Chapter 2
Continuity

2.1 Compactness

An *open interval* is a set of reals of the form $(a, b) = \{x : a < x < b\}$. As in §1.4, we are allowing $a = -\infty$ or $b = \infty$ or both. A *compact interval* is a set of reals of the form $[a, b] = \{x : a \leq x \leq b\}$, where a, b are real. The *length* of $[a, b]$ is $b - a$. Recall (§1.5) that a sequence *subconverges* to L if it has a subsequence converging to L.

Theorem 2.1.1. *Let $[a, b]$ be a compact interval and let (x_n) be any sequence in $[a, b]$. Then (x_n) subconverges to some x in $[a, b]$.*

To derive this result, assume, first, that $a = 0$ and $b = 1$. Divide the interval $I = [a, b] = [0, 1]$ into 10 subintervals (of the same length), and call them $I_0, \cdots, I_9$, ordering them from left to right (Figure 2.1). Pick one of them, say I_{d_1}, containing infinitely many terms of (x_n) (that is, $\{n : x_n \in I_{d_1}\}$ is infinite) and pick one of the terms of the sequence in I_{d_1} and call it x'_1. Then the length of I_{d_1} is $1/10$. Now divide I_{d_1} into 10 subintervals again ordered left to right and called $I_{d_1 0}, \cdots, I_{d_1 9}$. Pick one of them, say $I_{d_1 d_2}$, containing infinitely many terms of the sequence, and pick one of the terms (beyond x'_1) in the sequence in $I_{d_1 d_2}$ and call it x'_2. The length of $I_{d_1 d_2}$ is $1/100$. Continuing in this manner yields $I \supset I_{d_1} \supset I_{d_1 d_2} \supset I_{d_1 d_2 d_3} \supset \cdots$ and a subsequence (x'_n) where the length of $I_{d_1 d_2 \cdots d_n}$ is 10^{-n} and $x'_n \in I_{d_1 d_2 \cdots d_n}$ for all $n \geq 1$. But, by construction, the real

$$x = .d_1 d_2 d_3 \cdots$$

lies in all the intervals $I_{d_1 d_2 \cdots d_n}$, $n \geq 1$. Hence $|x'_n - x| \leq 10^{-n} \to 0$. Because (x'_n) is a subsequence of (x_n), this derives the result if $[a, b] = [0, 1]$. If this is not so the same argument works. The only difference is that the limiting point now obtained is $a + x(b - a)$. □

Thus this theorem is equivalent to, more or less, the existence of decimal expansions. If $[a, b]$ is replaced by an open interval (a, b), the theorem is false

O. Hijab, *Introduction to Calculus and Classical Analysis*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-1-4419-9488-2_2,

Fig. 2.1 The intervals $I_{d_1 d_2 \cdots d_n}$.

as it stands, hence the theorem needs to be modified. A useful modification is the following.

Theorem 2.1.2. *If (x_n) is a sequence of reals in (a, b), then (x_n) subconverges to some $a < x < b$ or to a or to b.*

To see this, because $a = \inf(a, b)$, there is (Theorem 1.5.4) a sequence (c_n) in (a, b) satisfying $c_n \to a$. Similarly, because $b = \sup(a, b)$, there is a sequence (d_n) in (a, b) satisfying $d_n \to b$. Now either there is an $m \geq 1$ with (x_n) in $[c_m, d_m]$ or not. If so, the result follows from Theorem 2.1.1. If not, for every $m \geq 1$, there is an x_{n_m} not in $[c_m, d_m]$. Let (y_m) be the subsequence of (x_{n_m}) obtained by restricting attention to terms satisfying $x_{n_m} > d_m$, and let (z_m) be the subsequence of (x_{n_m}) obtained by restricting attention to terms satisfying $x_{n_m} < c_m$. Then at least one of the sequences (y_m) or (z_m) is infinite, so either $y_m \to b$ or $z_m \to a$ (or both) as $m \to \infty$. Thus (x_n) subconverges to a or to b. $\square$

Note this result holds even when $a = -\infty$ or $b = \infty$.

Exercises

2.1.1. Let (a_n) and (b_n) be sequences. We say (a_n, b_n) *subconverges* to (a, b) if there is a sequence of naturals (n_k) such that (a_{n_k}) converges to a and (b_{n_k}) converges to b. Show that if (a_n) and (b_n) are bounded, then (a_n, b_n) subconverges to some (a, b).

2.1.2. In the derivation of the first theorem, suppose that the intervals are chosen, at each stage, to be the leftmost interval containing infinitely many terms. In other words, suppose that I_{d_1} is the leftmost of the intervals I_j containing infinitely many terms, $I_{d_1 d_2}$ is the leftmost of the intervals $I_{d_1 j}$ containing infinitely many terms, and so on. In this case, show that the limiting point obtained is x_*.

2.2 Continuous Limits

Let (a, b) be an open interval, and let $a < c < b$. The interval (a, b), *punctured* at c, is the set $(a, b) \setminus \{c\} = \{x : a < x < b, x \neq c\}$.

Let f be a function defined on an interval (a, b) punctured at c, $a < c < b$. We say L is the *limit* of $f(x)$ as x *approaches* c, and we write

$$\lim_{x \to c} f(x) = L$$

or $f(x) \to L$ as $x \to c$, if, for every sequence $(x_n) \subset (a, b)$ satisfying $x_n \neq c$ for all $n \geq 1$ and converging to c, $f(x_n) \to L$.

For example, let $f(x) = x^2$, and let $(a, b) = \mathbf{R}$. If $x_n \to c$, then (§1.5), $x_n^2 \to c^2$. This holds true no matter what sequence (x_n) is chosen, as long as $x_n \to c$. Hence in this case, $\lim_{x \to c} f(x) = c^2$.

Going back to the general definition, suppose that f is also defined at c. Then *the value* $f(c)$ *has no bearing on* $\lim_{x \to c} f(x)$ (Figure 2.2). For example, if $f(x) = 0$ for $x \neq 0$ and $f(0)$ is defined arbitrarily, then $\lim_{x \to 0} f(x) = 0$. For a more dramatic example of this phenomenon, see Exercise **2.2.1**.

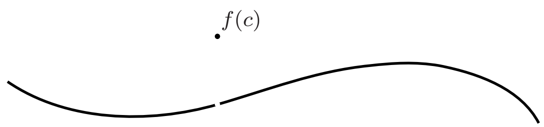

Fig. 2.2 The value $f(c)$ has no bearing on the limit at c.

Of course, not every function has limits. For example, set $f(x) = 1$ if $x \in \mathbf{Q}$ and $f(x) = 0$ if $x \in \mathbf{R} \setminus \mathbf{Q}$. Choose any c in $(a, b) = \mathbf{R}$. Because (§1.4) there is a rational and an irrational between any two reals, for each $n \geq 1$ we can find $r_n \in \mathbf{Q}$ and $i_n \in \mathbf{R} \setminus \mathbf{Q}$ with $c < r_n < c + 1/n$ and $c < i_n < c + 1/n$. Thus $r_n \to c$ and $i_n \to c$, but $f(r_n) = 1$ and $f(i_n) = 0$ for all $n \geq 1$. Hence f has no limit anywhere on $\mathbf{R}$.

Let (x_n) be a sequence approaching b. If $x_n < b$ for all $n \geq 1$, we write $x_n \to b-$. Let f be defined on (a, b). We say L is the *limit* of $f(x)$ as x *approaches* b *from the left*, and we write

$$\lim_{x \to b-} f(x) = L,$$

if $x_n \to b-$ implies $f(x_n) \to L$. In this case, we also write $f(b-) = L$. If $b = \infty$, we drop the minus and write $\lim_{x \to \infty} f(x) = L$, $f(\infty) = L$.

Let (x_n) be a sequence approaching a. If $x_n > a$ for all $n \geq 1$, we write $x_n \to a+$. Let f be defined on (a, b). We say L is the *limit* of $f(x)$ as x *approaches* a *from the right*, and we write

$$\lim_{x \to a+} f(x) = L,$$

if $x_n \to a+$ implies $f(x_n) \to L$. In this case, we also write $f(a+) = L$. If $a = -\infty$, we drop the plus and write $\lim_{x \to -\infty} f(x) = L$, $f(-\infty) = L$.

