f(x_k)}{x - x_k} =$

$$\frac{\Big((x_k - x) + \sum_{i \neq k} |x - x_i|\Big) - \Big((x_k - x_k) + \sum_{i \neq k} |x - x_i|\Big)}{x - x_k} = -1.$$

Thus the left and right limits at x_k are different and f is not differentiable at the points x_k, $k = 1, 2, 3, \ldots$. On the other hand, if a is not equal to any of the x_k, then $(f(x) - f(a))/(x - a)$ has the same sign on both sides of a, and the derivative exists at a. (Either 1 or -1.) ♦

◇ **4.7-2.** Does the mean value theorem apply to $f(x) = \sqrt{x}$ on [0, 1]? Does it apply to $g(x) = \sqrt{|x|}$ on $[-1, 1]$?

Solution. The function $f(x) = \sqrt{x}$ is continuous on the closed interval $[0, 1]$. (See Exercise 4.3-4.) and differentiable on the interior $]0, 1[$. (To see this, you can use the Inverse Function Theorem 4.7.15. f is the inverse of the differentiable function $x \mapsto x^2$.) So the mean value theorem does apply.

The function $g(x) = \sqrt{|x|}$ is continuous on $[-1, 1]$. (Again, see Exercise 4.3-4.) But it is not differentiable at the point 0 in the interior of that interval. So the mean value theorem does not directly apply. The theorem can be applied separately on the intervals $[-1, 0]$ and $[0, 1]$. With some care about how the function joins together at 0, the same conclusion can be drawn even though the theorem cannot be applied directly. ♦

◇ **4.7-3.** Let f be a nonconstant polynomial such that $f(0) = f(1)$. Prove that f has a local minimum or a local maximum point somewhere in the open interval $]0, 1[$.

Sketch. f attains both maximum and minimum. (Why?) If both are at ends, then f is constant. (Why?) ◊

Solution. Polynomials are continuous everywhere, so the function f is certainly continuous on the compact domain $[0, 1]$. By the maximum-minimum theorem, it must attain both its maximum and its minimum on that set. Since $f(0) = f(1)$, the only way that both of these can be at the ends is for the maximum and minimum to be the same. f would have to be constant, but it has been assumed to be nonconstant. At least one of the maximum or minimum must occur in the interior and thus be a local extremum. ♦

◇ **4.7-4.** A rubber cube of incompressible material is pulled on all faces with a force T. The material stretches by a factor ν in two directions and contracts by a factor ν^{-2} in the other. By balancing forces, one can establish *Rivlin's equation:*

$$\nu^3 - \frac{T\nu^2}{2\alpha} + 1 = 0,$$

where α is a strictly positive constant (analogous to the spring constant for a spring). Show that Rivlin's equation has one (real) solution if $T < 6\sqrt{2}\alpha$ and has three solutions if $T > 6\sqrt{2}\alpha$.

Solution. Let $f(x) = x^3 - (T/2\alpha)x^2 + 1$. Then $f'(x) = 3x^2 - (T/\alpha)x = x(3x-(T/\alpha))$. So $f'(x) = 0$ at $x = 0$ and at $x = T/3\alpha$. $f''(x) = 6x-(T/\alpha)$. So $f''(0) = -T/\alpha < 0$ and $f''(T/3\alpha) = T/\alpha > 0$. There is a local maximum at $x = 0$ and a local minimum at $x = T/3\alpha$. Since $f(0) = 1$, we conclude that there is one negative real root. Since $f(T/3\alpha) = 1 - (T/3\alpha)^3/2$, we conclude that there are no other real roots if $(T/3\alpha)^3/2 > 1$, exactly one if $(T/3\alpha)^3/2 = 1$, and 2 if $(T/3\alpha)^3/2 < 1$. So things are critical when $T = 3\alpha\sqrt[3]{2}$. There is one real root when T is smaller than this and three when T is larger. ♦

◇ **4.7-5.** Let f be continuous on $[3, 5]$ and differentiable on $]3, 5[$, and suppose that $f(3) = 6$ and $f(5) = 10$. Prove that, for some point x_0 in the open interval $]3, 5[$, the tangent line to the graph of f at x_0 passes through the origin. Illustrate your result with a sketch.

Suggestion. Consider the function $f(x)/x$. ◇

Solution. The equation of the line tangent to the graph of f at the point $(x_0, f(x_0)$ is $y = f(x_0) + f'(x_0)(x - x_0)$. For this to pass through the origin we must have $0 = f(x_0) - xf'(x_0)$. A medium sized amount of meditation and a bit of inspiration might remind one that this looks like the numerator of the derivative of the function $g(x) = f(x)/x$ computed by the quotient rule. This function is continuous and differentiable on $[3, 5]$, and $g(3) = g(5) = 2$. So Rolle's Theorem applies and says that there is a point x_0 in the interval at which $g'(x_0) = 0$. But $g'(x) = (xf(x) - f(x))/x^2$. For $g'(x_0) = 0$, we must have $x_0f'(x_0) = f(x_0)$. This is exactly the condition we needed. See Figure 4-9. ♦

4.8 Integration of Functions of One Variable

◇ **4.8-1.** Show directly from the definition that $\int_a^b dx = b - a$.

Sketch. The upper and lower sums are all equal to $b - a$. ◇

FIGURE 4-9.

Solution. Let $P = \{x_0 < x_1 < x_2 < \cdots < x_n\}$ be any partition of the interval $[a, b]$. The function we are integrating is $f(x) = 1$ for all x. So on each subinterval $[x_{j-1}, x_j]$ we have $M_j = \sup\{f(x) \mid x \in [x_{j-1}, x_j]\} = 1 = \sup\{f(x) \mid x \in [x_{j-1}, x_j]\} = m_j$. So

$$L(f, P) = \sum_{j=1}^{n} m_j(x_j - x_{j-1}) = \sum_{j=1}^{n} (x_j - x_{j-1}) = b - a,$$
$$U(f, P) = \sum_{j=1}^{n} M_j(x_j - x_{j-1}) = \sum_{j=1}^{n} (x_j - x_{j-1}) = b - a.$$

Thus

$$b - a = L(f, P) \leq \underline{\int_a^b} dx \leq \overline{\int_a^b} dx \leq U(f, P) = b - a.$$

Since the outer ends of this string of inequalities are equal, all terms in it must be equal. Thus $f(x) = 1$ is integrable on $[a, b]$ and

$$\int_a^b dx = \underline{\int_a^b} dx = \overline{\int_a^b} dx = b - a. \qquad \blacklozenge$$

$\diamond$ **4.8-2.** Evaluate $\int_0^3 (x + 5)\, dx$ using the definition of the integral.

Suggestion. Use the formula $\sum_{k=1}^{n} = n(n + 1)/2$ to evaluate the sums. $\lozenge$

Solution. Let $f(x) = x+5$. Partition the interval $[0, 3]$ into n equal pieces by $P = \{0, 3/n, 6/n, \ldots, 3n/n = 3\}$. That is, $x_j = 3j/n$, $j = 0, 1, 2, \ldots, n$.

The function f is increasing, so on each subinterval we have

$$m_j = \inf\{f(x) \mid x \in [x_{j-1}, x_j]\} = x_{j-1} + 5 = \frac{3(j-1)}{n} + 5,$$
$$M_j = \sup\{f(x) \mid x \in [x_{j-1}, x_j]\} = x_j + 5 = \frac{3j}{n} + 5.$$

So

$$\begin{aligned}
L(f,P) &= \sum_{j=1}^{n} m_j(x_j - x_{j-1}) = \sum_{j=1}^{n} \left(\frac{3(j-1)}{n} + 5\right)\frac{3}{n} \\
&= \frac{9}{n^2}\sum_{j=1}^{n}(j-1) + \frac{3}{n}\sum_{j=1}^{n} 5 = \frac{9}{n^2}\sum_{k=0}^{n-1} k + \frac{3}{n}\sum_{j=1}^{n} 5 \\
&= \frac{9}{n^2}\frac{(n-1)(n-1+1)}{2} + \frac{3}{n}5n = \frac{39}{2} - \frac{9}{2n}, \\
U(f,P) &= \sum_{j=1}^{n} M_j(x_j - x_{j-1}) = \sum_{j=1}^{n}\left(\frac{3j}{n} + 5\right)\frac{3}{n} \\
&= \frac{9}{n^2}\sum_{j=1}^{n} j + \frac{3}{n}\sum_{j=1}^{n} 5 \\
&= \frac{9}{n^2}\frac{(n)(n+1)}{2} + \frac{3}{n}5n = \frac{39}{2} + \frac{9}{2n}.
\end{aligned}$$

Thus

$$\frac{39}{2} - \frac{9}{2n} = L(f,P) \le \underline{\int_a^b} dx \le \overline{\int_a^b} dx \le U(f,P) = \frac{39}{2} + \frac{9}{2n}.$$

The upper and lower integrals differ by no more than $9/n$. Since this is true for every positive integer n, we conclude that they must be equal and that f is integrable on $[0,3]$. Letting $n \to \infty$, we find

$$\int_a^b dx = \underline{\int_a^b} dx = \overline{\int_a^b} dx = \frac{39}{2}. \qquad \blacklozenge$$

◇ **4.8-3.** Show directly from the definition that for $f : [a,b] \to \mathbb{R}$ and P any partition of $[a,b]$, $U(f,P) \ge L(f,P)$.

Sketch. This follows from the fact that for any $S \subseteq \mathbb{R}$ we have $\sup S \ge \inf S$ applied to the set $S = \{f(x) \mid x \in [x_j, x_{j+1}]\}$. ◇

Solution. Let $P = \{a = x_0 < x_1 < x_2 < \cdots < x_n = b\}$ be any partition of $[a, b]$. For each $j = 1, 2, \ldots, n$, we have

$$m_j = \inf\{f(x) \mid x \in [x_{j-1}, x_j]\} \le \sup\{f(x) \mid x \in [x_{j-1}, x_j]\} = M_j.$$

Since $x_j > x_{j-1}$, we have $x_j - x_{j-1} > 0$, and $m_j(x_j - x_{j-1}) \le M_j(x_j - x_{j-1})$. Adding, we obtain

$$L(f, P) = \sum_{j=1}^{n} m_j(x_j - x_{j-1}) \le \sum_{j=1}^{n} M_j(x_j - x_{j-1}) = U(f, P)$$

as claimed. ♦

◇ **4.8-4.** Let $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ be integrable and $f \le M$. Prove that

$$\int_a^b f(x)\, dx \le (b - a)M.$$

Solution. Let $P = \{a = x_0 < x_1 < x_2 < \cdots < x_n = b\}$ be any partition of $[a, b]$. Since $f(x) \le M$ for all x, we have for each $j = 1, 2, \ldots, n$,

$$M_j = \sup\{f(x) \mid x \in [x_{j-1}, x_j]\} \le M.$$

Since $x_j > x_{j-1}$, we have $x_j - x_{j-1} > 0$, and $M_j(x_j - x_{j-1}) \le M(x_j - x_{j-1})$. Adding, we obtain

$$U(f, P) = \sum_{j=1}^{n} M_j(x_j - x_{j-1}) \le \sum_{j=1}^{n} M(x_j - x_{j-1}) = M(b - a).$$

Since f is integrable, we know that $\int_a^b f(x)\, dx$ exists. So

$$\int_a^b f(x)\, dx = \overline{\int_a^b} f(x)\, dx \le U(f, P) \le M(b - a)$$

as claimed. ♦

◇ **4.8-5.** Evaluate

$$\int_0^5 x e^{3x^2}\, dx.$$

Answer. $(1/6)(e^{75} - 1)$. ◇

Solution. We use the Fundamental Theorem of Calculus. This is the bit of theory which underlies the method of integration by substitution or "change of variable". Let $u = g(x) = 3x^2$. Then $\dfrac{du}{dx} = g'(x) = 6x$. So

$$\begin{aligned}
\int_0^5 xe^{3x^2}\,dx &= \frac{1}{6}\int_0^5 e^{3x^2}\,6x\,dx \\
&= \frac{1}{6}\int_0^5 e^{g(x)}g'(x)\,dx \\
&= \frac{1}{6}\int_0^5 \frac{d}{dx}\left(e^{g(x)}\right)\,dx \\
&= \frac{1}{6}e^{g(x)}\Big|_{x=0}^{5} \\
&= \frac{1}{6}e^{3x^2}\Big|_{x=0}^{5} \\
&= \frac{1}{6}e^{75} - \frac{1}{6}e^{0} = \frac{e^{75}-1}{6}.
\end{aligned}$$

$$\boxed{\int_0^5 xe^{3x^2}\,dx = \frac{e^{75}-1}{6} \approx 6.222\times 10^{31}.}$$

♦

◊ **4.8-6.** Evaluate

$$\int_0^1 (x+2)^9\,dx.$$

Suggestion. Try a substitution and the Fundamental Theorem of Calculus. ◊

Solution. We use the Fundamental Theorem of Calculus. Let $u = g(x) = x+2$. Then $\dfrac{du}{dx} = g'(x) = 1$. So

$$\begin{aligned}
\int_0^1 (x+2)^9\,dx = \int_0^1 (x+2)^9\,1\,dx &= \frac{1}{10}\int_0^1 10(g(x))^9 g'(x)\,dx \\
&= \frac{1}{10}\int_0^1 \frac{d}{dx}\left((g(x))^{10}\right)\,dx = \frac{1}{10}(g(x))^{10}\Big|_{x=0}^{1} \\
&= \frac{1}{10}(x+2)^{10}\Big|_{x=0}^{1} = \frac{1}{10}3^{10} - \frac{1}{10}2^{10} = \frac{3^{10}-2^{10}}{10}.
\end{aligned}$$

$$\boxed{\int_0^1 (x+2)^9\,dx = \frac{3^{10}-2^{10}}{10} = \frac{58025}{10}.}$$ ♦

◇ **4.8-7.** Let $f:[0,1]\to\mathbb{R}$, $f(x)=1$ if $x=1/n$, n an integer, and $f(x)=0$ otherwise.

(a) Prove that f is integrable.

(b) Show that $\int_0^1 f(x)\,dx = 0$.

Sketch. $L(f,P)=0$ for every partition of $[0,1]$. (Why?) Now take a partition with the first interval $[x_0,x_1]=[0,\sqrt{2}/n]$ and the others of length no more than $1/n^2$. Show that $U(f,P)\le(\sqrt{2}/n)+(n-1)/n^2$. ◇

Solution. Let $P=\{0=x_0<x_1<x_2<\cdots<x_m=1\}$ be any partition of $[0,1]$. Since $f(x)=0$ except at the isolated points 1, $1/2$, $1/3$, $\ldots$, there are points in every subinterval where $f(x)=0$. Since $f(x)$ is never negative, we have $m_j=\inf\{f(x)\mid x\in[x_{j-1},x_j]\}=0$ for each $j=1,2,3,\ldots$. So

$$0 = L(f,P) \le \underline{\int_0^1} f(x)\,dx.$$

Now let n be any integer larger that 3 and $Q=\{x_0<x_1<x_2<\cdots<x_m\}$ be a partition of $[0,1]$ with $x_0=0$, $x_1=\sqrt{2}/n$, $x_m=1$, and $x_j-x_{j-1}<1/n^2$ for $j=2,3,\ldots,m$. The points $1/n, 1/(n+1), 1/(n+2),\ldots$ are all located in $[x_0,x_1]$. So $M_1=\sup\{f(x)\mid x\in[x_0,x_1]\}=1$, and this subinterval contributes an amount $M_1(x_1-x_0)=\sqrt{2}/n$ to the upper sum. Outside this interval, f is nonzero only at the points $1,1/2,1/3,\ldots,1/(n-1)$. There are $n-1$ of these points, and the subintervals in which they occur each have length no more than $1/n^2$. So the total contribution to the upper integral from these subintervals is no larger than $(n-1)/n^2$. Thus

$$\overline{\int_0^1} f(x)\,dx \le U(f,Q) \le \frac{\sqrt{2}}{n}+\frac{n-1}{n^2}.$$

Since this tends to 0 as n increases, and the inequality holds for every $n>3$, we conclude that

$$0 = L(f,P) \le \underline{\int_0^1} f(x)\,dx \le \overline{\int_0^1} f(x)\,dx \le U(f,Q) \le 0.$$

The upper and lower integrals must both be 0. Since they are equal, f is integrable on $[0, 1]$, and

$$\int_0^1 f(x)\,dx = \underline{\int_0^1} f(x)\,dx = \overline{\int_0^1} f(x)\,dx = 0.$$ ♦

◇ **4.8-8.** Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable and $|f(x)| \le M$. Let $F(x) = \int_a^x f(t)\,dt$. Prove that $|F(y) - F(x)| \le M\,|y - x|$. Deduce that F is continuous. Does this check with Example 4.8.10?

