Chapter 3

Continuous Functions

In this chapter, we define continuous functions and study their properties.

3.1. Continuity

According to the definition introduced by Cauchy, and developed by Weierstrass, continuous functions are functions that take nearby values at nearby points.

Definition 3.1. Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, and suppose that $c \in A$. Then f is continuous at c if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|x - c| < \delta \text{ and } x \in A \text{ implies that } |f(x) - f(c)| < \epsilon.$$

A function $f : A \to \mathbb{R}$ is continuous on a set $B \subset A$ if it is continuous at every point in B, and continuous if it is continuous at every point of its domain A.

The definition of continuity at a point may be stated in terms of neighborhoods as follows.

Definition 3.2. A function $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, is continuous at $c \in A$ if for every neighborhood V of $f(c)$ there is a neighborhood U of c such that

$$x \in A \cap U \text{ implies that } f(x) \in V.$$

The ϵ-δ definition corresponds to the case when V is an ϵ-neighborhood of $f(c)$ and U is a δ-neighborhood of c. We leave it as an exercise to prove that these definitions are equivalent.

Note that c must belong to the domain A of f in order to define the continuity of f at c. If c is an isolated point of A, then the continuity condition holds automatically since, for sufficiently small $\delta > 0$, the only point $x \in A$ with $|x - c| < \delta$ is $x = c$, and then $0 = |f(x) - f(c)| < \epsilon$. Thus, a function is continuous at every isolated point of its domain, and isolated points are not of much interest.

If $c \in A$ is an accumulation point of A, then continuity of f at c is equivalent to the condition that

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

meaning that the limit of f as $x \to c$ exists and is equal to the value of f at c.

Example 3.3. If $f : (a, b) \to \mathbb{R}$ is defined on an open interval, then f is continuous on (a, b) if and only if

$$\lim_{x \to c} f(x) = f(c) \qquad \text{for every } a < c < b$$

since every point of (a, b) is an accumulation point.

Example 3.4. If $f : [a, b] \to \mathbb{R}$ is defined on a closed, bounded interval, then f is continuous on $[a, b]$ if and only if

$$\begin{aligned} &\lim_{x \to c} f(x) = f(c) &&\text{for every } a < c < b, \\ &\lim_{x \to a^+} f(x) = f(a), &&\lim_{x \to b^-} f(x) = f(b). \end{aligned}$$

Example 3.5. Suppose that

$$A = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots\right\}$$

and $f : A \to \mathbb{R}$ is defined by

$$f(0) = y_0, \qquad f\left(\frac{1}{n}\right) = y_n$$

for some values $y_0, y_n \in \mathbb{R}$. Then $1/n$ is an isolated point of A for every $n \in \mathbb{N}$, so f is continuous at $1/n$ for every choice of y_n. The remaining point $0 \in A$ is an accumulation point of A, and the condition for f to be continuous at 0 is that

$$\lim_{n \to \infty} y_n = y_0.$$

As for limits, we can give an equivalent sequential definition of continuity, which follows immediately from Theorem 2.4.

Theorem 3.6. If $f : A \to \mathbb{R}$ and $c \in A$ is an accumulation point of A, then f is continuous at c if and only if

$$\lim_{n \to \infty} f(x_n) = f(c)$$

for every sequence (x_n) in A such that $x_n \to c$ as $n \to \infty$.

In particular, f is discontinuous at $c \in A$ if there is sequence (x_n) in the domain A of f such that $x_n \to c$ but $f(x_n) \not\to f(c)$.

Let's consider some examples of continuous and discontinuous functions to illustrate the definition.

Example 3.7. The function $f : [0, \infty) \to \mathbb{R}$ defined by $f(x) = \sqrt{x}$ is continuous on $[0, \infty)$. To prove that f is continuous at $c > 0$, we note that for $0 \le x < \infty$,

$$|f(x) - f(c)| = \left|\sqrt{x} - \sqrt{c}\right| = \left|\frac{x - c}{\sqrt{x} + \sqrt{c}}\right| \le \frac{1}{\sqrt{c}}|x - c|,$$

so given $\epsilon > 0$, we can choose $\delta = \sqrt{c}\epsilon > 0$ in the definition of continuity. To prove that f is continuous at 0, we note that if $0 \leq x < \delta$ where $\delta = \epsilon^2 > 0$, then

$$|f(x) - f(0)| = \sqrt{x} < \epsilon.$$

Example 3.8. The function $\sin : \mathbb{R} \to \mathbb{R}$ is continuous on $\mathbb{R}$. To prove this, we use the trigonometric identity for the difference of sines and the inequality $|\sin x| \leq |x|$:

$$\begin{aligned} |\sin x - \sin c| &= \left| 2\cos\left(\frac{x+c}{2}\right) \sin\left(\frac{x-c}{2}\right) \right| \\ &\leq 2 \left| \sin\left(\frac{x-c}{2}\right) \right| \\ &\leq |x - c|. \end{aligned}$$

It follows that we can take $\delta = \epsilon$ in the definition of continuity for every $c \in \mathbb{R}$.

Example 3.9. The sign function $\operatorname{sgn} : \mathbb{R} \to \mathbb{R}$, defined by

$$\operatorname{sgn} x = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0, \end{cases}$$

is not continuous at 0 since $\lim_{x\to 0} \operatorname{sgn} x$ does not exist (see Example 2.6). The left and right limits of sgn at 0,

$$\lim_{x\to 0^-} f(x) = -1, \qquad \lim_{x\to 0^+} f(x) = 1,$$

do exist, but they are unequal. We say that f has a jump discontinuity at 0.