Suppose $f(b-) = L$ and (x_n) is a sequence approaching b such that $x_n < b$ for all but finitely many $n \geq 1$. Then we may modify finitely many terms in (x_n) so that $x_n < b$ for all $n \geq 1$; because modifying a finite number of terms does not affect convergence, we have $f(x_n) \to L$. Similarly, if $f(b+) = L$ and (x_n) is a sequence approaching b such that $x_n > b$ for all but finitely many $n \geq 1$, we have $f(x_n) \to L$.

Of course, L above is either a real or $\pm\infty$.

Theorem 2.2.1. *Let f be defined on an interval (a, b) punctured at c, $a < c < b$. Then $\lim_{x\to c} f(x)$ exists and equals L iff $f(c+)$ and $f(c-)$ both exist and equal L.*

If $\lim_{x\to c} f(x) = L$, then $f(x_n) \to L$ for any sequence $x_n \to c$, whether the sequence is to the right, the left, or neither. Hence $f(c-) = L$ and $f(c+) = L$.

Conversely, suppose that $f(c-) = f(c+) = L$ and let $x_n \to c$ with $x_n \neq c$ for all $n \geq 1$. We have to show that $f(x_n) \to L$.

Let (y_n) denote the terms in (x_n) that are greater than c, and let (z_n) denote the terms in (x_n) that are less than c, arranged in their given order. If (y_n) is finite, then all but finitely many terms of (x_n) are less than c, thus $f(x_n) \to L$. If (z_n) is finite, then all but finitely many terms of (x_n) are greater than c, thus $f(x_n) \to L$. Hence we may assume both (y_n) and (z_n) are infinite sequences with $y_n \to c+$ and $z_n \to c-$. Because $f(c+) = L$, it follows that $f(y_n) \to L$; because $f(c-) = L$, it follows that $f(z_n) \to L$.

Let f^* and f_* denote the upper and lower limits of the sequence $(f(x_n))$, and set $f_n^* = \sup\{f(x_k) : k \geq n\}$. Then $f_n^* \searrow f^*$. Hence for any subsequence $(f_{k_n}^*)$, we have $f_{k_n}^* \searrow f^*$. The goal is to show that $f^* = L = f_*$.

Because $f(y_n) \to L$, its upper sequence converges to L, $\sup_{i\geq n} f(y_i) \searrow L$; because $f(z_n) \to L$, its upper sequence converges to L, $\sup_{i\geq n} f(z_i) \searrow L$.

For each $m \geq 1$, let x_{k_m} denote the term in (x_n) corresponding to y_m, if the term y_m appears after the term z_m in (x_n). Otherwise, if z_m appears after y_m, let x_{k_m} denote the term in (x_n) corresponding to z_m. In other words, if $y_n = x_{i_n}$ and $z_n = x_{j_n}$, $x_{k_n} = x_{\max(i_n, j_n)}$. Thus for each $n \geq 1$, if $k \geq k_n$, we must have x_k equal y_i or z_i with $i \geq n$, so

$$\{x_k : k \geq k_n\} \subset \{y_i : i \geq n\} \cup \{z_i : i \geq n\}.$$

Hence

$$f_{k_n}^* = \sup_{k\geq k_n} f(x_k) \leq \max\left[\sup_{i\geq n} f(y_i), \sup_{i\geq n} f(z_i)\right], \quad n \geq 1.$$

Now both sequences on the right are decreasing in $n \geq 1$ to L, and the sequence on the left decreases to f^* as $n \nearrow \infty$. Thus $f^* \leq L$. Now let $g = -f$. Because $g(c+) = g(c-) = -L$, by what we have just learned, we conclude that the upper limit of $(g(x_n))$ is $\leq -L$. But the upper limit of $(g(x_n))$ equals minus the lower limit f_* of $(f(x_n))$. Hence $f_* \geq L$, so $f^* = f_* = L$. $\square$

Because continuous limits are defined in terms of limits of sequences, they enjoy the same arithmetic and ordering properties. For example,

$$\lim_{x\to a}[f(x)+g(x)] = \lim_{x\to a} f(x) + \lim_{x\to a} g(x),$$
$$\lim_{x\to a}[f(x)\cdot g(x)] = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x).$$

These properties are used without comment.

A function f is *increasing (decreasing)* if $x \le x'$ implies $f(x) \le f(x')$ ($f(x) \ge f(x')$, respectively), for all x, x' in the domain of f. The function f is *strictly increasing (strictly decreasing)* if $x < x'$ implies $f(x) < f(x')$ ($f(x) > f(x')$, respectively), for all x, x' in the domain of f. If f is increasing or decreasing, we say f is *monotone*. If f is strictly increasing or strictly decreasing, we say f is *strictly monotone*.

In the exercises, the concept of a partition (Figure 2.3) is needed. If (a, b) is an open interval, a *partition of* (a, b) is a choice of points $(x_1, x_2, \cdots, x_n)$ in (a, b), arranged in increasing order. When choosing a partition, we write $a = x_0 < x_1 < \cdots < x_{n-1} < x_n < x_{n+1} = b$, denoting the endpoints a and b by x_0 and x_{n+1} respectively (even when they are infinite). We use the same notation for compact intervals; that is, a *partition of* $[a, b]$, by definition, is a partition of (a, b).

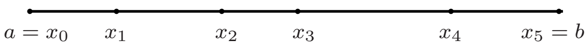

Fig. 2.3 A partition of (a, b).

Exercises

2.2.1. Define $f : \mathbf{R} \to \mathbf{R}$ by setting $f(m/n) = 1/n$, for $m/n \in \mathbf{Q}$ with no common factor in m and $n > 0$, and $f(x) = 0$, $x \notin \mathbf{Q}$. Show that $\lim_{x\to c} f(x) = 0$ for all $c \in \mathbf{R}$.

2.2.2. Let f be increasing on (a, b). Then $f(a+)$ (exists and) equals $\inf\{f(x) : a < x < b\}$, and $f(b-)$ equals $\sup\{f(x) : a < x < b\}$.

2.2.3. If f is monotone on (a, b), then $f(c+)$ and $f(c-)$ exist, and $f(c)$ is between $f(c-)$ and $f(c+)$, for all $c \in (a, b)$. Show also that, for each $\delta > 0$, there are, at most, countably many points $c \in (a, b)$ where $|f(c+) - f(c-)| \ge \delta$. Conclude that there are, at most, countably many points c in (a, b) at which $f(c+) \ne f(c-)$.

2.2.4. If $f : (a, b) \to \mathbf{R}$ let I_n be the sup of the sums

$$|f(x_2) - f(x_1)| + |f(x_3) - f(x_2)| + \cdots + |f(x_n) - f(x_{n-1})| \tag{2.2.1}$$

over all partitions $a < x_1 < x_2 < \cdots < x_n < b$ of (a, b) consisting of n points, and let $I = \sup\{I_n : n \geq 2\}$. We say that f is of *bounded variation* on (a, b) if I is finite. Show that bounded variation on (a, b) implies bounded on (a, b). The sum in (2.2.1) is the *variation* of f corresponding to the partition $a < x_1 < x_2 < \cdots < x_n < b$, whereas I, the sup of all such sums over all partitions consisting of arbitrarily many points, is the *total variation* of f over (a, b).

2.2.5. If f is bounded increasing on an interval (a, b), then f is of bounded variation on (a, b). If $f = g - h$ with g, h bounded increasing on (a, b), then f is of bounded variation on (a, b).

2.2.6. Let f be of bounded variation on (a, b), and, for $a < x < b$, let $v(x)$ denote the sup of the sums (2.2.1) over all partitions $a =< x_1 < x_2 < \cdots < x_n = x < b$ with $x_n = x$ fixed. Show that $a < x < y < b$ implies $v(x) + |f(y) - f(x)| \leq v(y)$, hence. $v : (a, b) \to \mathbf{R}$ and $v - f : (a, b) \to \mathbf{R}$ are bounded increasing. Conclude that f is of bounded variation iff f is the difference of two bounded increasing functions.

2.2.7. Show that the f in Exercise **2.2.1** is not of bounded variation on $(0, 2)$ (remember that $\sum 1/n = \infty$).