Solution. Using Proposition 4.8.5(iv), we have

$$|F(y) - F(x)| = \left| \int_a^y f(t)\,dt - \int_a^x f(t)\,dt \right| = \left| \int_x^y f(t)\,dt \right|.$$

If $x < y$, this is $\left|\int_x^y f(t)\,dt\right|$. If $y < x$, then it is $\left|-\int_y^x f(t)\,dt\right|$. In either case, it is $\left|\displaystyle\int_{\min(x,y)}^{\max(x,y)} f(t)\,dt\right|$. From the observation at the top of page 207 of the text following Proposition 4.8.5, we have

$$|F(y) - F(x)| = \left| \int_{\min(x,y)}^{\max(x,y)} f(t)\,dt \right| \le \int_{\min(x,y)}^{\max(x,y)} |f(t)|\,dt.$$

But $|f(t)| \le M$ for all t. With this, 4.8.5(iii), and the result of Exercise 4.8-4, we have

$$\begin{aligned} |F(y) - F(x)| &\le \int_{\min(x,y)}^{\max(x,y)} |f(t)|\,dt \le \int_{\min(x,y)}^{\max(x,y)} M\,dt \\ &\le M(\max(x, y) - \min(x, y)) = M\,|x - y| \end{aligned}$$

as claimed.

This inequality holds for all x and y in $[a, b]$. So F satisfies a Lipschitz condition on $[a, b]$ with constant M. This implies f is uniformly continuous on $[a, b]$. If $M = 0$, then F must be constant and is certainly uniformly continuous. If $M > 0$ and $\varepsilon > 0$, put $\delta = \varepsilon/M$. If x and y are in $[a, b]$ and $|x - y| < \delta$, we have $|F(x) - F(y)| \le M\,|x - y| < M\delta = \varepsilon$. So F is uniformly continuous on $[a, b]$.

In Example 4.8.10 we had

$$f(x) = \begin{cases} 0, & 0 \le x \le 1 \\ 1, & 1 < x \le 2 \end{cases} \quad \text{and } F(x) = \begin{cases} 0, & 0 \le x \le 1 \\ x - 1, & 1 < x \le 2 \end{cases}.$$

Although f is not continuous on $[0, 2]$, the indefinite integral $F(x)$ is. See Figure 4.8-3 of the text. ♦

Exercises for Chapter 4

◇ **4E-1.** (a) Prove directly that the function $1/x^2$ is continuous on $]0, \infty[$.

(b) A constant function $f : A \to \mathbb{R}^m$ is a function such that $f(x) = f(y)$ for all $x, y \in A$. Show that f is continuous.

(c) Is the function $f(y) = 1/(y^4 + y^2 + 1)$ continuous? Justify your answer.

Sketch. (a) $\left|(1/x^2) - (1/x_0)^2\right| \le |x - x_0|\,(|x| + |x_0|)/x^2x_0^2$. If $\varepsilon > 0$, try $\delta = \min(x_0/2, \varepsilon x_0^3/10)$.

(b) If $\varepsilon > 0$, let δ be anything larger than 0.

(c) Yes. ♢

Solution. (a) If $f(x) = 1/x^2$, suppose $x_0 > 0$ and compute $|f(x) - f(x_0)|$:

$$|f(x) - f(x_0)| = \left|\frac{1}{x^2} - \frac{1}{x_0^2}\right| = \left|\frac{x^2 - x_0^2}{x^2x_0^2}\right| = \left|\frac{x + x_0}{x^2x_0^2}\right| |x - x_0|\,.$$

First we impose a condition on x to keep the first factor from becoming large and to keep it away from 0. If we require that $|x - x_0| < x_0/2$, then $-x_0/2 < x - x_0 < x_0/2$, so $x_0/2 < x < 3x_0/2$. So $x^2x_0^2 > x_0^4/4$ and $3x_0/2 < x + x_0 < 5x_0/2$. So, with this restriction on x we will have

$$|f(x) - f(x_0)| = \left|\frac{x + x_0}{x^2x_0^2}\right| |x - x_0| \le \frac{5x_0/2}{x_0^4/4}\,|x - x_0| = \frac{10}{x_0^3}\,|x - x_0|\,.$$

So, if $\varepsilon > 0$ and we let $\delta = \min(x_0/2, x_0^3\varepsilon/10)$, then

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| \le \frac{10}{x_0^3}\,|x - x_0| < \frac{10}{x_0^3}\frac{x_0^3\varepsilon}{10} = \varepsilon.$$

Thus f is continuous at x_0. Since x_0 was an arbitrary point in $]0, \infty[$, f is continuous on that domain.

(b) If $f : A \to \mathbb{R}^n$ is a constant function, then there is a point $v_0 \in \mathbb{R}^n$ such that $f(x) = f(y) = v_0$ for all x and y in A. Let $x_0 \in A$. If $\varepsilon > 0$, we can let δ be any positive number. If $x \in A$, we have

$$\|\, f(x) - f(x_0)\,\| = \|\, v_0 - v_0\,\| = 0 < \varepsilon.$$

So f is continuous at x_0. Since x_0 was arbitrary in A, f is continuous on A.

(c) The numerator of the function $f(y) = 1/(y^4 + y^2 + 1)$ is the constant 1 and so is continuous everywhere (part (b)). The denominator is a polynomial and so is also continuous everywhere. (Sums and products of continuous functions are continuous.) So the quotient is continuous except where the denominator is 0. If we allow only real numbers, then y^4 and y^2 must both be nonnegative. So the denominator is at least as large as 1 and is never 0. So f is continuous on all of $\mathbb{R}$.

If we allow complex numbers, then the function would be discontinuous at the four points at which the denominator is 0. ♦

◇ **4E-2.** (a) Prove that if $f : A \to \mathbb{R}^m$ is continuous and $B \subset A$, then the restriction $f|B$ is continuous.

(b) Find a function $g : A \to \mathbb{R}$ and a set $B \subset A$ such that $g|B$ is continuous but g is continuous at no point of A.

Solution. (a) Suppose $\varepsilon > 0$ and $x_0 \in B$. Then $x_0 \in A$, so there is a $\delta > 0$ such that $\| f(x) - f(x_0) \| < \varepsilon$ whenever $x \in A$ and $\| x - x_0 \| < \delta$. If $x \in B$, then it is in A, so

$$(x \in B \text{ and } \| x - x_0 \| < \delta) \implies \| f(x) - f(x_0) \| .$$

So f is continuous at x_0. Since x_0 was arbitrary in B, f is continuous on B.

(b) Let $A = \mathbb{R}$, $B = \mathbb{Q}$, and define $g : A \to \mathbb{R}$ by $g(x) = 1$ if $x \in \mathbb{Q}$ and $g(x) = 0$ if $x \notin \mathbb{Q}$. The restriction of g to $B = \mathbb{Q}$ is constantly equal to 1 on B. So it is continuous on B. (See Exercise 4E-1(b).) But, if $x_0 \in \mathbb{R}$, then there are rational and irrational points in every short interval around x_0. So g takes the values 1 and 0 in every such interval. The values of $g(x)$ cannot be forced close to any single value by restricting to a short interval around x_0. So, as a function on $\mathbb{R}$, g is not continuous at x_0. This is true for every $x_0 \in \mathbb{R}$. ♦

◇ **4E-3.** (a) If $f : \mathbb{R} \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}$ is connected, is $f^{-1}(K)$ necessarily connected?

(b) Show that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous on all of $\mathbb{R}^n$ and $B \subset \mathbb{R}^n$ is bounded, then $f(B)$ is bounded.

Sketch. (a) No. Let $f(x) = \sin x$ and $K = \{1\}$.

(b) f is continuous on all of $\mathbb{R}^n$, so f is continuous on $\mathrm{cl}(B)$ which is compact. $f(\mathrm{cl}(B))$ is compact and thus bounded. So, since $f(B) \subseteq f(\mathrm{cl}(B))$, $f(B)$ is also bounded. ◇

Solution. (a) The continuous image of a connected set must be connected, but not necessarily a *preimage*. For example. Let $f(x) = \sin x$ for all

$x \in \mathbb{R}$, and let $K = \{1\}$. The one point set K is certainly connected, but $f^{-1}(K) = \{(4n+1)\pi/2 \mid n \in \mathbb{Z}\}$. This discrete set of points is not connected.

For an easier example, let $f(x) = x^2$ and $K = \{1\}$. Then f is continuous and $f^{-1}(K) = \{-1, 1\}$. This two point set is not connected.

(b) Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous and B is a bounded subset of $\mathbb{R}^n$. Since B is bounded, there is a radius R such that $\| v \| < R$ for every $v \in B$. If $w \in \mathrm{cl}(B)$, then there is a $v \in B$ with $v \in D(w, 1)$. So $\| w \| \leq \| w - v \| + \| v \| \leq 1 + R$. So $\mathrm{cl}(B)$ is a bounded set in $\mathbb{R}^n$. It is also closed, so it is compact. The function f is continuous on all of $\mathbb{R}^n$, so it is continuous on $\mathrm{cl}(B)$. (See Exercise 4E-2(a).) So the image $f(\mathrm{cl}(B))$ is compact by Theorem 4.2.2. Since it is compact, it must be bounded. Since $B \subseteq \mathrm{cl}(B)$, we have $f(B) \subseteq f(\mathrm{cl}(B))$. Since it is a subset of a bounded set, the image $f(B)$ is bounded. ♦

◇ **4E-4.** Discuss why it is necessary to impose the condition $x \neq x_0$ in the definition of $\lim_{x \to x_0} f(x)$ by considering what would happen in the case $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 0$ if $x \neq 0$ and $f(0) = 1$, and how one would define the derivative.

Solution. The function described has a limit of 0 as x tends to 0 even though $f(0) = 1$. The use of the condition $0 < |x - a| < \delta$ in the definition of limit allows this since it means that the limit depends on the behavior of the function near the point but *not* on the value of the function at the point. With this we can have a limit even if $f(a)$ does not exist, and we can have a limit different from $f(a)$ if $f(a)$ does exist. This allows us to use the limit in the definition of continuity simply by demanding that the limit be equal to $f(a)$. If we allowed $|x - a| = 0$ in the definition, then the limit would have to equal $f(a)$ no matter what.

The restriction is even more important in the definition of derivative as the limit of a difference quotient. The whole point of that definition is that we want to consider what is happening to the ratio as the denominator gets closer and closer to 0 without being forced to look directly at what happens when it is 0. ♦

◇ **4E-5.** Show that $f : A \to \mathbb{R}^m$ is continuous at x_0 iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that $\| x - x_0 \| \leq \delta$ implies $\| f(x) - f(x_0) \| \leq \varepsilon$. Can we replace $\varepsilon > 0$ or $\delta > 0$ by $\varepsilon \geq 0$ or $\delta \geq 0$?

Sketch. If f is continuous, the condition holds by Definition 4.1.1 and Definition 4.1.2. For the converse, use the hypothesis for $\varepsilon/2 > 0$, getting $\| f(x) - f(x_0) \| \leq \varepsilon/2 < \varepsilon$. We cannot replace "$> 0$" by "$\geq 0$". ◇

Solution. If f is continuous at $x_0 \in A$ and $\varepsilon > 0$, then there is a $\delta_0 > 0$ such that

$$(x \in A \text{ and } \| x - x_0 \| < \delta_0) \implies \| f(x) - f(x_0) \| < \varepsilon.$$

Put $\delta = \delta_0/2$. Then

$$\begin{aligned} (x \in A \text{ and } \| x - x_0 \| \le \delta) &\implies (x \in A \text{ and } \| x - x_0 \| < \delta_0) \\ &\implies \| f(x) - f(x_0) \| < \varepsilon \\ &\implies \| f(x) - f(x_0) \| \le \varepsilon. \end{aligned}$$

In the other direction, suppose the stated condition holds and that $\varepsilon > 0$. Then $\varepsilon/2 > 0$, and by hypothesis there is a $\delta > 0$ such that

$$\begin{aligned} (x \in A \text{ and } \| x - x_0 \| < \delta) &\implies (x \in A \text{ and } \| x - x_0 \| \le \delta) \\ &\implies \| f(x) - f(x_0) \| \le \varepsilon/2 \\ &\implies \| f(x) - f(x_0) \| < \varepsilon. \end{aligned}$$

So the modified condition is equivalent to continuity at x_0 as claimed.

If we allowed $\varepsilon = 0$ in the target, but still required $\delta > 0$, then we would be requiring that f be constant in a neighborhood of x_0. We certainly do not want to require this for continuity. If we allow $\delta = 0$, then every function defined at x_0 would be continuous at x_0. This is even more useless. ♦

◇ **4E-6.** (a) Let $\{c_k\}$ be a sequence in $\mathbb{R}$. Show that $c_k \to c$ iff every subsequence of c_k has a further subsequence that converges to c.

(b) Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function. Prove that f is continuous if and only if the graph of f is a closed subset of $\mathbb{R}^2$. What if f is unbounded?

Solution. (a) If $\langle c_k \rangle_1^\infty$ is a sequence in $\mathbb{R}$ and $c_k \to c \in \mathbb{R}$, then every subsequence also converges to c. For the converse, suppose that $\langle c_k \rangle_1^\infty$ is a sequence in $\mathbb{R}$, that $c \in \mathbb{R}$, and that every subsequence has a sub-subsequence which converges to c. We want to show that the whole sequence converges to c. If it did not, there would be an $\varepsilon > 0$ for which no N would work in the definition of convergence of a sequence. In particular, we could select an index $k(1)$ such that $\left| c_{k(1)} - c \right| > \varepsilon$. Then, having picked indices $k(1) < k(2) < \cdots < k(n)$ with $\left| c_{k(j)} - c \right| > \varepsilon$ for $j = 1, 2, \ldots, n$, there would have to be an index $k(n+1) > k(n)$ with $\left| c_{k(n+1)} - c \right| > \varepsilon$ also. This inductively generates indices $k(1) < k(2) < k(3) < \ldots$ such that $\left| c_{k(j)} - c \right| > \varepsilon$ for each $j = 1, 2, 3, \ldots$. Contrary to hypothesis, the subsequence $c_{k(1)}, c_{k(2)}, \ldots$ could not have any sub-subsequence converging to c. This contradiction shows that we must have $c_k \to c$ as claimed.

(b) Let $f : \mathbb{R} \to \mathbb{R}s$ be a bounded function and define the graph of f as a subset of $\mathbb{R}$ by $\Gamma = \{(x, f(x)) \mid x \in \mathbb{R}\}$. First let f be continuous, and suppose $(x, y) \in \mathrm{cl}(\Gamma)$. Then there is a sequence $v_k = (x_k, y_k) \in \Gamma$ with $v_k \to (x, y)$. A sequence in $\mathbb{R}^2$ converges to a limit w if and only if both coordinate sequences converge to the corresponding coordinates of w. So $x_k \to x$ and $y_k \to y$. Since $x_k \to x$ and f is continuous at x, we must have $f(x_k) \to f(x)$. But (x_k, y_k) is in the graph of f, so $y_k = f(x_k)$. Thus $y_k \to y$ and $y_k \to f(x)$. Since limits of sequences in $\mathbb{R}$ are unique, we must have $y = f(x)$. So $(x, y) \in \Gamma$. Thus $\mathrm{cl}(\Gamma) \subseteq \Gamma$ and Γ is closed.
For the converse, suppose that Γ is a closed subset of $\mathbb{R}^2$. We want to show that f is continuous on $\mathbb{R}$. So let $a \in \mathbb{R}$ and $\langle x_n \rangle_1^\infty$ be a sequence in $\mathbb{R}$ converging to a. We want to show that $f(x_n)$ must converge to $f(a)$. To do this we show that every subsequence has a sub-subsequence converging to $f(a)$ and appeal to part (a). So, let $c_1, c_2, c_3, \ldots$ be a subsequence of the images $f(x_n)$. Since f is a bounded function, the sequence $\langle c_k \rangle_1^\infty$ is a bounded sequence. By the Bolzano-Weierstrass property of $\mathbb{R}$, it must have a subsequence converging to some point $y \in \mathbb{R}$. But $(x_k, c_k) \in \Gamma$ for each k, and $x_k \to a$. So, along the subsequence we have $x_{k(j)} \to a$ and $c_{k(j)} = f(x_{k(j)}) \to y$. So $(x_{k(j)}, c_{k(j)}) \to (a, y)$. Since each of the points $(x_{k(j)}, c_{k(j)})$ is in the closed set Γ, the limit (a, y) must be in Γ also. So $y = f(a)$. We have shown that every subsequence of the points $f(x_k)$ has a sub-subsequence converging to $f(a)$. By the result of part (a), we must have $f(x_k) \to f(a)$. Since this works for every sequence converging to a, we have f continuous at a, and, since a was an arbitrary point in $\mathbb{R}$, we have f continuous on $\mathbb{R}$.
If f is not bounded, the first half of the proof just given still works. We can still conclude that if f is continuous, then the graph must be closed. In the second half of the proof we used the hypothesis of boundedness of f, and it is, in fact, required. Consider the function defined on $\mathbb{R}$ by $f(x) = 1/x$ for $x \neq 0$ and $f(0) = 0$. Then the graph of f consists of the hyperbola $\{(x, y) \in \mathbb{R}^2 \mid xy = 1\}$ together with the origin $\{(0, 0)\}$. This is a closed subset of $\mathbb{R}^2$, but f is not continuous. ♦

◇ **4E-7.** Consider a compact set $B \subset \mathbb{R}^n$ and let $f : B \to \mathbb{R}^m$ be continuous and one-to-one. Then prove that $f^{-1} : f(B) \to B$ is continuous. Show by example that this may fail if B is connected but not compact. (To find a counterexample, it is necessary to take $m > 1$.)

