7 Continuity

Looking back at the ten functions that we used at the beginning of Section 6 to motivate the definition of the limit, we see that

- some of them — f_2, f_3, f_6, f_7, f_8, f_9 and f_{10} — did not approach a limit near the particular a's under consideration,
- while the rest of them — f_1, f_4 and f_5 — did.

These last three are definitely "nicer" near the particular a's under consideration than the first seven. But even among these last three, there is a further split:

- two of them — f_4 and f_5 — either have the property that the function is not defined *at* a, or that the function is defined, but the function value at a is different from the limit that the function is approaching near a,
- while the third — f_1 — has the function defined at a, *and* the function value equally the limit that the function is approaching near a.

This last is definitely "very nice" behavior near a; we capture precisely what's going on with the central definition of this section, that of continuity of a function at a point.

Definition of f being continuous at a A function f is *continuous* at a if

- f is defined at and near a (meaning there is $\Delta > 0$ such that all of $(a - \Delta, a + \Delta)$ is in Domain(f)), and
- $\lim_{x \to a} f(x) = f(a)$.

The sense of the definition is that near a, small changes in the input to f lead to only small changes in the output, or (quite informally), "near a, the graph of f can be drawn with taking pen off paper".

Unpacking the ε-δ definition of the limit, the continuity of a function f (that is defined at and near a) at a can be expressed as follows:

- for all $\varepsilon > 0$
- there is $\delta > 0$
- such that $|x - a| < \delta$
- implies $|f(x) - f(a)| < \varepsilon$.

Note that this is just the definition of $\lim_{x \to a} f(x) = f(a)$ with $0 < |x - a| < \delta$ changed to just $|x - a| < \delta$, or $x \in (a - \delta, a + \delta)$; we can make this change because at the one new value of x that is introduced into consideration, namely $x = a$, we certainly have $|f(x) - f(a)| < \varepsilon$ for all $\varepsilon > 0$, since in fact we have $|f(x) - f(a)| = 0$ at $x = a$. This ε-δ statement is often taken as the definition of continuity of f at a.

7.1 A collection of continuous functions

Here we build up a a large collection of functions that are continuous at all points of their domains. We have done most of the work for this already, when we discussed limits.

Constant function Let $f : \mathbb{R} \to \mathbb{R}$ be the constant function $f(x) = c$ (where $c \in \mathbb{R}$ is some constant). Since we have already established that $\lim_{x\to a} f(x) = c = f(a)$ for all a, we immediately get that f is continuous at all points in its domain.

Linear function Let $g : \mathbb{R} \to \mathbb{R}$ be the linear function $g(x) = x$. Since we have already established that $\lim_{x\to a} g(x) = a = g(a)$ for all a, we immediately get that g is continuous at all points in its domain.

Sums, products and quotients of continuous functions Suppose that f and g are both continuous at a. Then

- $f + g$ is continuous at a (proof: $f + g$ is certainly defined at and near a, if both f and g are, and by the sum/product/reciprocal theorem for limits,
$$\lim_{x\to a}(f + g)(x) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = f(a) + g(a) = (f + g)(a));$$
- fg is continuous at a (proof: fg is certainly defined at and near a, if both f and g are, and by the sum/product/reciprocal theorem for limits,
$$\lim_{x\to a}(fg)(x) = \lim_{x\to a} f(x) \lim_{x\to a} g(x) = f(a)g(a) = (fg)(a));$$
- as long as $g(a) \neq 0$, $1/g$ is continuous at a (proof: that $1/g$ is defined at and near a follows from Claim 6.2, and for the limit part of the continuity definition, we have from the reciprocal part of sum/product/reciprocal theorem for limits that
$$\lim_{x\to a}(1/g)(x) = 1/\lim_{x\to a} g(x) = 1/g(a) = (1/g)(a));$$
- as long as $g(a) \neq 0$, f/g is continuous at a (proof: combine the last two parts).

Polynomials If P is a polynomial function, then P is continuous at all reals. For any *particular* polynomial, this follows by lots of applications of the the observations above about sums and products of continuous functions, together with the continuity of the constant and linear functions (to get things started); for polynomials *in general* this follows from the same ingredients as for the particular case, together with lots of applications of prove by induction.

Rational functions If R is a rational function, then R is continuous at all points in its domain; so in particular, if $R = P/Q$ where P, Q are polynomials and Q is not the constantly zero polynomial, then R is continuous at all reals x for which $Q(x)$ is not 0. This is an application of the continuity of polynomials, as well as the reciprocal part of the sum/product/reciprocal observation.

This gives us already a large collection of continuous functions. The list becomes even larger when we include the trigonometric functions:

Working assumption The functions sin and cos are continuous at all reals.[92]

This is a reasonable assumption; if we move only slightly along the unit circle from a point $(x, y) = (\cos\theta, \sin\theta)$, the coordinates of our position only move slightly, strongly suggesting that sin and cos are both continuous.

Armed with this working assumption, we can for example immediately say (appealing to our previous observations) that

$$f(x) = \frac{(x^2+1)\sin x - x(\cos x)^2}{2(x+1)\sin x}$$

is continuous, as long as $x \neq -1$ or $x \neq n\pi$ for $n \in \mathbb{Z}$ (i.e., it's continuous as long as it's defined); indeed, f is nothing more than a combination of known continuous functions, with the means of combination being addition, subtraction, multiplication and division, all of which we have discussed vis a vis continuity.

What about a superficially similar looking function like $f(x) = \sin(1/x)$? This is clearly not continuous at $x = 0$ (it is not even defined there), but it seems quite clear that is it continuous at all other x. None of the situations we have discussed so far apply to this particular function, though, because it is constructed from simpler functions not by addition, subtraction, multiplication and division, but rather by composition.

We could try to compute $\lim_{x\to a} \sin(1/x)$ and see if it is equal to $\sin(1/a)$, but that would almost certainly be quite messy. Instread, we appeal to one more general result about continuity:

Theorem 7.1. *If f, g are functions, and if g is continuous at a and f is continuous at $g(a)$ (so in particular, g is defined at and near a, and f is defined at and near $g(a)$, then $(f \circ g)$ is continuous at a.*

Proof: Unlike previous proofs involving continuity, this one will be quite subtle. Already we have to work a little to verify that $(f \circ g)$ is defined at and near a. That it is defined *at a* is obvious. To see that it is defined near a, note that f is continuous at $g(a)$, so there is some $\Delta' > 0$ such that f is defined at all points in the interval $(g(a) - \Delta', g(a) + \Delta')$. We want to show that there is a $\Delta > 0$ such that for all $x \in (a - \Delta, a + \Delta)$, we have $g(x) \in (g(a) - \Delta', g(a) + \Delta')$ (so that then for all $x \in (a - \Delta, a + \Delta)$, we have that $(f \circ g)(x)$ is defined). But this follows from the continuity of g at a: apply the ε-δ definition of continuity, with Δ' as the input tolerance ε, and take the output δ to be Δ.