2.3 Continuous Functions

Let f be defined on (a, b), and choose $a < c < b$. We say that f *is continuous at* c if

$$\lim_{x \to c} f(x) = f(c).$$

If f is continuous at every real c in (a, b), then we say that f *is continuous on* (a, b) or, if (a, b) is understood from the context, f *is continuous.*

Recalling the definition of $\lim_{x \to c}$, we see that f is continuous at c iff, for all sequences (x_n) satisfying $x_n \to c$ and $x_n \neq c$, $n \geq 1$, $f(x_n) \to f(c)$. In fact, f is continuous at c iff $x_n \to c$ implies $f(x_n) \to f(c)$; that is, the condition $x_n \neq c$, $n \geq 1$, is superfluous. To see this, suppose that f is continuous at c, and suppose that $x_n \to c$, but $f(x_n) \not\to f(c)$. Because $f(x_n) \not\to f(c)$, by Exercise **1.5.8**, there is an $\epsilon > 0$ and a subsequence (x'_n), such that $|f(x'_n) - f(c)| \geq \epsilon$ and $x'_n \to c$, for $n \geq 1$. But, then $f(x'_n) \neq f(c)$ for all $n \geq 1$, hence $x'_n \neq c$ for all $n \geq 1$. Because $x'_n \to c$, by the continuity at c, we obtain $f(x'_n) \to f(c)$, contradicting $|f(x'_n) - f(c)| \geq \epsilon$. Thus f *is continuous at* c *iff* $x_n \to c$ *implies* $f(x_n) \to f(c)$.

In the previous section we saw that $f(x) = x^2$ is continuous at c. Because this works for any c, f is continuous. Repeating this argument, one can show that $f(x) = x^4$ is continuous, because $x^4 = x^2x^2$. A simpler example is to choose a real k and to set $f(x) = k$ for all x. Here $f(x_n) = k$, and $f(c) = k$ for all sequences (x_n) and all c, so f is continuous. Another example is $f : (0, \infty) \to \mathbf{R}$ given by $f(x) = 1/x$. By the division property of sequences, $x_n \to c$ implies $1/x_n \to 1/c$ for $c > 0$, so f is continuous.

Functions can be continuous at various points and not continuous at other points. For example, the function f in Exercise **2.2.1** is continuous at every irrational c and not continuous at every rational c. On the other hand, the function $f : \mathbf{R} \to \mathbf{R}$, given by (§2.2)

$$f(x) = \begin{cases} 1, & x \in \mathbf{Q} \\ 0, & x \notin \mathbf{Q}, \end{cases}$$

is continuous at no point.

Continuous functions have very simple arithmetic and ordering properties. If f and g are defined on (a, b) and k is real, we have functions $f + g$, kf, fg, $\max(f, g)$, $\min(f, g)$ defined on (a, b) by setting, for $a < x < b$,

$$\begin{aligned} (f+g)(x) &= f(x) + g(x), \\ (kf)(x) &= kf(x), \\ (fg)(x) &= f(x)g(x), \\ \max(f,g)(x) &= \max[f(x), g(x)], \\ \min(f,g)(x) &= \min[f(x), g(x)]. \end{aligned}$$

If g is *nonzero* on (a, b) (i.e. $g(x) \neq 0$ for all $a < x < b$) , define f/g by setting

$$(f/g)(x) = \frac{f(x)}{g(x)}, \quad a < x < b.$$

Theorem 2.3.1. *If f and g are continuous, then so are $f + g$, kf, fg, $\max(f, g)$, and $\min(f, g)$. Moreover, if g is nonzero, then f/g is continuous.*

This is an immediate consequence of the arithmetic and ordering properties of sequences. If $a < c < b$ and $x_n \to c$, then $f(x_n) \to f(c)$, and $g(x_n) \to g(c)$. Hence $f(x_n) + g(x_n) \to f(c) + g(c)$, $kf(x_n) \to kf(c)$, $f(x_n)g(x_n) \to f(c)g(c)$, $\max[f(x_n), g(x_n)] \to \max[f(c), g(c)]$, and $\min[f(x_n), g(x_n)] \to \min[f(c), g(c)]$. If $g(c) \neq 0$, then $f(x_n)/g(x_n) \to f(c)/g(c)$. □

For example, we see immediately that $f(x) = |x|$ is continuous on $\mathbf{R}$ because $|x| = \max(x, -x)$.

Let us prove by induction that, for all $k \geq 1$, the *monomials* $f_k(x) = x^k$ are continuous (on $\mathbf{R}$). For $k = 1$, this is so because $x_n \to c$ implies $f_1(x_n) = x_n \to c = f_1(c)$. Assuming that this is true for k, $f_{k+1} = f_k f_1$ because $x^{k+1} = x^k x$. Hence the result follows from the arithmetic properties of continuous functions.

A *polynomial* $f : \mathbf{R} \to \mathbf{R}$ is a linear combination of monomials; that is, a polynomial has the form

$$f(x) = a_0 x^d + a_1 x^{d-1} + a_2 x^{d-2} + \cdots + a_{d-1} x + a_d.$$

If $a_0 \neq 0$, we call d the *degree* of f. The reals $a_0, a_1, \cdots, a_d$, are the *coefficients* of the polynomial.

Let f be a polynomial of degree $d > 0$, and let $a \in \mathbf{R}$. Then there is a polynomial g of degree $d-1$ satisfying[1]

$$\frac{f(x) - f(a)}{x - a} = g(x), \quad x \neq a. \tag{2.3.1}$$

To see this, because every polynomial is a linear combination of monomials, it is enough to check (2.3.1) on monomials. But, for $f(x) = x^n$,

$$\frac{x^n - a^n}{x - a} = x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}, \quad x \neq a, \tag{2.3.2}$$

which can be checked[2] by cross-multiplying. This establishes (2.3.1).

Because a monomial is continuous and a polynomial is a linear combination of monomials, by induction on the degree, we obtain the following.

Theorem 2.3.2. *Every polynomial f is continuous on $\mathbf{R}$. Moreover, if d is its degree, there are, at most, d real numbers x satisfying $f(x) = 0$.*

A real x satisfying $f(x) = 0$ is called a *zero* or a *root* of f. Thus every polynomial f has, at most, d roots. To see this, proceed by induction on the degree of f. If $d = 1$, $f(x) = a_0 x + a_1$, so f has one root $x = -a_1/a_0$. Now suppose that every dth degree polynomial has, at most, d roots, and let f be a polynomial of degree $d + 1$. We have to show that the number of roots of f is at most $d + 1$. If f has no roots, we are done. Otherwise, let a be a root: $f(a) = 0$. Then by (2.3.1) there is a polynomial g of degree d such that $f(x) = (x - a)g(x)$. Thus any root $b \neq a$ of f must satisfy $g(b) = 0$. By the inductive hypothesis, g has, at most, d roots; hence f has, at most, $d + 1$ roots. $\square$

A polynomial may have no roots (e.g., $f(x) = x^2 + 1$). However, every polynomial of odd degree has at least one root (Exercise **2.3.1**).

A *rational function* is a quotient $f = p/q$ of two polynomials. The natural domain of f is $\mathbf{R} \setminus Z(q)$, where $Z(q)$ denotes the set of roots of q. Because $Z(q)$ is a finite set, the natural domain of f is a finite union of open intervals. We conclude that *every rational function is continuous where it is defined.*

Let $f : (a, b) \to \mathbf{R}$. If f is not continuous at $c \in (a, b)$, we say that f is *discontinuous* at c. There are "mild" discontinuities, and there are "wild" discontinuities. The mildest situation (Figure 2.4) is when the limits $f(c+)$ and $f(c-)$ exist and are equal, but not equal to $f(c)$. This can be easily remedied by modifying the value of $f(c)$ to equal $f(c+) = f(c-)$. With this

---

[1] g also depends on a.

[2] (2.3.2) with $x = 1$ was used to sum the geometric series in §1.6.

modification, the resulting function then is continuous at c. Because of this, such a point c is called a *removable discontinuity*. For example, the function f in Exercise **2.2.1** has removable discontinuities at every rational.

The next level of complexity is when $f(c+)$ and $f(c-)$ exist but may or may not be equal. In this case, we say that f has a *jump discontinuity* (Figure 2.4) or a *mild discontinuity* at c. For example, every monotone function has (at worst) jump discontinuities. In fact, every function of bounded variation has (at worst) jump discontinuities (Exercise **2.3.18**). The (amount of) *jump* at c, a real number, is $f(c+) - f(c-)$. In particular, a jump discontinuity of jump zero is nothing more than a removable discontinuity.