Sketch. Suppose C is a closed subset of B. Then C is compact. (Why?) So $f(C)$ is closed. (Why?) Thus f^{-1} is continuous. (Why?) For a counterexample with $n = 2$ consider $f : [0, 2\pi[\to \mathbb{R}^2$ given by $f(t) = (\sin t, \cos t)$. ◇

Solution. FIRST PROOF: We use the characterization of continuity in terms of closed sets. To show that $f^{-1} : f(B) \to B$ is continuous on $f(B)$, we need to show that if C is a closed subset of the metric space B, then $(f^{-1})^{-1}(C)$ is closed relative to $f(B)$. Since B is a compact subset of $\mathbb{R}^n$, it is closed, and a subset C of it is closed relative to B if and only if it is closed in $\mathbb{R}^n$. Since it is a closed subset of the compact set B it is closed. (In $\mathbb{R}^n$ this follows since it is closed and bounded. However, it is true more generally. See Lemma 2 to the proof of the Bolzano-Weierstrass Theorem, 3.1.3, at the end of Chapter 3: *A closed subset of a compact space is compact.*) Since C is a compact subset of B and f is continuous on B and hence on C, the image $f(C)$ is compact. Since it is a compact subset of a metric space, it is closed. (See Lemma 1 to the proof of 3.1.3.) But since f is one-to-one, $f(C) = (f^{-1})^{-1}(C)$. Thus $(f^{-1})^{-1}(C)$ is closed for every closed subset C of $f(B)$. The inverse f^{-1} is thus a continuous function from $f(B)$ to B.

SECOND PROOF: Here is a proof using sequences. Suppose $y \in f(B)$ and $\langle y_k \rangle_1^\infty$ is a sequence in $f(B)$ with $y_k \to y$. We want to show that $x_k = f^{-1}(y_k) \to x = f^{-1}(y)$ in B. Since B is compact, there is a subsequence $x_{k(1)}, x_{k(2)}, x_{k(3)}, \ldots$ converging to some point $\hat{x} \in B$. Since f is continuous on B, we must have $y_{k(j)} = f(x_{k(j)}) \to f(\hat{x})$. But $y_{k(j)} \to y$. Since limits are unique in the metric space $f(B)$, we must have $y = f(\hat{x})$. But $y = f(x)$ and f is one-to-one, so $x = \hat{x}$. Not only does this argument show that there must be some subsequence of the x_k converging to x, it shows that x is the only possible limit of a subsequence. Since B is compact, every subsequence would have to have a sub-subsequence converging to something, and the only possible "something" is x. Thus $x_k \to x$ as needed.

If the domain is not compact, for example the half-open interval $B = [0, 2\pi[$, then we can get a counterexample. The map $f : [0, 2\pi[\to \mathbb{R}^2$ given by $f(t) = (\sin t, \cos t)$ takes $[0, 2\pi[$ onto the unit circle. The point $(0, 1)$ has preimages near 0 and near 2π. So the inverse function is not continuous at $(0, 1)$.

It turns out that a continuous map from a half-open interval one-to-one into $\mathbb{R}$ must have a continuous inverse. Challenge: Prove it. ♦

◇ **4E-8.** Define maps $s : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and $m : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ as addition and scalar multiplication defined by $s(x, y) = x + y$ and $m(\lambda, x) = \lambda x$. Show that these mappings are continuous.

Solution. If $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$. Then $s(x, y) = x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$, and $m(\lambda, x) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$.

We compute

$$\begin{aligned}\|\, s(x,y) - s(u,v) \,\|_{\mathbb{R}^n} &= \|\, (x+y) - (u+v) \,\|_{\mathbb{R}^n} = \|\, (x-u) + (y-v) \,\|_{\mathbb{R}^n} \\ &\le \|\, x-u \,\|_{\mathbb{R}^n} + \|\, y-v \,\|_{\mathbb{R}^n} \\ &\le 2\sqrt{\|\, x-u \,\|^2_{\mathbb{R}^n} + \|\, y-v \,\|^2_{\mathbb{R}^n}} \\ &\le 2\,\|\, (x-u, y-v) \,\|_{\mathbb{R}^n\times\mathbb{R}^n} = 2\,\|\, (x,y) - (u,v) \,\|_{\mathbb{R}^n\times\mathbb{R}^n}\,.\end{aligned}$$

So, if $\|\, (x,y) - (u,v) \,\|_{\mathbb{R}^n\times\mathbb{R}^n} < \varepsilon/2$, then $\|\, s(x,y) - s(u,v) \,\|_{\mathbb{R}^n} < \varepsilon$. So s is continuous.

Fix $(\mu, u) \in \mathbb{R} \times \mathbb{R}^n$. We have

$$\|\, (\lambda, x) - (\mu, u) \,\|^2_{\mathbb{R}^n\times\mathbb{R}^n} = \|\, (\lambda - \mu, x-u) \,\|^2_{\mathbb{R}^n\times\mathbb{R}^n} = (\lambda-\mu)^2 + \|\, x-u \,\|^2_{\mathbb{R}^n}\,.$$

Thus if $\|\, (\lambda, x) - (\mu, u) \,\|_{\mathbb{R}^n\times\mathbb{R}^n} < \delta$ then $|\lambda - \mu| < \delta$ and $\|\, x-u \,\|_{\mathbb{R}^n} < \delta$.

$$\begin{aligned}\|\, m(\lambda,x) - m(\mu,u) \,\|_{\mathbb{R}^n} &= \|\, \lambda x - \mu u \,\|_{\mathbb{R}^n} = \|\, \lambda x - \lambda u + \lambda u - \mu u \,\|_{\mathbb{R}^n} \\ &\le \|\, \lambda x - \lambda u \,\|_{\mathbb{R}^n} + \|\, \lambda u - \mu u \,\|_{\mathbb{R}^n} = |\lambda|\,\|\, x-u \,\|_{\mathbb{R}^n} + |\lambda - \mu|\,\|\, u \,\|_{\mathbb{R}^n} \\ &\le |\lambda|\,\delta + \delta\,\|\, u \,\|_{\mathbb{R}^n} \le (|\mu| + \delta)\delta + +\delta\,\|\, u \,\|_{\mathbb{R}^n}\,.\end{aligned}$$

If we require that $\delta < 1$ and $\delta < \varepsilon/2(|\mu|+1)$ and $\delta < \varepsilon/2(\|\, u \,\| + 1)$, we have

$$\begin{aligned}\|\, m(\lambda,x) - m(\mu,u) \,\|_{\mathbb{R}^n} &\le (|\mu| + \delta)\delta + \delta\,\|\, u \,\|_{\mathbb{R}^n} \\ &\le (|\mu|+1)\frac{\varepsilon}{2(|\mu|+1)} + \frac{\varepsilon}{2(\|\, u \,\|+1)}\,\|\, u \,\| < \varepsilon.\end{aligned}$$

Thus m is continuous. ♦

◇ **4E-9.** Prove the following "gluing lemma": *Let $f : [a,b[\to \mathbb{R}^m$ and $g : [b,c] \to \mathbb{R}^m$ be continuous. Define $h : [a,c] \to \mathbb{R}^m$ by $h = f$ on $[a,b]$ and $h = g$ on $[b,c]$. If $f(b) = g(b)$, then h is continuous.* Generalize this result to sets A, B in a metric space.

Suggestion. Show that If F is closed in $\mathbb{R}^m$, then $h^{-1}(F) = f^{-1}(F) \cup g^{-1}(F)$ and so is closed. ◇

Solution. Put $A = [a,b]$ and $B = [b,c]$. Let $y_0 = f(b) = g(b)$. Then the function h is well-defined since it makes no difference whether we use f or g to define h at the point b in $A \cap B$. and suppose F is a closed set in $\mathbb{R}^m$. Then

$$\begin{aligned}h^{-1}(F) &= h^{-1}(F) \cap (A \cup B) = (h^{-1}(F) \cap A) \cup (h^{-1}(F) \cap B) \\ &= (f^{-1}(F) \cap A) \cup (g^{-1}(F) \cap B)\end{aligned}$$

Since f and g are continuous and F, $[a,b]$, and $[b,c]$ are closed, this set is closed. The inverse image of every closed set is closed, so h is continuous.

For the generalization, suppose A and B are closed sets in $\mathbb{R}^n$ and that $f : A \to \mathbb{R}^m$ and $g : B \to \mathbb{R}^m$ are continuous and that $f(x) = g(x)$ for $x \in A \cap B$. Define h on $A \cup B$ by putting $h(x) = f(x)$ for $x \in A$ and $h(x) = g(x)$ for $x \in B$. After observing that h is well-defined since f and g agree on the intersection, the proof of continuity is the same.

Note that some sort of assumption needs to be made about the sets A and B. Otherwise we could take something like $A = \mathbb{Q}$ and $B = \mathbb{R} \setminus \mathbb{Q}$. If we put $f(x) = 1$ for all x in A and $g(x) = 0$ for all x in B, then, since $A \cap B = \emptyset$, h is still well-defined, but h is not continuous at any point of $\mathbb{R} = A \cup B$. ♦

◇ **4E-10.** Show that $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$, is continuous iff for every set $B \subset A$, $f(\mathrm{cl}(B) \cap A) \subset \mathrm{cl}(f(B))$.

Solution. Suppose f is continuous on A and $B \subseteq A$. Then $B \subseteq f^{-1}(f(B)) \subseteq f^{-1}(\mathrm{cl}(f(B)))$. So $f^{-1}(\mathrm{cl}(f(B))) = A \cap F$ for some closed set F since f is continuous on A and $\mathrm{cl}(f(B))$ is closed. So $B \subseteq F$ and $\mathrm{cl}(B) \subseteq F$. Thus

$$\mathrm{cl}(B) \cap A \subseteq F \cap A = f^{-1}(\mathrm{cl}(f(B))).$$

This implies that $f(\mathrm{cl}(B) \cap A) \subseteq f(f^{-1}\mathrm{cl}(f(B))) = \mathrm{cl}(f(B))$ as required.

For the converse, suppose $f(\mathrm{cl}(B) \cap A) \subseteq \mathrm{cl}(f(B))$ for every subset B of A. Let C be any closed set in $\mathbb{R}^m$ and put $B = f^{-1}(C)$. The hypothesis gives

$$f(\mathrm{cl}(f^{-1}(C)) \cap A) \subseteq \mathrm{cl}(f(f^{-1}(C))) \subseteq \mathrm{cl}(C) = C.$$

So

$$\mathrm{cl}(f{-}1(C)) \cap A \subseteq f^{-1}(f(\mathrm{cl}(f{-}1(C)) \cap A)) \subseteq f^{-1}(C) \subseteq \mathrm{cl}(f{-}1(C)) \cap A.$$

We must have $f^{-1}(C) = \mathrm{cl}(f{-}1(C)) \cap A$. So $f^{-1}(C)$ is closed relative to A. Since this is true for every closed set C in $\mathbb{R}^m$, f is continuous on A. ♦

◇ **4E-11.** (a) For $f :]a,b[\to \mathbb{R}$, show that if f is continuous, then its graph Γ is path-connected. Argue intuitively that if the graph of f is path-connected, then f is continuous. (The latter is true, but it is a little more difficult to prove.)

(b) For $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$, show that for $n \geq 2$, connectedness of the graph does not imply continuity. [Hint: For $f : \mathbb{R}^2 \to \mathbb{R}$, cut a slit in the graph.]

(c) Discuss (b) for $m = n = 1$. [Hint: On $\mathbb{R}$, consider $f(x) = 0$ if $x = 0$ and $f(x) = \sin(1/x)$ if $x > 0$.]

Sketch. (a) If $(c, f(c))$ and $(d, f(d))$ are on the graph, put $\gamma(t) = (t, f(t))$. ♢

Solution. (a) If $(c, f(c))$ and $(d, f(d))$ are two points on the graph, then either $a < c < d < b$ or $a < d < c < b$. We may as well assume the first. The map $\gamma : [c, d] \to \mathbb{R}^2$ given by $\gamma(t) = (t, f(t))$ is a continuous path in the graph joining $(c, f(c))$ to $d, f(d))$ by Worked Example 4WE-1 since both coordinate functions are continuous. Thus the graph is path-connected.

An intuitive argument for the converse is that the only paths possible in the graph must be of the form $\gamma(t) = (u(t), f(u(t)))$. For this to be continuous, u and f should be continuous.

(b) Represent a point $v \in \mathbb{R}^2$ in polar coordinates $v = (r, \vartheta)$ where $r = \| v \|$ and $0 \le \vartheta < 2\pi$ is the polar angle counterclockwise from the positive horizontal axis. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(v) = r\vartheta$ and $f(0, 0) = 0$. Then the graph of f is connected since it is path-connected. But f is not continuous since it has a jump discontinuity across the positive x-axis.

(c) Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \sin(1/x)$ if $x \neq 0$ and $f(0) = 0$. Although the graph of f is not path-connected, it is connected since any open set containing $(0, 0)$ would also contain points in the path-connected portion of the graph. The function is not continuous since it attain all values in the interval $[-1, 1]$ in every neighborhood of 0. ♦

◇ **4E-12.** (a) A map $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is called *Lipschitz on A* if there is a constant $L \ge 0$ such that $\| f(x) - f(y) \| \le L \| x - y \|$, for all $x, y \in A$. Show that a Lipschitz map is uniformly continuous.

(b) Find a bounded continuous function $f : \mathbb{R} \to \mathbb{R}$ that is not uniformly continuous and hence is not Lipschitz.

(c) Is the sum (product) of two Lipschitz functions again a Lipschitz function?

(d) Is the sum (product) of two uniformly continuous functions again uniformly continuous?

(e) Let f be defined and have a continuous derivative on $]a - \varepsilon, b + \varepsilon[$ for some $\varepsilon > 0$. Show that f is a Lipschitz function on $[a, b]$.

Solution. (a) Suppose $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and there is a constant $L \ge 0$ such that $\| f(x) - f(y) \| \le L \| x - y \|$ for all x, $y \in A$. If $L = 0$, then $\| f(x) - f(y) \| = 0$ for all x and y in A. f must be a constant function on A and so is certainly uniformly continuous. If $L > 0$ and $\varepsilon > 0$, put $\delta = \varepsilon/L$. If x and y are in A and $\| x - y \| < \delta$ we have

$$\| f(x) - f(y) \| \le L \| x - y \| < L\delta = \varepsilon.$$

So f is uniformly continuous on A.

(b) Let $f(x) = \sin(x^2)$ then f is continuous and $|f(x)| \leq 1$ for all x in $\mathbb{R}$. Let $x_n = \sqrt{2\pi n}$ and $y_n = \sqrt{2\pi n + (\pi/2)}$. Then $|f(y_n) - f(x_n)| = 1$, but

$$\begin{aligned} |y_n - x_n| &= \sqrt{2\pi n + (\pi/2)} - \sqrt{2\pi n} \\ &= \frac{2\pi n + (\pi/2) - 2\pi n}{\sqrt{2\pi n + (\pi/2)} + \sqrt{2\pi n}} \\ &\leq \frac{\pi}{4\sqrt{2\pi n}}. \end{aligned}$$

This tends to 0 as n increases. No matter how small a positive number δ is specified, we can always find x_n and y_n closer together than δ with $|f(x_n) - f(y_n)| = 1$. So f is not uniformly continuous.

(c) Suppose $\| f(x) - f(y) \| \leq L \| x - y \|$ and $\| g(x) - g(y) \| \leq M \| x - y \|$ for all x and y in A. Put $h(x) = f(x) + g(x)$. Then

$$\begin{aligned} \| h(x) - h(y) \| &= \| (f(x) + g(x)) - (f(y) + g(y)) \| \\ &= \| (f(x) - f(y)) + (g(x) - g(y)) \| \\ &\leq \| f(x) - f(y) \| + \| g(x) - g(y) \| \\ &\leq L \| x - y \| + M \| x - y \| = (L + M) \| x - y \| . \end{aligned}$$

So the sum is a Lipschitz function with constant no larger than $L + M$. The product of two Lipschitz functions need not be a Lipschitz function. If we put $f(x) = g(x) = x$ for all $x \in \mathbb{R}$, then f and g are both Lipschitz with constant 1. But, if we put $x_n = n + (1/n)$ and $y_n = n$, then $|x_n - y_n| = 1/n$, but $\left|x_n^2 - y_n^2\right| = 2 + (1/n^2) > 2 = (2n)\,|x_n - y_n|$. So there can be no Lipschitz constant for the function $x \mapsto x^2$ which works on all of $\mathbb{R}$.