[92] This is a "working assumption" rather than a theorem; we haven't yet formally defined the trigonometric functions, and without a precise and formal definition of the functions, there is no point in even attempting a proof of continuity.

Next we move on to showing that $(f \circ g)(x) \to f(g(a))$ as $x \to a$. Given $\varepsilon > 0$, we want to say that if x is sufficiently close to a then $|f(g(x)) - f(g(a))| < \varepsilon$.

Here's the informal idea: by choosing x close enough to a, we can make $g(x)$ close to $g(a)$ (since g is continuous at a). But then, since $g(x)$ is close to $g(a)$, we must have $f(g(x))$ close to $f(g(a))$ (since f is continuous at $g(a)$).

Formally: given $\varepsilon > 0$ there is $\delta' > 0$ such that $|X - g(a)| < \delta'$ implies $|f(X) - f(g(a))| < \varepsilon$ (this is applying the definition of the continuity of f at $g(a)$, with input ε).

Now use that δ' as the input for the definition of g being continuous at a, i.e., for $g(x) \to g(a)$ as $x \to a$: we get that there is some $\delta > 0$ such that $|x - a| < \delta$ implies $|g(x) - g(a)| < \delta'$, which, by definition of δ', implies $|f(g(x)) - f(g(a))| < \varepsilon$.[93] $\square$

From this theorem, we can conclude that any function that is built from known continuous functions (such as polynomial and rational functions, or sin and cos) using addition, subtraction, multiplication, division and composition, is continuous at every point in its domain. So, for example, all of

- $\sin(1/x)$
- $x \sin(1/x)$
- $\sin^3(2x^2 + \cos x) - \frac{3x}{\cos^2 x - \sin(\sin x)}$

are all continuous wherever they are defined.

What about *discontinuous* functions? It's easy to come up with examples of functions that are discontinuous at sporadic points:

- $f(x) = x/|x|$ is discontinuous at $x = 0$ (it's not defined at 0, but even if we augment the definition of f to give it a value, it will still be discontinuous at 0, since $\lim_{x \to 0} f(x)$ does not exist);

- $f(x) = [x]$[94] is defined for all reals, but is discontinuous at infinitely many places, specifically at the infinitely many integers. Indeed, for any integer t there are values of x arbitrarily close to t for which $f(x) = t$ (any x slightly larger than t), and values of x arbitrarily close to t for which $f(x) = t - 1$ (any x slightly smaller than t), so it's an easy exercise that $\lim_{x \to t} f(x)$ doesn't exist;

- $f(x) = [1/x]$ is defined for all reals other than 0. Arbitrarily close to 0, it is discontinuous at infinitely many points (so there is a "clustering" of discontinuities close to 0). Indeed, f is easily seen to be discontinuous at 1 (across which it jumps from 2 to 1), at 1/2 (across which it jumps from 3 to 2), and more generally at $\pm 1/k$ for *every* integer k.

[93] There were only two things we could have used in this proof: the continuity of f at $g(a)$ and the continuity of g at a. The only question was, which one to use *first*? Using the continuity of g at a first would have lead us nowhere.

[94] "$[x]$" is the *floor*, or *integer part*, of x — the largest integer that is less than or equal to x. So for example $[2.1] = [2.9] = [2] = 2$ and $[-0.5] = [-.001] = [-1] = -1$.

There are even easy examples of functions that has $\mathbb{R}$ as its domain, and is discontinuous *everywhere*. One such is the *Dirichlet function* f_{10} defined earlier:

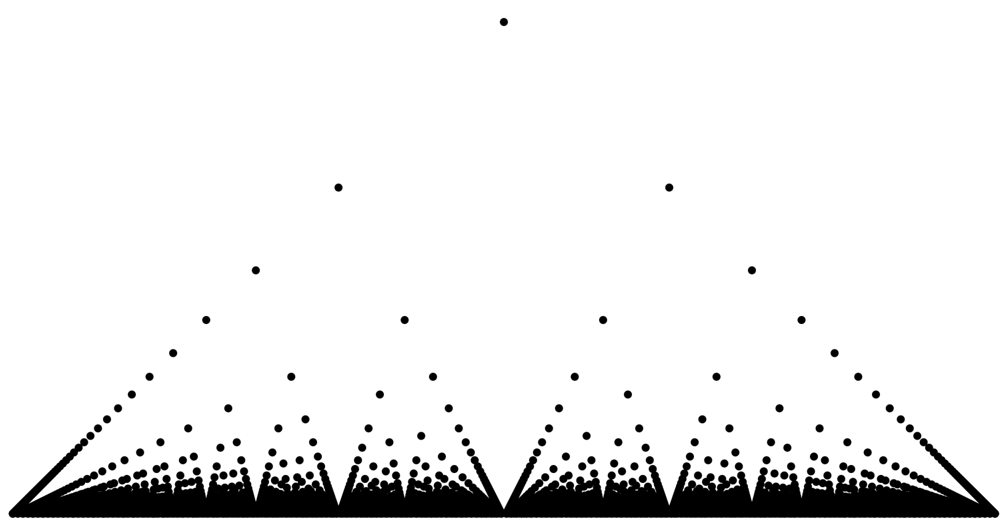
$$f_{10}(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Indeed, fix $a \in \mathbb{R}$. We claim that $\lim_{x \to a} f_{10}(x)$ does not exist. Let L be given. It must be the case that at least one of $|0 - L|, |1 - L|$ is greater than, say, $1/10$. Suppose $|1 - L| > 1/10$. Take $\varepsilon = 1/10$. Given any $\delta > 0$, in the interval $(a - \delta, a + \delta)$ there must be[95] some irrational x (other than a, which may or may not be irrational; but we don't consider a when checking for a limit existing or not). We have $f_{10}(x) = 1$, so $|f_{10}(x) - L| > 1/10 = \varepsilon$. If on the other hand $|0 - L| > 1/10$, again take $\varepsilon = 1/10$. Given any $\delta > 0$, in the interval $(a - \delta, a + \delta)$ there must be[96] some rational x (other than a, which may or may not be rational). We have $f_{10}(x) = 0$, so $|f_{10}(x) - L| > 1/10 = \varepsilon$. In either case we have the necessary witness to $\lim_{x \to a} f_{10}(x) \neq L$, and since L was arbitrary, the limit does not exist.