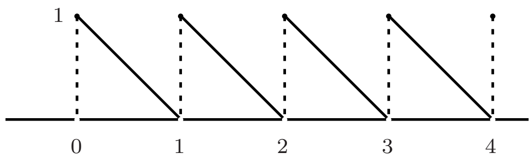

Fig. 2.4 A jump of 1 at each integer.

Any discontinuity that is not a jump is called a *wild discontinuity* (Figure 2.5). If f has a wild discontinuity at c, then from above, f cannot be of bounded variation on any open interval surrounding c. The converse of this statement is false. It is possible for f to have mild discontinuities but not be of bounded variation (Exercise **2.2.7**).

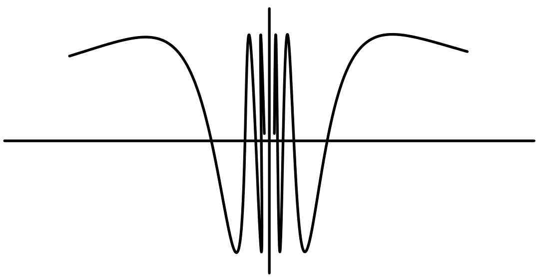

Fig. 2.5 A wild discontinuity.

An alternate and useful description of continuity is in terms of a modulus of continuity. Let $f : (a, b) \to \mathbf{R}$, and fix $a < c < b$. For $\delta > 0$, let

$$\mu_c(\delta) = \sup\{|f(x) - f(c)| : |x - c| < \delta, a < x < b\}.$$

Because the sup, here, is possibly that of an unbounded set, we may have $\mu_c(\delta) = \infty$. The function $\mu_c : (0, \infty) \to [0, \infty) \cup \{\infty\}$ is the *modulus of continuity of f at c* (Figure 2.6).

For example, let $f : (1, 10) \to \mathbf{R}$ be given by $f(x) = x^2$ and pick $c = 9$. Because x^2 is monotone over any interval not containing zero, the maximum value of $|x^2 - 81|$ over any interval not containing zero is obtained by plugging in the endpoints. Hence $\mu_9(\delta)$ is obtained by plugging in $x = 9 \pm \delta$, leading to $\mu_9(\delta) = \delta(\delta + 18)$. In fact, this is correct only if $0 < \delta \le 1$. If $1 \le \delta \le 8$, the interval under consideration is $(9-\delta, 9+\delta) \cap (1, 10) = (9-\delta, 10)$. Here plugging in the endpoints leads to $\mu_9(\delta) = \max(19, 18\delta - \delta^2)$. If $\delta \ge 8$, then $(9-\delta, 9+\delta)$ contains $(1, 10)$ and, hence $\mu_9(\delta) = 80$. Summarizing, for $f(x) = x^2$, $c = 9$, and $(a, b) = (1, 10)$,

$$\mu_c(\delta) = \begin{cases} \delta(\delta + 18), & 0 < \delta \le 1, \\ \max(19, 18\delta - \delta^2), & 1 \le \delta \le 8, \\ 80, & \delta \ge 8. \end{cases}$$

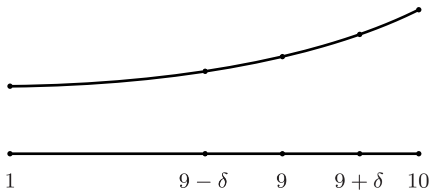

Fig. 2.6 Computing the modulus of continuity.

Going back to the general definition, note that $\mu_c(\delta)$ is an increasing function of δ, and, hence $\mu_c(0+)$ exists (Exercise **2.2.2**).

Theorem 2.3.3. *Let $f : (a, b) \to \mathbf{R}$, and choose $c \in (a, b)$. The following are equivalent.*

A. *f is continuous at c.*
B. $\mu_c(0+) = 0$.
C. *For all $\epsilon > 0$, there exists $\delta > 0$, such that*

$$|x - c| < \delta \textit{ implies } |f(x) - f(c)| < \epsilon.$$

That **A** implies **B** is left as Exercise **2.3.2**. Now assume **B**, and suppose that $\epsilon > 0$ is given. Because $\mu_c(0+) = 0$, there exists a $\delta > 0$ with $\mu_c(\delta) < \epsilon$. Then by definition of μ_c, $|x - c| < \delta$ implies $|f(x) - f(c)| \le \mu_c(\delta) < \epsilon$, which establishes **C**. Now assume the ϵ–δ criterion **C**, and let $x_n \to c$. Then for all but a finite number of terms, $|x_n - c| < \delta$. Hence for all but a finite number of terms, $f(c)-\epsilon < f(x_n) < f(c)+\epsilon$. Let $y_n = f(x_n)$, $n \ge 1$. By the ordering properties of sup and inf, $f(c) - \epsilon \le y_{n*} \le y_n^* \le f(c) + \epsilon$. By the ordering properties of sequences, $f(c) - \epsilon \le y_* \le y^* \le f(c) + \epsilon$. Because $\epsilon > 0$ is arbitrary, $y^* = y_* = f(c)$. Thus $y_n = f(x_n) \to f(c)$. Because (x_n) was any sequence converging to c, $\lim_{x\to c} f(x) = f(c)$ which establishes **A**. $\square$

Thus in practice, one needs to compute $\mu_c(\delta)$ only for δ small enough, because it is the behavior of μ_c near zero that counts. For example, to check continuity of $f(x) = x^2$ at $c = 9$, it is enough to note that $\mu_9(\delta) = \delta(\delta + 18)$ for small enough δ, which clearly approaches zero as $\delta \to 0+$.

To check the continuity of $f(x) = x^2$ at $c = 9$ using the ϵ–δ criterion **C**, given $\epsilon > 0$, it is enough to exhibit a $\delta > 0$ with $\mu_9(\delta) < \epsilon$. Such a δ is the lesser of $\epsilon/20$ and 1, $\delta = \min(\epsilon/20, 1)$. To see this, first, note that $\delta(\delta+18) \le 19$ for this δ. Then $\epsilon \le 19$ implies $\delta(\delta+18) \le (\epsilon/20)(1+18) = (19/20)\epsilon < \epsilon$, whereas $\epsilon > 19$ implies $\delta(\delta + 18) < \epsilon$. Hence in either case, $\mu_9(\delta) < \epsilon$, establishing **C**.

Now we turn to the mapping properties of a continuous function. First, we define one-sided continuity. Let f be defined on $(a, b]$. We say that f is *continuous at b from the left* if $f(b-) = f(b)$. In addition, if f is continuous on (a, b), we say that f is *continuous on* $(a, b]$. Let f be defined on $[a, b)$. We say that f is *continuous at a from the right* if $f(a+) = f(a)$. In addition, if f is continuous on (a, b), we say that f is *continuous on* $[a, b)$.

Note by Theorem 2.2.1 that a function f is continuous at a particular point c iff f is continuous at c from the right and continuous at c from the left.

Let f be defined on $[a, b]$. We say that f is *continuous on* $[a, b]$ if f is continuous on $[a, b)$ and $(a, b]$. Checking the definitions, we see f is continuous on A if, for every $c \in A$ and every sequence $(x_n) \subset A$ converging to c, $f(x_n) \to f(c)$, whether A is (a, b), $(a, b]$, $[a, b)$, or $[a, b]$.

Theorem 2.3.4. *Let f be continuous on a compact interval $[a, b]$. Then $f([a, b])$ is a compact interval $[m, M]$.*

Thus a continuous function maps compact intervals to compact intervals. Of course, it may not be the case that $f([a, b])$ equals $[f(a), f(b)]$. For example, if $f(x) = x^2$, $f([-2, 2]) = [0, 4]$ and $[f(-2), f(2)] = \{4\}$. We derive two consequences of this theorem.

Let $f([a, b]) = [m, M]$. Then we have two reals c and d in $[a, b]$, such that $f(c) = m$ and $f(d) = M$. In other words, the sup is attained in

$$M = \sup\{f(x) : a \le x \le b\} = \max\{f(x) : a \le x \le b\}$$

and the inf is attained in

$$m = \inf\{f(x) : a \le x \le b\} = \min\{f(x) : a \le x \le b\}.$$

More succintly, M is a max and m is a min for the set $f([a, b])$. Thus *a continuous function on a compact interval attains its sup and its inf.* Of course, this is not generally true on noncompact intervals because $f(x) = 1/x$ has no max on $(0, 1]$.