(d) If f and g are uniformly continuous on A and $\varepsilon > 0$, then there are constants $\delta_1 > 0$ and $\delta_2 > 0$ such that for x and y in A we have

$$\begin{aligned} \| x - y \| < \delta_1 &\implies \| f(x) - f(y) \| < \varepsilon/2, \\ \| x - y \| < \delta_2 &\implies \| g(x) - g(y) \| < \varepsilon/2. \end{aligned}$$

If x and y are n A and $\| x - y \| < \delta = \min(\delta_1, \delta_2)$, then

$$\begin{aligned} \| (f + g)(x) - (f + g)(y) \| &= \| (f(x) - f(y)) + (g(x) - g(y)) \| \\ &\leq \| f(x) - f(y) \| + \| g(x) - g(y) \| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Thus the sum of two uniformly continuous functions must be uniformly continuous.

The function $x \mapsto x$ is certainly uniformly continuous on $\mathbb{R}$ since $\delta = \varepsilon$ will work everywhere. However, the product of this function with itself is $x \mapsto x^2$. The computation in part (c) shows that this is not uniformly continuous. So the product of two uniformly continuous functions might not be uniformly continuous.

(e) Since f' is continuous on the compact set $[a, b]$, it is bounded on that set. There is a constant M such that $|f'(c)| \le M$ for every c in $[a, b]$. If $a \le x < y \le b$, we know from the mean value theorem that there is a point c between x and y with $f(y) - f(x) = f'(c)(y - x)$. So $|f(y) - f(x)| \le M\,|y - x|$. Thus f satisfies a Lipschitz condition on $[a, b]$ with constant M. ♦

◇ **4E-13.** Let f be a bounded continuous function $f : \mathbb{R}^n \to \mathbb{R}$. Prove that $f(U)$ is open for all open sets $U \subset \mathbb{R}^n$ iff for all nonempty open sets $V \subset \mathbb{R}^n$,

$$\inf_{x \in V} f(x) < f(y) < \sup_{x \in V} f(x)$$

for all $y \in V$.

Sketch. If $\inf(f(V))$ or $\sup(f(V))$ were in $f(V)$, then $f(V)$ could not be open since it could not contain an interval around either of these points. ◇

Solution. First suppose that $f(u)$ is an open subset of $\mathbb{R}$ for every open $U \subseteq \mathbb{R}^n$, and let V be a nonempty subset of $\mathbb{R}^n$. If $\inf(f(V)) = a = f(y)$ for some $y \in V$, then $a = f(y)$ cannot be an interior point of $f(V)$ since $f(V)$ can contain no points smaller than a. So $f(V)$ would not be open contrary to hypothesis. Similarly, if $\sup(f(V)) = b = f(y)$ for some $y \in V$, then b would be in $f(V)$ but could not be an interior point of $f(V)$ since $f(V)$ could contain no points larger than b. Again $f(V)$ would not be open contrary to hypothesis. Since we always have $\inf_{x \in V} f(x) \le f(y) \le \sup_{x \in V} f(x)$ for all y in V and we have just shown that neither equality can occur, we must have $\inf_{x \in V} f(x) < f(y) < \sup_{x \in V} f(x)$ as claimed.

For the converse, suppose that for every open set V in $\mathbb{R}^n$ we have $\inf_{x \in V} f(x) < f(y) < \sup_{x \in V} f(x)$ for every y in V, and let U be an open subset of $\mathbb{R}^n$. We want to show that $f(U)$ is open. So let $y = f(x) \in f(U)$. Since $x \in U$ and U is open, there is an $r > 0$ with $D(x, 2r) \subseteq U$. So $\mathrm{cl}(D(x, r)) \subseteq U$. Let $K = \mathrm{cl}(D(x, r))$. The closed disk K is a closed bounded set in $\mathbb{R}^n$ and so is compact. Since f is continuous, the image, $f(K)$, is a connected, compact set in $\mathbb{R}$. So it must be a closed, bounded interval containing y. $y \in f(K) = [a, b]$. Let V be the open disk $D(x, r)$. We have $y \in f(V) \subseteq f(K) = [a, b]$. By hypothesis,

$$a = \inf(f(K)) \le \inf(f(V)) < y < \sup(f(V)) \le \sup(F(K)) = b.$$

Let ρ be the smaller of $(b-y)/2$ and $(y-a)/2$. Then $]a+\rho, b-\rho[$ is an open interval in $f(U)$ containing y. Since this can be done for any y in $f(U)$, the set $F(U)$ is open as claimed. ♦

◇ **4E-14.** (a) Find a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\lim_{x\to 0}\lim_{y\to 0} f(x,y) \quad \text{and} \quad \lim_{y\to 0}\ \lim_{x\to 0} f(x,y)$$

exist but are not equal.

(b) Find a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that the two limits in (a) exist and are equal but f is not continuous. [Hint: $f(x,y) = xy/(x^2+y^2)$ with $f = 0$ at $(0,0)$.]

(c) Find a function $f : \mathbb{R}^2 \to \mathbb{R}$ that is continuous on every line through the origin but is not continuous. [Hint: Consider the function given in polar coordinates by $r\tan(\theta/4)$, $0 \le r < \infty$, $0 \le \theta < 2\pi$.]

Solution. (a) Let $f(0,0) = 0$, and for other points, put $f(x,y) = x^2/(x^2+y^2)$. For fixed nonzero x we have $\lim_{y\to 0} f(x,y) = x^2/x^2 = 1$. So

$$\lim_{x\to 0}\left(\lim_{y\to 0} f(x,y)\right) = \lim_{x\to 1}(0) = 1.$$

For fixed, nonzero y, we have $\lim_{x\to 0} f(x,y) = 0/y^2 = 0$. So

$$\lim_{y\to 0}\left(\lim_{x\to 0} f(x,y)\right) = \lim_{y\to 0}(0) = 0.$$

(b) Let $f(0,0) = 0$, and for other points, put $f(x,y) = xy/(x^2+y^2)$. For fixed nonzero x we have $\lim_{y\to 0} f(x,y) = 0/x^2 = 0$. So

$$\lim_{x\to 0}\left(\lim_{y\to 0} f(x,y)\right) = \lim_{x\to 0}(0) = 0.$$

For fixed, nonzero y, we have $\lim_{x\to 0} f(x,y) = 0/y^2 = 0$. So

$$\lim_{y\to 0}\left(\lim_{x\to 0} f(x,y)\right) = \lim_{y\to 0}(0) = 0.$$

But, if we look at the values of f along the line $y = x$, we find $f(x,x) = x^2(x^2+x^2) = 1/2$. Since there are such points in every neighborhood of the origin, and $f(0,0) = 0$, f is not continuous at the origin.

(c) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given in polar coordinates by $r\tan(\vartheta/4)$, $0 \le r < \infty$ and $0 \le \vartheta < 2\pi$. Any line through the origin can be parameterized as $\gamma(t) = (t\cos\vartheta_0, t\sin\vartheta_0)$ for $-\infty < t < \infty$ and fixed ϑ_0 with $-\pi/2 < \theta_0 \le \pi/2$. Along such a line, the values of f are given by $t\tan(\vartheta_0/4)$ which is a continuous function of t. So f is continuous along every line through the origin with the value 0 at the origin. However, if you fix $r_0 > 0$ and move counterclockwise around the circle of radius r_0, the values of f start at 0 on the horizontal axis at $(r_0, 0)$. As you move around the circle, the values are $r_0 \tan(\vartheta/4$. As you get close to $(r_0, 0)$ from below the horizontal axis. $\vartheta/4 \to \pi/2$, so the values of f tend to $+\infty$. Thus f is not continuous across the positive horizontal axis. ♦

◇ **4E-15.** Let $f_1, \dots, f_N$ be functions from $A \subset \mathbb{R}^n$ to $\mathbb{R}$. Let m_i be the maximum of f_i, that is, $m_i = \sup(f_i(A))$. Let $f = \sum f_i$ and $m = \sup(f(A))$. Show that $m \le \sum m_i$. Give an example where equality fails.

Suggestion. For an example with inequality, try $f_1(x) = x$ and $f_2(x) = 1 - x$ on $[0, 1]$. ♢

Solution. If $x \in A$, then $f_k(x) \le \sup\{f_k(x) \mid x \in A\} = m_k$ for each $k = 1, 2, \dots N$. So

$$\begin{aligned} f(x) &= f_1(x) + f_2(x) + \cdots + f_N(x) \\ &\le m_1 + f_2(x) + \cdots + f_N(x) \\ &\le m_1 + m_2 + \cdots + f_N(x) \\ &\vdots \\ &\le m_1 + m_2 + \cdots + m_N. \end{aligned}$$

This holds for every $x \in A$, so $m = \sup\{f(x) \mid x \in A\} \le m_1 + m_2 + \cdots + m_N$ as claimed.

The inequality can be strict. Let $A = [0, 1] \subseteq \mathbb{R}$, and set $f_1(x) = x$ and $f_2(x) = 1 - x$. Then $m_1 = \sup\{x \mid x \in [0, 1]\} = 1$ and $m_2 = \sup\{1 - x \mid x \in [0, 1]\} = 1$. So $m_1 + m_2 = 2$. But $f(x) = f_1(x) + f_2(x) = x + (1 - x) = 1$ for all x. So $m = \sup\{f(x) \mid x \in [0, 1]\} = 1$. This is strictly less than 2. ♦

◇ **4E-16.** Consider a function $f : A \times B \to \mathbb{R}^m$, where $A \subset \mathbb{R}$, $B \subset \mathbb{R}^p$. Call f *separately continuous* if for each fixed $x_0 \in A$, the map $g(y) = f(x_0, y)$ is continuous and for $y_0 \in B$, $h(x) = f(x, y_0)$ is continuous. Say that f is continuous on A *uniformly with respect to* B if for each $\varepsilon > 0$ and $x_0 \in A$, there is a $\delta > 0$ such that $\| x - x_0 \| < \delta$ implies $\| f(x, y) - f(x_0, y) \| < \varepsilon$ for all $y \in B$. Show that if f is separately continuous and is continuous on A uniformly with respect to B, then f is continuous.

Solution. Let $\varepsilon > 0$ and $(x_0, y_0) \in A \times B \subseteq \mathbb{R} \times \mathbb{R}^2$. Since f is separately continuous in x and y, there is a $\delta_1 > 0$ such that $\| y - y_0 \| <$

$\delta_1 \implies \| f(x_0, y) - f(x_0, y_0) \| < \varepsilon/2$. Since f is continuous on A uniformly with respect to B, there is a $\delta_2 > 0$ such that $|x - x_0| < \delta_2 \implies \| f(x, y) - f(x_0, y) \| < \varepsilon/2$ for every $y \in B$. Let $\delta = \min(\delta_1, \delta_2)$.

If $\| (x, y) - (x_0, y_0) \| < \delta$, then

$$\| x - x_0 \| \le \sqrt{(x - x_0)^2 + \| y - y_0 \|^2} = \| (x, y) - (x_0, y_0) \| < \delta \le \delta_1,$$
$$\| y - y_0 \| \le \sqrt{(x - x_0)^2 + \| y - y_0 \|^2} = \| (x, y) - (x_0, y_0) \| < \delta \le \delta_2.$$

So

$$\begin{aligned}\| f(x, y) - f(x_0, y_0) \| &= \| f(x, y) - f(x_0, y) + f(x_0, y) - f(x_0, y_0) \| \\ &\le \| f(x, y) - f(x_0, y) \| + \| f(x_0, y) - f(x_0, y_0) \| \\ &\le \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

Thus f is continuous on $A \times B$ as claimed. ♦

◇ **4E-17.** Demonstrate that multilinear maps on Euclidean space are not necessarily uniformly continuous.

Suggestion. Let $f(x, y) = xy$. Show that f is not uniformly continuous by showing that $|f(a, a) - f(b, b)| = (|a + b| / \sqrt{2}) \| (a, a) - (b, b) \|$. ◇

Solution. For $(x, y) \in \mathbb{R} \times \mathbb{R}$, let $f(x, y) = xy \in \mathbb{R}$. Then f is bilinear since

$$f(as + bt, y) = (as + bt)y = ast + bty = af(s, y) + bf(t, y) \quad \text{and}$$
$$f(x, as + bt) = x(as + bt) = xas + xbt = axs + bxt = af(x, s) + bf(x, t).$$

To show that f is not uniformly continuous, let $\varepsilon = 2$, and suppose $\delta > 0$. Pick an integer n with $0 < 2/n < \delta$. If we let $a = n$ and $b = n + (1/n)$, then $\| (a, a) - (b, b) \| = \sqrt{(a - b)^2 + (a - b)^2} = \sqrt{2}\, |a - b| = \sqrt{2}/n < \delta$. But, for the images, $|f(a, a) - f(b, b)| = \left| n^2 - (n + (1/n))^2 \right| = 2 + (1/n^2) > 2$. So, no matter how small a positive δ is specified, we can always find points closer together than δ whose images are farther apart than 2. So f is not uniformly continuous on $\mathbb{R} \times \mathbb{R}$. ♦

◇ **4E-18.** Let $A \subset M$ be connected and let $f : A \to \mathbb{R}$ be continuous with $f(x) \ne 0$ for all $x \in A$. Show that $f(x) > 0$ for all $x \in A$ or else $f(x) < 0$ for all $x \in A$.

Solution. If there were a points x_1 and x_2 with $f(x_1) < 0$ and $f(x_2) > 0$, then by the intermediate value theorem, 4.5.1, there would be a point z in A with $f(z) = 0$. By hypothesis, this does not happen. So $f(x)$ must have the same sign for all x in A. ♦

◇ **4E-19.** Find a continuous map $f : \mathbb{R}^n \to \mathbb{R}^m$ and a closed set $A \subset \mathbb{R}^n$ such that $f(A)$ is not closed. In fact, do this when $f : \mathbb{R}^2 \to \mathbb{R}$ is the projection on the x–axis.

Sketch. $A = \{(x, y) \in \mathbb{R}^2 \mid xy = 1\}$; $f(A) = \{x \in \mathbb{R} \mid x \neq 0\}$. ◊

Solution. For $(x, y) \in \mathbb{R}^2$, let $f(x, y) = \pi_1(x, y) = x$. Then f is a continuous map from $\mathbb{R}^2$ to $\mathbb{R}$ (Exercise 4.1-1(b)). Let $A = \{(x, y) \in \mathbb{R}^2 \mid xy = 1\}$. This hyperbola is a closed set in $\mathbb{R}^2$, but $f(A) = \mathbb{R}\backslash\{0\}$ which is not closed.

For a second example, let $A \subseteq \mathbb{R}^2$ be the graph of $y = \tan x$ for $-\pi/2 < x < \pi/2$. The curve A is a closed set in $\mathbb{R}^2$, but $f(A)$ is the open interval $]-\pi/2, \pi/2[$ which is not closed. ♦

◇ **4E-20.** Give an alternative proof of Theorem 4.4.1 using sequential compactness.

Solution. We are asked to use sequential compactness to prove the maximum-minimum theorem, 4.4.1.

Theorem. *Let (M, d) be a metric space, $A \subseteq m$, and $f : A \to \mathbb{R}$ be continuous. Let $K \subseteq A$ be compact. Then f is bounded on K. That is, $B = \{f(x) \mid x \in K\} \subseteq \mathbb{R}$ is a bounded set. Furthermore, there exist points $x_0, x_1 \in K$ such that $f(x_0) = \inf(B)$ and $f(x_1) = \sup(B)$.*

First we want to show that $B = f(K)$ is a bounded set. If it were not, the we could pick a sequence of points $\langle y_n \rangle_1^\infty$ in $f(K)$ with $|y_n| > n$ for each n. For each n there is a point $x_n \in K$ with $f(x_n) = y_n$. Since K is sequentially compact, there should be a convergent subsequence $x_{n(k)} \to x \in K$ as $k \to \infty$. Since f is continuous on K we should have $f(x_{n(k)}) \to f(x)$. But this leads to trouble since $\left|f(x_{n(k)})\right| = \left|y_{n(k)}\right| > n(k) \to \infty$ as $k \to \infty$. So $|f(x)|$ would have to be infinite. This is not possible, so $f(K)$ must be a bounded set.

Since $f(K)$ is a bounded subset of $\mathbb{R}$, $a = \inf(f(K))$ and $b = \sup(f(K))$ exist as finite real numbers. There must be y_n and z_n in $f(K)$ with

$$a \leq y_n < a + \frac{1}{n} \qquad \text{and} \qquad b - \frac{1}{n} < z_n \leq b.$$

Since these are in $f(K)$, there must be points s_n and t_n in K with $f(s_n) = y_n$ and $f(t_n) = z_n$. Since K is sequentially compact, there must be subsequences converging to points in K

$$s_{n(k)} \to x_0 \in K \qquad \text{and} \qquad t_{n(k)} \to x_1 \in K$$

as $k \to \infty$. Since f is continuous on K, we have

$$y_{n(k)} = f(s_{n(k)}) \to f(x_0) \in K \qquad \text{and} \qquad z_{n(k)} = f(t_{n(k)}) \to f(x_1) \in K$$

as $k \to \infty$. But $y_{n(k)} \to a$ and $z_{n(k)} \to b$. Since limits are unique, we must have $a = f(x_0)$ and $b = f(x_1)$ as desired. ♦

◇ **4E-21.** Which of the following functions on $\mathbb{R}$ are uniformly continuous?

