4. Limits of functions

We will now apply the concept of limits to functions defined on subsets of the real line.

For further use we introduce the following notation:

$$(a, \infty) = \{ x \in \mathbb{R} \mid a < x \}$$

$$[a, \infty) = \{x \in \mathbb{R} \mid a \leq x\}$$

$$(-\infty, a) = \{x \in \mathbb{R} \mid x < a\}$$

$$(-\infty, a] = \{x \in \mathbb{R} \mid x \leq a\}$$

and also

$$(-\infty, \infty) = \mathbb{R}$$

These intervals are called **unbounded intervals**. The intervals defined earlier will be referred to as **bounded intervals**.

Before we start the discussion of functions more seriously some remarks:

It is a common bad practice to write $f(x)$ for both the *function* and the *value* of the function for some specific $x \in \mathbb{R}$. You must be careful to know which one is meant. For example, if we say $f(x) = \sqrt{x}$ is continuous, we cannot mean the *value* as continuity is mostly about behaviour of a function around a certain place, so that would be nonsensical. But what about "suppose $f(x) > 0$"? Do we mean $f(x) > 0$ for all x or for a given x?

This is occasionally confusing, but rarely poses serious problems. We will follow the bad practice, and often assume both the domain and codomain of a function to be understood implicitly.

For example, we may say things like: "Show $x^2 - \frac{x}{x^3+2}$ is continuous at $x_0 = 0$.", without explicitly stating that the domain of that function is $\mathbb{R} \setminus \{-\sqrt[3]{2}\}$.

It gets trickier in situations as the following: What is the domain of $\frac{x^2-1}{x+1}$? You will receive different answers depending on whom you ask. The "correct" (because my) answer is that this is an ill-posed question. A "rule" or "formula" has no domain on its own. However, one can talk about a *maximum subset of* $\mathbb{R}$ where this formula defines a function. In that case that subset would be $\mathbb{R} \setminus \{-1\}$, even though some would argue that $\frac{x^2-1}{x+1} = x - 1$ and hence it is defined everywhere. This is a valid point. It depends on whether you view $\frac{x^2-1}{x+1}$ as a formal, well, "formula", in which case you cannot evaluate it at $x = -1$, and hence the domain cannot contain 1, or as an element of some set (here, for instance, what is called a "polynomial ring") with some arithmetic operations, where you can say the equation $\frac{x^2-1}{x+1} = x - 1$ is true (but you then need to say what kind of object x is). There is also a method of continuous extension, which would also allow to argue that $\frac{x^2-1}{x+1}$ should define a function everywhere. Again, this is not a serious issue but mainly a matter of personal preference (yes, such things exist even in mathematics).

Convention

We will follow the following ill-defined convention: when we say let $f(x) = \frac{x^3}{x}$ we mean the function defined on all of $\mathbb{R}$ except the points where the denominator is 0.

4.1 Real valued functions

If X is any set, a real valued function on X is a function $f: X \to \mathbb{R}$. We write $\mathcal{F}(X)$ for the set of real valued functions on X.

If $c \in \mathbb{R}$, the function $c_X: X \to \mathbb{R}$ with $c_X(x) = c$ for all x is the constant function associated with c.

If $X \neq \emptyset$, then $c_X = d_X$ if and only if $c = d$. In this case we may identify the constant functions with $\mathbb{R}$ and think of $\mathbb{R}$ as a subset of $\mathcal{F}(X)$.

If $X = \{x\}$ has exactly one element, then $\mathcal{F}(X) = \{c_X \mid c \in \mathbb{R}\}$.

If $X = \emptyset$, then there is exactly one function $\emptyset \to \mathbb{R}$ (its graph is the empty set). We could denote this function by 0. But in this case $c_X = 0$ for all $c \in \mathbb{R}$.

4.1.1 Arithmetic with functions

Let X be a set that we will keep fixed throughout this section.

For $f, g \in \mathcal{F}(X)$ we define $f + g, f \cdot g \in \mathcal{F}(X)$ as follows: both are functions with domain X and codomain $\mathbb{R}$. And for all $x \in X$:

$$(f+g)(x) = f(x) + g(x)$$
$$(f \cdot g)(x) = f(x)g(x)$$

We refer to the $+$ operation as the **addition**, and the $\cdot$ operation as the **multiplication** on $\mathcal{F}(X)$.

We can define a third operation, the so-called **scalar multiplication**. It assigns to $c \in \mathbb{R}$ and $f \in \mathcal{F}(X)$ a function $cf: X \to \mathbb{R}$ defined by

$$(cf)(x) = cf(x)$$

Note that $cf = c_X \cdot f$.

Formally, we have defined 3 operations:

$$+: \mathcal{F}(X) \times \mathcal{F}(X) \to \mathcal{F}(X)$$
$$\cdot: \mathcal{F}(X) \times \mathcal{F}(X) \to \mathcal{F}(X)$$

and the scalar multiplication $\mathbb{R} \times \mathcal{F}(X) \to \mathcal{F}(X)$, $(c, f) \mapsto cf$.

Proposition.

The addition, multiplication, and scalar multiplication have the following properties (all quantifiers range over $\mathcal{F}(X)$, unless explicitly stated otherwise):

V1 $\forall f, g, h: (f+g) + h = f + (g+h)$
V2 $\forall f, g: f + g = g + f$

V3 $\exists 0 : \forall f : f + 0 = 0 + f$.
V4 $\forall f,\ \exists f_1 : f + f_1 = 0$
V5 $\forall c, d \in \mathbb{R}, \forall f : c(df) = (cd)f$
V6 $\forall f : 1f = f$
V7 $\forall c, d \in \mathbb{R}, \forall f : (c + d)f = cf + df$
V8 $\forall c \in \mathbb{R}, \forall f, g : c(f + g) = cf + cg$
A9 $\forall f, g, h : f(gh) = (fg)h$
A10 $\forall f, g, h : f(g + h) = fg + fh$
A11 $\forall f, g, h : (f + g)h = fh + gh$
A12 $\forall c \in \mathbb{R}, \forall f, g : (cf)g = c(fg) = f(cg)$
CA $\forall f, g : fg = gf$
UA $\exists e : \forall f : ef = f = fe$

EOP.

The proof of this proposition is straight-forward. All statements follow directly from similar statements about the arithmetic of numbers.

Proof (Sketch). We prove V1, V4, and A10, and leave the rest as an exercise.

Most statements are essentially of the form $a = b$ where a, b are specific functions with the same domain and codomain. So all there is to show is that $a(x) = b(x)$ for all $x \in X$.

But first note that the element 0 mentioned is of course the constant function 0_X, and the element e is the function 1_X. If $X \neq \emptyset$, then $0 \neq e$.

To show V1 observe

$$\big((f + g) + h\big)(x) = (f + g)(x) + h(x) = \big(f(x) + g(x)\big) + h(x) = f(x) + \big(g(x) + h(x)\big)$$

$$= f(x) + (g + h)(x) = (f + (g + h))(x)$$

For V4, define f_1 by $f_1(x) = -f(x)$. Then $f + f_1 = 0_X$.

For A10,

$$\big(f(g + h)\big)(x) = f(x)(g + h)(x) = f(x)\big(g(x) + h(x)\big)$$

$$= f(x)g(x) + f(x)h(x) = (fg + fh)(x)$$

QED.

Remark
A set V together with an operation called addition and a scalar multiplication $\mathbb{R} \times V \to V$ that satisfy V1-V8 is called a **vector space** (over $\mathbb{R}$), or $\mathbb{R}$-vector space, (or for us, just vector space). This should indicate that there are other vector spaces, and indeed, the concept carries over to any field F, where $\mathbb{R}$ is replaced by F everywhere.
A vector space that also has a multiplication operation that satisfies A9-A12 is called an **associative algebra**. You may guess that if we omit A9, it is then called an algebra. If it also satisfies CA it is called a **commutative (associative) algebra**, and if it satisfies UA it is a **unital (associative) algebra**. Thus, $\mathcal{F}(X)$ is a commutative associative unital algebra.
Algebras appear frequently in physics and functional analysis (but also in-um-algebra).