A second consequence is: suppose that L is an intermediate value between $f(a)$ and $f(b)$. Then there must be a c, $a < c < b$, satisfying $f(c) = L$. This follows because $f(a)$ and $f(b)$ are two reals in $f([a, b])$, and $f([a, b])$ is an interval. Thus *a continuous function on a compact interval attains every intermediate value.* This is the *intermediate value property*.

On the other hand, the two consequences, the existence of the max and the min and the intermediate value property, combine to yield the theorem. To see this, let $m = f(c)$ and $M = f(d)$ denote the max and the min, with $c, d \in [a, b]$. If $m = M$, f is constant, hence $f([a, b]) = [m, M]$. If $m < M$ and $m < L < M$, apply the intermediate value property to conclude that there is an x between c and d with $f(x) = L$. Hence $f([a, b]) = [m, M]$. Thus to derive the theorem, it is enough to derive the two consequences.

For the first, let $M = \sup f([a, b])$. By Theorem 1.5.4, there is a sequence (x_n) in $[a, b]$ such that $f(x_n) \to M$. But by Theorem 2.1.1, (x_n) subconverges to some $c \in [a, b]$. By continuity, $(f(x_n))$ subconverges to $f(c)$. Because $(f(x_n))$ also converges to M, $M = f(c)$, so f has a max. Similarly for the min.

For the second, suppose that $f(a) < f(b)$, and let L be an intermediate value, $f(a) < L < f(b)$. We proceed as in the construction of $\sqrt{2}$ in §1.4. Let $S = \{x \in [a, b] : f(x) < L\}$, and let $c = \sup S$. S is nonempty because $a \in S$, and S is clearly bounded. By Theorem 1.5.4, select a sequence (x_n) in S converging to c, $x_n \to c$. By continuity, it follows that $f(x_n) \to f(c)$. Because $f(x_n) < L$ for all $n \geq 1$, we obtain $f(c) \leq L$. On the other hand, $c + 1/n$ is not in S, hence $f(c + 1/n) \geq L$. Because $c + 1/n \to c$, we obtain $f(c) \geq L$. Thus $f(c) = L$. The case $f(a) > f(b)$ is similar or is established by applying the preceding to $-f$. □

From this theorem, it follows that a continuous function maps open intervals to intervals. However, they need not be open. For example, with $f(x) = x^2$, $f((-2, 2)) = [0, 4)$. However, a function that is continuous and strictly monotone maps open intervals to open intervals (Exercise **2.3.3**).

The above theorem is the result of compactness mixed with continuity. This mixture yields other dividends. Let $f : (a, b) \to \mathbf{R}$ be given, and fix a subset $A \subset (a, b)$. For $\delta > 0$, set

$$\mu_A(\delta) = \sup\{\mu_c(\delta) : c \in A\}.$$

This is the *uniform modulus of continuity of f on A*. Because $\mu_c(\delta)$ is an increasing function of δ for each $c \in A$, it follows that $\mu_A(\delta)$ is an increasing function of δ, and hence $\mu_A(0+)$ exists. We say f is *uniformly continuous on A* if $\mu_A(0+) = 0$. When $A = (a, b)$ equals the whole domain of the function, we delete the subscript A and write $\mu(\delta)$ for the uniform modulus of continuity of f on its domain.

Whereas continuity is a property pertaining to the behavior of a function at (or near) a given point c, uniform continuity is a property pertaining to the behavior of f near a given set A. Moreover, because $\mu_c(\delta) \leq \mu_A(\delta)$, uniform continuity on A implies continuity at every point $c \in A$.

Inserting the definition of $\mu_c(\delta)$ in $\mu_A(\delta)$ yields

$$\mu_A(\delta) = \sup\{|f(x) - f(c)| : |x - c| < \delta, a < x < b, c \in A\},$$

where, now, the sup is over both x and c.

For example, for $f(x) = x^2$, the uniform modulus $\mu_A(\delta)$ over $A = (1, 10)$ equals the sup of $|x^2 - y^2|$ over all $1 < x < y < 10$ with $y - x < \delta$. But this is largest when $y = x + \delta$, hence $\mu_A(\delta)$ is the sup of $\delta^2 + 2x\delta$ over $1 < x < 10 - \delta$ which yields $\mu_A(\delta) = 20\delta - \delta^2$. In fact, this is correct only if $0 < \delta \leq 9$. For $\delta = 9$, the sup is already over all of $(1, 10)$, hence cannot get any bigger. Hence $\mu_A(\delta) = 99$ for $\delta \geq 9$. Summarizing, for $f(x) = x^2$ and $A = (1, 10)$,

$$\mu_A(\delta) = \begin{cases} 20\delta - \delta^2, & 0 < \delta \leq 9, \\ 99, & \delta \geq 9. \end{cases}$$

Because f is uniformly continuous on A if $\mu_A(0+) = 0$, in practice one needs to compute $\mu_A(\delta)$ only for δ small enough. For example, to check uniform continuity of $f(x) = x^2$ over $A = (1, 10)$, it is enough to note that $\mu_A(\delta) = 20\delta - \delta^2$ for small enough δ, which clearly approaches zero as $\delta \to 0+$.

Now let $f : (a, b) \to \mathbf{R}$ be continuous, and fix $A \subset (a, b)$. What additional conditions on f are needed to guarantee uniform continuity on A? When A is a finite set $\{c_1, \cdots, c_N\}$,

$$\mu_A(\delta) = \max\left[\mu_{c_1}(\delta), \cdots, \mu_{c_N}(\delta)\right],$$

and, hence f is necessarily uniformly continuous on A.

When A is an infinite set, this need not be so. For example, with $f(x) = x^2$ and $B = (0, \infty)$, $\mu_B(\delta)$ equals the sup of $\mu_c(\delta) = 2c\delta + \delta^2$ over $0 < c < \infty$, or $\mu_B(\delta) = \infty$, for each $\delta > 0$. Hence f is not uniformly continuous on B.

It turns out that continuity on a compact interval is sufficient for uniform continuity.

Theorem 2.3.5. *If $f : [a, b] \to \mathbf{R}$ is continuous, then f is uniformly continuous on (a, b).*

To see this, suppose that $\mu(0+) = \mu_{(a,b)}(0+) > 0$, and set $\epsilon = \mu(0+)/2$. Because μ is increasing, $\mu(1/n) \geq 2\epsilon$, $n \geq 1$. Hence for each $n \geq 1$, by the definition of the sup in the definition of $\mu(1/n)$, there is a $c_n \in (a, b)$ with $\mu_{c_n}(1/n) > \epsilon$. Now by the definition of the sup in $\mu_{c_n}(1/n)$, for each $n \geq 1$, there is an $x_n \in (a, b)$ with $|x_n - c_n| < 1/n$ and $|f(x_n) - f(c_n)| > \epsilon$. By compactness, (x_n) subconverges to some $x \in [a, b]$. Because $|x_n - c_n| < 1/n$ for all $n \geq 1$, (c_n) subconverges to the same x. Hence by continuity,

$(|f(x_n) - f(c_n)|)$ subconverges to $|f(x) - f(x)| = 0$, which contradicts the fact that this last sequence is bounded below by $\epsilon > 0$. □

The conclusion may be false if f is continuous on (a, b) but not on $[a, b]$ (see Exercise **2.3.23**). One way to understand the difference between continuity and uniform continuity is as follows.

Let f be a continuous function defined on an interval (a, b), and pick $c \in (a, b)$. Then by definition of μ_c, $|f(x) - f(c)| \leq \mu_c(\delta)$ whenever x lies in the interval $(c - \delta, c + \delta)$. Setting $g(x) = f(c)$ for $x \in (c - \delta, c + \delta)$, we see that, for any error tolerance ϵ, by choosing δ satisfying $\mu_c(\delta) < \epsilon$, we obtain a constant function g approximating f to within ϵ, at least in the interval $(c-\delta, c+\delta)$. Of course, in general, we do not expect to approximate f closely by one and the same constant function over the whole interval (a, b). Instead, we use piecewise constant functions.