A rather more interesting example is the *Stars over Babylon* function.[97] We define it here just on the open interval $(0, 1)$:

$$f(x) = \begin{cases} 1/q & \text{if } x \text{ is rational, } x = p/q,\ p, q \in \mathbb{N},\ p, q \text{ have no common factors} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Here's the graph of the Stars over Babylon function:



It takes the value $1/2$ at $1/2$; at $1/3$ and $2/3$ it takes the value $1/3$; at $1/4$ and $3/4$ it takes the value $1/4$ (but not at $2/4$; that was already covered by $1/2$); at $1/5$, $2/5$, $3/5$ and $4/3$ it

[95] Musn't there be?

[96] Again, musn't there be?

[97] So named by John Conway, for it's unusual graph; it is also called *Thomae's* function, or the *popcorn* function.

takes the value $1/5$; at $1/6$ and $5/6$ it takes the value $1/6$ (but not at $2/6$, $3/6$ or $4/6$; these were already covered by $1/3$, $1/2$ and $2/3$); et cetera.

We claim that for all $a \in (0,1)$, f approaches a limit near a, and specifically f approaches the limit 0. Indeed, given $a \in (0,1)$, and given $\varepsilon > 0$, we want to find a $\delta > 0$ such that $0 < |x-a| < \delta$ implies $|f(x)| < \varepsilon$.

Now there are only *finitely many* $x \in (0,1)$ with $f(x) \geq \varepsilon$, namely

$$1/2, 1/3, 2/3, 1/4, 3/4, \ldots, 1/n, \ldots, (n-1)/n$$

where $1/n$ is the largest natural number with $1/n \geq \varepsilon$. There are certainly no more than n^2 of these numbers; call them $x_1, x_2, \ldots, x_m$, written in increasing order. As long as none of these numbers satisfy $0 < |x-a| < \delta$, then for x satisfying this bound we have $|f(x)| < \varepsilon$.

So, let δ be any positive number that is smaller than

- the distance from a to 0
- the distance from a to 1 and
- the distance from a to the closest of the x_i to a (other than a itself, which may or may not be one of the x_i; but we don't care, because we don't consider a when checking for a limit existing or not).

If $0 < |x-a| < \delta$, then, because of the first two clauses above, we have that $x \in (0,1)$, so in the domain of f; and, because of the third clause, the only number in $(a-\delta, a+\delta)$ that could be among the x_i's is a itself; so, combining, if $0 < |x-a| < \delta$ then x is *not* among the x_i's, so $|f(x)| < \varepsilon$.

This completes the proof that $\lim_{x \to a} f(x) = 0$. An interesting consequence brings us back to the topic at hand, continuity: since $f(x) = 0$ exactly when x is irrational,

Stars over Babylon is continuous at all irrationals, discontinuous at all rationals.

7.2 Continuity on an interval

Continuity at a *point* can say something about a function on an *interval*. Indeed, we have the following extremely useful fact about functions:

Claim 7.2. *Suppose f is continuous at a, and that $f(a) \neq 0$. Then there is some interval around a on which f is non-zero. Specifically, there is a $\delta > 0$ such that*

- *if $f(a) > 0$, then for all $x \in (a-\delta, a+\delta)$, $f(x) > f(a)/2$, and*
- *if $f(a) < 0$, then for all $x \in (a-\delta, a+\delta)$, $f(x) < f(a)/2$.*

We won't give a proof of this, as it is an immediate corollary of Claim 6.2, taking $M = f(a)$.

This moves us nicely along to our next main point, which is thinking about what can be said about a function that is known to be continuous not just at a point, but on an entire interval. We start with open intervals.

- Say that $f : (a, b) \to \mathbb{R}$ is *continuous on* (a, b) if it is continuous at all $c \in (a, b)$;
- say that $f : (-\infty, b) \to \mathbb{R}$ is *continuous on* $(-\infty, b)$ if it is continuous at all $c \in (-\infty, b)$;
- say that $f : (a, \infty) \to \mathbb{R}$ is *continuous on* (a, ∞) if it is continuous at all $c \in (a, \infty)$.

So, for example, the function $f(x) = 1/(x-1)(x-2)$ is continuous on the intervals $(-\infty, 1)$, $(1, 2)$ and $(2, \infty)$.

For functions defined on closed intervals, we have to be more careful, because we cannot talk about continuity at the end-points of the interval. Instead we introduce notions of one-sided continuity, using our previous notions of one-sided limits:

Definition of f being *right continuous* or *continuous from above* at a: A function f is right continuous or continuous from above at a if $\lim_{x\to a^+} f(x) = f(a)$.

Definition of f being *left continuous* or *continuous from below* at b: A function f is left continuous or continuous from below at b if $\lim_{x\to b^-} f(x) = f(b)$.

- Say that $f : [a, b] \to \mathbb{R}$ is *continuous on* $[a, b]$ if it is continuous at all $c \in (a, b)$, is right continuous at a and is left continuous at b;
- say that $f : (-\infty, b] \to \mathbb{R}$ is *continuous on* $(-\infty, b]$ if it is continuous at all $c \in (-\infty, b)$ and is left continuous at b;
- say that $f : [a, \infty) \to \mathbb{R}$ is *continuous on* $[a, \infty)$ if it is continuous at all $c \in (a, \infty)$ and is right continuous at a;
- say that $f : [a, b) \to \mathbb{R}$ is *continuous on* $[a, b)$ if it is continuous at all $c \in (a, b)$ and is right continuous at a;
- say that $f : (a, b] \to \mathbb{R}$ is *continuous on* $(a, b]$ if it is continuous at all $c \in (a, b)$ and is left continuous at b.

So, for example (easy exercises),

- the function that is defined by $f(x) = x/|x|$ away from 0 and is defined to be 1 at 0 is continuous on the intervals $(-\infty, 0)$ and $[0, \infty)$; the function, and
- $f(x) = [x]$ is continuous on all intervals of the form $[k, k+1)$, $k \in \mathbb{Z}$.