The point made here is simply that one can study algebras independently of the definitions of the operations, just like one can study fields independently of the definition of addition and multiplication, by just using these rules.
Just as for fields, it follows for example, that the 0 element in an algebra is always unique, that the element f_1 in V4 is always unique, denoted $-f$, and also equal to $(-1)f$.
It follows that $0f = 0$, where the left 0 is the 0 in $\mathbb{R}$ and the right 0 is the 0 in the algebra. It is also true that $0 \cdot f = 0$ where all 0s are now the algebra 0. If an algebra has an identity element e as in UA, then $cf = (ce)f$ for all $c \in \mathbb{R}$. EOR.

Definition
A subalgebra of $\mathcal{F}(X)$ is a nonempty subset that is closed under addition, multiplication, and scalar-multiplication. EOD.

A subalgebra is then an algebra in its own right because it satisfies V1-V8, A9-A12, and, here CA.

Example
As we will see (very soon), if $D \subset \mathbb{R}$ is a subset, then the set $\mathcal{C}(D)$ of continuous functions on D is a subalgebra of $\mathcal{F}(D)$. If I is an interval, then the set $\mathcal{D}(I)$ of differentiable functions on I is a subalgebra of both $\mathcal{C}(I)$ and $\mathcal{F}(I)$. EOE.

Exercise
Let X be nonempty and $x_0 \in X$. Let $\mathfrak{m}_{x_0} = \{f \in \mathcal{F}(X) \mid f(x_0) = 0\}$. Show that $\mathfrak{m}_{x_0}$ is a subalgebra. Show that the function $e: X \to \mathbb{R}$ defined by

$$e(x) = \begin{cases} 1 & x \neq x_0 \\ 0 & x = x_0 \end{cases}$$

is an identity element for $\mathfrak{m}_{x_0}$. Show that e is not a multiplicative identity for $\mathcal{F}(X)$. EOE.

Exercise
Compare the rules V1-UA to the field axioms. Which one carry over, which ones don't. EOE.

4.1.2 Quotients

You will have noticed that there is no mentioning of multiplicative inverses in $\mathcal{F}(X)$.

Let $f \in \mathcal{F}(X)$. Note that $f \neq 0$ means that there is at least one $x \in X$ such that $f(x) \neq 0$. It does not mean that $f(x) \neq 0$ for all $x \in X$.

If $f, g \in \mathcal{F}(X)$ such that $g(x) \neq 0$ for **all** $x \in X$, we may define the function

$$\frac{f}{g}: X \to \mathbb{R}$$

by $\frac{f}{g}(x) := \frac{f(x)}{g(x)}$ for $x \in X$.

If $f = 1_X$ we also sometimes write g^{-1} for the function $\frac{1}{g}$.

In general, if $f, g \in \mathcal{F}(X)$, let $D = \{\, x \in X \mid g(x) \neq 0 \,\}$. Then $D \subset X$ may be a very small subset (it is empty if and only if $g = 0$). We can form the quotient $\frac{f|_D}{g|_D} \in \mathcal{F}(D)$. We are often sloppy, though, and denote this function also by $\frac{f}{g}$.

Warning
Note that there is in general no cancellation for the multiplication:
Consider $X = \mathbb{R}$. Let f, g be defined as

$$f(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$$
$$g(x) = \begin{cases} x & x < 0 \\ 0 & x \geq 0 \end{cases}$$

Then $fg = 0$, but neither f nor g is equal to 0.
So for any function $h \in \mathcal{F}(X)$ we have $(f + h)g = hg$, but certainly $f + h \neq h$. EOW.

Definition
If $f, g: \mathbb{R} \to \mathbb{R}$ are polynomial functions with $g \neq 0$, then $\frac{f}{g}$ is defined on $D = \{x \in \mathbb{R} \mid g(x) \neq 0\}$.
Such a function is called a **rational function** on D. EOD.

4.2 General concepts associated to functions

4.2.1 Composition of functions

Definition
Let $f: X \to Y$ and $g: Z \to W$. If $f(Y) \subseteq Z$, then it makes sense to define the **composition of** g **with** f, as $g \circ f: X \to W$ (also pronounced as g "after" f), defined by $g \circ f(x) = g(f(x))$. EOD.

Example

1. Consider the functions $f(x) = x^2$ on $\mathbb{R}$ and the function $g(x) = \sqrt{x}$ defined on $\mathbb{R}_{\geq 0}$. Since f has image in $\mathbb{R}_{\geq 0}$, the composition $g \circ f$ is defined on all of $\mathbb{R}$ and results in $g \circ f(x) = \sqrt{x^2} = |x|$. But f is defined everywhere, so $f \circ g$ is also defined (with domain $\mathbb{R}_{\geq 0}$), and we get $f \circ g(x) = \left(\sqrt{x}\right)^2 = x$. Note that while it is tempting to view $f \circ g$ as a function on all of $\mathbb{R}$, since we can evaluate its formula everywhere, this makes no sense. $f \circ g$ is by definition a function on the domain of g. There are many ways to extend the function $f \circ g$ to all of $\mathbb{R}$.
2. Consider the function $f(x) = \sqrt{2x + 5}$ defined for $x \geq -\frac{5}{2}$. It is the composition of the function $h: \mathbb{R}_{\geq 0} \to \mathbb{R}$, $h(x) = \sqrt{x}$ with the function $g: \left[-\frac{5}{2}, \infty\right) \to \mathbb{R}$, defined by $g(x) = 2x + 5$.

Note there are many ways of writing a given function f as a composition of other functions. In fact, there are infinitely many ways. For example, let $\mathrm{id}_Y: Y \to Y$ be the identity map, that is the function $y \mapsto y$. Then $f: X \to Y$ is equal to $\mathrm{id}_Y \circ f = \mathrm{id}_Y \circ \mathrm{id}_Y \circ f = \cdots$.

Some functions can be composed with themselves: if $f: X \to X$ is a function, then $f \circ f$ is defined. Depending on the context, this is written as f^2. Then recursively, f^n is defined as

$$\underbrace{f \circ f \circ \cdots \circ f}_{n \text{ terms}}$$

However, **we will not use the notation** f^n for this function, but rather write $f^{\circ n}$, to distinguish it from the function g defined by $g(x) = f(x)^n$ (if this is defined, e.g. if f takes values in $\mathbb{R}$) which is often denoted by f^n as well.

4.2.2 Structure preserving functions

Studying arbitrary functions from $\mathbb{R} \to \mathbb{R}$ is a futile (and not very interesting) exercise. Arbitrary functions don't "use" any additional structure of $\mathbb{R}$. Calculus is about functions that actually make use of the fact that we have a lot of additional structure on $\mathbb{R}$ besides just being a set.

It is an ordered complete field. Calculus uses all of these properties to analyze "reasonable" functions. Luckily, many if not most functions arising in applications are reasonable in that sense (however ill defined).

4.2.3 Additive maps

A function $f: \mathbb{R} \to \mathbb{R}$ that preserves addition is called **additive**. To be precise, f is additive, if and only if for all $x, y \in \mathbb{R}$ we have $f(x + y) = f(x) + f(y)$. Again, without further information, we can't say a lot about such functions in general.

Typical examples are functions of the form $f(x) = ax$ for a fixed real number a.

4.2.4 Multiplicative maps

Similarly, there are **multiplicative** functions $f: \mathbb{R} \to \mathbb{R}$, ie. those that satisfy that for all x, y we have $f(xy) = f(x)f(y)$.

Here, a typical example is $f(x) = x^n$ where n is a natural number.

4.2.5 Field homomorphisms

Combining the previous two properties f is called a **field homomorphism** if $f(x + y) = f(x) + f(y)$ and $f(xy) = f(x)f(y)$ and in addition $f(1) = 1$ (this is automatic unless $f(x) = 0$ for all x).

Any field homomorphism is injective. If it is also surjective it is called a **field automorphism**. These are important in Galois[1] theory, and they can be defined for an arbitrary field F (you just need to adjust the domain and codomain to F, and the definition of "additive" or "multiplicative" carries over verbatim).

4.2.6 Order preserving maps

The previous examples only used the algebraic structure of $\mathbb{R}$ and are not at the centre of our interest.