We say $g : (a, b) \to \mathbf{R}$ is *piecewise constant* if there is a partition $a = x_0 < x_1 < \cdots < x_n < x_{n+1} = b$, such that g restricted to (x_{i-1}, x_i) is constant for $i = 1, \cdots, n + 1$ (in this definition, the values of g at the points x_i are not restricted in any way). The *mesh* δ of the partition $a = x_0 < x_1 < \cdots < x_{n+1} = b$, by definition, is the largest length of the subintervals, $\delta = \max_{1 \leq i \leq n+1} |x_i - x_{i-1}|$. Note that an interval has partitions of arbitrarily small mesh iff the interval is bounded.

Let $f : [a, b] \to \mathbf{R}$ be continuous. Then from above, f is uniformly continuous on (a, b). Given a partition $a = x_0 < x_1 < \cdots < x_{n+1} = b$ with mesh δ, choose $x_i^\#$ in (x_{i-1}, x_i) arbitrarily, $i = 1, \cdots, n + 1$. Then by definition of μ, $|f(x) - f(x_i^\#)| \leq \mu(\delta)$ for $x \in (x_{i-1}, x_i)$. If we set $g(x) = f(x_i^\#)$ for $x \in (x_{i-1}, x_i)$, $i = 1, \cdots, n+1$, and $g(x_i) = f(x_i)$, $i = 0, 1, \cdots, n+1$, we obtain a piecewise constant function $g : [a, b] \to \mathbf{R}$ satisfying $|f(x)-g(x)| \leq \mu(\delta)$ for every $x \in [a, b]$. Because f is uniformly continuous, $\mu(0+) = 0$. Hence for any error tolerance $\epsilon > 0$, we can find a mesh δ, such that $\mu(\delta) < \epsilon$. We have derived the following (Figure 2.7).

Theorem 2.3.6. *If $f : [a, b] \to \mathbf{R}$ is continuous, then for each $\epsilon > 0$, there is a piecewise constant function $f_\epsilon : [a, b] \to \mathbf{R}$, such that*

$$|f(x) - f_\epsilon(x)| \leq \epsilon, \quad a \leq x \leq b. \quad \square$$

If f is continuous on an open interval, this result may be false. For example, $f(x) = 1/x$, $0 < x < 1$, cannot be approximated as above by a piecewise constant function (unless infinitely many subintervals are used) precisely because f "shoots up to ∞" near 0.

Let us turn to the continuity of compositions (§1.1). Suppose that $f : (a, b) \to \mathbf{R}$ and $g : (c, d) \to \mathbf{R}$ are given with the range of f lying in the domain of g, $f[(a, b)] \subset (c, d)$. Then the *composition* $g \circ f : (a, b) \to \mathbf{R}$ is given by

$$(g \circ f)(x) = g[f(x)], \quad a < x < b.$$

Theorem 2.3.7. *If f and g are continuous, so is $g \circ f$.*

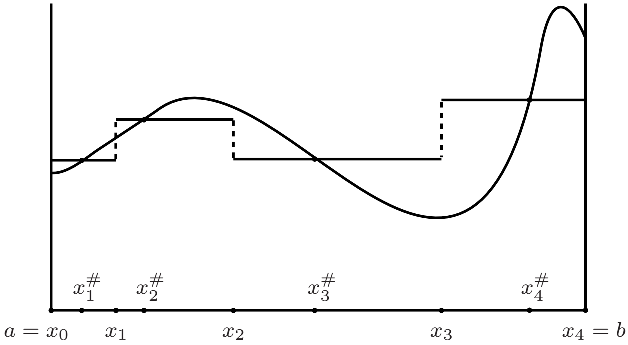

Fig. 2.7 Piecewise constant approximation.

Because f is continuous, $x_n \to c$ implies $f(x_n) \to f(c)$. Because g is continuous, $(g \circ f)(x_n) = g[f(x_n)] \to g[f(c)] = (g \circ f)(c)$. □

This result can be written

$$\lim_{x \to c} g[f(x)] = g\left[\lim_{x \to c} f(x)\right].$$

Because $g(x) = |x|$ is continuous, this implies

$$\lim_{x \to c} |f(x)| = \left|\lim_{x \to c} f(x)\right|.$$

The final issue is the invertibility of continuous functions. Let $f : [a, b] \to [m, M]$ be a continuous function. When is there an inverse (§1.1) $g : [m, M] \to [a, b]$? If it exists, is the inverse g necessarily continuous? It turns out that the answers to these questions are related to the monotonicity properties (§2.2) of the continuous function. For example, if f is continuous and increasing on $[a, b]$ and $A \subset [a, b]$, $\sup f(A) = f(\sup A)$, and $\inf f(A) = f(\inf A)$ (Exercise **2.3.4**). It follows that the upper and lower limits of $(f(x_n))$ are $f(x^*)$ and $f(x_*)$, respectively, where x^*, x_* are the upper and lower limits of (x_n) (Exercise **2.3.5**).

Theorem 2.3.8 (Inverse Function Theorem). *Let $f : [a, b] \to \mathbf{R}$ be continuous. Then f is injective iff f is strictly monotone. In this case, let $[m, M] = f([a, b])$. Then the inverse $g : [m, M] \to [a, b]$ is continuous and strictly monotone.*

If f is strictly monotone and $x \neq x'$, then $x < x'$ or $x > x'$ which implies $f(x) < f(x')$ or $f(x) > f(x')$, hence f is injective.

Conversely, suppose that f is injective and $f(a) < f(b)$. We claim that f is strictly increasing (Figure 2.8). To see this, suppose not and choose $a \leq x < x' \leq b$ with $f(x) > f(x')$. There are two possibilities: either $f(a) < f(x)$ or $f(a) \geq f(x)$. In the first case, we can choose L in $(f(a), f(x)) \cap (f(x'), f(x))$.

By the intermediate value property there are c, d with $a < c < x < d < x'$ with $f(c) = L = f(d)$. Because f is injective, this cannot happen, ruling out the first case. In the second case we must have $f(x') < f(b)$, hence $x' < b$, so we choose L in $(f(x'), f(x)) \cap (f(x'), f(b))$. By the intermediate value property, there are c, d with $x < c < x' < d < b$ with $f(c) = L = f(d)$. Because f is injective, this cannot happen, ruling out the second case. Thus f is strictly increasing. If $f(a) > f(b)$, applying what we just learned to $-f$ yields $-f$ strictly increasing or f strictly decreasing. Thus in either case, f is strictly monotone.

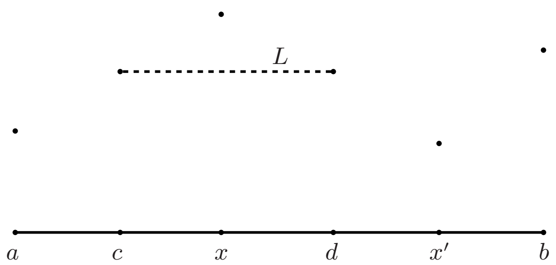

Fig. 2.8 Derivation of the IFT when $f(a) < f(b)$.

Clearly strict monotonicity of f implies that of g. Now assume that f is strictly increasing, the case with f strictly decreasing being entirely similar. We have to show that g is continuous. Suppose that $(y_n) \subset [m, M]$ with $y_n \to y$. Let $x = g(y)$, let $x_n = g(y_n)$, $n \geq 1$, and let x^* and x_* denote the upper and lower limits of (x_n). We have to show $g(y_n) = x_n \to x = g(y)$. Because f is continuous and increasing, $f(x^*)$ and $f(x_*)$ are the upper and lower limits of $y_n = f(x_n)$ (Exercise **2.3.5**). Hence $f(x^*) = y = f(x_*)$. Hence by injectivity, $x^* = x = x_*$. □