(a) $f(x) = 1/(x^2+1)$.
(b) $f(x) = \cos^3 x$.
(c) $f(x) = x^2/(x^2+2)$.
(d) $f(x) = x \sin x$.

Answer. (a) Yes.
(b) Yes.
(c) Yes.
(d) No. ◇

Solution. Parts (a), (b), and (c) can all be handled using Example 4.6.4. Each is differentiable everywhere on $\mathbb{R}$.

(a) If $f(x) = 1/(x^2+1)$, then $f'(x) = -2x/(x^2+1)^2$. The denominator is always at least as large as 1. So, if $|x| \le 1$, then $|f'(x)| \le 2$. If $|x| > 1$, then $|f'(x)| \le \left|2x/x^4\right| = 2/\left|x^3\right| < 2$. So $|f'(x)| \le 2$ for all $x \in \mathbb{R}$, and f is uniformly continuous on $\mathbb{R}$ by Example 4.6.4.
(b) If $f(x) = \cos^3 x$, then $|f'(x)| = \left|-3\cos^2 x \sin x\right| \le 3$ for all real x since $|\sin x| \le 1$ and $|\cos x| \le 1$ for all real x. Again, f is uniformly continuous on $\mathbb{R}$ by Example 4.6.4.
(c) If $f(x) = x^2/(x^2+2)$, then

$$|f'(x)| = \left|\frac{4x}{x^4+4x^2+4}\right| \le \left|\frac{x}{x^2+1}\right|.$$

The denominator is always at least as large as 1, so, if $|x| \le 1$, then $|f'(x)| \le 1$. If $x > 1$, then $|f'(x)| \le \left|x/x^2\right| = 1/|x| < 1$. So $|f'(x)| \le 1$ for all $x \in \mathbb{R}$, and f is uniformly continuous on $\mathbb{R}$ by Example 4.6.4.
(d) The function $f(x) = x \sin x$ is continuous on $\mathbb{R}$ but not absolutely continuous. Let $\varepsilon = 1$. We will show that no $\delta > 0$ can satisfy the definition of uniform continuity everywhere in $\mathbb{R}$. Let a be any number with $0 < a < \min(\delta, \pi/4)$. Then $\sin a > 0$. Pick an integer k large enough so that $2k\pi \sin a > 1$. Let $x = 2k\pi$ and $y = 2k\pi + a$. Then $|x - y| = a < \delta$, but

$$\begin{aligned} |f(y) - f(x)| &= |(2k\pi + a)\sin(2k\pi + a) - \sin(2k\pi)| \\ &= (2k\pi + a)\sin(a) \\ &> 1. \end{aligned}$$

No single $\delta > 0$ can work for $\varepsilon = 1$ in the definition of uniform continuity everywhere in $\mathbb{R}$. So f is not uniformly continuous on $\mathbb{R}$. ♦

◇ **4E-22.** Give an alternative proof of the uniform continuity theorem using the Bolzano-Weierstrass Theorem as follows. First, show that if f is not uniformly continuous, there is an $\varepsilon > 0$ and there are sequences x_n, y_n such that $\rho(x_n, y_n) < 1/n$ and $\rho\left(f(x_n), f(y_n)\right) \geq \varepsilon$. Pass to convergent subsequences and obtain a contradiction to the continuity of f.

Solution. We are asked to prove that a continuous function on a compact set is uniformly continuous on that set by using the Bolzano-Weierstrass Theorem which says that a compact set is sequentially compact. So, suppose K is a compact subset of a metric space M with metric d and f is a continuous function from K into a metric space N with metric ρ. If f were not uniformly continuous, then there would be an $\varepsilon > 0$ for which no $\delta > 0$ would work in the definition of uniform continuity. In particular, $\delta = 1/n$ would not work. So there would be points x_n and y_n with $d(x_n, y_n) < 1/n$ and $\rho(f(x_n), f(y_n)) > \varepsilon$. Since K is a compact subset of the metric space M, it is sequentially compact by the Bolzano-Weierstrass Theorem (3.1.3). So there are indices $n(1) < n(2) < n(3) < \ldots$ and a point $z \in K$ such that $x_{n(k)} \to z$ as $k \to \infty$. Since $n(k) \to \infty$ and $d(x_{n(k)}, y_{n(k)})) < 1/n(k)$, we can compute

$$d(y_{n(k)}, z) \leq d(y_{n(k)}, x_{n(k)}) + d(x_{n(k)}, z) < \frac{1}{n(k)} + d(x_{n(k)}, z) \to 0.$$

So $y_{n(k)} \to z$ also. Since f is continuous on K, we should have $f(x_{n(k)}) \to f(z)$ and $f(y_{n(k)}) \to f(z)$ as $k \to \infty$. We can select k large enough so that $\rho(f(x_{n(k)}), f(z)) < \varepsilon/2$ and $\rho(f(y_{n(k)}), f(z)) < \varepsilon/2$. But this would give

$$\begin{aligned} \varepsilon &< \rho(f(x_{n(k)}), f(y_{n(k)})) \\ &\leq \rho(f(x_{n(k)}), f(z)) + \rho(f(y_{n(k)}), f(z)) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

This impossibility shows that f must, in fact, be uniformly continuous on K. ♦

◇ **4E-23.** Let X be a compact metric space and $f : X \to X$ an isometry; that is, $d\left(f(x), f(y)\right) = d(x, y)$ for all $x, y \in X$. Show that f is a bijection.

Sketch. To show "onto" suppose $y_1 \in \mathcal{X} \backslash f(\mathcal{X})$ and consider the sequence $y_2 = f(y_1)$, $y_3 = f(y_2)$, $\ldots$. ◇

Solution. If $f(x) = f(y)$, then $0 = d(f(x), f(y)) = d(x, y)$, so $x = y$. Thus f is one-to-one. If $\varepsilon > 0$, let $\delta = \varepsilon$. If $d(x, y) < \delta$, then $d(f(x), f(y)) = d(x, y) < \delta = \varepsilon$, so f is continuous, in fact uniformly continuous, on $\mathcal{X}$. It remains to show that f maps $\mathcal{X}$ onto $\mathcal{X}$. Since $\mathcal{X}$ is compact and f is continuous, the image, $f(\mathcal{X})$ is a compact subset of the metric space $\mathcal{X}$. So it must be closed. Its complement, $\mathcal{X} \setminus f(\mathcal{X})$ must be open. If there were a point x in $\mathcal{X} \setminus f(\mathcal{X})$, then there would be a radius $r > 0$ such that $D(x, r) \subseteq \mathcal{X} \setminus f(\mathcal{X})$. That is, $y \in f(\mathcal{X})$ implies $d(y, x) > r$. Consider the sequence defined by $x_0 = x$ and $x_{n+1} = f(x_n)$ for $n = 0, 1, 2, \ldots$. For positive integer k, let f^k denote the composition of f with itself k times. If n and p are positive integers, then

$$d(x_{n+p}, x_n) = d(f^n \circ f^p(x), f^n(x)) = d(f^p(x), x) > r$$

since $f^p(x) \in f(\mathcal{X})$. The points in the sequence are pairwise separated by distances of at least r. This would prevent any subsequence from converging. But $\mathcal{X}$ is sequentially compact by the Bolzano-Weierstrass Theorem. So there should be a convergent subsequence. This contradiction shows that there can be no such starting point x for our proposed sequence. The complement $\mathcal{X} \setminus f(\mathcal{X})$ must be empty. So $f(\mathcal{X}) = \mathcal{X}$ and f maps $\mathcal{X}$ onto $\mathcal{X}$ as claimed. ♦

◇ **4E-24.** Let $f : A \subset M \to N$.

(a) Prove that f is uniformly continuous on A iff for every pair of sequences x_k, y_k of A such that $d(x_k, y_k) \to 0$, we have $\rho\left(f(x_k), f(y_k)\right) \to 0$.

(b) Let f be uniformly continuous, and let x_k be a Cauchy sequence of A. Show that $f(x_k)$ is a Cauchy sequence.

(c) Let f be uniformly continuous and N be complete. Show that f has a unique extension to a continuous function on $\mathrm{cl}(A)$.

Solution. (a) Suppose f is uniformly continuous on A and that $\langle x_n \rangle_1^\infty$ and $\langle y_n \rangle_1^\infty$ are sequences in A such that $d(x_n, y_n) \to 0$. Let $\varepsilon > 0$. Since f is uniformly continuous on A, there is a $\delta > 0$ such that $\rho(f(x), f(y)) < \varepsilon$ whenever x and y are in A and $d(x, y) < \delta$. Since $d(x_n, y_n) \to 0$, there is a K such that $d(x_k, y_k) < \delta$ whenever $k \geq K$. So

$$k \geq K \implies d(x_k, y_k) < \delta \implies \rho(f(x_k), f(y_k)) < \varepsilon.$$

So $\rho(f(x_k), f(y_k)) \to 0$.

Now suppose $\rho(f(x_k), f(y_k)) \to 0$ for every pair of sequences $\langle x_k \rangle_1^\infty$ and $\langle y_k \rangle_1^\infty$ in A with $d(x_k, y_k) \to 0$. If f were not uniformly continuous on A, there would be an $\varepsilon > 0$ for which no $\delta > 0$ would work in the definition of uniform continuity on A. In particular, $\delta = 1/k$ would fail for each positive integer k. There would be points x_k and y_k in A with $d(x_k, y_k) < 1/k$ but $\rho(f(x_k), f(y_k)) > \varepsilon$. We would have $d(x_k, y_k) \to 0$ but $\rho(f(x_k), f(y_k)) \not\to 0$. By hypothesis this should not happen. So f must be uniformly continuous on A.

(b) Suppose f is uniformly continuous on A and that $\langle x_n \rangle_1^\infty$ is a Cauchy sequence in A. Let $\varepsilon > 0$. Since f is uniformly continuous on A, there is a $\delta > 0$ such that $\rho(f(x), f(y)) < \varepsilon$ whenever x and y are in A and $d(x, y) < \delta$. Since $\langle x_n \rangle_1^\infty$ is a Cauchy sequence in A, there is an index K such that $d(x_k, x_j) < \delta$ whenever $k \geq K$ and $j \geq K$. We have

$$(k \geq K \text{ and } j \geq K) \implies d(x_k x_j) < \delta \implies \rho(f(x_k), f(x_j)) < \varepsilon.$$

So $\langle f(x_k) \rangle_{k=1}^\infty$ is a Cauchy sequence in N as claimed.

(c) We are asked to prove the following theorem about the extension of functions.

Theorem. *If f is a uniformly continuous function from a subset A of a metric space M into a complete metric space N, then f has a unique continuous extension to* $\mathrm{cl}(A)$. *That is, there is a unique function $g : \mathrm{cl}(A) \to N$ such that G is continuous on* $\mathrm{cl}(A)$ *and $g(x) = f(X)$ for every x in A.*

Proof: It is easy to see that there can be no more than one such function. If $x \in \mathrm{cl}(A)$, then there is a sequence $\langle x_k \rangle_1^\infty$ of points in A such that $x_k \to x$. If g is to be continuous on $\mathrm{cl}(A)$, we must have $g(x_k) \to g(x)$. But g is to be an extension of f and each of the points x_k is in A. So $g(x_k) = f(x_k)$ for each k. So, the only possible value for $g(x)$ is $\lim_{k\to\infty} f(x_k)$. In order to use this to define the function g we need to show that the limit exists and that it does not depend on what sequence is selected from A converging to x.

ONE: *If $\langle x_k \rangle_1^\infty$ is a sequence in A and $x_k \to x \in \mathrm{cl}(A)$, then there is a point $z \in N$ with $f(x_k) \to z$.*

Pf: Since $x_k \to x$ in M, the sequence $\langle x_k \rangle_1^\infty$ is a Cauchy sequence in M. By the result of part (b), $\langle f(x_k) \rangle_{k=1}^\infty$ is a Cauchy sequence in N. Since N is complete, there is a point z in N with $f(x_k) \to z$ as claimed.

TWO: *If $\langle x_k \rangle_1^\infty$ and $\langle y_k \rangle_1^\infty$ are sequences in A with $x_k \to x$ and $y_k \to x$ in* $\mathrm{cl}(A)$, *then* $\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} f(y_k)$ *in N.*

Pf: Both the limits exist by ONE. Let $a = \lim_{k\to\infty} f(x_k)$ and $b = \lim_{k\to\infty} f(y_k)$. We want to show that $a = b$. Let $\varepsilon > 0$. Pick K_1 and K_2 such that

$$k \geq K_1 \implies \rho(f(x_k), a) < \frac{\varepsilon}{3} \quad \text{and} \quad k \geq K_2 \implies \rho(f(y_k), b) < \frac{\varepsilon}{3}.$$

Since f is uniformly continuous on A, we can pick $\delta > 0$ such that $\rho(f(x), f(y)) < \varepsilon/3$ whenever x and y are in A and $d(x, y) < \delta$. Pick K_3 and K_4 such that

$$k \geq K_3 \implies d(x_k, x) < \frac{\delta}{2} \quad \text{and} \quad k \geq K_4 \implies d(y_k, x) < \frac{\delta}{2}.$$

Fix an index k larger than $max(K_1, K_2, K_3, K_4)$. Then $d(x_k, y_k) \leq d(x_k, x) + d(x, y_k) < \delta/2 + \delta/2 = \delta$. So $\rho(f(x_k)f(y_k)) < \varepsilon/3$ and

$$\rho(a,b) \leq \rho(a, f(x_k)) + \rho(f(x_k)f(y_k)) + \rho(f(y_k), b) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Since this is true for every $\varepsilon > 0$, we must have $\rho(a,b) = 0$. So $a = b$. This completes the proof of the second claim.

Using the two results just established, we can unambiguously define a function $g : \mathrm{cl}(A) \to N$ by the following process: For $x \in \mathrm{cl}(A)$,

(1) Select a sequence $\langle x_k \rangle_1^\infty$ in A with $x_k \to x$.

(2) Set $g(x) = \lim_{k\to\infty} f(x_k)$.

The limit in step (2) exists by the first claim and is independent of which particular sequence is used to define it by the second. Thus this process does define a function on $\mathrm{cl}(A)$. If $x \in A$, then the defining sequence could have been selected as $x_k = x$ for all k. For this sequence we certainly have $f(x_k) \to f(x)$. So $g(x) = f(x)$ for $x \in A$. Thus g is an extension of f to $\mathrm{cl}(A)$. It remains to show that g is continuous on $\mathrm{cl}(A)$.

Suppose $x \in \mathrm{cl}(A)$ and $\langle z_k \rangle_1^\infty$ is a sequence in $\mathrm{cl}(A)$ with $z_k \to x$. We need to tie the sequence $\langle z_k \rangle_1^\infty$ to a closely related sequence actually in A. Since $g(z_k)$ is defined as the limit of the images of any sequence in A converging to z_k, we can, for each positive integer k, select a point $x_k \in A$ with $d(x_k, z_k) < 1/k$ and $\rho(f(x_k), g(z_k)) < 1/k$. Since the z_k converge to x and the x_k are progressively closer to the z_k, we also have $x_k \to x$. More precisely:

$$d(x_k, x) \leq d(x_k, z_k) + d(z_k, x) < \frac{1}{k} + d(z_k, x) \to 0.$$

Since $\langle x_k \rangle_1^\infty$ is a sequence of points in A converging to x, we have $f(x_k) \to g(x)$. But then

$$\begin{aligned}\rho(g(z_k), g(x)) &\leq \rho(g(z_k), f(x_k)) + \rho(f(x_k), g(x)) \\ &\leq \frac{1}{k} + \rho(f(x_k, g(x)) \\ &\to 0.\end{aligned}$$

So $g(z_k) \to g(x)$. This works for every sequence $\langle z_k \rangle_1^\infty$ in $\mathrm{cl}(A)$ converging to x. So g is continuous at x. Since x was arbitrary in $\mathrm{cl}(A)$, the function g is continuous on the set $\mathrm{cl}(A)$.

There is actually something to be done here. Simply getting an extension to $\mathrm{cl}(A)$ is trivial. The extension could be defined in any manner desired on $\mathrm{cl}(A) \setminus A$. The trick is to get a *continuous* extension. As we have seen, there cannot be more than one. But there might not be any.

For an easy example, consider $A = [0,2] \setminus \{1\} \subseteq \mathbb{R}$. On A define f by

$$f(x) = \begin{cases} x, & \text{for } 0 \le x < 1 \\ x - 1, & \text{for } 1 < x \le 2 \end{cases}.$$

See the figure.

FIGURE 4-10. A function without continuous extension to the closure of the domain.

Then f is continuous on A. The closure of A is $[0,2]$. The limits as x tends to 1 from the two portions of A are different. The limit is 1 from the left and 0 from the right. So there is no way to consistently define the function at 1 in such a way as to make it continuous on the whole interval from 0 to 2.