Definition
Let $I \subseteq \mathbb{R}$ be a subset and $f: I \to \mathbb{R}$ a function. We say f is **monotone increasing**, if whenever $x < y \in I$ then $f(x) \leq f(y)$. It is **strictly monotone increasing** if $f(x) < f(y)$ for all such x, y.
f is called **monotone decreasing**, if whenever $x < y \in I$, then $f(x) \geq f(y)$. It is **strictly monotone decreasing** if $f(x) > f(y)$ for all such x, y.
f is **(strictly) monotone** if it is either **(strictly) monotone increasing** or **decreasing**. EOD.

There are many examples of monotone functions.

By some UFO the functions $f(x) = x^n$ are all (strictly) monotone increasing ($n \in \mathbb{N}$) on $\mathbb{R}_{\geq 0}$.

Exercise
If n is an odd natural number, show that $f(x) = x^n$ defines a strictly increasing function on all of $\mathbb{R}$.
EOE.

[1] Évariste Galois (1811 – 1832). He died at age 20 in a duel.

4.2.7 Even/odd functions

If I is an interval **symmetric about** 0 (that is for all $x \in I$ also $-x \in I$ and $0 \in I$, for example $(-a, a)$), a real valued function f on I is called **even** if for all $x \in I$, $f(-x) = f(x)$. A real valued function is called **odd**, if $f(-x) = -f(x)$ for all $x \in I$.

Exercise
Show that a polynomial function f on $\mathbb{R}$ is even if and only if its coefficients of all odd powers of x are zero. Show that f is odd, if all its coefficients of even powers of x are zero. Here x is the polynomial function $x \mapsto x$ for all $x \in \mathbb{R}$. EOE.

Exercise
Show that if I is any symmetric interval as above then any f defined on I can be written uniquely as $f = f_e + f_o$ where f_e is even and f_o is odd. EOE.

4.2.8 Continuous functions

For us, the most important structure preserving maps are the so-called **continuous functions**. However, it is not immediately apparent what structure they are preserving. Loosely speaking they preserve the notion of "closeness" between points: if x is "close enough" to y then $f(x)$ is "close" to $f(y)$. We will spend some time to make this precise in later chapters.

The general idea is easily described: consider the function $f(x) = x^2$ defined on $\mathbb{R}$. When showing that square roots exist, we used the fact that if $x^2 > a$ and y is sufficiently close to x, then $y^2 > a$. We similarly showed that if $x^2 < b$ then $y^2 < b$ as long as y is close enough to x. Thus, if $x^2 \in (a, b)$, we showed that also $y^2 \in (a, b)$, as long as y is close enough to x. This is, in a nutshell the definition of continuity.

4.3 The topology of $\mathbb{R}$

In mathematics, *topology* refers to an area that considers itself with spaces on which a "topology" has been defined. This means there are certain sets that are called "open" which are thought of as "neighborhoods" of its elements.

Definition
A subset $U \subseteq \mathbb{R}$ is called **open** if for every $x \in U$, there is $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subseteq U$.
A subset $A \subseteq \mathbb{R}$ is **closed**, if for every *convergent* sequence $a_n \in A$ we have $\lim a_n \in A$.
Let $\mathfrak{U}$ be the set of *all* open subsets of $\mathbb{R}$. EOD.

Example

1. Every open interval is open.
2. Every closed interval is closed.
3. $\emptyset$ and $\mathbb{R}$ both are both open and closed.

EOE.

Exercise
Show that a subset is closed if and only if its complement is open. EOE.

Lemma

1. For finitely many $U_1, U_2, \ldots, U_n \in \mathfrak{U}$, we have $U_1 \cap U_2 \cap \ldots \cap U_n \in \mathfrak{U}$.

2. Let U_i ($i \in I$) be a *family*[2] of open subsets. Then $\bigcup_{i \in I} U_i \in \mathfrak{U}$.

EOL.

A set of subsets with these two properties that contains also $\emptyset$ and $\mathbb{R}$ is called a *topology* on $\mathbb{R}$.

Exercise
Show that arbitrary intersections and finite unions of closed sets are closed. EOE.

We will occasionally use the following terminology:

Let $S \subseteq \mathbb{R}$ be any subset.

1. An **accumulation point** of S is an element $x \in \mathbb{R}$ such that there is a sequence (a_n) with $a_n \in S$ and $a_n \neq x$ (this is important) such that $\lim a_n = x$.
   $L \in \{\pm\infty\}$ is called an **improper accumulation point**[3] of S if there is a sequence (a_n) in S such that $\lim a_n = L$.
2. An **interior** point of S is an element $x \in S$ such that there is $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subseteq S$. The **interior** of S, denoted S°, is the set of all its interior points. It is a subset of S.
3. A **boundary** point of S is an element $x \in \mathbb{R}$ such that for every $\varepsilon > 0$, both $S \cap (x - \varepsilon, x + \varepsilon)$ and $S^c \cap (x - \varepsilon, x + \varepsilon)$ are not empty. The **boundary** of S, denoted ∂S, is the set of all its boundary points. (Here $S^c = \mathbb{R} \setminus S$.)
4. The **closure** of S, denoted $\bar{S}$, is the union of S with the set of all its accumulation points.
5. A point $x \in S$ is called **isolated** if there is $\varepsilon > 0$ such that $S \cap (x - \varepsilon, x + \varepsilon) = \{x\}$. S is called **discrete** if every element is isolated.
6. A subset $U \subseteq S$ is called **relative open** if it is of the form $U = V \cap S$ where V is open. Equivalently, U is relative open, if for every $x \in U$ there is $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \cap S \subseteq U$.
7. A subset $A \subseteq S$ is called relative closed if it is of the form $A = S \cap B$ where B is closed. Equivalently, A is relative closed if for all sequences $a_n \in A$ which have a limit $a \in S$, the limit is in A.

Lemma
A subset $U \subset \mathbb{R}$ is open if and only if U^c is closed. EOL.

Proof. Exercise. QED.

Lemma
The closure $\bar{A}$ of a subset $A \subset \mathbb{R}$ is closed. EOL.

Proof. We must show that $U = \bar{A}^c$ is open. If not, then for every $\varepsilon > 0$, there is an element in $(x - \varepsilon, x + \varepsilon) \cap \bar{A}$. Thus for every n there is $a_n \in \bar{A}$ such that $|x - a_n| < \frac{1}{2n}$. If $a_n \in A$ let $b_n = a_n$. If $a_n \notin A$, then a_n is an accumulatin point of A, so there is $b_n \in A$ such that $|a_n - b_n| < \frac{1}{2n}$.

[2] A *family* of mathematical objects is a set of these objects *indexed* by some set I. Formally, a family is a function $f: I \to X$ where X is some set, such that $f(i)$ is an object of the desired type.

[3] This is not necessarily standard terminology, but it will be convenient later on.

Then $|x - b_n| \leq |x - a_n| + |a_n - b_n| < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}$. Thus $b_n \to x$ and x is an accumulation point of A (since $x \in A^c$, and $b_n \in A$ for all n, $x \neq b_n$). This is a contradiction, as x is now an accumulation point of A. QED.

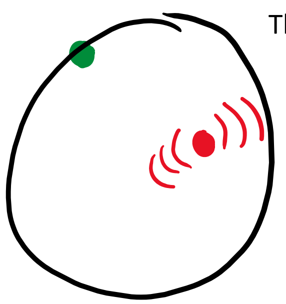
Exercise

1. Show that $\bar{S} = S^\circ \cup \partial S$.
2. Show that $S^\circ \cap \partial S = \emptyset$.
3. Show that if A is closed and $S \subseteq A$, then $\bar{S} \subseteq A$.
4. Conclude that $\bar{S} = \bigcap_{A \subseteq \mathbb{R} \text{ closed}} A$. (We say $\bar{S}$ is the "smallest" closed subset containing S.
5. Show that S is open if and only if $S^\circ = S$.
6. Show that S is closed if and only if $\bar{S} = S$.
7. Show that $[a, b]^\circ = (a, b)$ and $\partial(a, b) = \{a, b\}$.
8. Show that $\mathbb{N}$ is both closed and discrete.
9. Show that $\left\{ \frac{1}{n} \,\middle|\, n \in \mathbb{N} \right\}$ is discrete, but not closed.

EOE.