As an application, note that $f(x) = x^2$ is strictly increasing on $[0, n]$, hence has an inverse $g_n(x) = \sqrt{x}$ on $[0, n^2]$, for each $n \geq 1$. By uniqueness of inverses (Exercise **1.1.4**), the functions g_n, $n \geq 1$, agree wherever their domains overlap, hence yield a single, continuous, strictly monotone $g : [0, \infty) \to [0, \infty)$ satisfying $g(x) = \sqrt{x}$, $x \geq 0$. Similarly, for each $n \geq 1$, $f(x) = x^n$ is strictly increasing on $[0, \infty)$. Thus every positive real x has a unique positive nth root $x^{1/n}$, and, moreover, the function $g(x) = x^{1/n}$ is continuous on $[0, \infty)$. By composition, it follows that $f(x) = x^{m/n} = (x^m)^{1/n}$ is continuous and strictly monotone on $[0, \infty)$ for all naturals m, n. Because $x^{-a} = 1/x^a$ for $a \in \mathbf{Q}$, we see that the *power functions* $f(x) = x^r$ are defined, strictly increasing, and continuous on $(0, \infty)$ for all rationals r. Moreover, $x^{r+s} = x^r x^s$, $(x^r)^s = x^{rs}$ for r, s rational, and, for $r > 0$ rational, $x^r \to 0$ as $x \to 0$ and $x^r \to \infty$ as $x \to \infty$. The following limit is important: for $x > 0$,

$$\lim_{n \nearrow \infty} x^{1/n} = 1. \tag{2.3.3}$$

To derive this, assume $x \geq 1$. Then $x \leq xx^{1/n} = x^{(n+1)/n}$, so $x^{1/(n+1)} \leq x^{1/n}$, so the sequence $(x^{1/n})$ is decreasing and bounded below by 1, hence its limit $L \geq 1$ exists. Because $L \leq x^{1/2n}$, $L^2 \leq x^{2/2n} = x^{1/n}$, hence $L^2 \leq L$ or $L \leq 1$. We conclude that $L = 1$. If $0 < x < 1$, then $1/x > 1$, so $x^{1/n} = 1/(1/x)^{1/n} \to 1$ as $n \nearrow \infty$.

Any function that can be obtained from polynomials or rational functions by arithmetic operations and/or the taking of roots is called a *(constructible) algebraic function*. For example,

$$f(x) = \frac{1}{\sqrt{x(1-x)}}, \quad 0 < x < 1,$$

is an algebraic function.

We now know what a^b means for any $a > 0$ and $b \in \mathbf{Q}$. But what if $b \notin \mathbf{Q}$? What does $2^{\sqrt{2}}$ mean? To answer this, fix $a > 1$ and $b > 0$, and let

$$c = \sup\{a^r : 0 < r < b, r \in \mathbf{Q}\}.$$

Let us check that when b is rational, $c = a^b$. Because $r < s$ implies $a^r < a^s$, $a^r \leq a^b$ when $r < b$. Hence $c \leq a^b$. Similarly, $c \geq a^{b-1/n} = a^b/a^{1/n}$ for all $n \geq 1$. Let $n \nearrow \infty$ and use (2.3.3) to get $c \geq a^b$. Hence $c = a^b$ when b is rational. Thus it is consistent to *define*, for any $a > 1$ and *real* $b > 0$,

$$a^b = \sup\{a^r : 0 < r < b, r \in \mathbf{Q}\},$$

$a^0 = 1$, and $a^{-b} = 1/a^b$. For all b real, we define $1^b = 1$, whereas for $0 < a < 1$, we define $a^b = 1/(1/a)^b$. This defines $a^b > 0$ for all positive real a and all real b. Moreover (Exercise **2.3.7**),

$$a^b = \inf\{a^s : s > b, s \in \mathbf{Q}\}.$$

Theorem 2.3.9. *The power function a^b satisfies*

A. *For $a > 1$ and $0 < b < c$ real, $1 < a^b < a^c$.*
B. *For $0 < a < 1$ and $0 < b < c$ real, $a^b > a^c$.*
C. *For $0 < a < b$ and $c > 0$ real, $a^c b^c = (ab)^c$, $(b/a)^c = b^c/a^c$, and $a^c < b^c$.*
D. *For $a > 0$ and b, c real, $a^{b+c} = a^b a^c$.*
E. *For $a > 0$, b, c real, $a^{bc} = \left(a^b\right)^c$.*

Because $A \subset B$ implies $\sup A \leq \sup B$, $a^b \leq a^c$ when $a > 1$ and $b < c$. Because, for any $b < c$, there is an $r \in \mathbf{Q} \cap (b, c)$, $a^b < a^c$, thus the first assertion. Because, for $0 < a < 1$, $a^b = 1/(1/a)^b$, applying the first assertion to $1/a$ yields $(1/a)^b < (1/a)^c$ or $a^b > a^c$, yielding the second assertion. For the third, assume $a > 1$. If $0 < r < c$ is in $\mathbf{Q}$, then $a^r < a^c$ and $b^r < b^c$ yields $(ab)^r = a^r b^r < a^c b^c$. Taking the sup over $r < c$ yields $(ab)^c \leq a^c b^c$. If $r < c$ and $s < c$ are positive rationals, let t denote their max. Then $a^r b^s \leq a^t b^t = (ab)^t < (ab)^c$. Taking the sup of this last inequality over all

$0 < r < c$, first, then over all $0 < s < c$ yields $a^c b^c \le (ab)^c$. Hence $(ab)^c = a^c b^c$ for $b > a > 1$. Using this, we obtain $(b/a)^c a^c = b^c$ or $(b/a)^c = b^c/a^c$. Because $b/a > 1$ implies $(b/a)^c > 1$, we also obtain $a^c < b^c$. The cases $a < b < 1$ and $a < 1 < b$ follow from the case $b > a > 1$. This establishes the third. For the fourth, the case $0 < a < 1$ follows from the case $a > 1$, so assume $a > 1$, $b > 0$, and $c > 0$. If $r < b$ and $s < c$ are positive rationals, then $a^{b+c} \ge a^{r+s} = a^r a^s$. Taking the sups over r and s yields $a^{b+c} \ge a^b a^c$. If $r < b + c$ is rational, let $d = (b+c-r)/3 > 0$. Pick rationals t and s with $b > t > b-d$, $c > s > c-d$. Then $t+s > b+c-2d > r$, so $a^r < a^{t+s} = a^t a^s \le a^b a^c$. Taking the sup over all such r, we obtain $a^{b+c} \le a^b a^c$. This establishes the fourth when b and c are positive. The cases $b \le 0$ or $c \le 0$ follow from the positive case. The fifth involves approximating b and c by rationals, and we leave it to the reader. □

As an application, we define the power function with an irrational exponent. This is a nonalgebraic or *transcendental* function. Some of the transcendental functions in this book are the power function x^a (when a is irrational), the exponential function a^x, the logarithm $\log_a x$, the trigonometric functions and their inverses, and the gamma function. The trigonometric functions are discussed in §3.5, the gamma function in §5.1, whereas the power, exponential, and logarithm functions are discussed below.

Theorem 2.3.10. *Let a be real, and let $f(x) = x^a$ on $(0, \infty)$. For $a > 0$, f is strictly increasing and continuous with $f(0+) = 0$ and $f(\infty) = \infty$. For $a < 0$, f is strictly decreasing and continuous with $f(0+) = \infty$ and $f(\infty) = 0$.*

Because $x^{-a} = 1/x^a$, the second part follows from the first, so assume $a > 0$. Let r, s be positive rationals with $r < a < s$, and let $x_n \to c$. We have to show that $x_n^a \to c^a$. But the sequence (x_n^a) lies between (x_n^r) and (x_n^s). Because we already know that the rational power function is continuous, we conclude that the upper and lower limits L^*, L_*, of (x_n^a) satisfy $c^r \le L_* \le L^* \le c^s$. Taking the sup over all r rational and the inf over all s rational, with $r < a < s$, gives $L^* = L_* = c^a$. Thus f is continuous. Also because $x^r \to \infty$ as $x \to \infty$ and $x^r \le x^a$ for $r < a$, $f(\infty) = \infty$. Because $x^a \le x^s$ for $s > a$ and $x^s \to 0$ as $x \to 0+$, $f(0+) = 0$. □

Now we vary b and fix a in a^b.

Theorem 2.3.11. *Fix $a > 1$. Then the function $f(x) = a^x$, $x \in \mathbf{R}$, is strictly increasing and continuous. Moreover,*

$$f(x + x') = f(x)f(x'), \tag{2.3.4}$$

$f(-\infty) = 0$, $f(0) = 1$, and $f(\infty) = \infty$.