Example 3.10. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1/x & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is not continuous at 0 since $\lim_{x\to 0} f(x)$ does not exist (see Example 2.7). Neither the left or right limits of f at 0 exist either, and we say that f has an essential discontinuity at 0.

Example 3.11. The function $f : \mathbb{R} \to \mathbb{R}$, defined by

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

is continuous at $c \neq 0$ (see Example 3.20 below) but discontinuous at 0 because $\lim_{x\to 0} f(x)$ does not exist (see Example 2.8).

Example 3.12. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x\sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

is continuous at every point of $\mathbb{R}$. (See Figure 1. The continuity at $c \neq 0$ is proved in Example 3.21 below. To prove continuity at 0, note that for $x \neq 0$,

$$|f(x) - f(0)| = |x \sin(1/x)| \leq |x|,$$

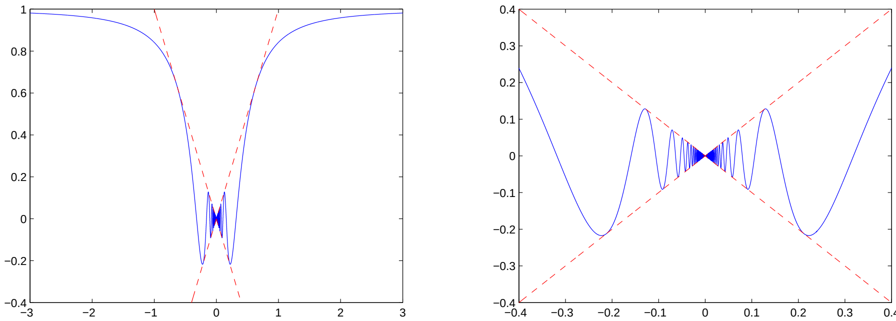

Figure 1. A plot of the function $y = x\sin(1/x)$ and a detail near the origin with the lines $y = \pm x$ shown in red.

so $f(x) \to f(0)$ as $x \to 0$. If we had defined $f(0)$ to be any value other than 0, then f would not be continuous at 0. In that case, f would have a removable discontinuity at 0.

Example 3.13. The Dirichlet function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

is discontinuous at every $c \in \mathbb{R}$. If $c \notin \mathbb{Q}$, choose a sequence (x_n) of rational numbers such that $x_n \to c$ (possible since $\mathbb{Q}$ is dense in $\mathbb{R}$). Then $x_n \to c$ and $f(x_n) \to 1$ but $f(c) = 0$. If $c \in \mathbb{Q}$, choose a sequence (x_n) of irrational numbers such that $x_n \to c$; for example if $c = p/q$, we can take

$$x_n = \frac{p}{q} + \frac{\sqrt{2}}{n},$$

since $x_n \in \mathbb{Q}$ would imply that $\sqrt{2} \in \mathbb{Q}$. Then $x_n \to c$ and $f(x_n) \to 0$ but $f(c) = 1$. In fact, taking a rational sequence (x_n) and an irrational sequence $(\tilde{x}_n)$ that converge to c, we see that $\lim_{x\to c} f(x)$ does not exist for any $c \in \mathbb{R}$.

Example 3.14. The Thomae function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1/q & \text{if } x = p/q \text{ where } p \text{ and } q > 0 \text{ are relatively prime}, \\ 0 & \text{if } x \notin \mathbb{Q} \text{ or } x = 0 \end{cases}$$

is continuous at 0 and every irrational number and discontinuous at every nonzero rational number. See Figure 2 for a plot.

We can give a rough classification of a discontinuity of a function $f : A \to \mathbb{R}$ at an accumulation point $c \in A$ as follows.

(1) *Removable discontinuity*: $\lim_{x\to c} f(x) = L$ exists but $L \neq f(c)$, in which case we can make f continuous at c by redefining $f(c) = L$ (see Example 3.12).

(2) *Jump discontinuity*: $\lim_{x\to c} f(x)$ doesn't exist, but both the left and right limits $\lim_{x\to c^-} f(x)$, $\lim_{x\to c^+} f(x)$ exist and are different (see Example 3.9).

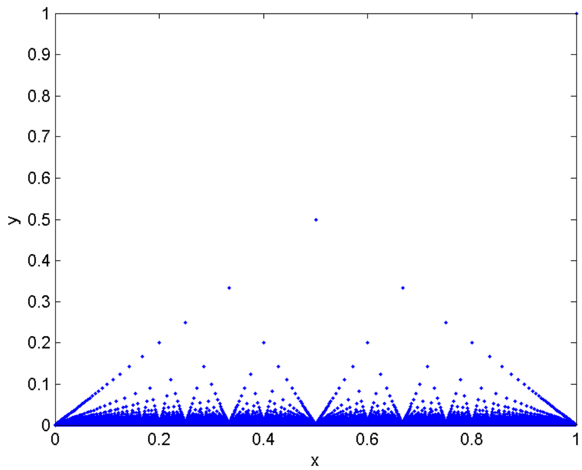

Figure 2. A plot of the Thomae function on $[0, 1]$

(3) *Essential discontinuity*: $\lim_{x\to c} f(x)$ doesn't exist and at least one of the left or right limits $\lim_{x\to c^-} f(x)$, $\lim_{x\to c^+} f(x)$ doesn't exist (see Examples 3.10, 3.11, 3.13).

3.2. Properties of continuous functions

The basic properties of continuous functions follow from those of limits.