The idea of "open" vs "closed" is best illustrated in pictures using subsets of $\mathbb{R}^2$:

Consider a set bounded by the black line. The set is "open" if for every element in the set you can "wiggle" the element a little bit without leaving the set.

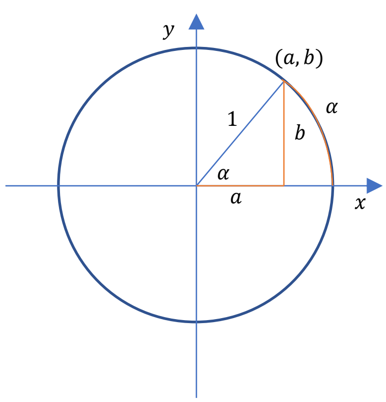


This applies to the red dot in the picture.

If you "wiggle" the green dot, however, you may leave the set.

In this picture, the interior is the actual interior bounded by the black line, not including it.

The boundary is the black line, and the closure is the set including its boundary.



4.4 Trigonometric functions

An important class of functions is inspired by Euclidean[4] geometry.

Recall that for the point (a, b) on a circle with radius 1 centered at the origin, the Pythagorean[5] Theorem (known to the Babylonians and in India way earlier in some form) states that $1 = a^2 + b^2$.

In geometry we also define $\cos \alpha := \frac{a}{1}$ (pronounced *cosine*) and $\sin \alpha = \frac{b}{1}$ (pronounced *sine*). In other words, $(a, b) = (\cos \alpha, \sin \alpha)$.

The *angle* α is a real number, and therefore we obtain two functions defined on $\mathbb{R}$. There are many problems with this definition:

1. What exactly is a *circle*?

[4] Euclid (of Alexandria) (fl. 300 BC)
[5] Pythagoras (c. 570 – c. 495 BC)

2. What is an *angle*?

We get away without knowing what an angle is by simply *defining* the angle α as the **length of the arc** from the point $(1,0)$ to the point (a, b). Which raises the next question:

3. What is the length of an arc?

All these issues can be resolved, but it can become quite involved.

A circle of radius one is defined as the set of points (a, b) where $a^2 + b^2 = 1$. Whether or not this set actually forms a circle is a philosophical question. We have developed enough theory to know that for any $x \in [0,1]$ there is a unique nonnegative $y \in [0,1]$ such that $x^2 + y^2 = 1$. Indeed, $y = \sqrt{1 - x^2}$. These points (x, y) then evidently fill out the arc from $(1,0)$ to $(0,1)$ (a quarter circle).

More problematic is 2. It mixes notions from geometry with those of calculus. Euclidean geometry follows a set of axioms that are not strictly speaking part of set theory. They would need a proof. However, without proper definitions of *points, lines, angles* there can't be such a proof within set theory. With the proper definitions, calculus (and linear algebra) can provide a *model* of Euclidean geometry (that is, a set of objects satisfying the axioms of Euclidean geometry). This is a lengthy process and would lead us too far astray.

Finally, 3. can be solved in calculus using *limits* (or, if you want, *integration*). Again, this is beyond our means right now. We will instead do the following:

We pretend to know what the arc-length is, and then heuristically use our geometric intuition to derive properties of the trigonometric functions. Later, we will then independently show that there are precisely two functions with the desired properties.

We will therefore assume that there are two functions $\sin x$ and $\cos x$ defined on all of $\mathbb{R}$ such that

T1 $\sin x$ is odd and $\cos x$ is even.
T2 For all $x, y \in \mathbb{R}$, $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$
T3 For all $x, y \in \mathbb{R}$, $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$
T4 For all sequences $x_n \to 0$ with $x_n \neq 0$ for all $n \in \mathbb{N}$, we have $\lim_{n\to\infty} \frac{\sin x_n}{x_n} = 1$.
T5 For all sequences $x_n \to 0$, we have $\cos x_n \to 1$.

Most of these properties (not T2 and T3) should be intuitively clear (if one uses the geometric definition). For example T4 reflects the fact that an approximation of the arc length for small arc length is a straight line, so if the point (a, b) wanders to (1,0), a will be close to 1.

As mentioned we will show later that there are two unique functions satisfying these properties.

Remark
We can give the definition of $\sin x$ and $\cos x$ right now, but we won't yet be able to prove all the properties as claimed above:

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$$

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}$$

EOR.

4.4.1 First properties of the trigonometric functions

T1 together with T2 implies

$$\sin(x-y) = \sin(x)\cos(y) - \sin(y)\cos(x)$$

Likewise, T1 and T3 give

Equation 4-1

$$\cos(x-y) = \cos(x)\cos(y) + \sin(x)\sin(y)$$

This uses that $\sin(-y) = -\sin(y)$ and $\cos(-y) = \cos(y)$ for all y.

By T5, $\cos(0) = 1$: indeed, the sequence $x_n = 0$ has limit 0.

Putting $x = y$ in Equation 4-1 then gives

$$1 = \cos^2(x) + \sin^2(x)$$

(Here $\sin^2(x) = (\sin(x))^2$ and $\cos^2(x) = (\cos(x))^2$).

Also, if we note that $x = \frac{x+y}{2} + \frac{x-y}{2}$ for any x, we obtain the formulas

$$\sin(x) = \sin\left(\frac{x+y}{2} + \frac{x-y}{2}\right) = \sin\left(\frac{x+y}{2}\right)\cos\left(\frac{x+y}{2}\right) + \sin\left(\frac{x-y}{2}\right)\cos\left(\frac{x-y}{2}\right)$$

$$\cos(x) = \cos\left(\frac{x+y}{2}\right)\cos\left(\frac{x-y}{2}\right) - \sin\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right)$$

4.5 Continuous functions

Arguably the most natural class of functions on $\mathbb{R}$ is the one of *continuous* ones. Intuitively, a function is continuous if we can draw its graph “without lifting the pen”, that is, it has no “jumps”. The precise definition is as follows:

Definition

Let $f: D \to \mathbb{R}$ be a function and $x_0 \in D$. We say f is **continuous at** x_0, if for every sequence $x_n \in D$ with $\lim x_n = x_0$ we have $f(x_n) \to f(x_0)$.
We say f is **continuous** (on D) if f is continuous at all $x_0 \in D$. We sometimes write $\mathcal{C}(D)$ or $\mathcal{C}^0(D)$ for the set of continuous functions on D. EOD.

Remark

Why is this a reasonable or useful definition? It encapsulates the following idea: if x is “close” to x_0, then $f(x)$ should be “close” to $f(x_0)$. What is a “jump”? For example, consider a function $f(x)$ with $f(x) \leq 0$ for $x \leq 0$, and $f(x) \geq 1$ for $x > 0$. In this case, if we choose $x_n = \frac{(-1)^n}{n}$, then $x_n \to 0$ but $f(x_n)$ does not converge (why?). EOR.

Example

1. Any polynomial function is continuous everywhere by ULT.
2. Any rational function is continuous everywhere it is defined by ULT.
3. The function $f: \mathbb{R}_{\geq 0} \to \mathbb{R}$, $f(x) = \sqrt{x}$ is continuous. (We have shown that if $a_n \geq 0$ converges to a, then $\sqrt{a_n} \to \sqrt{a}$.)
4. The Heaviside[6] function $H: \mathbb{R} \to \mathbb{R}$ is defined by
$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$
It is continuous everywhere, except at $x = 0$.

EOE.