For a less artificial example, represent a point in the plane by polar coordinates r and ϑ where $r \ge 0$ and $0 \le \vartheta < 2\pi$ with ϑ being the counterclockwise angle from the positive horizontal axis. Let $A = \mathbb{R}^2 \setminus$ nonnegative horizontal axis. Then the map $(r, \vartheta) \mapsto \vartheta$ is continuous on A, but there is no way to extend it continuously to the whole plane. It makes a jump of 2π as you cross the positive horizontal axis. ♦

◇ **4E-25.** Let $f : \,]0,1[\, \to \mathbb{R}$ be differentiable and let $f'(x)$ be bounded. Show that $\lim_{x\to 0^+} f(x)$ and $\lim_{x \to 1^-} f(x)$ exist. Do this both directly and by applying Exercise 4E-24(c). Give a counterexample if $f'(x)$ is not bounded.

Sketch. $f(x) = \sin(1/x)$ is a counterexample with f' not bounded. ◇

Solution. First notice that as in Example 4.6.4, if $|f'(x)| \le M$ for all x in $]0,1[$, then, if s and t are in $[0,1[$, the mean value theorem says that there is a point c between s and t where $|f(s) - f(t)| = |f'(c)(s-t)| \le M\,|s-t|$. Thus f satisfies a Lipschitz condition on $]0,1[$ and is uniformly continuous there.

Claim. *If $\langle s_k \rangle_1^\infty$ is a Cauchy sequence in $]0,1[$, then $\langle f(s_k) \rangle$ is a Cauchy sequence in $\mathbb{R}$.*

To see this, note that if $\varepsilon > 0$ there is an index K such that $|s_k - s_j| < \varepsilon/M$. So, for such k and j we have $|f(s_k) - f(s_j)| \le M\,|s_k - s_j| < M\varepsilon/M = \varepsilon$.

Suppose $\langle x_k \rangle_1^\infty$ is a sequence in $]0,1[$ increasing monotonically to a limit of 1. Then $\langle x_k \rangle_1^\infty$ is a Cauchy sequence in $]0,1[$, so $\langle f(x_k) \rangle$ is a Cauchy sequence in $\mathbb{R}$. There must be a number $\lambda \in \mathbb{R}$ with $f(x_k) \to \lambda$. Fix an index k with

$$1 - \frac{\varepsilon}{2M} < x_k < 1 \quad \text{and} \quad |f(x_k) - \lambda| < \frac{\varepsilon}{2}.$$

Put $\delta = \varepsilon/2M$. If $1 - \delta < t < 1$, then

$$\begin{aligned} |f(t) - \lambda| &\le |f(t) - f(x_k)| + |f(x_k) - \lambda| \\ &\le M\,|t - x_k| + \frac{\varepsilon}{2} < M\frac{\varepsilon}{2M} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

So $\lim_{t\to 1^-} f(t)$ exists and is equal to λ.

The argument at the other end is similar. Suppose $\langle y_k \rangle_1^\infty$ is a sequence in $]0,1[$ decreasing monotonically to a limit of 0. Then $\langle y_k \rangle_1^\infty$ is a Cauchy sequence in $]0,1[$, so $\langle f(y_k) \rangle$ is a Cauchy sequence in $\mathbb{R}$. There must be a number $\mu \in \mathbb{R}$ with $f(y_k) \to \mu$. Fix an index k with

$$0 < y_k < \frac{\varepsilon}{2M} \quad \text{and} \quad |f(y_k) - \mu| < \frac{\varepsilon}{2}.$$

Put $\delta = \varepsilon/2M$. If $0 < t < \delta$, then

$$\begin{aligned} |f(t) - \lambda| &\le |f(t) - f(y_k)| + |f(y_k) - \mu| \\ &\le M\,|t - y_k| + \frac{\varepsilon}{2} < M\frac{\varepsilon}{2M} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

So $\lim t \to 0^+ f(t)$ exists and is equal to μ.

Another approach is to appeal to the more general result of Exercise 4E-24(c). Since f is uniformly continuous on the open interval $]0,1[$, it has a continuous extension to the closure of that domain which is the closed interval $[0,1]$. There is a unique function g from $[0,1]$ into $\mathbb{R}$ which is continuous on $[0,1]$ and for which $g(x) = f(x)$ for all $x \in]0,1[$. Since g is continuous on $[0,1]$, we have

$$g(0) = \lim_{x\to 0^+} g(x) = \lim_{x\to 0^+} f(x) \quad \text{and} \quad g(1) = \lim_{x\to 1^-} g(x) = \lim_{x\to 1^-} f(x).$$

In particular, both of those limits exist.

The function $f(x) = \sin(1/x)$ is continuous and differentiable on the open interval $]0, 1[$. The derivative, $f'(x) = -(\sin(1/x))/x^2$, is not bounded on that interval. Indeed, $f'(2/((4k+1)\pi)) = -(4k+1)^2\pi^2/4$ for integer k. The limit as x tends to 0 does not exist. The function f takes on every value from -1 to 1 in every neighborhood of 0. ♦

◇ **4E-26.** Let $f :]a, b] \to \mathbb{R}$ be continuous and such that $f'(x)$ exists on $]a, b[$ and $\lim_{x\to a^+} f'(x)$ exists. Prove that f is uniformly continuous.

Suggestion. Use the limit of the derivative at a to get uniform continuity on a short interval $]a, a+2d]$. Use Theorem 4.6.2 to get uniform continuity on an overlapping interval $[a+d, b]$. Then combine the two results. ◊

Solution. Let $\varepsilon > 0$, and suppose $\lim_{x\to a^+} f(x) = \lambda$. There is a d such that $0 < 2d < b-a$ and $|f'(x)| \le |\lambda|+1$ for $a < x < a+2d$. As is Example 4.6.4, f is uniformly continuous on $]a, a+2d]$, and there is a $\delta_1 > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever x and y are in $]a, a+2d]$ and $|x-y| < \delta_1$. Since f is continuous on $]a, b]$, it is continuous on the subinterval $[a+d, b]$. Since that interval is compact, f is uniformly continuous on it by the uniform continuity theorem, 4.6.2. There is a $\delta_2 > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever x and y are in $[a+d, b]$ and $|x-y| < \delta_2$. Now we take advantage of the overlap we have carefully arranged between our two subdomains. If x and y are in $]a, b]$ and $|x-y| < \min(\delta_1, \delta_2, d/2)$, then either they are both in $]a, a+2d]$ or they are both in $[a+d, b]$, or both. If they are both in $]a, a+2d]$, then $|f(x) - f(y)| < \varepsilon$ since $|x-y| < \delta_1$. If they are both in $]a, a+2d]$, then $|f(x) - f(y)| < \varepsilon$ since $|x-y| < \delta_2$. In any case, $|f(x) - f(y)| < \varepsilon$ whenever x and y are in $]a, b]$ ans $|x-y| < \delta = \min(\delta_1, \delta_2, d/2)$. So f is uniformly continuous on $]a, b]$ as claimed. ♦

◇ **4E-27.** Find the sum of the series $\sum_{k=4}^{\infty}(3/4)^k$.

Answer. 81/64. ◊

Solution. Use the known sum for a geometric series with ratio smaller than 1 to compute

$$\begin{aligned}
\sum_{k=4}^{\infty}\left(\frac{3}{4}\right)^k &= \left(\frac{3}{4}\right)^4 + \left(\frac{3}{4}\right)^5 + \left(\frac{3}{4}\right)^6 + \dots \\
&= \left(\frac{3}{4}\right)^4 \left(1 + \left(\frac{3}{4}\right) + \left(\frac{3}{4}\right)^2 + \dots\right) \\
&= \left(\frac{3}{4}\right)^4 \frac{1}{1-(3/4)} = \left(\frac{3}{4}\right)^3 \frac{1}{4-3} = \frac{3^4}{4^3} = \frac{81}{64}.
\end{aligned}$$

$$\boxed{\sum_{k=4}^{\infty}\left(\frac{3}{4}\right)^k = \frac{81}{64}.}$$ ♦

◇ **4E-28.** Let $f :]0,1[\to \mathbb{R}$ be uniformly continuous. Must f be bounded?

Answer. Yes. ♢

Solution. If f were not bounded on $]0,1[$, we could inductively select a sequence of points $\langle x_k \rangle_1^\infty$ in $]0,1[$ such that $|f(x_{k+1})| > |f(x_k)| + 1$ for each k. In particular, we would have $|f(x_k) - f(x_j)| > 1$ whenever $k \neq j$. But the points x_k are all in the compact interval $[0,1]$, so there should be a subsequence converging to some point in $[0,1]$. This subsequence would have to be a Cauchy sequence, so no matter how small a positive number δ were specified, we could get points x_k and x_j in the subsequence with $|x_k - x_j| < \delta$ and $|f(x_k) - f(x_j)| > 1$. This contradicts the uniform continuity of f on $]0,1[$. So the image $f(]0,1[)$ must, in fact, be bounded.

If we knew the result of Exercise 4E-24(c), then we would know that f has a unique continuous extension to the closure, $[0,1]$. There is a continuous $g : [0,1] \to \mathbb{R}$ such that $g(x) = f(x)$ for all x in $]0,1[$. Since g is continuous on the compact domain $[0,1]$, the image $g([0,1])$ is compact and hence bounded. So $f(]0,1[) = g(]0,1[) \subseteq g([0,1])$ is bounded. ♦

◇ **4E-29.** Let $f : \mathbb{R} \to \mathbb{R}$ satisfy $|f(x) - f(y)| \leq |x-y|^2$. Prove that f is a constant. [Hint: What is $f'(x)$?]

Suggestion. Divide by $x - y$ and let y tend to x to show that $f'(x) = 0$. ♢

Solution. Suppose $x_0 \in \mathbb{R}$. Then for $x \neq x_0$ we have $|f(x) - f(x_0)| \leq |x - x_0|^2$, so

$$0 \leq \left| \frac{f(x) - f(x_0)}{x - x_0} \right| \leq |x - x_0| \, .$$

Letting $x \to x_0$, we find that $\lim_{x \to x_0} (f(x) - f(x_0))/(x - x_0) = 0$. So $f'(x_0)$ exists and is equal to 0 for every $x_0 \in \mathbb{R}$. If $t \in \mathbb{R}$, there is, by the mean value theorem, a point x_0 between 0 and t such that $|f(t) - f(0)| = |f'(x_0)(t-0)| = |0(t-0)| = 0$. So $f(t) = f(0)$ for all $t \in \mathbb{R}$. Thus f is a constant function as claimed. ♦

◇ **4E-30.** (a) Let $f : [0, \infty[\to \mathbb{R}$, $f(x) = \sqrt{x}$. Prove that f is uniformly continuous.

(b) Let $k > 0$ and $f(x) = (x - x^k)/\log x$ for $0 < x < 1$ and $f(0) = 0$, $f(1) = 1 - k$. Show that $f : [0,1] \to \mathbb{R}$ is continuous. Is f uniformly continuous?

Suggestion. (a) Use Theorem 4.6.2 to show that f is uniformly continuous on $[0,3]$ and Example 4.6.4 to show that it is uniformly continuous on $[1,\infty[$. Then combine these results.

(b) Use L'Hôpital's Rule. ◊

Solution. (a) Let $\varepsilon > 0$. We know that $f(x) = \sqrt{x}$ is continuous on $[0,\infty[$, so it is certainly continuous on the compact domain $[0,3]$. By the uniform continuity theorem, 4.6.2, it is uniformly continuous on that set. There is a $\delta_1 > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever x and y are in $[0,3]$ and $|x - y| < \delta_1$.

We also know that f is differentiable for $x > 0$ with $f'(x) = 1/(2\sqrt{x})$. So $|f'(x)| \leq 1/2$ for $x \geq 1$. As in Example 4.6.4, we can use the mean value theorem to conclude that if $\delta_2 = 2\varepsilon$, and x and y are in $[1,\infty[$ with $|x - y| < \delta_2$, then there is a point c between x and y such that $|f(x) - f(y)| = |f'(c)(x - y)| < (1/2)(2\varepsilon) = \varepsilon$.

Now take advantage of the overlap of our two domains. If x and y are in $[0,\infty[$ and $|x - y| < \delta = \min(1, \delta_1, \delta_2)$, then either x and y are both in $[0,3]$ or both are in $[1,\infty[$ or both. If they are both in $[0,3]$, then $|f(x) - f(y)| < \varepsilon$ since $|x - y| < \delta_1$. If they are both in $[1,\infty[$, then $|f(x) - f(y)| < \varepsilon$ since $|x - y| < \delta_2$. In either case, $|f(x) - f(y)| < \varepsilon$. So f is uniformly continuous on $[0,\infty[$ as claimed.

(b) Suppose k is a positive integer and $f(x) = (x - x^k)/\log x$ for $0 < x < 1$, $f(0) = 0$, and $f(1) = 1 - k$. The numerator, $x - x^k$, is continuous for all x. The denominator, $\log x$, is continuous for $x > 0$. So f is continuous on $x > 0$ except possibly at $x = 1$ where the denominator is 0, However, the numerator is also 0 at $x = 1$. To apply L'Hôpital's Rule, we consider the ratio of the derivatives

$$\frac{1 - kx^{k-1}}{1/x} = x - kx^k \to 1 - k = f(1) \quad \text{as} \quad x \to 1.$$

By L'Hôpital's Rule, $\lim_{x\to 1}(x - x^k)/\log x = \lim_{x\to 1} f(x)$ also exists and is equal to $f(1)$. So f is continuous at 1. As $x \to 0^+$, the numerator of $f(x)$ tends to 0 and the denominator to $-\infty$. So $\lim_{x\to 0^+} f(x) = 0 = f(0)$. So f is continuous from the right at 0. So f is continuous on $[0,\infty[$ and on the smaller domain $[0,1]$. Since the latter is compact, f is uniformly continuous on it by the uniform continuity theorem, 4.6.2. ♦

$\diamond$ **4E-31.** Let $f(x) = x^{1/(x-1)}$ for $x \neq 1$. How should $f(1)$ be defined to make f continuous at $x = 1$?

Answer. $f(1) = e$. $\Diamond$

Solution. For $x \neq 1$, $f(x) = x^{1/x-1} = e^{(\log x)/(x-1)}$. This is continuous on $x > 0$ except possibly at $x = 1$. We have $\log f(x) = (\log x)/(x-1)$. Both numerator and denominator tend to 0 as x approaches 1, so we can try L'Hôpital's Rule. For the ratio of derivatives we have

$$\frac{1/x}{1} = \frac{1}{x} \to 1 \quad \text{as} \quad x \to 1.$$

By L'Hôpital's Rule, $\lim_{x\to 1}(\log x)/(x-1)$ also exists and is 1. So

$$\lim_{x\to 1} f(x) = \lim_{x\to 1} e^{(\log x)/(x-1)} \text{ exists and is } e^1.$$

So, if we put $f(1) = e$, f will be continuous at 1. $\blacklozenge$

$\diamond$ **4E-32.** Let $A \subset \mathbb{R}^n$ be open, $x_0 \in A$, $r_0 > 0$ and $B_{r_0} = \{x \in \mathbb{R}^n \mid \| x - x_0 \| \le r_0\}$. Suppose that $B_{r_0} \subset A$. Prove that there is an $r > r_0$ such that $B_r \subset A$.

Solution. The closed disk B_{r_0} is a closed bounded set in $\mathbb{R}^n$ and so is compact. Since A is open, there is for each $x \in A$ a radius $\rho(x) > 0$ such that $D(x, \rho(x)) \subseteq A$, and we surely have $B_{r_0} \subseteq \bigcup\{D(x, \rho(x)/2) \mid x \in B_{r_0}\}$. This open cover of the compact set B_{r_0} must have a finite subcover, so there are points $x_1, x_2, \ldots, x_N$ in B_{r_0} with $B_{r_0} \subseteq \bigcup_{k=1}^N D(x_k, \rho(x_k)/2)$. Let $\rho = \min\{\rho(x_k)/2 \mid 1 \le k \le N\}$ and $r = r_0 + \rho$. If $y \in B_r$, then $\| y - x_0 \| \le r_0 + \rho$. Since $r_0 < r$, $B_{r_0} \subseteq B_r$. So if $\| y - x_0 \| \le r_0$, then y is certainly in A. If not, then on the straight line between y and x_0 there is a point $x \in B_{r_0}$ with $\| x - x_0 \| = r_0$ and $\| y - x \| \le \rho$. There is an index k with $x \in D(x_k, \rho(x_k)/2)$. So

$$\| y - x_k \| \le \| y - x \| + \| x - x_k \| \le \rho + \| x - x_k \| < \rho + \frac{\rho(x_k)}{2} \le \rho(x_k).$$

So $y \in A$. Thus $B_r \subseteq A$ as required. See Figure 4-11.

$\blacklozenge$

$\diamond$ **4E-33.** A set $A \subset \mathbb{R}^n$ is called *relatively compact* when $\mathrm{cl}(A)$ is compact. Prove that A is relatively compact iff every sequence in A has a subsequence that converges to a point in $\mathbb{R}^n$.

Figure 4-11. An open set in $\mathbb{R}^n$ containing a closed disk contains a bigger closed disk.