Theorem 3.15. If $f, g : A \to \mathbb{R}$ are continuous at $c \in A$ and $k \in \mathbb{R}$, then kf, $f+g$, and fg are continuous at c. Moreover, if $g(c) \neq 0$ then f/g is continuous at c.

Proof. This result follows immediately Theorem 2.22. □

A polynomial function is a function $P : \mathbb{R} \to \mathbb{R}$ of the form

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where $a_0, a_1, a_2, \ldots, a_n$ are real coefficients. A rational function R is a ratio of polynomials P, Q

$$R(x) = \frac{P(x)}{Q(x)}.$$

The domain of R is the set of points in $\mathbb{R}$ such that $Q \neq 0$.

Corollary 3.16. Every polynomial function is continuous on $\mathbb{R}$ and every rational function is continuous on its domain.

Proof. The constant function $f(x) = 1$ and the identity function $g(x) = x$ are continuous on $\mathbb{R}$. Repeated application of Theorem 3.15 for scalar multiples, sums, and products implies that every polynomial is continuous on $\mathbb{R}$. It also follows that a rational function $R = P/Q$ is continuous at every point where $Q \neq 0$. □

Example 3.17. The function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \frac{x + 3x^3 + 5x^5}{1 + x^2 + x^4}$$

is continuous on $\mathbb{R}$ since it is a rational function whose denominator never vanishes.

In addition to forming sums, products and quotients, another way to build up more complicated functions from simpler functions is by composition.

We recall that if $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ where $f(A) \subset B$, meaning that the domain of g contains the range of f, then we define the composition $g \circ f : A \to \mathbb{R}$ by

$$(g \circ f)(x) = g\left(f(x)\right).$$

The next theorem states that the composition of continuous functions is continuous. Note carefully the points at which we assume f and g are continuous.

Theorem 3.18. *Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ where $f(A) \subset B$. If f is continuous at $c \in A$ and g is continuous at $f(c) \in B$, then $g \circ f : A \to \mathbb{R}$ is continuous at c.*

Proof. Let $\epsilon > 0$ be given. Since g is continuous at $f(c)$, there exists $\eta > 0$ such that

$$|y - f(c)| < \eta \text{ and } y \in B \text{ implies that } |g(y) - g\left(f(c)\right)| < \epsilon.$$

Next, since f is continuous at c, there exists $\delta > 0$ such that

$$|x - c| < \delta \text{ and } x \in A \text{ implies that } |f(x) - f(c)| < \eta.$$

Combing these inequalities, we get that

$$|x - c| < \delta \text{ and } x \in A \text{ implies that } |g\left(f(x)\right) - g\left(f(c)\right)| < \epsilon,$$

which proves that $g \circ f$ is continuous at c. □

Corollary 3.19. *Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ where $f(A) \subset B$. If f is continuous on A and g is continuous on $f(A)$, then $g \circ f$ is continuous on A.*

Example 3.20. The function

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

is continuous on $\mathbb{R} \setminus \{0\}$, since it is the composition of $x \mapsto 1/x$, which is continuous on $\mathbb{R} \setminus \{0\}$, and $y \mapsto \sin y$, which is continuous on $\mathbb{R}$.

Example 3.21. The function

$$f(x) = \begin{cases} x\sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

is continuous on $\mathbb{R} \setminus \{0\}$ since it is a product of functions that are continuous on $\mathbb{R} \setminus \{0\}$. As shown in Example 3.12, f is also continuous at 0, so f is continuous on $\mathbb{R}$.

3.3. Uniform continuity

Uniform continuity is a subtle but powerful strengthening of continuity.

Definition 3.22. Let $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$. Then f is uniformly continuous on A if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|x - y| < \delta \text{ and } x, y \in A \text{ implies that } |f(x) - f(y)| < \epsilon.$$

The key point of this definition is that δ depends only on ϵ, not on x, y. A uniformly continuous function on A is continuous at every point of A, but the converse is not true, as we explain next.

If a function is continuous on A, then given $\epsilon > 0$ there exists $\delta(c) > 0$ for every $c \in A$ such that

$$|x - c| < \delta(c) \text{ and } x \in A \text{ implies that } |f(x) - f(c)| < \epsilon.$$

In general, $\delta(c)$ depends on both ϵ and c, but we don't show the ϵ-dependence explicitly since we're thinking of ϵ as fixed. If

$$\inf_{c \in A} \delta(c) = 0$$

however we choose $\delta(c)$, then no $\delta_0 > 0$ depending only on ϵ works simultaneously for all $c \in A$. In that case, the function is continuous on A but not uniformly continuous.

Before giving examples, we state a sequential condition for uniform continuity to fail.

Proposition 3.23. A function $f : A \to \mathbb{R}$ is not uniformly continuous on A if and only if there exists $\epsilon_0 > 0$ and sequences (x_n), (y_n) in A such that

$$\lim_{n \to \infty} |x_n - y_n| = 0 \text{ and } |f(x_n) - f(y_n)| \geq \epsilon_0 \text{ for all } n \in \mathbb{N}.$$

Proof. If f is not uniformly continuous, then there exists $\epsilon_0 > 0$ such that for every $\delta > 0$ there are points $x, y \in A$ with $|x - y| < \delta$ and $|f(x) - f(y)| \geq \epsilon_0$. Choosing $x_n, y_n \in A$ to be any such points for $\delta = 1/n$, we get the required sequences.