An important fact about right and left continuity is that the process of checking continuity at a is equivalent to the process of checking right and left continuity; this is an immediate corollary of Theorem 6.6:

Claim 7.3. *f is continuous at a if and only if it is both right continuous and left continuous at a.*

A quick corollary of this gives us another way to form new continuous functions from old: splicing. Suppose that f and g are defined on (a, b), and $c \in (a, b)$ has $f(c) = g(c)$. Define a new function $h : (a, b) \to \mathbb{R}$ by[98]

$$h(x) = \begin{cases} f(x) & \text{if } x \leq c \\ g(x) & \text{if } x \geq c \end{cases}$$

Corollary 7.4. *(of Claim 7.3) If f and g are both continuous at c, then h is continuous at c (and so if f, g are both continuous on (a, b), so is h).*

Proof: On $(a, c]$ h agrees with f. f is continuous at c, so is left continuous at c, and so h is left continuous at c. On $[c, b)$ h agrees with g, so right continuity of h at c follows similarly from continuity of g at c. Since h is both right and left continuous at c, it is continuous at c. □

As an example, consider the function $h(x) = |x|$. This is a splice of $f(x) = -x$ and $g(x) = x$, the splicing done at 0 (where f and g agree). Both f and g are continuous on $\mathbb{R}$, so h is continuous on $\mathbb{R}$.

7.3 The Intermediate Value Theorem

An "obvious" fact about continuous functions is that if f is continuous on $[a, b]$, with $f(a) < 0$ and $f(b) > 0$, then there must be some $c \in (a, b)$ such that $f(c) = 0$; a continuous function cannot "jump" over the x-axis.

But is this really obvious? We think of continuity at a point as meaning that the graph of the function near that point can be drawn without taking pen off paper, but the Stars over Babylon function, which is continuous at each irrational, but whose's graph near any irrational certainly *can't* be drawn without taking pen off paper, show us that we have to be careful with that intuition. The issue here, of course, is that when we say that a continuous function cannot jump over the x-axis, we are thinking about functions which are continuous at *all* points in an interval.

Here is a stronger argument for the "obvious" fact not necessarily being so obvious. Suppose that when specifying the number system we work with, we had stopped with axiom P12. Just using axioms P1-P12, we have a very nice set of numbers that we can work with — the rational numbers $\mathbb{Q}$ — inside which all usual arithmetic operations can be performed.

[98] Notice that there's no problem with the overlap of clauses here, since $f(c) = g(c)$.

If we agreed to just do our mathematics in $\mathbb{Q}$, we could still define functions, and still define the notion of a function approaching a limit, and still define the notion of a function being continuous — all of those definitions relied only on arithmetic operations (addition, subtraction, multiplication, division, comparing magnitudes) that make perfect sense in $\mathbb{Q}$. All the theorems we have proven about functions, limits and continuity would still be true.

Unfortunately, the "obvious" fact would *not* be true! The function $f : \mathbb{Q} \to \mathbb{Q}$ given by $f(x) = x^2 - 2$ is a continuous function, in the $\mathbb{Q}$-world, has $f(0) = -2 < 0$ and $f(2) = 2 > 0$, but in the $\mathbb{Q}$-world there is *no* $x \in (0, 2)$ with $x^2 = 2$ (as we have proven earlier), and so there is *no* $x \in (0, 2)$ with $f(x) = 0$: f goes from negative to positive without ever equalling 0.

So, if our "obvious" fact is true, it is as much a fact about real numbers as it is a fact about continuity, and it's proof will necessarily involve an appeal to the one axiom we introduced after P1-P12, namely the completeness axiom.

The "obvious" fact is indeed true in the $\mathbb{R}$-world, and goes under a special name:

Theorem 7.5. *(Intermediate Value Theorem, or IVT) Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function defined on a closed interval. If $f(a) < 0$ and $f(b) > 0$ then there is some $c \in (a, b)$ (so $a < c < b$) with $f(c) = 0$.*

We'll defer the proof for a while, and first make some remarks. The first remark to make is on the necessity of the hypothesis.[99]

- *Is IVT still true if f is not continuous on all of $[a, b]$?* No. Consider

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

  Viewed as, for example, a function on the closed interval $[-2, 2]$, f is continuous at all points on the interval $[-2, 2]$ *except* at 0. Also, $f(-2) < 0$ while $f(2) > 0$. But there is no $x \in (-2, 2)$ with $f(x) = 0$.

- *What if f is continuous on all of (a, b), just not at a and/or b*? Still No. Consider

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

  Viewed as, for example, a function on the closed interval $[0, 1]$, f is continuous at all points on the interval $(0, 1)$. It's also left continuous at 1. The only place where (right) continuity fails is at 0. Also, $f(0) < 0$ while $f(1) > 0$. But there is no $x \in (0, 1)$ with $f(x) = 0$.

[99] Most important theorem come with hypotheses — conditions that must be satisfied in order for the theorem to be valid (for the IVT, the hypothesis is that f is *continuous on the whole closed interval* $[a, b]$). Most of the theorems we will see have been refined over time to the point where the hypotheses being assumed are the bare minimum necessary to make the theorem true. As such, it should be possible to come up with counterexamples to the conclusions of these theorems, whenever the hypothesis are even slightly weakened. You should get into the habit of questioning the hypotheses of every big theorem we see, specifically asking yourself "is this still true if I weaken any of the hypotheses?". Usually, it will not be true anymore.

A second remark is that the IVT quickly gives us the existence of a unique square root of any positive number:

Claim 7.6. *For each $a \geq 0$ there is a unique number $a' \geq 0$ such that $(a')^2 = a$. We refer to this number as the* square root *of a, and write it either as $\sqrt{a}$ or as $a^{1/2}$.*

Proof: If $a = 0$ then we take $a' = 0$. This is the unique possibility, since as we have earlier proven, if $a' \neq 0$ then $(a')^2 > 0$, so $(a')^2 \neq 0$.

Suppose $a > 0$. Consider the function $f_a : [0, a+1] \to \mathbb{R}$ given by $f_a(x) = x^2 - a$. This is a continuous function at all points on the interval, as we have previously proven. Also $f_a(0) = -a < 0$ and $f_a(a+1) = (a+1)^2 - a = a^2 + a + 1 > 0$. So by IVT, there is $a' \in (0, a+1)$ with $f_a(a') = 0$, that is, with $(a')^2 = a$.