Example

1. The Dirichlet function $\chi_{\mathbb{Q}}$ is nowhere continuous.
Indeed, let $x_0 \in \mathbb{R}$. If $x_0 \in \mathbb{Q}$, there is a sequence $x_n \to x_0$ such that $x_n \notin \mathbb{Q}$. For example, choose $x_n = x_0 + \frac{\sqrt{2}}{n}$ (why is this irrational?). Then $\chi_{\mathbb{Q}}(x_n) = 0$ for all n, so $\lim \chi_{\mathbb{Q}}(x_n) = 0 \neq \chi_{\mathbb{Q}}(x_0) = 1$.
If $x_0 \notin \mathbb{Q}$, then every interval $\left(x_0, x_0 + \frac{1}{n}\right)$ contains a rational number x_n. Then $\lim \chi_{\mathbb{Q}}(x_n) = 1$, but $\chi_{\mathbb{Q}}(x_0) = 0$.
2. Intuition is not always the best guide in determining whether a function is continuous:
Let f be defined on $(0, \infty)$ as follows:
$$f(x) = \vdash \begin{cases} \frac{1}{n} & x > 0 \in \mathbb{Q}, x = \frac{m}{n}, m \in \mathbb{N}_0, n \in \mathbb{N}, \gcd(m, n) = 1 \\ 0 & x > 0 \notin \mathbb{Q} \end{cases}$$
Then f is continuous precisely at all irrational $x \geq 0$.
This is somewhat involved: Suppose first that $x_0 > 0 \in \mathbb{Q}$. Then $f(x_0) > 0$. But $x_n = x_0 + \frac{\sqrt{2}}{n}$ is a sequence of irrational numbers with limit x_0. $f(x_n) = 0$ for all n and hence $\lim f(x_n) = 0 \neq f(x_0)$. If on the other hand, $x_0 > 0 \notin \mathbb{Q}$, we argue as follows: for a given $\varepsilon > 0$ such that $x_0 - \varepsilon > 0$, there are at most finitely many rational numbers r with a given denominator n such that $\frac{1}{n} < \varepsilon$ and $r \in (x_0 - \varepsilon, x_0 + \varepsilon)$
3. Let f be defined on $\mathbb{R}$ as $f(x) = [x]$ where $[x] = \max\{z \in \mathbb{Z} \mid z \leq x\}$. Then f is continuous precisely at all $x \notin \mathbb{Z}$. (Exercise)

The idea of avoiding jumps is better captured by the second (equivalent) definition of continuity:

Lemma ($\varepsilon - \delta$-Definition of continuity)
$f: D \to \mathbb{R}$ is continuous at $x_0 \in D$ if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in D$ with $|x - x_0| < \delta$, we have $|f(x) - f(x_0)| < \varepsilon$. EOL.

Proof. Suppose f is continuous at x_0 and let $\varepsilon > 0$. If for every $\delta > 0$, there exists $x \in D$ with $|f(x) - f(x_0)| \geq \varepsilon$ but $|x - x_0| < \delta$, then this applies to $\delta = \frac{1}{n}$. Let $x_n \in D$ with $|x_n - x_0| < \frac{1}{n}$ and $|f(x_n) - f(x_0)| \geq \varepsilon$. Then $x_n \to x_0$ and $f(x_n) \to f(x_0)$ because f is continuous at x_0. This is a contradiction.

[6] Oliver Heaviside (1850 – 1925)

Conversely, let $x_n \to x_0$ be a sequence in D. We must show that $f(x_n) \to f(x_0)$. Let $\varepsilon > 0$. Then there exists $\delta > 0$ such that $|f(x) - f(x_0)| < \varepsilon$ as long as $x \in D$ and $|x - x_0| < \delta$. There is n_0 such that $|x_n - x_0| < \delta$ for all $n > n_0$. But then $|f(x_n) - f(x_0)| < \varepsilon$ for all $n > n_0$. By definition this means $\lim_{n\to\infty} f(x_n) = f(x_0)$ as needed. QED.

Example

The exponential function $E: \mathbb{R} \to \mathbb{R}$ defined earlier as $E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ is continuous at $x_0 = 0$.

Note that $E(0) = 1$. For $m \in \mathbb{N}$, let $E_m(x) = \sum_{n=0}^{m} \frac{x^n}{n!}$.

Then $E_m(x) - 1 = \sum_{n=1}^{m} \frac{x^n}{n!} = x \sum_{n=1}^{m} \frac{x^{n-1}}{n!} = x \sum_{n=0}^{m-1} \frac{x^n}{(n+1)!}$. Note that

$$|E_m(x) - 1| = |x| \left| \sum_{n=0}^{m-1} \frac{x^n}{(n+1)!} \right| \leq |x| \sum_{n=0}^{m-1} \frac{|x^n|}{(n+1)!}$$

by the triangle inequality. We are only interested in values of x close to zero, so we may assume that $|x| < 1$. Then

$$|x| \sum_{n=0}^{m-1} \frac{|x^n|}{(n+1)!} \leq |x| \sum_{n=0}^{m-1} \frac{|x^n|}{n!} \leq |x| \sum_{n=0}^{m} \frac{1}{n!} \leq |x| E(1)$$

This holds because $E_m(1)$ is a monotone increasing sequence, so $E_m(1) \leq E(1)$.

If $\varepsilon > 0$, and $\delta > 0$ satisfies that $\delta < \frac{\varepsilon}{2E(1)}$. Then for all x with $|x| < \min\{1, \delta\}$ we get

$$|E_m(x) - 1| < \frac{\varepsilon}{2}$$

As this holds for all $m \in \mathbb{N}$, in the limit we find that $|E(x) - 1| \leq \varepsilon < \varepsilon$. This shows that E is continuous at $x = 0$. EOE.

Exercise

Show that a function $f: D \to \mathbb{R}$ is continuous at $x_0 \in \mathbb{R}$ if and only if for every open interval I containing $f(x_0)$, the preimage $f^{-1}(I) = \{\, x \in D \mid f(x) \in I \,\}$ contains $(x_0 - \delta, x_0 + \delta) \cap D$ for some $\delta > 0$.

Show that f is continuous if and only if for every open interval $I \subseteq \mathbb{R}$, the preimage $f^{-1}(I)$ is relative open in D. EOE.

4.6 Properties of continuous functions

4.6.1 Operations with continuous functions

Lemma (Compositions of continous functions are continuous)

Let $f: D \to \mathbb{R}$ and $g: E \to \mathbb{R}$ be functions such that $f(D) \subseteq E$ (then $g \circ f$ is defined on D). Let $x_0 \in D$, and $y_0 = f(x_0)$.

If f is continuous at x_0 and g is continuous at y_0, then $g \circ f$ is continuous at x_0. EOL.

Proof. Let (x_n) be any sequence in D converging to x_0. By continuity of f, $f(x_n)$ is a sequence in E converging to $y_0 = f(x_0)$. By the continuity of g, $g\big(f(x_n)\big)$ converges to $g(y_0) = g \circ f(x_0)$. QED.

Arithmetic operations on functions are defined pointwise. It should therefore be no surprise that these operations respect continuity.

Theorem (Useful Theorem for Continuous Functions; UTCF)

Let f, g be functions defined on some domain D continuous at $x_0 \in D$.

1. Both $f + g$ and fg are continuous at x_0.
2. If c, d are real numbers, then also $cf + dg$ are continuous at x_0.
3. If $g(x) \neq 0$ for all $x \in D$, then $g^{-1} = \frac{1}{g}$ is continuous at x_0, and as a consequence $\frac{f}{g}$ is continuous at x_0.
4. The functions $|f|$, $\min\{f, g\}$, and $\max\{f, g\}$ are continuous at x_0.

EOT.

Example

Applying induction (or the theorem repeatedly) we conclude that any polynomial function is continuous on all of $\mathbb{R}$. Applying 3. we conclude that a rational function $\frac{f}{g}$ where f, g are polynomials is continuous at every point x where $g(x) \neq 0$. EOE.

4.6.2 Examples of continuous functions

In the following the domains of the functions mentioned are always the "obvious" ones. E.g. the domain of $f(x) = \sqrt{x}$ will be $x \geq 0$.

Polynomial functions

Recall from Chapter 2 that a polynomial function f on $\mathbb{R}$ is a function determined by some coefficients $a_0, a_1, \dots, a_n$ such that $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$.
Any such function is continuous everywhere. Indeed, we have shown in Chapter 3 that if $x_n \to x_0$ is a sequence in $\mathbb{R}$, then also $x_n^m \to x_0^m$ for any $m \in \mathbb{N}$. This immediately shows that monomials (functions of the form x^m) are continuous. By UTFC 1. and 2. (and induction on the number of summands) we conclude that any linear combination of monomials (which is exactly what a polynomial is) is also continuous everywhere.

Rational functions

If f, g are polynomial functions with $g \neq 0$ and $D = \{x \in \mathbb{R} \mid g(x) \neq 0\}$, then $r(x) = \frac{f(x)}{g(x)}$ defines a rational function on D. It is everywhere continuous (UTFC).