Conversely, if the sequential condition holds, then for every $\delta > 0$ there exists $n \in \mathbb{N}$ such that $|x_n - y_n| < \delta$ and $|f(x_n) - f(y_n)| \geq \epsilon_0$. It follows that the uniform continuity condition in Definition 3.22 cannot hold for any $\delta > 0$ if $\epsilon = \epsilon_0$, so f is not uniformly continuous. □

Example 3.24. Example 3.8 shows that the sine function is uniformly continuous on $\mathbb{R}$, since we can take $\delta = \epsilon$ for every $x, y \in \mathbb{R}$.

Example 3.25. Define $f : [0, 1] \to \mathbb{R}$ by $f(x) = x^2$. Then f is uniformly continuous on $[0, 1]$. To prove this, note that for all $x, y \in [0, 1]$ we have

$$\left|x^2 - y^2\right| = |x + y| \, |x - y| \leq 2|x - y|,$$

so we can take $\delta = \epsilon/2$ in the definition of uniform continuity. Similarly, $f(x) = x^2$ is uniformly continuous on any bounded set.

Example 3.26. The function $f(x) = x^2$ is continuous but not uniformly continuous on $\mathbb{R}$. We have already proved that f is continuous on $\mathbb{R}$ (it's a polynomial). To prove that f is not uniformly continuous, let

$$x_n = n, \qquad y_n = n + \frac{1}{n}.$$

Then

$$\lim_{n\to\infty} |x_n - y_n| = \lim_{n\to\infty} \frac{1}{n} = 0,$$

but

$$|f(x_n) - f(y_n)| = \left(n + \frac{1}{n}\right)^2 - n^2 = 2 + \frac{1}{n^2} \geq 2 \qquad \text{for every } n \in \mathbb{N}.$$

It follows from Proposition 3.23 that f is not uniformly continuous on $\mathbb{R}$. The problem here is that, for given $\epsilon > 0$, we need to make $\delta(c)$ smaller as c gets larger to prove the continuity of f at c, and $\delta(c) \to 0$ as $c \to \infty$.

Example 3.27. The function $f : (0, 1] \to \mathbb{R}$ defined by

$$f(x) = \frac{1}{x}$$

is continuous but not uniformly continuous on $(0, 1]$. We have already proved that f is continuous on $(0, 1]$ (it's a rational function whose denominator x is nonzero in $(0, 1]$). To prove that f is not uniformly continuous, define $x_n, y_n \in (0, 1]$ for $n \in \mathbb{N}$ by

$$x_n = \frac{1}{n}, \qquad y_n = \frac{1}{n+1}.$$

Then $x_n \to 0$, $y_n \to 0$, and $|x_n - y_n| \to 0$ as $n \to \infty$, but

$$|f(x_n) - f(y_n)| = (n + 1) - n = 1 \qquad \text{for every } n \in \mathbb{N}.$$

It follows from Proposition 3.23 that f is not uniformly continuous on $(0, 1]$. The problem here is that, for given $\epsilon > 0$, we need to make $\delta(c)$ smaller as c gets closer to 0 to prove the continuity of f at c, and $\delta(c) \to 0$ as $c \to 0^+$.

The non-uniformly continuous functions in the last two examples were unbounded. However, even bounded continuous functions can fail to be uniformly continuous if they oscillate arbitrarily quickly.

Example 3.28. Define $f : (0, 1] \to \mathbb{R}$ by

$$f(x) = \sin\left(\frac{1}{x}\right)$$

Then f is continuous on $(0, 1]$ but it isn't uniformly continuous on $(0, 1]$. To prove this, define $x_n, y_n \in (0, 1]$ for $n \in \mathbb{N}$ by

$$x_n = \frac{1}{2n\pi}, \qquad y_n = \frac{1}{2n\pi + \pi/2}.$$

Then $x_n \to 0$, $y_n \to 0$, and $|x_n - y_n| \to 0$ as $n \to \infty$, but

$$|f(x_n) - f(y_n)| = \sin\left(2n\pi + \frac{\pi}{2}\right) - \sin 2n\pi = 1 \qquad \text{for all } n \in \mathbb{N}.$$

It isn't a coincidence that these examples of non-uniformly continuous functions have a domain that is either unbounded or not closed. We will prove in Section 3.5 that a continuous function on a closed, bounded set is uniformly continuous.

3.4. Continuous functions and open sets

Let $f : A \to \mathbb{R}$ be a function. Recall that if $B \subset A$, the set

$$f(B) = \{y \in \mathbb{R} : y = f(x) \text{ for some } x \in B\}$$

is called the image of B under f, and if $C \subset \mathbb{R}$, the set

$$f^{-1}(C) = \{x \in A : f(x) \in C\}$$

is called the inverse image or preimage of C under f. Note that $f^{-1}(C)$ is a well-defined set even if the function f does not have an inverse.

Example 3.29. Suppose $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2$. If $I = (1, 4)$, then

$$f(I) = (1, 16), \qquad f^{-1}(I) = (-2, -1) \cup (1, 2).$$

Note that we get two intervals in the preimage because f is two-to-one on $f^{-1}(I)$. If $J = (-1, 1)$, then

$$f(J) = [0, 1), \qquad f^{-1}(J) = (-1, 1).$$

In the previous example, the preimages of the open sets I, J under the continuous function f are open, but the image of J under f isn't open. Thus, a continuous function needn't map open sets to open sets. As we will show, however, the inverse image of an open set under a continuous function is always open. This property is the topological definition of a continuous function; it is a global definition in the sense that it implies that the function is continuous at every point of its domain.