To prove that this a' is the *unique* possibility for the positive square root of a, note that if $0 \leq a'' < a'$ then $0 \leq (a'')^2 < (a')^2$ (this was something we proved earlier), so $(a'')^2 \neq a$, while if $0 \leq a' < a''$ then $0 \leq (a')^2 < (a'')^2$, so again $(a'')^2 \neq a$. Hence a' is indeed unique. $\square$

We can go further, with essentially no extra difficulty:

Claim 7.7. *Fix $n \geq 2$ a natural number. For each $a \geq 0$ there is a unique number $a' \geq 0$ such that $(a')^n = a$. We refer to this number as the n*th root *of a, and write it either as $\sqrt[n]{a}$ or as $a^{1/n}$.*

Proof: If $a = 0$ then we take $a' = 0$. This is the unique possibility, since if $a' \neq 0$ then $(a')^n \neq 0$.

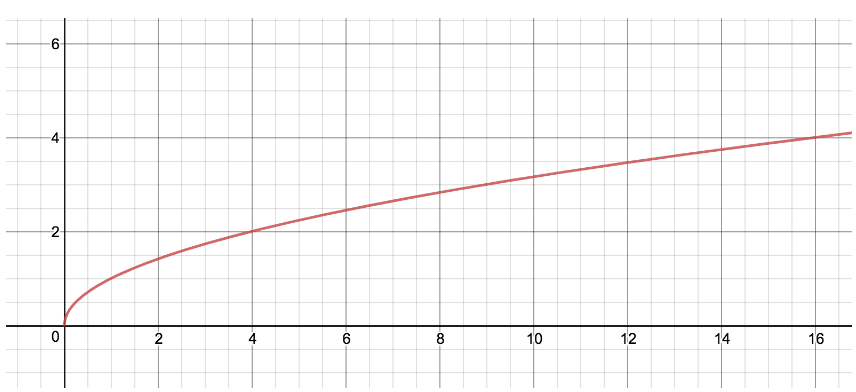
Suppose $a > 0$. Consider the function $f_a : [0, a+1] \to \mathbb{R}$ given by $f_a(x) = x^n - a$. This is a continuous function. Also $f_a(0) = -a < 0$ and (using the binomial theorem)

$$f_a(a+1) = (a+1)^n - a = a^n + \binom{n}{n-1} a^{n-1} + \cdots + \binom{n}{n-k} a^{n-k} + \cdots + \left(\binom{n}{1} - 1\right) a + 1 > 0.$$

So by IVT, there is $a' \in (0, a+1)$ with $f_a(a') = 0$, that is, with $(a')^n = a$.

To prove that this a' is the *unique* possibility for the positive nth root of a, note that if $0 \leq a'' < a'$ then $0 \leq (a'')^n < (a')^n$ while if $0 \leq a' < a''$ then $0 \leq (a')^n < (a'')^n$. $\square$

Define, for natural numbers $n \geq 2$, a function $f_n : [0, \infty) \to [0, \infty)$ by $x \mapsto x^{1/n}$. The graph of the function f_2 is shown below; it looks like it is continuous on its whole domain, and we would strongly expect f_n to be continuous on all of $[0, \infty)$, too.



Claim 7.8. *For all $n \geq 2$, $n \in \mathbb{N}$, the function f_n is continuous on $[0, \infty)$.*

Proof: As a warm=up, we deal with $n = 2$. Fix $a > 0$. Given $\varepsilon > 0$ we want to find $\delta > 0$ such that $|x - a| < \delta$ implies $|x^{1/2} - a^{1/2}| < \varepsilon$.

As usual, we try to manipulate $|x^{1/2} - a^{1/2}|$ to make an $|x - a|$ pop out. The good manipulation here is to multiply above and below by $|x^{1/2} + a^{1/2}|$, and use the difference-of-two-squares factorization, $X^2 - Y^2 = (X - Y)(X + Y)$, to get

$$\begin{aligned} |x^{1/2} - a^{1/2}| &= |x^{1/2} - a^{1/2}| \frac{|x^{1/2} + a^{1/2}|}{|x^{1/2} + a^{1/2}|} \\ &= \frac{|x^{1/2} - a^{1/2}||x^{1/2} + a^{1/2}|}{|x^{1/2} + a^{1/2}|} \\ &= \frac{|(x^{1/2} - a^{1/2})(x^{1/2} + a^{1/2}|}{|x^{1/2} + a^{1/2}|} \\ &= \frac{|x - a|}{|x^{1/2} + a^{1/2}|} \\ &= \frac{|x - a|}{x^{1/2} + a^{1/2}}, \end{aligned}$$

the last equality valid since $x^{1/2} \geq 0, a^{1/2} > 0$.

Now $x^{1/2} \geq 0$ so $x^{1/2} + a^{1/2} \geq a^{1/2}$ and sp

$$|x^{1/2} - a^{1/2}| \leq \frac{|x - a|}{a^{1/2}}.$$

Choose any δ at least as small as the minimum of a (to make sure that $|x - a| < \delta$ implies $x > 0$, so x is in the domain of f_2) and $a^{1/2}\varepsilon$. Then $|x - a| < \delta$ implies

$$|x^{1/2} - a^{1/2}| \leq \frac{|x - a|}{a^{1/2}} < \varepsilon.$$

That proves continuity of f_2 at all $a > 0$; right continuity at 0 (i.e., $\lim_{x \to a^+} x^{1/2} = 0$) is left as an exercise.

For the case of general n, we replace $X^2 - Y^2 = (X - Y)(X + Y)$ with

$$X^n - Y^n = (X - Y)(X^{n-1} + X^{n-2}Y + \cdots + XY^{n-2} + Y^{n-1}).$$

In the case $a > 0$, repeating the same argument as in the case $n = 2$ leads to

$$|x^{1/n} - a^{1/n}| = \frac{|x - a|}{(x^{1/n})^{n-1} + (x^{1/n})^{n-2}(a^{1/n}) + \cdots + (x^{1/n})(a^{1/n})^{n-2} + (a^{1/n})^{n-1}} \leq \frac{|x - a|}{(a^{1/n})^{n-1}},$$

and so continuity of f_n at $a > 0$ follows as before, this time taking any $\delta > 0$ at least as small as the minimum of a and $(a^{1/n})^{n-1}\varepsilon$. Again, right continuity at 0 is left as an exercise. $\square$

We know that $a^{1/2}$ cannot make sense (i.e., cannot be defined) for $a < 0$: if there was a real number $a^{1/2}$ for negative a, we would have $(a^{1/2})^2 \geq 0$ (since squares of reals are always positive), but also $(a^{1/2})^2 = a < 0$, a contradiction. By the same argument, we don't expect $a^{1/n}$ to make sense for negative a for any even natural number n.