From Theorem 2.3.9, we know that f is strictly increasing. Because $a^n \nearrow \infty$ as $n \nearrow \infty$, $f(\infty) = \infty$. Because $f(-x) = 1/f(x)$, $f(-\infty) = 0$. Continuity remains to be shown. If $x_n \searrow c$, then (a^{x_n}) is decreasing and $a^{x_n} \ge a^c$, so its limit L is $\ge a^c$. On the other hand, for $d > 0$, the sequence is eventually below

$a^{c+d} = a^c a^d$, hence $L \leq a^c a^d$. Choosing $d = 1/n$, we obtain $a^c \leq L \leq a^c a^{1/n}$. Let $n \nearrow \infty$ to get $L = a^c$. Thus $a^{x_n} \searrow a^c$. If $x_n \to c+$ is not necessarily decreasing, then $x_n^* \searrow c$, hence $a^{x_n^*} \to a^c$. But $x_n^* \geq x_n$ for all $n \geq 1$, hence $a^{x_n^*} \geq a^{x_n} \geq a^c$, so $a^{x_n} \to a^c$; similarly, from the left. □

The function $f(x) = a^x$ is the *exponential function with base* $a > 1$. In fact, the exponential is the unique continuous function f on $\mathbf{R}$ satisfying the functional equation (2.3.4) and $f(1) = a$.

By the inverse function theorem, f has an inverse g on any compact interval, hence on $\mathbf{R}$. We call g the *logarithm with base* $a > 1$, and write $g(x) = \log_a x$. By definition of inverse, $a^{\log_a x} = x$, for $x > 0$, and $\log_a(a^x) = x$, for $x \in \mathbf{R}$.

Theorem 2.3.12. *The inverse of the exponential $f(x) = a^x$ with base $a > 1$ is the logarithm with base $a > 1$, $g(x) = \log_a x$. The logarithm is continuous and strictly increasing on $(0, \infty)$. The domain of $\log_a$ is $(0, \infty)$, the range is $\mathbf{R}$, $\log_a(0+) = -\infty$, $\log_a 1 = 0$, $\log_a \infty = \infty$, and*

$$\log_a(bc) = \log_a b + \log_a c,$$
$$\log_a(b^c) = c \log_a b,$$

for $b > 0$, $c > 0$.

This follows immediately from the properties of the exponential function with base $a > 1$. □

Exercises

2.3.1. If f is a polynomial of odd degree, then $f(\pm\infty) = \pm\infty$ or $f(\pm\infty) = \mp\infty$, and there is at least one real c with $f(c) = 0$.

2.3.2. If f is continuous at c, then $\mu_c(0+) = 0$.

2.3.3. If $f : (a, b) \to \mathbf{R}$ is continuous, then $f((a, b))$ is an interval. In addition, if f is strictly monotone, $f((a, b))$ is an open interval.

2.3.4. If f is continuous and increasing on $[a, b]$ and $A \subset [a, b]$, then $\sup f(A) = f(\sup A)$, and $\inf f(A) = f(\inf A)$.

2.3.5. With f as in Exercise **2.3.4**, let x^* and x_* be the upper and lower limits of a sequence (x_n). Then $f(x^*)$ and $f(x_*)$ are the upper and lower limits of $(f(x_n))$.

2.3.6. With $r, s \in \mathbf{Q}$ and $x > 0$, show that $(x^r)^s = x^{rs}$ and $x^{r+s} = x^r x^s$.

2.3.7. Show that $a^b = \inf\{a^s : s > b, s \in \mathbf{Q}\}$.

2.3.8. With b and c real and $a > 0$, show that $(a^b)^c = a^{bc}$.

2.3.9. Fix $a > 0$. If $f : \mathbf{R} \to \mathbf{R}$ is continuous, $f(1) = a$, and $f(x + x') = f(x)f(x')$ for $x, x' \in \mathbf{R}$, then $f(x) = a^x$.

2.3.10. Use the ϵ–δ criterion to show that $f(x) = 1/x$ is continuous at $x = 1$.

2.3.11. A real x is *algebraic* if x is a root of a polynomial of degree $d \geq 1$,

$$a_0 x^d + a_1 x^{d-1} + \cdots + a_{d-1} x + a_d = 0,$$

with rational coefficients $a_0, a_1, \cdots, a_d$. A real is *transcendental* if it is not algebraic. For example, every rational is algebraic. Show that the set of algebraic numbers is countable (§1.7). Conclude that the set of transcendental numbers is uncountable.

2.3.12. Let a be an algebraic number. If $f(a) = 0$ for some polynomial f with rational coefficients, but $g(a) \neq 0$ for any polynomial g with rational coefficients of lesser degree, then f is a *minimal polynomial* for a, and the degree of f is the *algebraic order* of a. Now suppose that a is algebraic of order $d \geq 2$. Show that all the roots of a minimal polynomial f are irrational.

2.3.13. Suppose that the algebraic order of a is $d \geq 2$. Then there is a $c > 0$, such that

$$\left| a - \frac{m}{n} \right| \geq \frac{c}{n^d}, \qquad n, m \geq 1.$$

(See Exercise **1.4.9**. Here you'll need the modulus of continuity μ_a at a of $g(x) = f(x)/(x - a)$, where f is a minimal polynomial of a.)

2.3.14. Use the previous exercise to show that

$$.1100010\cdots010\cdots = \frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^6} + \cdots = \sum_{n=1}^{\infty} \frac{1}{10^{n!}}$$

is transcendental.

2.3.15. For $s > 1$ real, $\sum_{n=1}^{\infty} n^{-s}$ converges.

2.3.16. If $a > 1$, $b > 0$, and $c > 0$, then $b^{\log_a c} = c^{\log_a b}$, and

$$\sum_{n=1}^{\infty} \frac{1}{5^{\log_3 n}}$$

converges.

2.3.17. Give an example of an $f : [0, 1] \to [0, 1]$ that is invertible but not monotone.

2.3.18. Let f be of bounded variation (Exercise **2.2.4**) on (a, b). Then the set of points at which f is not continuous is at most countable. Moreover, every discontinuity, at worst, is a jump.

2.3.19. Let $f : (a, b) \to \mathbf{R}$ be continuous and let $M = \sup\{f(x) : a < x < b\}$. Assume $f(a+)$ exists with $f(a+) < M$ and $f(b-)$ exists with $f(b-) < M$. Then the sup is attained,

$$\sup\{f(x) : a < x < b\} = \max\{f(x) : a < x < b\}.$$

Use Theorem 2.1.2.

2.3.20. If $f : \mathbf{R} \to \mathbf{R}$ satisfies

$$\lim_{x \to \pm\infty} \frac{f(x)}{|x|} = +\infty,$$

we say that f is *superlinear*. If f is superlinear and continuous, then the sup is attained in

$$g(y) = \sup_{-\infty < x < \infty} [xy - f(x)] = \max_{-\infty < x < \infty} [xy - f(x)],$$

and g is superlinear. Use Exercise **2.3.19**.

2.3.21. If $f : \mathbf{R} \to \mathbf{R}$ is superlinear and continuous and g is as above, then g is also continuous. (Modify the logic of the previous solution.)

2.3.22. Let $f(x) = 1 + \lfloor x \rfloor - x$, $x \in \mathbf{R}$, where $\lfloor x \rfloor$ denotes the greatest integer $\leq x$ (Figure 2.4). Compute

$$\lim_{n \nearrow \infty} \left(\lim_{m \nearrow \infty} [f(n!x)]^m \right)$$

for $x \in \mathbf{Q}$ and for $x \notin \mathbf{Q}$.

2.3.23. Let $f(x) = 1/x$, $0 < x < 1$. Compute $\mu_c(\delta)$ explicitly for $0 < c < 1$ and $\delta > 0$. With $I = (0, 1)$, show that $\mu_I(\delta) = \infty$ for all $\delta > 0$. Conclude that f is not uniformly continuous on $(0, 1)$. (There are two cases, $c \leq \delta$ and $c > \delta$.)

2.3.24. Let $f : \mathbf{R} \to \mathbf{R}$ be continuous, and suppose that $f(\infty)$ and $f(-\infty)$ exist and are finite. Show that f is uniformly continuous on $\mathbf{R}$.

2.3.25. Use $\sqrt{2}^{\sqrt{2}}$ to show that there are irrationals a, b, such that a^b is rational.

http://www.springer.com/978-1-4419-9487-5

Introduction to Calculus and Classical Analysis
Hijab, O.
2011, XII, 364 p., Hardcover
ISBN: 978-1-4419-9487-5