Recall from Section 1.2 that a subset B of a set $A \subset \mathbb{R}$ is relatively open in A, or open in A, if $B = A \cap U$ where U is open in $\mathbb{R}$. Moreover, as stated in Proposition 1.22, B is relatively open in A if and only if every point $x \in B$ has a relative neighborhood $C = A \cap V$ such that $C \subset B$, where V is a neighborhood of x in $\mathbb{R}$.

Theorem 3.30. A function $f : A \to \mathbb{R}$ is continuous on A if and only if $f^{-1}(V)$ is open in A for every set V that is open in $\mathbb{R}$.

Proof. First assume that f is continuous on A, and suppose that $c \in f^{-1}(V)$. Then $f(c) \in V$ and since V is open it contains an ϵ-neighborhood

$$V_\epsilon\left(f(c)\right) = \left(f(c) - \epsilon, f(c) + \epsilon\right)$$

of $f(c)$. Since f is continuous at c, there is a δ-neighborhood

$$U_\delta(c) = (c - \delta, c + \delta)$$

of c such that

$$f\left(A \cap U_\delta(c)\right) \subset V_\epsilon\left(f(c)\right).$$

This statement just says that if $|x - c| < \delta$ and $x \in A$, then $|f(x) - f(c)| < \epsilon$. It follows that

$$A \cap U_\delta(c) \subset f^{-1}(V),$$

meaning that $f^{-1}(V)$ contains a relative neighborhood of c. Therefore $f^{-1}(V)$ is relatively open in A.

Conversely, assume that $f^{-1}(V)$ is open in A for every open V in $\mathbb{R}$, and let $c \in A$. Then the preimage of the ϵ-neighborhood $(f(c)-\epsilon, f(c)+\epsilon)$ is open in A, so it contains a relative δ-neighborhood $A \cap (c-\delta, c+\delta)$. It follows that $|f(x)-f(c)| < \epsilon$ if $|x-c| < \delta$ and $x \in A$, which means that f is continuous at c. □

3.5. Continuous functions on compact sets

Continuous functions on compact sets have especially nice properties. For example, they are bounded and attain their maximum and minimum values, and they are uniformly continuous. Since a closed, bounded interval is compact, these results apply, in particular, to continuous functions $f : [a, b] \to \mathbb{R}$.

First we prove that the continuous image of a compact set is compact.

Theorem 3.31. If $f : K \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}$ is compact, then $f(K)$ is compact.

Proof. We show that $f(K)$ is sequentially compact. Let (y_n) be a sequence in $f(K)$. Then $y_n = f(x_n)$ for some $x_n \in K$. Since K is compact, the sequence (x_n) has a convergent subsequence (x_{n_i}) such that

$$\lim_{i\to\infty} x_{n_i} = x$$

where $x \in K$. Since f is continuous on K,

$$\lim_{i\to\infty} f(x_{n_i}) = f(x).$$

Writing $y = f(x)$, we have $y \in f(K)$ and

$$\lim_{i\to\infty} y_{n_i} = y.$$

Therefore every sequence (y_n) in $f(K)$ has a convergent subsequence whose limit belongs to $f(K)$, so $f(K)$ is compact.

Let us also give an alternative proof based on the Heine-Borel property. Suppose that $\{V_i : i \in I\}$ is an open cover of $f(K)$. Since f is continuous, Theorem 3.30 implies that $f^{-1}(V_i)$ is open in K, so $\{f^{-1}(V_i) : i \in I\}$ is an open cover of K. Since K is compact, there is a finite subcover

$$\left\{f^{-1}(V_{i_1}), f^{-1}(V_{i_2}), \dots, f^{-1}(V_{i_N})\right\}$$

of K, and it follows that

$$\{V_{i_1}, V_{i_2}, \dots, V_{i_N}\}$$

is a finite subcover of the original open cover of $f(K)$. This proves that $f(K)$ is compact. □

Note that compactness is essential here; it is not true, in general, that a continuous function maps closed sets to closed sets.

Example 3.32. Define $f : [0,\infty) \to \mathbb{R}$ by

$$f(x) = \frac{1}{1+x^2}.$$

Then $[0,\infty)$ is closed but $f([0,\infty)) = (0,1]$ is not.

The following result is the most important property of continuous functions on compact sets.

Theorem 3.33 (Weierstrass extreme value)**.** If $f : K \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}$ is compact, then f is bounded on K and f attains its maximum and minimum values on K.

Proof. Since $f(K)$ is compact, Theorem 1.40 implies that it is bounded, which means that f is bounded on K. Proposition 1.41 implies that the maximum M and minimum m of $f(K)$ belong to $f(K)$. Therefore there are points $x, y \in K$ such that $f(x) = M$, $f(y) = m$, and f attains its maximum and minimum on K. □

Example 3.34. Define $f : [0,1] \to \mathbb{R}$ by

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

Then f is unbounded on $[0,1]$ and has no maximum value (f does, however, have a minimum value of 0 attained at $x = 0$). In this example, $[0,1]$ is compact but f is discontinuous at 0, which shows that a discontinuous function on a compact set needn't be bounded.

Example 3.35. Define $f : (0,1] \to \mathbb{R}$ by $f(x) = 1/x$. Then f is unbounded on $(0,1]$ with no maximum value (f does, however, have a minimum value of 1 attained at $x = 1$). In this example, f is continuous but the half-open interval $(0,1]$ isn't compact, which shows that a continuous function on a non-compact set needn't be bounded.