But for odd n, we do expect that $a^{1/n}$ should make sense for negative a, and that is indeed the case.

Claim 7.9. *Fix $n \geq 3$ an odd natural number. For each $a \in \mathbb{R}$ there is a unique number $a' \in \mathbb{R}$ such that $(a')^n = a$. We refer to this number as the* nth root *of a, and write it either as $\sqrt[n]{a}$ or as $a^{1/n}$.*

Extending the function f_n defined above to all real numbers, we have that $f_n : \mathbb{R} \to \mathbb{R}$ given by $x \mapsto x^{1/n}$ is continuous for all reals.

We will not prove this, but rather leave it as an exercise. The main point is that if we define, for odd integer n and for *any* real a, the (continuous) function $f_a : \mathbb{R} \to \mathbb{R}$ via $f_a(x) = x^n - a$, then we can find $a' < a''$ for which $f_a(a') < 0 < f_a(a'')$. Once we have found a', a'' (which is a little tricky), the proof is very similar to the proofs we've already seen.

But in fact we will prove something more general than the existence of a', a''. From the section on graphing function, we have a sense that if $P(x)$ is an odd-degree polynomial of degree n, for which the coefficient of x^n is positive, then for all sufficiently negative numbers x we have $P(x) < 0$, while for all sufficiently positive x we have $P(x) > 0$. Since P is continuous, that would say (applying the IVT on any interval $[a', a'']$ where a' is negative and satisfies $P(a') < 0$, and a'' is positive and satisfies $P(a'') > 0$) that there is some $a \in \mathbb{R}$ with $P(a) = 0$ (and in particular applying this to $P(x) = x^n - a$ yields an nth root of a for every real a).

Claim 7.10. *Let $P(x) = x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$ be a polynomial, with n* odd. *There are numbers x_1 and x_2 such that $P(x) < 0$ for all $x \leq x_1$, and $P(x_2) > 0$ for all $x \geq x_2$. As a consequence (via IVT) there is a real number c such that $P(c) = 0$.*

Proof: The idea is that for large x the term x^n "dominates" the rest of the polynomial — if x is sufficiently negative, then x^n is very negative, so much so that it remains negative after $a_1x^{n-1} + \cdots + a_{n-1}x + a_n$ is added to it; while if x is sufficiently positive, then x^n is very positive, so much so that it remains positive after $a_1x^{n-1} + \cdots + a_{n-1}x + a_n$ (which may itself be negative) is added to it.

To formalize this, we use the triangle inequality to bound $|a_1x^{n-1} + \cdots + a_{n-1}x + a_n|$. Setting $M = |a_1| + |a_2| + \cdots + |a_n| + 1$ (the $+1$ at the end to make sure that $M > 1$), and considering only those x for which $|x| > 1$ (so that $1 < |x| < |x|^2 < |x|^3 < \cdots$), we have

$$\begin{aligned} |a_1x^{n-1} + \cdots + a_{n-1}x + a_n| &\leq |a_1x^{n-1}| + \cdots + |a_{n-1}x| + |a_n| \\ &= |a_1||x|^{n-1} + \cdots + |a_{n-1}||x| + |a_n| \\ &\leq |a_1||x^{n-1}| + \cdots + |a_{n-1}||x|^{n-1} + |a_n||x|^{n-1} \\ &< M|x|^{n-1}. \end{aligned}$$

It follows that for any x satisfying $|x| > 1$,

$$x^n - M|x|^{n-1} < P(x) < x^n + M|x|^{n-1}$$

Now take $x_2 = 2M$ (note $|x_2| > 1$). For $x \geq x_2$ (so in particular $x > 0$) we have

$$P(x) > x^n - M|x|^{n-1} = x^n - Mx^{n-1} = x^{n-1}(x - M) \geq 2^{n-1}M^n > 0$$

(using $x - M \geq M$ in the next-to-last inequality).

On the other hand, taking $x_1 = -2M$ (note $|x_1| > 1$) we have that for $x \leq x_1$,

$$P(x) < x^n + M|x|^{n-1} = x^n + Mx^{n-1} = x^{n-1}(x + M) \leq -2^{n-1}M^n < 0$$

(note that in the first equality above, we use $|x|^{n-1} = x^{n-1}$, valid since $n - 1$ is even). $\square$

If the coefficient of x^n in P is not 1, but some positive real $a_0 > 0$, then an almost identically proof works to demonstrate the same conclusion ($P(x)$ is negative for all sufficiently negative x, and positive for all sufficiently positive x, and so $P(c) = 0$ for some c); and if the coefficient of x^n in P is instead some negative real $a_0 < 0$ then, applying the theorem just proven to the polynomial $-P$, we find that $P(x)$ is positive for all sufficiently negative x, and negative for all sufficiently positive x, and so again by the IVT $P(c) = 0$ for some c. In other words:

> every odd degree polynomial has a real root.

Note that no such claim can be proved for *even* n; for example, the polynomial $P(x) = x^2 + 1$ never takes the value 0. We will return to even degree polynomials when we discuss the Extreme Value Theorem.

We now turn to the proof of IVT. As we have already observed, necessarily the proof will involve the completeness axiom. The informal idea of the proof is: "the first point along the interval $[a, b]$ where f stops being negative, must be a point at which f is zero". We will formalize this by considering the set of numbers x such that f is negative on the entire closed interval from a to x. This set is non-empty (a is in it), and is bounded above (b is an upper bound), so by completeness (P13), the set has a least upper bound. We'll argue that that least upper bound is strictly between a and b, and that that function evaluates to 0 at that point.

Proof (of Intermediate Value Theorem): Let $A \subseteq [a, b]$ be

$$\{x \in [a, b] : f \text{ is negative on } [a, x]\}.$$

We have $a \in A$ (since $f(a) < 0$), so A is not empty. We have that b is an upper bound for a (since $f(b) > 0$), so by the completeness axiom (P13), A has a least upper bound, call it c. Recall that this means that

- c is an upper bound for A ($x \leq c$ for all $x \in A$), and that

- c is the least such number (if c' is any other upper bound then $c' \geq c$).

We will argue that $a < c < b$, and that $f(c) = 0$. That $c > a$ follows from left continuity of f at a, and $f(a) < 0$ (the proof that if f is *continuous* and negative at a, then there's some $\delta > 0$ such that f is negative on all of $(a - \delta, a + \delta)$, can easily be modified to show that if f is *right continuous* and negative at a, then there's some $\delta > 0$ such that f is negative on all of $[a, a + \delta)$, so certainly $a + \delta/2 \in A$). Similarly, that $c < b$ follows from right continuity of f at b, and $f(b) > 0$ (there's $\delta > 0$ such that f is positive on all of $(b - \delta, b]$, so certainly $b - \delta/2$ is an upper bound for A).