Root functions

Let $n \in \mathbb{N}$ and $f(x) = \sqrt[n]{x}$, defined for $x \geq 0$.
Then f is continuous everywhere. (We have shown this already earlier.)
Recall that for any natural n, and any $z \neq z_0$ we have

$$\frac{z^n - z_0^n}{z - z_0} = \sum_{k=0}^{n-1} z^k z_0^{n-k-1}$$

Let $x_m \to x_0 > 0$ be any sequence with $x_m \geq 0$. Substitute $z = \sqrt[n]{x_m}$ and $z_0 = \sqrt[n]{x_0}$ above to conclude that for all m with $x_m \neq x_0$ we have

$$\sqrt[n]{x_m} - \sqrt[n]{x_0} = \frac{x_m - x_0}{B}$$

where $B = \sum_{k=0}^{n-1} \sqrt[n]{x_m}^{\,k} \sqrt[n]{x_0}^{\,n-k-1} \geq \sqrt[n]{x_0}^{\,n-1}$ (this holds even if $n = 1$).
Thus

$$\left|\sqrt[n]{x_m} - \sqrt[n]{x_0}\right| = \left|\frac{x_m - x_0}{B}\right| \leq \frac{1}{\sqrt[n]{x_0}^{\,n-1}} |x_m - x_0| \to 0$$

for $m \to \infty$. We conclude that f is continuous at all $x_0 > 0$.

If $x_0 = 0$ and $x_m \geq 0$ is a zero sequence, then also $\sqrt[n]{x_m} \to 0$. Indeed, we have seen that if $x < y$, then $\sqrt[n]{x} < \sqrt[n]{y}$ (see UFO8' in 1.4). If $\varepsilon > 0$, there is m_0 such that for all $m > m_0$ we have $x_m < \sqrt[n]{\varepsilon}$. But then $\sqrt[n]{x_m} < \varepsilon$ for all such m. It follows f is also continuous at $x_0 = 0$.

Exercise
Let $f: \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = \sqrt[3]{x}$ if $x \geq 0$, and $f(x) = -\sqrt[3]{-x}$ if $x < 0$. (So $f(x)$ is the unique solution z to the equation $z^3 = x$.)
Show that f is continuous. EOE.

Exercise
Prove that

$$\frac{(\sqrt{x} - 5) \cdot \sqrt{|7 - x|}}{x^2 - \sqrt[3]{x - 9} + |x - x^2 + 4x^4|}$$

is continuous wherever it is defined. EOE.

4.6.3 The Intermediate Value Theorem

The IVT is one of the most important properties of continuous functions. It again encapsulates the idea that continuous functions cannot make "jumps."

The crucial ingredient is the following special case:

Theorem (Bolzano's Theorem)
Let $f: [a, b] \to \mathbb{R}$ be continuous and $f(a) < 0 < f(b)$ or $f(a) > 0 > f(b)$. Then there exists $x_0 \in (a, b)$ such that $f(x_0) = 0$. EOT.

Proof. We first assume $f(a) < 0 < f(b)$. Let $A = \{x \in [a, b] \mid f(x) < 0\}$. Then A is bounded and nonempty ($a \in A$). Therefore, $x_0 \coloneqq \sup A$ exists. If $x \in A$, then there is $\delta > 0$ such that $[x, x + \delta) \subseteq A$. Indeed, if $f(x) < 0$, there is $\delta > 0$ such that for all $y \in (x - \delta, x + \delta) \cap [a, b]$, we have $|f(x) - f(y)| < \frac{|f(x)|}{2}$. This means $f(y) < 0$ for all such y. Since $x < b$, we may choose $\delta > 0$ small enough such that $x + \delta < b$, and then $[x, x + \delta) \subseteq A$.

This shows that $x_0 \notin A$ and $f(x_0) \geq 0$. On the other hand there is a sequence $x_n \in A$ such that $x_n \to x_0$, and consequently $f(x_n) \to f(x_0)$. This forces $f(x_0) \leq 0$, and together we get $f(x_0) = 0$.

If $f(a) > 0 > f(b)$ we may apply the above to the function $g = -f$ which is continuous and satisfies $g(a) < 0 < g(b)$. QED.

Corollary (Intermediate Value Theorme; IVT)
Let $f: [a, b] \to \mathbb{R}$ be continuous. Then f takes any value between $f(a)$ and $f(b)$. EOC.

Proof. If $f(a) = f(b)$ there is nothing to show. If $f(a) < f(b)$, and let $c \in (f(a), f(b))$. Consider $g: [a, b] \to \mathbb{R}$ defined by $g(x) = f(x) - c$. Then g is continuous and $g(a) < 0 < g(b)$. Therefore, by the theorem there is x_0 such that $g(x_0) = 0$, or equivalently $f(x_0) = c$.

If $f(a) > f(b)$, the same reasoning for any $c \in (f(b). f(a))$ works (with $g(a) > 0 > g(b)$). QED.

We have not yet discussed enough material to give meaningful examples. But imagine we are given a polynomial $f(x) = x^3 + ax^2 + bx + c$ for some $a. b, c \in \mathbb{R}$. We will see soon that for $x > 0$ large enough, we must have $f(x) > 0$. Likewise, for $x < 0$ small enough, we must have $f(x) < 0$. Therefore, there must be $x_0 \in \mathbb{R}$ such that $f(x_0) = 0$. In other words, such a polynomial always has a root. As the example $x^2 + 1$ shows the fact that f has degree 3 (or, more generally, has odd degree) is important.

Another application of Bolzano's Theorem is the following fixe-point result:

Lemma
Let $f: [a, b] \to [a, b]$ be a continuous function ($a \leq b$). Then f has a fixed point: there is $x_0 \in [a, b]$ such that $f(x_0) = x_0$. EOL.

Proof. If g denotes the function $f - x$ on $[a, b]$. Then g is continuous. Moreover, $g(a) \geq 0 \geq g(b)$.

If $g(a) = 0$ or $g(b) = 0$, then a or b is the desired fixed point. Otherwise $g(a) > 0 > g(b)$, and by Bolzano's Theorem there is $x_0 \in (a, b)$ such that $g(x_0) = f(x_0) - x_0 = 0$. QED.

4.6.4 The maximum principle

Definition
For a function $f: D \to \mathbb{R}$ we introduce the following terminology:

1. f is called **bounded above**, if there is $B \in \mathbb{R}$ such that $f(x) \leq B$ for all $x \in D$.
2. f is called **bounded below**, if there is $B \in \mathbb{R}$ such that $f(x) \leq B$ for all $x \in D$.
3. f is called **bounded**, if there are A, B such that $A \leq f(x) \leq B$, or equivalently, if there is C such that $|f(x)| \leq C$ for all $x \in D$.
4. $\sup f := \sup f(D)$, the **supremum** of f, is the supremum of its range.
5. $\inf f := \inf f(D)$, the **infimum** of f, is the infimum of its range.
6. If $\sup f \in f(D)$, then it is called the **maximum** of f, and denoted $\max f$.
7. If $\inf f \in f(D)$, then it is called the **minimum** of f, and denoted $\min f$.

If $E \subset D$ is a subset, it is also common to write $\sup_{x \in E} f(x)$, $\inf_{x \in E} f(x)$, $\min_{x \in E} f(x)$, $\max_{x \in E} f(x)$, instead of $\sup f|_E$, $\inf f|_E$, $\min f|_E$, $\max f|_E$, respectively. EOD.

Note that $\sup f \neq \infty$ if and only if f is bounded above. Similarly, $\inf f \neq -\infty$ if and only if f is bounded below.

A function does not have to have a minimum or maximum. For example, if f is defined on any open interval I as $f(x) = x$, then f does not have a minimum or maximum on that interval.

$$\sup_{x \in (0,2)} \frac{1}{x} = \infty, \inf_{x \in (0,2)} \frac{1}{x} = \frac{1}{2}$$

If we include 2 in the domain, then

$$\min_{x \in (0,2]} \frac{1}{x} = \frac{1}{2}$$

A consequence of the IVT is referred to as the minimum/maximum principle (or just maximum principle). The general statement is as follows:

Theorem (Minimum/Maximum Principle, general version; MMP)
Let f be defined on $[a, b]$ and continuous. Then there are $c \leq d$ such that $f([a, b]) = [c, d]$. In particular, $c = \min f$ and $d = \max f$. EOT.