Example 3.36. Define $f : (0,1) \to \mathbb{R}$ by $f(x) = x$. Then

$$\inf_{x\in(0,1)} f(x) = 0, \qquad \sup_{x\in(0,1)} f(x) = 1$$

but $f(x) \neq 0$, $f(x) \neq 1$ for any $0 < x < 1$. Thus, even if a continuous function on a non-compact set is bounded, it needn't attain its supremum or infimum.

Example 3.37. Define $f : [0, 2/\pi] \to \mathbb{R}$ by

$$f(x) = \begin{cases} x + x\sin(1/x) & \text{if } 0 < x \leq 2/\pi, \\ 0 & \text{if } x = 0. \end{cases}$$

(See Figure 3.) Then f is continuous on the compact interval $[0, 2/\pi]$, so by Theorem 3.33 it attains its maximum and minimum. For $0 \leq x \leq 2/\pi$, we have $0 \leq f(x) \leq 1/\pi$ since $|\sin 1/x| \leq 1$. Thus, the minimum value of f is 0, attained at $x = 0$. It is also attained at infinitely many other interior points in the interval,

$$x_n = \frac{1}{2n\pi + 3\pi/2}, \qquad n = 0, 1, 2, 3, \ldots,$$

where $\sin(1/x_n) = -1$. The maximum value of f is $1/\pi$, attained at $x = 2/\pi$.

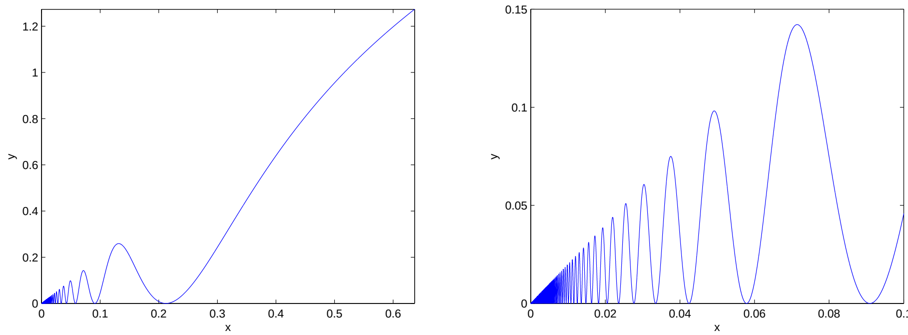

Figure 3. A plot of the function $y = x + x \sin(1/x)$ on $[0, 2/\pi]$ and a detail near the origin.

Finally, we prove that continuous functions on compact sets are uniformly continuous

Theorem 3.38. If $f : K \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}$ is compact, then f is uniformly continuous on K.

Proof. Suppose for contradiction that f is not uniformly continuous on K. Then from Proposition 3.23 there exists $\epsilon_0 > 0$ and sequences (x_n), (y_n) in K such that

$$\lim_{n\to\infty} |x_n - y_n| = 0 \text{ and } |f(x_n) - f(y_n)| \geq \epsilon_0 \text{ for every } n \in \mathbb{N}.$$

Since K is compact, there is a convergent subsequence (x_{n_i}) of (x_n) such that

$$\lim_{i\to\infty} x_{n_i} = x \in K.$$

Moreover, since $(x_n - y_n) \to 0$ as $n \to \infty$, it follows that

$$\lim_{i\to\infty} y_{n_i} = \lim_{i\to\infty} [x_{n_i} - (x_{n_i} - y_{n_i})] = \lim_{i\to\infty} x_{n_i} - \lim_{i\to\infty} (x_{n_i} - y_{n_i}) = x,$$

so (y_{n_i}) also converges to x. Then, since f is continuous on K,

$$\lim_{i\to\infty} |f(x_{n_i}) - f(y_{n_i})| = \left| \lim_{i\to\infty} f(x_{n_i}) - \lim_{i\to\infty} f(y_{n_i}) \right| = |f(x) - f(x)| = 0,$$

but this contradicts the non-uniform continuity condition

$$|f(x_{n_i}) - f(y_{n_i})| \geq \epsilon_0.$$

Therefore f is uniformly continuous. $\square$

Example 3.39. The function $f : [0, 2/\pi] \to \mathbb{R}$ defined in Example 3.37 is uniformly continuous on $[0, 2/\pi]$ since it is is continuous and $[0, 2/\pi]$ is compact.

3.6. The intermediate value theorem

The intermediate value theorem states that a continuous function on an interval takes on all values between any two of its values. We first prove a special case.

Theorem 3.40. Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function on a closed, bounded interval. If $f(a) < 0$ and $f(b) > 0$, or $f(a) > 0$ and $f(b) < 0$, then there is a point $a < c < b$ such that $f(c) = 0$.

Proof. Assume for definiteness that $f(a) < 0$ and $f(b) > 0$. (If $f(a) > 0$ and $f(b) < 0$, consider $-f$ instead of f.) The set

$$E = \{x \in [a, b] : f(x) < 0\}$$

is nonempty, since $a \in E$, and E is bounded from above by b. Let

$$c = \sup E \in [a, b],$$

which exists by the completeness of $\mathbb{R}$. We claim that $f(c) = 0$.

Suppose for contradiction that $f(c) \neq 0$. Since f is continuous at c, there exists $\delta > 0$ such that

$$|x - c| < \delta \text{ and } x \in [a, b] \text{ implies that } |f(x) - f(c)| < \frac{1}{2}|f(c)|.$$

If $f(c) < 0$, then $c \neq b$ and

$$f(x) = f(c) + f(x) - f(c) < f(c) - \frac{1}{2}f(c)$$

for all $x \in [a, b]$ such that $|x - c| < \delta$, so $f(x) < \frac{1}{2}f(c) < 0$. It follows that there are points $x \in E$ with $x > c$, which contradicts the fact that c is an upper bound of E.