Next we argue that $f(c) = 0$, by showing that assuming $f(c) > 0$ leads to a contradiction, and similarly assuming $f(c) < 0$ leads to a contradiction.

Suppose $f(c) > 0$. There's $\delta > 0$ such that f is positive on $(c - \delta, c + \delta)$, so $c - \delta/2$ is an upper bound for A — no number in $[c - \delta/2, c]$ can be in A, because f is positive at all these numbers — contradicting that c is the *least* upper bound for A.

Suppose $f(c) < 0$. There's $\delta > 0$ such that f is negative on $(c - \delta, c + \delta)$. In fact, f is negative on all of $[a, c + \delta)$ — if f was positive at any $c' < c$, c' would be an upper bound on A, contradicting that c is the *least* upper bound for A — and so $c + \delta/2 \in A$, contradicting that c is even an *upper bound* for A. □

There are a few obvious variants of the Intermediate Value Theorem that are worth bearing in mind, any require virtually no work to prove once we have the version we have already proven.

- If f is continuous on $[a, b]$, and if $f(a) > 0$, $f(b) < 0$, then there is some $c \in (a, b)$ with $f(c) = 0$. (To prove this, apply the IVT as we have proven it to the function $-f$; the c thus produced has $(-f)(c) = 0$ so $f(c) = 0$.)

- If f is continuous on $[a, b]$, with $f(a) \neq f(b)$, and if t is any number that lies between $f(a)$ and $f(b)$, then there is $c \in (a, b)$ with $f(c) = t$. (To prove this in the case where $f(a) < f(b)$, apply the IVT as we have proven it to the function $x \mapsto f(x) - t$, and to prove it in the case where $f(a) > f(b)$, apply the IVT as we have proven it to the function $x \mapsto t - f(x)$.)

- If f is a continuous function on an interval, and f takes on two different values, then it takes on all values between those two values.[100] (To prove this, let a and b be the two inputs on which f is seen to take on different values, where, without loss of generality, $a < b$, and apply the version of the IVT in the second bullet point above to f on the interval $[a, b]$.)

7.4 The Extreme Value Theorem

We begin this section with some definitions. In each of these definitions, we want to think about a function not necessarily on its whole natural domain, but rather on some specific

[100] This is often taken as the statement of the Intermediate Value Theorem.

subset of the domain. For example, we may wish to consider the function $x \mapsto 1/x$ not at being defined on all reals except 0, but rather being defined on all positive reals, or on the open interval $(0, 1)$. One way to do that is to artificially define the function only on the particular set of reals that we are interested in; but this is a little restrictive, as we may want to think about the same function defined on many different subsets of its natural domain. The approach taken in this definitions, while it may seem a little wordy at first, allows us this flexibility, and will be very useful in other situations too.

Definition of a function being bounded from above f is *bounded from above* on a subset S of Domain(f) if there is some number M such that $f(x) \leq M$ for all $x \in S$; M is an *upper bound* for the function *on* S.

Definition of a function being bounded from below f is *bounded from below* on S if there is some number m such that $m \leq f(x)$ for all $x \in S$; m is a *lower bound* for the function on S.

Definition of a function being bounded f is *bounded* on S if it is bounded from above **and** bounded from below on S.

Definition of a function achieving its maximum f *achieves* its maximum on S if there is a number $x_0 \in S$ such that $f(x) \leq f(x_0)$ for all $x \in S$. (Notice that this automatically implies that f is bounded from above on S: $f(x_0)$ is an upper bound.)

Definition of a function achieving its mimimum f achieves its minimum on S if there is a number $x_0 \in S$ such that $f(x_0) \leq f(x)$ for all $x \in S$. (Notice that this automatically implies that f is bounded from below on S: $f(x_0)$ is a lower bound.)

It's an easy exercise that f is bounded on S if and only if there is a *single* number M such that $|f(x)| < M$ for all $x \in S$.

Basically anything can happen vis a vis upper and lower bounds, depending on the specific choice of f and S. For example:

- $f(x) = 1/x$ is bounded on $[1, 2]$, and achieves both maximum and minimum;
- $f(x) = 1/x$ is bounded on $(1, 2)$, but achieves *neither* maximum *nor* minimum;
- $f(x) = 1/x$ is bounded on $[1, 2)$, does not achieve its maximum, but does achieve its minimum;
- $f(x) = 1/x$ is not bounded from above on $(0, 2)$, is bounded from below, and does not achieve its minimum;
- $f(x) = 1/x$ is not bounded from above or from below on its natural domain.

The second important theorem of continuity (IVT was the first) says that a continuous function *on a closed interval* is certain to be as well-behaved as possible with regards bounding.

Theorem 7.11. *(Extreme Value Theorem, or EVT for short) Suppose $f : [a,b] \to \mathbb{R}$ is continuous. Then*

- *f is bounded on $[a,b]$*[101]*, and*
- *f achieves both its maximum and minimum on $[a,b]$.*

We will see many applications of the EVT throughout this semester and next, but for the moment we just give one example. Recall that earlier we used the IVT to prove that if P is an *odd degree* polynomial then there must be c with $P(c) = 0$, and we observed that no such general statement could be made about *even degree* polynomials. Using the EVT, we can say something about the behavior of even degree polynomials.

Claim 7.12. *Let $P(x) = x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$ be a polynomial, with n* even*. There is a number $x^\star$ such that $P(x^\star)$ is the minimum of P on $(-\infty, \infty)$, that is, such that $p(x^\star) \leq p(x)$ for all real x.*

Proof: Here's the idea: we have $p(0) = a_n$. We'll try to find numbers $x_1 < 0 < x_2$ such that

$$P(x) > a_n \text{ for all } x \leq x_1 \text{ and for all } x \geq x_2. \ (\star)$$

We then apply the EVT on the interval $[x_1, x_2]$ to conclude that there is a number $x^\star \in [x_1, x_2]$ such that $P(x^\star) \leq P(x)$ for all $x \in [x_1, x_2]$. Now since $0 \in [x_1, x_2]$, we have $P(x^\star) \leq P(0) = a_n$, and so also $P(x^\star) \leq P(x)$ for all $x \in (-\infty, x_1]$ and $[x_2, \infty)$ (using $(\star)$). So, $P(x^\star) \leq P(x)$ for all $x \in (-\infty, \infty)$.