Sketch. If A is relatively compact, then $\operatorname{cl}(A)$ is compact. Convergent subsequences exist by the Bolzano-Weierstrass theorem. If every sequence in A has a subsequence convergent in $\mathbb{R}^n$, then start with a sequence in $\operatorname{cl}(A)$. Get a nearby sequence in A. Take a subsequence converging in $\mathbb{R}^n$, and show that the corresponding subsequence of the original sequence converges also, necessarily to a point in $\operatorname{cl}(A)$. ♢

Solution. If A is relatively compact, then $\operatorname{cl}(A)$ is compact. Every sequence in A is also in $\operatorname{cl}(A)$ and so must have a subsequence converging to some point in the sequentially compact set $\operatorname{cl}(A)$ by the Bolzano-Weierstrass theorem.

For the converse, suppose that every sequence in A has a subsequence converging in $\mathbb{R}^n$. To show that $\operatorname{cl}(A)$ is compact, we will show that it is sequentially compact. Let $\langle y_k \rangle_1^\infty$ be a sequence in $\operatorname{cl}(A)$. For each k we can select a point x_k in A with $\| x_k - y_k \| < 1/k$. By hypothesis, the sequence $\langle x_k \rangle_1^\infty$ has a subsequence $\langle x_{k(j)} \rangle_1^\infty$ converging to some point $x \in \mathbb{R}^n$. Since each of the points x_k is in A, the limit x of the subsequence is in $\operatorname{cl}(A)$. If $\varepsilon > 0$, we can pick an index J large enough so that $\left\| x - x_{n(j)} \right\| < \varepsilon/2$ and $1/k(j) < \varepsilon/2$ whenever $j \geq J$. For such j we have

$$\left\| y_{k(j)} - x \right\| \leq \left\| y_{k(j)} - x_{k(j)} \right\| + \left\| x_{k(j)} - x \right\| < \frac{1}{k(j)} + \frac{\varepsilon}{2} < \varepsilon.$$

So $y_{k(j)} \to x \in \operatorname{cl}(A)$. Every sequence in $\operatorname{cl}(A)$ has a subsequence converging to a point in $\operatorname{cl}(A)$. So $\operatorname{cl}(A)$ is a sequentially compact subset of $\mathbb{R}^n$. So $\operatorname{cl}(A)$ is compact. ♦

◇ **4E-34.** Assuming that the temperature on the surface of the earth is a continuous function, prove that on any great circle of the earth there are two antipodal points with the same temperature.

Solution. View the great circle as a circle of radius R in the xy-plane. If t is a real number, then $(R\cos t, R\sin t)$ and $(R\cos(t+\pi), R\sin(t+\pi))$ are antipodal points (at opposite ends of a diameter). Let $f(t)$ be the temperature at $(R\cos t, R\sin t)$, and $g(t) = f(t) - f(t+\pi)$. We want to show that there is a t_0 with $g(t_0) = 0$. But $g(t+\pi) = f(t+\pi) - f(t+2\pi) = f(t+\pi) - f(t) = -g(t)$. If $g(t)$ is not zero, then $g(t)$ and $g(t+\pi)$ have opposite sign. Since f is continuous, so is g, and the intermediate value theorem guarantees a point t_0 at which $g(t_0) = 0$ just as we need. ♦

◇ **4E-35.** Let $f : \mathbb{R} \to \mathbb{R}$ be increasing and bounded above. Prove that the limit $\lim_{x\to 0^+} f(x)$ exists.

Sketch. $A = f(\{x \in \mathbb{R} \mid x > 0\})$ is nonempty and bounded below by $f(0)$. Show that $\lim_{x\to 0^+} f(x) = \inf(A)$. ◊

Solution. Since f is increasing, $f(x) \geq f(0)$ for all $x > 0$. So $A = f(\{x : x > 0\})$ is a nonempty subset of $\mathbb{R}$ which is bounded below by $f(0)$. By completeness, it has a finite greatest lower bound $a = \inf\{f(x) \in \mathbb{R} \mid x > 0\}$. If $\varepsilon > 0$, there is a point $\delta > 0$ with $a \leq f(\delta) < a + \varepsilon$. Since f is increasing, we conclude that $a \leq f(x) \leq f(\delta) < a + \varepsilon$ whenever $0 < x < \delta$. So $\lim_{x\to 0^+} f(x)$ exists and is equal to a. ♦

◇ **4E-36.** Show that $\{(x, \sin(1/x)) \mid x > 0\} \cup (\{0\} \times [-1, 1])$ in $\mathbb{R}^2$ is connected but not path-connected.

Solution. Let $A = \{(x, \sin(1/x)) \mid x > 0\}$ and $B = \{0\} \times [-1, 1] = \{(0, y) \mid -1 \leq y \leq 1\}$. Let $C = A \cup B$. We are asked to show that C is connected but not path-connected.

Each of A and B are path-connected, and so connected, subsets of $\mathbb{R}^2$. Suppose U and V were open sets with C contained in their union and $U \cap V \cap C = \emptyset$. If $A \cap U$ and $A \cap V$ were both nonempty, then U and V would disconnect A. But A is connected. So one of these must be empty. Similarly, one of $B \cap U$ and $B \cap V$ must be empty. Say $B \cap V = \emptyset$ and $B \subseteq U$. Since U is open and $(0, 0) \in U$, the point $(1/2\pi n, 0)$ is also in U for large enough integer n. But these points are in A. So $A \cap U$ is not empty. So $A \cap V$ is empty, and $A \subseteq U$. So $C = A \cup B \subseteq U$. The sets U and V cannot disconnect C. So C must be connected.

Suppose $\gamma : [0, 1] \to C$ were a continuous path with $\gamma(0) = (1/2\pi, 0)$ and $\gamma(1) = (0, 0)$. Since γ is continuous on the compact domain $[0, 1]$, it would be uniformly continuous, and there would be a $\delta > 0$ such that

$0 \leq s \leq t < \delta$ implies $\| \gamma(s) - \gamma(t) \| < 1/2$. But this gets us into trouble. As the path moves from $(1/2\pi, 0)$ to $(0, 0)$, it must pass through the points $v_n = (1/2\pi n, 0)$ and $w_n = (1/(2\pi n + (\pi/2)), 1)$. We could inductively select preimages $0 = s_1 < t_1 < s_2 < t_2 < s_3 < \cdots \to 1$ such that $\gamma(s_n) = v_n$ and $\gamma(t_n) = w_n$. For large enough n, both s_n and t_n are within δ of 1, so their images should be separated by less than 1/2. But $\| v_n - w_n \| \geq 1$. So there can be no such path. ♦

◇ **4E-37.** Prove the following intermediate value theorem for derivatives: If f is differentiable at all points of $[a, b]$, and if $f'(a)$ and $f'(b)$ have opposite signs, then there is a point $x_0 \in \,]a, b[$ such that $f'(x_0) = 0$.

Sketch. Suppose $f'(a) < 0 < f'(b)$. Since f is continuous on $[a, b]$ (why?), it has a minimum at some x_0 in $[a, b]$. (Why?) $x_0 \neq a$ since $f(x) < f(a)$ for x slightly larger than a. (Why?) $x_0 \neq b$ since $f(x) < f(b)$ for x slightly smaller that b. (Why?) So $a < x_0 < b$. So $f'(x_0) = 0$. (Why?) The case of $f'(a) > 0 > f'(b)$ is similar. ◇

Solution. We know from Proposition 4.7.2 that f is continuous on the compact domain $[a, b]$. By the maximum-minimum theorem, 4.4.1, it attains a finite minimum, m, and a finite maximum, M, at points x_1 and x_2 in $[a, b]$.

CASE ONE: $\underline{f'(a) < 0 < f'(b)}$: Since $f'(b) > 0$, and it is the limit of the difference quotients at b, we must have $(f(x) - f(b))/(x - b) > 0$ for x slightly smaller than b. Since $x - b < 0$, we must have $f(x) < f(b)$ for such x. So the minimum doe not occur at b. Since $f'(a) < 0$, and it is the limit of the difference quotients at a, we must have $(f(x) - f(a))/(x - a) < 0$ for x slightly larger than a. Since $x - a > 0$, we must have $f(x) < f(a)$ for such x. So the minimum does not occur at a. Thus the minimum must occur at a point $x_0 \in]a, b[$. By Proposition 4.7.9, we must have $f'(x_0) = 0$.

CASE TWO: $\underline{f'(a) > 0 > f'(b)}$: Since $f'(b) < 0$, and it is the limit of the difference quotients at b, we must have $(f(x) - f(b))/(x - b) < 0$ for x slightly smaller than b. Since $x - b < 0$, we must have $f(x) > f(b)$ for such x. So the maximum does not occur at b. Since $f'(a) > 0$, and it is the limit of the difference quotients at a, we must have $(f(x) - f(a))/(x - a) > 0$ for x slightly larger than a. Since $x - a > 0$, we must have $f(x) > f(a)$ for such x. So the maximum does not occur at a. Thus the maximum must occur at a point $x_0 \in]a, b[$. By Proposition 4.7.9, we must have $f'(x_0) = 0$.

By modifying this result a bit, we can establish an apparently stronger result.

Theorem. *If f is differentiable at all points of $[a, b]$ and w is between $f'(a)$ and $f'(b)$, then there is at least one point c in $[a, b]$ at which $f'(c) = w$.*

That is, f' satisfies the same sort of intermediate value property as does a continuous function. This may not seem particularly striking until you notice that we do not require that f' be continuous. It might not be, but it cannot have a simple jump discontinuity, that is, a point at which the left and right limits exist but are different.

Corollary. *If f is differentiable at all points of $[a, b]$, then f' can have no simple jump discontinuities in $[a, b]$.*

A function which is a derivative at every point need not be continuous. But any discontinuities must be rather messy. For an example of the kind of thing which can happen, consider the function defined on $\mathbb{R}$ by $f(x) = x^2 \sin(1/x)$ for $x \neq 0$ and $f(0) = 0$. One can use L'Hôpital's rule to show that the limit of the difference quotients exists and is 0 at 0. So f is differentiable everywhere. For x not equal to 0, the derivative is $f'(x) = 2x \sin(1/x) - \cos(1/x)$. This oscillates wildly in every neighborhood of 0. See the figure. ♦

FIGURE 4-12. Intermediate Value Property for derivatives.

◇ **4E-38.** A real-valued function defined on $]a, b[$ is called *convex* when the following inequality holds for x, y in $]a, b[$ and t in $[0, 1]$:

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y).$$

If f has a continuous second derivative and $f'' > 0$, show that f is convex.

Solution. Fix x and y in $]a, b[$. We want to show that $f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$ for all t in $[0, 1]$. Let $g(t) = tf(x) + (1-t)f(y) - f(tx + (1-t)y)$. Our project is to show that $g(t) \ge 0$ everywhere in $[0, 1]$. Since f is twice differentiable, so is g.

$$\begin{aligned} g(t) &= tf(x) + (1-t)f(y) - f(tx + (1-t)y) \\ g'(t) &= f(x) - f(y) - f'(tx + (1-t)y)(x - y) \\ g''(t) &= -f''(tx + (1-t)y)(x - y)^2 < 0. \end{aligned}$$

The second derivative of g is negative everywhere in $[0,1]$. We are in a position to use the second derivative test about local extremes. Since f is continuous between x and y, g is continuous on the compact domain $[0,1]$ and must attain a finite minimum at some point t_1 and a finite maximum at some point t_2 in $[0,1]$ by the maximum-minimum theorem, 4.4.1. Since g'' exists and is negative everywhere in $]0,1[$, the second derivative test, Proposition 4.7.16, implies that any extreme occurring in the interior would have to be a maximum. Thus the minimum must occur at one of the ends. But

$$\begin{aligned} g(0) &= 0f(x) + 1f(y) - f(0x + 1y) = f(y) - f(y) = 0 \\ g(1) &= 1f(x) + 0f(y) - f(1x + 0y) = f(x) - f(x) = 0. \end{aligned}$$

Since the minimum must be $g(0)$ or $g(1)$, we conclude that $g(t) \geq 0$ for all t in $[0,1]$ just as we wanted.

Geometrically, the inequality we have just established says that the chord between any two points on the graph of f always lies above the graph. See the figure. ♦

FIGURE 4-13. A convex function.

◇ **4E-39.** Suppose f is continuous on $[a,b]$, $f(a) = f(b) = 0$, and $x^2 f''(x) + 4xf'(x) + 2f(x) \geq 0$ for $x \in\,]a,b[$. Prove that $f(x) \leq 0$ for x in $[a,b]$.

Suggestion. Consider the second derivative of $x^2 f(x)$. ◇

Solution. Let $g(x) = x^2 f(x)$. Since $x^2 \geq 0$, it is enough to show that $g(x) \leq 0$ for all x in $[a,b]$. At the ends we have

$$g(a) = a^2 f(a) = 0 \quad \text{and} \quad g(b) = b^2 f(b) = 0.$$

Since f is continuous on the compact domain $[a, b]$, so is g, and the maximum-minimum theorem, 4.4.1, says that g must attain its maximum at some point x_0 in $[a, b]$. If $g(x)$ were ever larger than 0, then $g(x_0) > 0$ and x_0 would have to be in the interior. From these we will get a contradiction. Since f is twice differentiable, so is g, and

$$\begin{aligned} g(x) &= x^2 f(x) \\ g'(x) &= 2xf(x) + x^2 f'(x) \\ g''(x) &= x^2 f''(x) + 4xf'(x) + 2f(x) \geq 0 \end{aligned}$$

for $x \in]a, b[$. With x_0 in the interior and $g(x_0) > 0$, the mean value theorem gives points c_1 and c_2 with $a \leq c_1 \leq x_0 \leq c_2 \leq b$ where

$$0 < g(x_0) - g(a) = g'(c_1)(x_0 - a) \quad \text{and} \quad 0 > g(b) - g(x_0) = g'(c_2)(b - x_0).$$

So $g'(c_1) > 0$ and $g'(c_2) < 0$. By the mean value theorem again there would be a point c_3 between c_1 and c_2 with

$$0 < g'(c_2) - g'(c_1) = g''(c_3)(c_2 - c_1).$$

This would imply that $g''(c_3) < 0$. But we know that $g''(x) \geq 0$ for all x. This contradiction shows that we must, in fact, have $g(x) \leq 0$ for all x. Since $g(x) = x^2 f(x)$ and $x^2 \geq 0$, we also have $f(x) \leq 0$ for all x in $[a, b]$. ♦

◇ **4E-40.** Calculate

$$\frac{d}{dt} \int_0^t \frac{dx}{1 + x^2}.$$

Solution. The integrand is continuous on $\mathbb{R}$, so the fundamental theorem of calculus says that $F(t) = \int_0^t \frac{dx}{1+x^2}$ is an antiderivative for $f(t) = 1/(1 + t^2)$.

$$\boxed{\frac{d}{dt} \left(\int_0^t \frac{dx}{1 + x^2} \right) = \frac{1}{1 + t^2}.}$$ ♦

◇ **4E-41.** Prove that

$$\sum_{k=1}^{n} k = \frac{n(n + 1)}{2}$$

and

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

These formulas were used in Worked Example 4WE-6.

Sketch. Both are reasonably straightforward by mathematical induction. For the first, get from step n to step $n+1$ by adding $n+1$ to both sides. For the second, add $(n+1)^2$. $\Diamond$

Solution. We proceed in both by mathematical induction.

Proposition. *For every positive integer $n \in \mathbb{N}$, we have*

$$(a) \quad \sum_{k=1}^{n} k = \frac{n(n+1)}{2} \quad \textit{and} \quad (b) \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

(a) *Initialization Step*: For $n = 1$, the formula becomes $1 = \sum_{k=1}^{1} k = 1(1+1)/2 = 2/2 = 1$ which is true.
Induction Step: Form the *Induction Hypothesis*. **IH**: Suppose the formula is true for $n = m$: $\displaystyle\sum_{k=1}^{m} k = \frac{m(m+1)}{2}$.
We want to show that this implies that the formula would also hold for $n = m+1$, so we compute:

$$\begin{aligned}
\sum_{k=1}^{m+1} k &= \sum_{k=1}^{m} k + (m+1) \\
&= \frac{m(m+1)}{2} + (m+1) \qquad \text{(Using the induction hypothesis)} \\
&= \frac{m(m+1) + 2(m+1)}{2} \\
&= \frac{(m+1)(m+2)}{2} \\
&= \frac{(m+1)((m+1)+1)}{2}.
\end{aligned}$$

This is exactly the desired formula for $n = m+1$. Truth of our formula for all $n \in \mathbb{N}$ follows by mathematical induction.

(b) *Initialization Step*: For $n = 1$, the formula becomes $1 = \sum_{k=1}^{1} k^2 = 1(1+1)(2 \cdot 1 + 1)/6 = (1 \cdot 2 \cdot 3)/6 = 1$ which is true.
Induction Step: Form the *Induction Hypothesis*. **IH**: Suppose the formula is true for $n = m$: $\displaystyle\sum_{k=1}^{m} k^2 = \frac{m(m+1)(2m+1)}{6}$.