If $f(c) > 0$, then $c \neq a$ and

$$f(x) = f(c) + f(x) - f(c) > f(c) - \frac{1}{2}f(c)$$

for all $x \in [a, b]$ such that $|x - c| < \delta$, so $f(x) > \frac{1}{2}f(c) > 0$. It follows that there exists $\eta > 0$ such that $c - \eta \geq a$ and

$$f(x) > 0 \text{ for } c - \eta \leq x \leq c.$$

In that case, $c - \eta < c$ is an upper bound for E, since c is an upper bound and $f(x) > 0$ for $c - \eta \leq x \leq c$, which contradicts the fact that c is the least upper bound. This proves that $f(c) = 0$. Finally, $c \neq a, b$ since f is nonzero at the endpoints, so $a < c < b$. $\square$

We give some examples to show that all of the hypotheses in this theorem are necessary.

Example 3.41. Let $K = [-2, -1] \cup [1, 2]$ and define $f : K \to \mathbb{R}$ by

$$f(x) = \begin{cases} -1 & \text{if } -2 \leq x \leq -1 \\ 1 & \text{if } 1 \leq x \leq 2 \end{cases}$$

Then $f(-2) < 0$ and $f(2) > 0$, but f doesn't vanish at any point in its domain. Thus, in general, Theorem 3.40 fails if the domain of f is not a connected interval $[a, b]$.

Example 3.42. Define $f : [-1, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} -1 & \text{if } -1 \le x < 0 \\ 1 & \text{if } 0 \le x \le 1 \end{cases}$$

Then $f(-1) < 0$ and $f(1) > 0$, but f doesn't vanish at any point in its domain. Here, f is defined on an interval but it is discontinuous at 0. Thus, in general, Theorem 3.40 fails for discontinuous functions.

Example 3.43. Define the continuous function $f : [1, 2] \to \mathbb{R}$ by

$$f(x) = x^2 - 2.$$

Then $f(1) < 0$ and $f(2) > 0$, so Theorem 3.40 implies that there exists $1 < c < 2$ such that $c^2 = 2$. Moreover, since $x^2 - 2$ is strictly increasing on $[0, \infty)$, there is a unique such positive number, so we have proved the existence of $\sqrt{2}$.

We can get more accurate approximations to $\sqrt{2}$ by repeatedly bisecting the interval $[1, 2]$. For example $f(3/2) = 1/4 > 0$ so $1 < \sqrt{2} < 3/2$, and $f(5/4) < 0$ so $5/4 < \sqrt{2} < 3/2$, and so on. This bisection method is a simple, but useful, algorithm for computing numerical approximations of solutions of $f(x) = 0$ where f is a continuous function.

Note that we used the existence of a supremum in the proof of Theorem 3.40. If we restrict $f(x) = x^2 - 2$ to rational numbers, $f : A \to \mathbb{Q}$ where $A = [1, 2] \cap \mathbb{Q}$, then f is continuous on A, $f(1) < 0$ and $f(2) > 0$, but $f(c) \ne 0$ for any $c \in A$ since $\sqrt{2}$ is irrational. This shows that the completeness of $\mathbb{R}$ is essential for Theorem 3.40 to hold. (Thus, in a sense, the theorem actually describes the completeness of the continuum $\mathbb{R}$ rather than the continuity of f!)

The general statement of the Intermediate Value Theorem follows immediately from this special case.

Theorem 3.44 (Intermediate value theorem)**.** Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function on a closed, bounded interval. Then for every d strictly between $f(a)$ and $f(b)$ there is a point $a < c < b$ such that $f(c) = d$.

Proof. Suppose, for definiteness, that $f(a) < f(b)$ and $f(a) < d < f(b)$. (If $f(a) > f(b)$ and $f(b) < d < f(a)$, apply the same proof to $-f$, and if $f(a) = f(b)$ there is nothing to prove.) Let $g(x) = f(x) - d$. Then $g(a) < 0$ and $g(b) > 0$, so Theorem 3.40 implies that $g(c) = 0$ for some $a < c < b$, meaning that $f(c) = d$. □

As one consequence of our previous results, we prove that a continuous function maps compact intervals to compact intervals.

Theorem 3.45. Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function on a closed, bounded interval. Then $f([a, b]) = [m, M]$ is a closed, bounded interval.

Proof. Theorem 3.33 implies that $m \le f(x) \le M$ for all $x \in [a, b]$, where m and M are the maximum and minimum values of f, so $f([a, b]) \subset [m, M]$. Moreover, there are points $c, d \in [a, b]$ such that $f(c) = m$, $f(d) = M$.

Let $J = [c, d]$ if $c \le d$ or $J = [d, c]$ if $d < c$. Then $J \subset [a, b]$, and Theorem 3.44 implies that f takes on all values in $[m, M]$ on J. It follows that $f([a, b]) \supset [m, M]$, so $f([a, b]) = [m, M]$. □

First we give an example to illustrate the theorem.

Example 3.46. Define $f : [-1, 1] \to \mathbb{R}$ by

$$f(x) = x - x^3.$$

Then, using calculus to compute the maximum and minimum of f, we find that

$$f([-1, 1]) = [-M, M], \qquad M = \frac{2}{3\sqrt{3}}.$$

This example illustrates that $f([a, b]) \neq [f(a), f(b)]$ unless f is increasing.