To find x_1, x_2, we use a very similar strategy to the one used in the proof of Claim 7.10, to show that if $M = |a_1| + \cdots + |a_n| + 1$ then there are numbers x_1, x_2 with $x_1 < 0 < x_2$ such that $P(x) \geq 2^{n-1}M^n$ for all $x \leq x_1$ and for all $x \geq x_2$ (the details of this step are left as an exercise).

Because M is positive and at least 1, and because n is at least 2, we have

$$2^{n-1}M^n \geq M \geq |a_n| + 1 \geq a_n + 1 > a_n,$$

and so we are done. □

We now turn to the proof of the Extreme Value Theorem. We begin with a preliminary observation, that if f is continuous at a point c then it is "locally bounded": there is a $\delta > 0$ such that f is bounded above, and below, on $(a - \delta, a + \delta)$. Indeed, apply the definition of continuity at c with $\varepsilon = 1$ in order to get such a δ, with specifically $f(c) - 1 < f(x) < f(c) + 1$ for all $x \in (a - \delta, a + \delta)$.

The intuition of the proof we give is that we can stringing together local boundedness at each point in the interval $[a,b]$ to get that f is bounded on $[a,b]$. We have to do it

[101] In other words, a continuous function on a closed interval cannot "blow up" to infinity (or negative infinity).

carefully, though, to avoid the upper bounds growing unbounded larger, and the lower bounds unbounded smaller. The approach will be similar to our approach to the IVT: this time, we find the longest closed interval starting at a on which f is bounded, and try to show that the interval goes all the way to b, by arguing that it it falls short of b, getting only as far as some $c < b$, *one* application of "local boundedness" allows us to stretch the interval a little further, contradicting that the interval stopped at c.

Proof (of Extreme Value Theorem): We start with the first statement, that a function f that is continuous on $[a, b]$ is bounded on $[a, b]$. We begin by showing that f is bounded from above. Let

$$A = \{x : a \leq x \leq b \text{ and } f \text{ is bounded above on } [a, x]\}.$$

We have that $a \in A$ and that b is an upper bound for A, so $\sup A := \alpha$ exists.

We cannot have $\alpha < b$. For suppose this was the case. Since f is continuous at α, it is bounded on $(\alpha - \delta, \alpha + \delta)$ for some $\delta > 0$. Now we consider two cases.

Case 1, $\alpha \in A$ Here f is bounded on $[a, \alpha]$ (by M_1, say) and also on $[\alpha - \delta/2, \alpha + \delta/2]$ (by M_2, say), so it is bounded on $[a, \alpha + \delta/2]$ (by $\max\{M_1, M_2\}$), so $\alpha + \delta/2 \in A$, contradicting that α is the least upper bound of A.

Case 2, $\alpha \notin A$ Here it must be that some $c \in (\alpha - \delta, \alpha)$ is in A; if not, $\alpha - \delta$ would be an upper bound for A, contradicting that α is the least upper bound of A. As in Case 1, f is bounded on $[a, c]$ and also on $[c, \alpha + \delta/2]$, so it is bounded on $[a, \alpha + \delta/2]$, again a contradiction.

We conclude that $\alpha = b$, so it seems like we are done; but, we wanted f bounded on $[a, b]$, and $\sup A = b$ doesn't instantly say this, because the supremum of a set doesn't have to be in the set.[102] So we have to work a little more.

Since f is right continuous at b, f is bounded on $(b - \delta, b]$ for some $\delta > 0$. If $b \notin A$, then, since $b = \sup A$, we must have $x_0 \in A$ for some $x_0 \in (b - \delta, b)$ (otherwise $b - \delta$ would work as an upper bound for A). So f is bounded on $[a, x_0]$ and also on $[x_0, b]$, so it is bounded on $[a, b]$, so $b \in A$, a contradiction. So in fact $b \in A$, and f is bounded from above on $[a, b]$.

and since f bounded on $[a, x_0]$ for some $x_0 \in (b - \delta, b)$ (our fact again — $b \notin A$), have f bdd on $[a, b]$.

A similar proof, using the equivalent form of the Completeness axiom introduced earlier (a non-empty set with a lower bound has a greatest lower bound) can be used to show that f is also bounded from below on $[a, b]$; or, we can just apply what we have just proven about upper bounds to the (continuous) function $-f$ defined on $[a, b]$ — $-f$ has some upper bound M on $[a, b]$, so $-M$ is a lower bound for f on $[a, b]$.

[102] A easy example: $\sup(0, 1) = 1$ which is not in $(0, 1)$. An example more relevant to this proof: consider $g(x) = 1/(1 - x)$ on $[0, 1)$, and $g(1) = 0$ at 1. If $A = \{x : g \text{ bounded on } [0, x]\}$, then $\sup A = 1$ but $1 \not{i}nA$. The problem here of course is that g is not *continuous* at 1

We now move on to the second part of the EVT: if $f : [a, b] \to \mathbb{R}$ is continuous, it achieves both its maximum and its minimum; there are $y, z \in [a, b]$ such that $f(z) \leq f(x) \leq f(y)$ for all $x \in [a, b]$. We just show that f achieves its maximum; the trick of applying this result to $-f$ will again work to show that f also achieves its minimum.

Consider $A = \{f(x) : x \in [a, b]\}$ (notice that now we are looking at a "vertical" set; a set of points along the y-axis of the graph of f). A is non-empty ($f(a) \in A$), and has an upper bound (by previous part of the EVT, that we have already proven). So $\sup A = \alpha$ exists. We have $f(x) \leq \alpha$ for all $x \in [a, b]$, so to complete the proof we just need to find a y such that $f(y) = \alpha$.

Suppose there is no such y. Then the function $g : [a, b] \to \mathbb{R}$ given by

$$g(x) = \frac{1}{\alpha - f(x)}$$

is continuous function (the denominator is never 0). So, again by the previous part of the EVT, g is bounded above on $[a, b]$, say by some $M > 0$. So on $[a, b]$ we have $1/(\alpha - f(x)) \leq M$, or $\alpha - f(x) \geq 1/M$, or $f(x) \leq \alpha - 1/M$. But this contradicts that $\alpha = \sup A$.

We conclude[103] that there must be a y with $f(y) = \alpha$, completing the proof of the theorem. $\square$

[103] Somewhat miraculously — the function g was quite a rabbit-out-of-a-hat in this proof.