We want to show that this implies that the formula would also hold for $n = m+1$, so we compute:

$$\begin{aligned}
\sum_{k=1}^{m+1} k^2 &= \sum_{k=1}^{m} k^2 + (m+1)^2 \\
&= \frac{m(m+1)(2m+1)}{6} + (m+1)^2 \qquad \text{(Using the } \mathbf{IH}\text{)} \\
&= \frac{m(m+1)(2m+1) + 6(m+1)^2}{6} \\
&= \frac{(m+1)(m(2m+1) + 6(m+1))}{6} \\
&= \frac{(m+1)(2m^2 + 7m + 6)}{6} \\
&= \frac{(m+1)(m+2)(2m+3)}{6} \\
&= \frac{(m+1)((m+1)+1)(2(m+1)+1)}{6}
\end{aligned}$$

This is exactly the desired formula for $n = m+1$. Truth of our formula for all $n \in \mathbb{N}$ follows by mathematical induction.

Less formal proofs, especially of the first formula, abound. One of the most popular involves writing out the terms twice, the second time in reverse order under the first.

$$\begin{array}{cccccccc}
1 & 2 & 3 & 4 & \ldots & (n-2) & (n-1) & n \\
n & (n-1) & (n-2) & (n-3) & \ldots & 3 & 2 & 1 \\
 & & & & \ldots & & & \\
\hline
n+1 & n+1 & n+1 & n+1 & \ldots & n+1 & n+1 & n+1
\end{array}$$

Each of the n columns sums to $n+1$, so the total of both rows is $n(n+1)$. This is twice the desired sum since all terms appear twice.

A graphical presentation of essentially the same idea is to stack bars of length $1, 2, 3, \ldots n$ into a triangle and then arrange two of these triangles into an $n \times (n+1)$ rectangle. The area of this rectangle is twice our sum. This is illustrated for $n = 5$ in Figure 4-14. ♦

◇ **4E-42.** For $x > 0$, define $L(x) = \int_1^x (1/t)\, dt$. Prove the following, using this definition:

(a) L is increasing in x.

(b) $L(xy) = L(x) + L(y)$.

FIGURE 4-14. $1 + 2 + 3 + 4 + 5 = (5 \cdot 6)/2$.

(c) $L'(x) = 1/x$.
(d) $L(1) = 0$.
(e) Properties (c) and (d) uniquely determine L. What is L?

Sketch. Do parts (c) and (d) before part (b). Then do part (b) by fixing $y > 0$ and considering $f(x) = L(xy) - L(x) - L(y)$ as a function of x. $\Diamond$

Solution. (c) The integrand $1/t$ is a continuous function of t for $t > 0$. So, by the fundamental theorem of calculus we have

$$\frac{d}{dx}\int_1^x \frac{1}{t}\,dt = \frac{1}{x}.$$

(d)

$$L(1) = \int_1^1 \frac{1}{t}\,dt = 0.$$

(a) From part (c) we know that L is differentiable on $x > 0$ and that $L'(x) = 1/x > 0$ on that half-line. So L is strictly increasing by Proposition 4.7.14.

(b) Fix an arbitrary $y > 0$. For $x > 0$ let $f(x) = L(xy) - L(x) - L(y)$. We want to show that $f(x) = 0$ for all $x > 0$. Using part (d) we have $f(1) = L(x) - L(x) - L(1) = L(x) - L(x) - 0 = 0$. Using part (c) and the chain rule, we have

$$\begin{aligned} f'(x) &= \frac{d}{dx}\Big(L(xy) - L(x) - L(y)\Big) \\ &= L'(xy)\frac{d}{dx}\Big(xy\Big) - L'(x) - \frac{d}{dx}\Big(L(y)\Big) \\ &= \frac{1}{xy}\cdot y - \frac{1}{x} - 0 \\ &= \frac{1}{x} - \frac{1}{x} - 0 = 0. \end{aligned}$$

From Proposition 4.1.14(v) we conclude that f is constant. Since $f(1) = 0$, we must have $f(x) = 0$ for all $x > 0$ as we wanted.

(e) Suppose we also had $K'(x) = 1/x$ for all $x > 0$ and $K(1) = 0$. Let $g(x) = L(x) - K(x)$ Then g is differentiable on $x > 0$, and $g'(x) = K'(x) - L'(x) = (1/x) - (1/x) = 0$. By Proposition 4.1.14(v) we know that g is constant on $x > 0$. Since $g(1) = K(1) - L(1) = 0 - 0 = 0$, we conclude that $0 = g(x) = K(x) - L(x)$ for all $x > 0$. So $K(x) = L(x)$ for all $x > 0$. Thus there is exactly one function on the positive real line which has these properties. It is the natural logarithm function. (See also §5.3.) ♦

◇ **4E-43.** Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and set $F(x) = \int_0^{x^2} f(y)\,dy$. Prove that $F'(x) = 2xf(x^2)$. Give a more general theorem.

Sketch. Use the chain rule and the fundamental theorem of calculus. If $F(x) = \int_0^{g(x)} f(y)\,dy$ and g is differentiable and f continuous, then $F'(x) = f(g(x))g'(x)$. ◊

Solution. If $g(x)$ is differentiable and $u = g(x)$ and $I(u) = \int_0^u f(y)\,dy$. Put $F(x) = I(g(x)) = \int_0^{g(x)} f(y)\,dy$. Since f is continuous, the fundamental theorem of calculus gives

$$\frac{dI}{du} = \frac{d}{du}\int_0^u f(y)\,dy = f(u).$$

So, by the chain rule, $F'(x) = I'(g(x))g'(x) = f(g(x))g'(x)$.

In particular, in our problem $g(x) = x^2$, so $g'(x) = 2x$, and $F'(x) = f(x^2)2x = 2xf(x^2)$ as claimed. ♦

◇ **4E-44.** Let $f : [0, 1] \to \mathbb{R}$ be Riemann integrable and suppose for every a, b with $0 \le a < b \le 1$ there is a c, $a < c < b$, with $f(c) = 0$. Prove $\int_0^1 f = 0$. Must f be zero? What if f is continuous?

Suggestion. Show that the upper and lower sums are both 0 for every partition of $[0, 1]$. Consider a function which is 0 except at finitely many points. ◊

Solution. Since f is integrable on $[0, 1]$, the upper and lower integrals are the same and are equal to the integral. Let $P = \{0 = x_0 < x_1 < x_2 < \cdots < x_n = 1\}$ be any partition of $[0, 1]$. For each subinterval $[x_{j-1}, x_j]$ there is a point c_j in it with $f(c_j) = 0$. So

$$m_j = \inf\{f(x) \mid x \in [x_{j-1}, x_j]\} \le 0 \le \sup\{f(x) \mid x \in [x_{j-1}, x_j]\} = M_j.$$

So

$$L(f,P) = \sum_{j=1}^{n} m_j(x_j - x_{j-1}) \le 0 \le \sum_{j=1}^{n} M_j(x_j - x_{j-1}) = U(f,P).$$

This is true for every partition of $[0,1]$. So

$$\begin{aligned}\int_0^1 f(x)\,dx = \underline{\int_0^1} f(x)\,dx &= \sup_{P \text{ a partition of } [0,1]} L(f,P) \le 0 \\ &\le \inf_{P \text{ a partition of } [0,1]} U(f,P) = \overline{\int_0^1} f(x)\,dx = \int_0^1 f(x)\,dx.\end{aligned}$$

So we must have $\int_0^1 f(x)\,dx = 0$.

The function f need not be identically 0. We could, for example, have $f(x) = 0$ for all but finitely many points at which $f(x) = 1$.

If f is continuous and satisfies the stated condition, then f must be identically 0. Let $x \in [0,1]$. By hypothesis there is, for each integer $n > 0$, at least one point c_n in $[0,1]$ with $x-(1/n) \le c_n \le x+(1/n)$ and $f(c_n) = 0$. Since $c_n \to 0$ and f is continuous, we must have $0 = f(c_n) \to f(x)$. So $f(x) = 0$. ♦

◇ **4E-45.** Prove the following *second mean value theorem. Let f and g be defined on $[a,b]$ with g continuous, $f \ge 0$, and f integrable. Then there is a point $x_0 \in\ ]a,b[$ such that*

$$\int_a^b f(x)g(x)\,dx = g(x_0)\int_a^b f(x)\,dx.$$

Sketch. Let $m = \inf(g([a,b]))$ and $M = \sup(g([a,b]))$. Then

$$m\int_a^b f(x)\,dx \le \int_a^b f(x)g(x)\,dx \le M\int_a^b f(x)\,dx.$$

(Why?) Since $t\int_a^b f(x)\,dx$ depends continuously on t, the intermediate value theorem gives t_0 in $[m,M]$ with

$$\int_a^b f(x)g(x)\,dx = t_0\int_a^b f(x)\,dx.$$

Now apply that theorem to g to get x_0 with $g(x_0) = t_0$. (Supply details.) ◇

Solution. Since g is continuous on the compact interval $[a,b]$, we know that $m = \inf(g([a,b]))$ and $M = \sup(g([a,b]))$ exist as finite real numbers and that there are points x_1 and x_2 in $[a, b]$ where $g(x_1) = m$ and $g(x_2) = M$. Since $m \le g(x) \le M$ and $f(x) \ge 0$, we have $mf(x) \le f(x)g(x) \le Mf(x)$ for all x in $[a, b]$. Assuming that f and fg are integrable on $[a, b]$, Proposition 4.8.5(iii) gives

$$\begin{aligned} m \int_a^b f(x)\,dx &= \int_a^b mf(x)\,dx \\ &\le \int_a^b f(x)g(x)\,dx \\ &\le \int_a^b Mf(x)\,dx \\ &= M \int_a^b f(x)\,dx. \end{aligned}$$

The function $h(t) = t \int_a^b f(x)\,dx$ is a continuous function of t in the interval $m \le t \le M$, and $\int_a^b f(x)g(x)\,dx$ is a number between $h(m)$ and $h(M)$. By the intermediate value theorem there is a number t_0 in $[m, M]$ with $h(t_0) = \int_a^b f(x)g(x)\,dx$. Since g is continuous between x_1 and x_2 and t_0 is between $m = g(x_1)$ and $M = g(x_2)$, another application of the intermediate value theorem gives a point x_0 between x_1 and x_2 with $g(x_0) = t_0$. So

$$\int_a^b f(x)g(x)\,dx = h(t_0) = t_0 \int_a^b f(x)\,dx = g(x_0) \int_a^b f(x)\,dx$$

as desired.

In the argument above we assumed that the function fg was integrable on $[a, b]$. We need the following modification of Theorem 4.8.4(i):

Theorem. *If $f : [a, b] \to \mathbb{R}$ is integrable on $[a, b]$ and $g : [a, b] \to \mathbb{R}$ is continuous, then the product $f \cdot g : [a, b] \to \mathbb{R}$ is integrable on $[a, b]$.*

Proof: Let $\varepsilon > 0$. Since f is integrable there are partitions P_1 and P_2 of $[a, b]$ with

$$U(f, P_1) - \varepsilon < \overline{\int_a^b} f = \int_a^b f = \underline{\int_a^b} f < L(f, P_2) + \varepsilon.$$

If Q is any common refinement of P_1 and P_2, then

$$L(f, P_2) \le L(f, Q) \le U(f, Q) \le U(f, P_1).$$

So $U(f, Q) - L(f, Q) < 2\varepsilon$. Since f is integrable on $[a, b]$, it is bounded, as is the continuous function g

$$|f(x)| \le \| f \|_\infty \quad \text{and} \quad |g(x)| \le \| g \|_\infty \quad \text{for all } x \in [a, b].$$

Also, since g is continuous on the compact domain $[a, b]$, it is uniformly continuous on it. So there is a $\delta > 0$ such that $|g(x) - g(t)| < \varepsilon$ whenever x and t are in $[a, b]$ and $|x - t| < \delta$. Let $Q = \{a = t_0 < t_1 < \cdots < t_n = b\}$ be a common refinement of P_1 and P_2 with $t_j - t_{j-1} < \delta$ for each j. If x and t are in the same subinterval of Q, then

$$\begin{aligned}
|f(x)g(x) - f(t)g(t)| &\leq |f(x)g(x) - f(t)g(x)| + |f(t)g(x) - f(t)g(t)| \\
&\leq |f(x) - f(t)|\,|g(x)| + |f(t)|\,|g(x) - g(t)| \\
&\leq (M(f, Q, j) - m(f, Q, j))\,\|\,g\,\|_\infty + \varepsilon\,\|\,f\,\|_\infty\,.
\end{aligned}$$

Since this holds for every x and t in the subinterval, we have

$$M(fg, Q, j) - m(fg, Q, j) \leq (M(f, Q, j) - m(f, Q, j))\,\|\,g\,\|_\infty + \varepsilon\,\|\,f\,\|_\infty\,.$$

So

$$\begin{aligned}
U(fg, Q) - L(fg, Q) &= \sum_{j=1}^{n} (M(fg, Q, j) - m(fg, Q, j))(t_j - t_{j-1}) \\
&\leq \sum_{j=1}^{n} ((M(f, Q, j) - m(f, Q, j))\,\|\,g\,\|_\infty \\
&\qquad + \varepsilon\,\|\,f\,\|_\infty)(t_j - t_{j-1}) \\
&\leq \|\,g\,\|_\infty \sum_{j=1}^{n} ((M(f, Q, j) - m(f, Q, j))(t_j - t_{j-1}) \\
&\qquad + \varepsilon\,\|\,f\,\|_\infty \sum_{j=1}^{n} (t_j - t_{j-1}) \\
&\leq \|\,g\,\|_\infty\,(U(f, Q) - L(f, Q)) + \varepsilon\,\|\,f\,\|_\infty\,(b - a) \\
&\leq \|\,g\,\|_\infty\,2\varepsilon + \varepsilon\,\|\,f\,\|_\infty\,(b - a) \\
&= (2\,\|\,g\,\|_\infty + \|\,f\,\|_\infty\,(b - a))\varepsilon.
\end{aligned}$$

So

$$\begin{aligned}
L(fg, Q) \leq \underline{\int_a^b} fg \leq \overline{\int_a^b} fg &\leq U(fg, Q) \\
&< L(fg, Q) + (2\,\|\,g\,\|_\infty + \|\,f\,\|_\infty\,(b - a))\varepsilon.
\end{aligned}$$

Since this can be done for any $\varepsilon > 0$, we must have $\underline{\int_a^b} fg = \overline{\int_a^b} fg$. So fg is integrable on $[a, b]$. ♦

◇ **4E-46.** (a) For complex-valued functions on an interval, prove the fundamental theorem of calculus.

(b) Evaluate $\int_0^\pi e^{ix}\,dx$ using (a).

Solution. (a) If $F(x) = U(x) + iV(x)$ is a complex valued function of a real variable x with $U(x)$ and $V(x)$ differentiable real valued functions, then

$$\begin{aligned}\lim_{h\to 0} \frac{F(x+h)-F(x)}{h} &= \lim_{h\to 0} \frac{U(x+h)-U(x)+i(V(x+h)-V(x))}{h} \\ &= \lim_{h\to 0} \frac{U(x+h)-U(x)}{h} + i \lim_{h\to 0} \frac{V(x+h)-V(x)}{h} \\ &= U'(x) + iV'(x).\end{aligned}$$

Now suppose $f; [a,b] \to \mathbb{C}$ is a continuous function and $f = u + iv$ with u and v real. The u and v are continuous functions. From the fundamental theorem of calculus for real valued functions, 4.8.8, each of u and v have antiderivatives on $]a,b[$ given by integrals. $u(x) = U'(x)$ and $v(x) = V'(x)$ where

$$U(x) = \int_a^x u(t)\,dt \quad \text{ and } \quad V(x) = \int_a^x v(t)\,dt.$$

Now we put together the computation of derivatives above and Worked Example 4WE-7.

$$\begin{aligned}\frac{d}{dx}\left(\int_a^x f(t)\,dt\right) &= \frac{d}{dx}\left(\int_a^x (u(t)+iv(t))\,dt\right) \\ &= \frac{d}{dx}\left(\int_a^x u(t)\,dt + i\int_a^x v(t))\,dt\right) \\ &= \frac{d}{dx}(U(x)+iV(x)) = U'(x)+iV'(x) \\ &= u(x)+iv(x) = f(x).\end{aligned}$$

Thus $F(x) = \int_a^x f(t)\,dt$ is an antiderivative for $f(x)$ on $]a,b[$.
If G is any other antiderivative then F and G differ by a constant. So

$$\int_a^b f(t)\,dt = F(b) - F(a) = G(b) - G(a).$$

(b) Let $f(x) = e^{ix} = \cos x + i\sin x$. Then $f'(x) = -\sin x + i\cos x = i(\cos x + \sin x) = ie^{ix} = if(x)$. So an antiderivative for $f(x)$ is given by $G(x) = (1/i)e^{ix}$. So we have

$$\int_0^\pi e^{ix}\,dx = G(\pi) - G(0) = \frac{e^{i\pi}}{i} - \frac{e^0}{i} = \frac{-1}{i} - \frac{1}{i} = \frac{-2}{i} = 2i. \quad \blacklozenge$$