Next we give some examples to show that the continuity of f and the connectedness and compactness of the interval $[a, b]$ are essential for Theorem 3.45 to hold.

Example 3.47. Let $\operatorname{sgn} : [-1, 1] \to \mathbb{R}$ be the sign function defined in Example 2.6. Then f is a discontinuous function on a compact interval $[-1, 1]$, but the range $f([-1, 1]) = \{-1, 0, 1\}$ consists of three isolated points and is not an interval.

Example 3.48. In Example 3.41, the function $f : K \to \mathbb{R}$ is continuous on a compact set K but $f(K) = \{-1, 1\}$ consists of two isolated points and is not an interval.

Example 3.49. The continuous function $f : [0, \infty) \to \mathbb{R}$ in Example 3.32 maps the unbounded, closed interval $[0, \infty)$ to a half-open interval $(0, 1]$.

The last example shows that a continuous function may map a closed but unbounded interval to an interval which isn't closed (or open). Nevertheless, it follows from the fact that a continuous function maps compact intervals to compact intervals that it maps intervals to intervals (where the intervals may be open, closed, half-open, bounded, or unbounded). We omit a detailed proof.

3.7. Monotonic functions

Monotonic functions have continuity properties that are not shared by general functions.

Definition 3.50. Let $I \subset \mathbb{R}$ be an interval. A function $f : I \to \mathbb{R}$ is increasing if

$$f(x_1) \leq f(x_2) \qquad \text{if } x_1, x_2 \in I \text{ and } x_1 < x_2,$$

strictly increasing if

$$f(x_1) < f(x_2) \qquad \text{if } x_1, x_2 \in I \text{ and } x_1 < x_2,$$

decreasing if

$$f(x_1) \geq f(x_2) \qquad \text{if } x_1, x_2 \in I \text{ and } x_1 < x_2,$$

and strictly decreasing if

$$f(x_1) > f(x_2) \qquad \text{if } x_1, x_2 \in I \text{ and } x_1 < x_2.$$

An increasing or decreasing function is called a monotonic function, and a strictly increasing or strictly decreasing function is called a strictly monotonic function.

A commonly used alternative (and, unfortunately, incompatible) terminology is "nondecreasing" for "increasing," "increasing" for "strictly increasing," "nonincreasing" for "decreasing," and "decreasing" for "strictly decreasing." According to our terminology, a constant function is both increasing and decreasing. Monotonic functions are also referred to as monotone functions.

Theorem 3.51. If $f : I \to \mathbb{R}$ is monotonic on an interval I, then the left and right limits of f,

$$\lim_{x\to c^-} f(x), \qquad \lim_{x\to c^+} f(x),$$

exist at every interior point c of I.

Proof. Assume for definiteness that f is increasing. (If f is decreasing, we can apply the same argument to $-f$ which is increasing). We will prove that

$$\lim_{x\to c^-} f(x) = \sup E, \qquad E = \{f(x) \in \mathbb{R} : x \in I \text{ and } x < c\}.$$

The set E is nonempty since c in an interior point of I, so there exists $x \in I$ with $x < c$, and E bounded from above by $f(c)$ since f is increasing. It follows that $L = \sup E \in \mathbb{R}$ exists. (Note that L may be strictly less than $f(c)$!)

Suppose that $\epsilon > 0$ is given. Since L is a least upper bound of E, there exists $y_0 \in E$ such that $L - \epsilon < y_0 \le L$, and therefore $x_0 \in I$ with $x_0 < c$ such that $f(x_0) = y_0$. Let $\delta = c - x_0 > 0$. If $c - \delta < x < c$, then $x_0 < x < c$ and therefore $f(x_0) \le f(x) \le L$ since f is increasing and L is an upper bound of E. It follows that

$$L - \epsilon < f(x) \le L \qquad \text{if } c - \delta < x < c,$$

which proves that $\lim_{x\to c^-} f(x) = L$.

A similar argument, or the same argument applied to $g(x) = -f(-x)$, shows that

$$\lim_{x\to c^+} f(x) = \inf\{f(x) \in \mathbb{R} : x \in I \text{ and } x > c\}.$$

We leave the details as an exercise. □

Similarly, if $I = [a, b]$ is a closed interval and f is monotonic on I, then the left limit $\lim_{x\to b^-} f(x)$ exists at the right endpoint, although it may not equal $f(b)$, and the right limit $\lim_{x\to a^+} f(x)$ exists at the left endpoint, although it may not equal $f(a)$.

Corollary 3.52. Every discontinuity of a monotonic function $f : I \to \mathbb{R}$ at an interior point of the interval I is a jump discontinuity.

Proof. If c is an interior point of I, then the left and right limits of f at c exist by the previous theorem. Moreover, assuming for definiteness that f is increasing, we have

$$f(x) \le f(c) \le f(y) \qquad \text{for all } x, y \in I \text{ with } x < c < y,$$

and since limits preserve inequalities

$$\lim_{x\to c^-} f(x) \le f(c) \le \lim_{x\to c^+} f(x).$$

If the left and right limits are equal, then the limit exists and is equal to the left and right limits, so

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

meaning that f is continuous at c. In particular, a monotonic function cannot have a removable discontinuity at an interior point of its domain (although it can have one at an endpoint of a closed interval). If the left and right limits are not equal, then f has a jump discontinuity at c, so f cannot have an essential discontinuity either. □

One can show that a monotonic function has, at most, a countable number of discontinuities, and it may have a countably infinite number, but we omit the proof. By contrast, the non-monotonic Dirichlet function has uncountably many discontinuities at every point of $\mathbb{R}$.