The theorem is a consequence of the IVT together with the Maximum Principle:

Lemma (Maximum Principle)
Let f be continuous on $[a, b]$. Then f is bounded. Moreover, there are $x_1, x_2 \in [a, b]$ such that $f(x_1) = \sup f$, and $f(x_2) = \inf f$. That is, f attains a maximum and a minimum. EOL.

Proof. Let $S = f([a, b])$ which is a nonempty subset of $\mathbb{R}$. Let $y_1 = \sup S$. Then there is a sequence $y_n \in S$ such that $\lim y_n = y_1$. For each n there is $x_n \in [a, b]$ with $f(x_n) = y_n$. The sequence (x_n) is bounded, and therefore has a convergent subsequence $\left(x_{n_k}\right)$. Let $x_1 = \lim\ x_{n_k}$. Then $x_1 \in [a, b]$ as $[a, b]$ is closed. We must have $\lim\limits_{k\to\infty} f\left(x_{n_k}\right) = f(x_1)$ because f is continuous. On the other hand $f\left(x_{n_k}\right) = y_{n_k}$ is a subsequence of (y_n) and therefore must have the same limit as y_n. It follows that $y_1 = f(x_1)$. In particular $y_1 \in \mathbb{R}$. By definition $y_1 = \max f = \sup f$.

Similar reasoning applied to the infimum of S (or the above applied to $-f$) shows that there is $x_2 \in [a, b]$ such that $f(x_2) = \inf S$.

It now follows that f is bounded. QED.

Combining the maximum principle with the IVT gives the general version:

Proof (of the theorem). Let $c = \min f$ and $d = \max f$. By the MP they both exist, and there are x_1, x_2 such that $f(x_1) = c$ and $f(x_2) = d$. By definition, $f([a, b]) \subseteq [c, d]$. By the IVT f attains any value in $[c, d]$ (already on the possibly smaller interval $[x_1, x_2]$ (if $x_1 \leq x_2$) or $[x_2, x_1]$ (if $x_2 < x_1$). QED.

Remark
It is crucial that the interval of definition $[a, b]$ is closed and bounded. A continuous function defined on all of $\mathbb{R}$ need not have a maximum or minimum (think $f(x) = x$). Likewise, a continuous function defined on $(0,1)$ need not have one either (think $f(x) = \frac{1}{x}$ on $(0,1)$, or $\tan x$ on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

4.6.5 Invertible functions and continuity

Review the definition of injective, surjective, and bijective (see Chapter 2.7.2).

Lemma
An arbitrary function $f: X \to Y$ (here X and Y are arbitrary sets) is *bijective* if and only if there is a function $g: Y \to X$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. EOL.

Proof. f is bijective if and only if $\forall y \in Y: \exists!\, x \in X: f(x) = y$.

For $y \in Y$, let $g(y) = x$ where $f(x) = y$. As x is unique, this determines a function $g: Y \to X$.

Then for all $x \in X$, we have $g\big(f(x)\big) = x$ by the definition of g. For all $y \in Y$, we have $f\big(g(y)\big) = y$, also by the definition of g.

Now suppose g exists. Then f is bijective. It is injective, since if $f(x) = f(x')$, then $x = g\big(f(x)\big) = g\big(f(x')\big) = x'$. It is surjective, since if $y \in Y$, then $f\big(g(y)\big) = y$. QED.

Exercise
Show the function g as in the lemma is unique if it exists. EOE.

If f is bijective, we write f^{-1} or $f^{\circ -1}$ for the unique function g as in the lemma.

It is called the **inverse function** to f.

Warning

f^{-1} as defined here must not be confused with $f^{-1} = \frac{1}{f}$ for a real valued function that is nonzero everywhere. Because of this, we sometimes will use the notation $f^{\circ -1}$ for the inverse function. EOW.

Exercise

Let $f: X \to Y$ be a function. Show that f is

1. *Injective* if and only if there is a function $g: Y \to X$ such that $g \circ f(x) = x$ for all $x \in X$.
2. *Surjective* if and only if there is a function $g: Y \to X$ such that $f \circ g(y) = y$ for all $y \in Y$.
3. In the situation of 1. show that g is surjective. In the situation of 2. show that g is injective.
4. Show that if f is bijective f^{-1} is again bijective and $(f^{-1})^{-1} = f$.

EOE.

A function $f: D \to \mathbb{R}$ is called **invertible** if it is one-to-one/injective. The reason is that in this situation, the function $\bar{f}: D \to f(D)$ is bijective and admits an inverse function.

We then have an inverse function $g: f(D) \to D$ such that $f\big(g(y)\big) = y$ and $g\big(f(x)\big) = x$.

Normally, we only consider invertible functions, where the range is a "reasonable" subset of $\mathbb{R}$ (such as an interval, or a union of intervals).

Any strictly monotone function $f: D \to \mathbb{R}$ where $D \subset \mathbb{R}$ is invertible.

Lemma (Continuous and invertible implies stricty monotone)

Let I be an interval and $f: I \to \mathbb{R}$ be a continuous injective function. Then f is strictly monotone increasing or decreasing. EOL.

Proof.

Step 1. Let $a < b \in I$. Then f attains a maximum and minimum on $[a, b]$ by MMP. Let $f(x_1)$ be the minimum and $f(x_2)$ be the maximum. Then $\{x_1, x_2\} = \{a, b\}$. That is, either $x_1 = a$ and $x_2 = b$, or $x_1 = b$ and $x_2 = a$.

To see this, suppose first that $f(a) < f(b)$. If $x_1 > a$, then $f(x_1) < f(a)$ (since $f(a) \neq f(x_1)$). Then $f(x_1) < f(a) < f(b)$, so by the IVT, f attains the value $f(a)$ also on the interval $[x_1, b]$. This contradicts that f is injective. Therefore $x_1 = a$. Similarly, if $x_2 \neq b$, then $x_2 < b$ and $f(b) < f(x_2)$. Then f attains the value $f(b)$ also on the interval $[a, x_2]$. Again, this contradicts that f is injective.

If $f(a) > f(b)$ we apply the discussion to $-f$ to conclude that $a = x_2$ and $b = x_1$.

Step 2. Let $a < b \in I$. Then $f|_{[a,b]}$ is strictly monotone.

Assume $f(a) < f(b)$. Let $c < d \in [a, b]$. Then $f(a) < f(d)$. Therefore, by Step 1, $f(a)$ is the minimum and $f(d)$ is the maximum on $[a, d]$. Thus, $f(c) < f(d)$.

If $f(a) > f(b)$, then $f(a) > f(d)$, and $f(d)$ is the minimum of f on $[a, d]$. Thus $f(c) > f(d)$. This shows Step 2.

Step 3. f is strictly monotone.

Let $a < b \in I$ and suppose $f(a) < f(b)$. We must show that f is strictly increasing, i.e. for all $c < d \in I$ we must have $f(c) < f(d)$. Let $u = \min\{a, c\}$ and $v = \max\{b, d\}$, then f is strictly monotone increasing on $[u, v]$ (because $a, b \in [u, v]$ and because of Step 2). Therefore $f(c) < f(d)$.

The reasoning is similar if $f(a) > f(b)$, but the f is strictly monotone deacreasing. QED.

Remark
Step 2 would do the trick whenever $I = [a, b]$ is a bounded closed interval. The reason that Step 3 is needed is that I may not be closed or not be bounded. EOR.

Note the converse is not true: a strictly monotone function need not be continuous (it is invertible, though). However, they do have a remarkable property regarding limits:

Lemma
Let I be a nonempty interval and $f: I \to \mathbb{R}$ be a strictly monotone function. If $x_n \in I$ is a sequence such that x_n *does not converge* to $x_0 \in I$, then $f(x_n)$ does not converge to $f(x_0)$.
Equivalently, if $f(x_n) \to f(x_0)$, we must have $x_n \to x_0$. EOL.

Proof. Suppose x_n does not converge to x_0. That is, there is $\varepsilon > 0$ such that for all $k \in \mathbb{N}$, there is $n_k > k$ such that $|x_n - x_0| \geq \varepsilon$. For infinitely many values of k, we must have $x_{n_k} > x_0$ or $x_{n_k} < x_0$. Replacing the subsequence x_{n_k} by a subsequence of itself, we can therefore assume that $x_{n_k} \geq x_0 + \varepsilon$ or $x_{n_k} \leq x_0 - \varepsilon$ for all k. And then $x_0 - \varepsilon \in I$ in the first, and $x_0 + \varepsilon \in I$ in the second case.

If f is increasing this means $f\left(x_{n_k}\right) \leq f(x_0 - \varepsilon) < f(x_0)$ in the first case or $f\left(x_{n_k}\right) \geq f(x_0 + \varepsilon) > f(x_0)$ in the second case. If f is decreasing tis means $f\left(x_{n_k}\right) \geq f(x_0 - \varepsilon) > f(x_0)$ or $f(x_{n_k})) \leq f(x_0 + \varepsilon) < f(x_0)$.

But then $f\left(x_{n_k}\right)$ cannot converge to $f(x_0)$. As it is a subsequence of $f(x_n)$, that sequence cannot converge to $f(x_0)$ as well. QED.

Exercise.
Where in the proof of the lemma did we use the fact that I is an interval. EOE.

The lemma has an immediate and maybe surprising consequence:

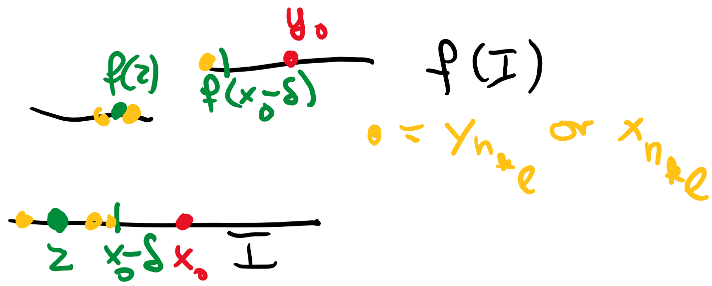
Theorem (Invertible Function Theorem; IFT1)
Let I be an interval and $f: I \to \mathbb{R}$ be a strictly monotone increasing function. Then f is invertible, and the inverse function $f^{-1}: f(I) \to I$ is continuous and strictly monotone increasing. An analogous statement holds for strictly decreasing functions. EOT.

Proof. First note that if f is strictly monotone, then f is injective and hence invertible. It is also immediate that f^{-1} is strictly monotone increasing: for if $y_1 < y_2$, then $f^{-1}(y_1) < f^{-1}(y_2)$ because f is strictly monotone increasing[7].

---

[7] We have used this kind of reasoning before when discussing some of the UFOs: for positive x, y we have $x^n < y^n$ only if $x < y$ (for if $x > y$, then $x^n > y^n$).

It remains to see why f^{-1} is continuous. Let $y_0 \in f(I)$ and let $y_n \to y_0$ be a sequence in $f(I)$ with limit y_0. If $x_n \coloneqq f^{-1}(y_n)$ does not converge to $x_0 \coloneqq f^{-1}(y_0)$, then by the lemma, $f(x_n) = y_n$ cannot converge to $y_0 = f(x_0)$. Thus $x_n \to x_0$, and we are done. QED.



Remark
It might be tempting to argue as follows: if f is strictly monotone on I, it is the inverse of the strictly monotone function $f^{-1}: f(I) \to \mathbb{R}$ and therefore must be continuous. However, this is not the case: we heavily relied on the fact that I is an interval. But $f(I)$ is not necessarily an interval, and therefore the theorem does not apply to f^{-1} in general. EOR.

Corollary
Let f be a continuous injective function defined on an interval I. Then $f(I)$ is also an interval and f^{-1} is continuous. If I is open, or closed and bounded, then so is $f(I)$. EOC.

Example

1. We can now conclude that for any $p \in \mathbb{N}$ the function $f: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined by $f(x) = \sqrt[p]{x}$ is continuous. Indeed, it is the inverse function to $x \mapsto x^p$ which is continuous and invertible.
2. We may conclude that $\ln = \exp^{-1}$ is continuous (and also strictly monotone increasing).
3. Let $f: [0,1] \cup (2,3] \to \mathbb{R}$ be defined as
$$f(x) = \begin{cases} x & ; x \in [0,1] \\ \frac{1}{2}x & ; x \in (2,3] \end{cases}$$
Then $g = f^{-1}$ is defined on $f([0,1] \cup (2,3]) = \left[0, \frac{3}{2}\right]$. It is strictly monotone but not continuous in $x = 1$. On $[0,1]$, $g(x)$ is equal to x. On $\left(1, \frac{3}{2}\right]$, $g(x) = 2x$.

EOE.

4.7 Exponential functions

A gap in our discussion so far is that we haven't defined exponential functions, that is functions of the form $f(x) = a^x$ where $a > 0$ and $x \in \mathbb{R}$.

There are several ways to introduce them. Here is one we will **not** follow:

1. Recall the definition of a^r where $r \in \mathbb{Q}$.
2. For $x \in \mathbb{R}$, choose sequence $r_n \to x$ with $r_n \in \mathbb{Q}$.
3. Show that a^{r_n} converges.
4. Show that the limit in 3. is independent of the chosen sequence r_n. For this it suffice sto show that if $s_n \to 0$ is a zero sequence of rational functions, then $a^{s_n} \to 1$.

We will adopt a different approach.

Definition
A function $f: \mathbb{R} \to \mathbb{R}$ is called an **exponential function** if

1. $f \neq 0$ (this means $f(x) \neq 0$ for **at least one** x).
2. f is continuous at $x = 0$.
3. $\forall x, y: f(x+y) = f(x)f(y)$.

EOD.

Exponential functions are important in many applications.

Lemma
Let f be an exponential function. Then

1. $f(0) = 1$.
2. $f(x) > 0$ for all $x \in \mathbb{R}$.
3. f is continuous everywhere.

EOL.

Proof.

1. There is $x_0 \in \mathbb{R}$ such that $f(x_0) \neq 0$. Then $f(x_0) = f(x_0 + 0) = f(x_0)f(0)$. It follows that $f(0) = 1$.
2. Suppose $f(x_0) = 0$. Then $f(x) = f\big((x - x_0) + x_0\big) = f(x - x_0)f(x_0) = 0\ \forall x$, contradicting that $f \neq 0$. Since $f(0) = 1 > 0$, the IVT guarantees that $f(x) > 0$ for all x, as if $f(x) < 0$, there is an $x_0 \in (x, 0)$ or $(0, x)$ with $f(x_0) = 0$. Note we could also argue that $f(x) = f\left(\frac{x}{2} + \frac{x}{2}\right) = f\left(\frac{x}{2}\right)^2$ is a square and hence nonnegative.
3. We must show that for all sequences $x_n \to x_0$, we have $f(x_n) \to f(x_0)$.
   But $h_n := x_n - x_0$ is a zero sequence and $f(h_n) \to f(0) = 1$. Then $f(x_n) = f(x_0 + h_n) = f(x_0)f(h_n) \to f(x_0)$ by ULT.

QED.

Lemma
Let f be an exponential function. Then for all x and all $r \in \mathbb{Q}$, we have
$$f(rx) = f(x)^r$$
EOL.

Proof. Let $r = \frac{n}{m}$ with $n \in \mathbb{Z}$ and $m \in \mathbb{N}$. We first show that $f(nx) = f(x)^n$. This is an elementary induction on $|n|$.

If $n = 0$, then $f(0x) = f(0) = 1 = f(x)^0$. If $n = 1$, then $f(x) = f(x)^1$ and $f(-x) = f(x)^{-1}$ because $1 = f(0) = f(x - x) = f(x)f(-x)$. This is the base case.

Suppose for $n \in \mathbb{N}$, we know that $f(nx) = f(x)^n$ and $f(-nx) = f(x)^{-n}$. Then

$$f\big((n+1)x\big) = f(nx + x) = f(nx)f(x) = f(x)^n f(x) = f(x)^{n+1}$$

and

$$f\Big(\big(-(n+1)\big)x\Big) = f(-(n+1)x) = f(-nx - x) = f(-nx)f(-x) = f(x)^{-n}f(x)^{-1} = f(x)^{-(n+1)}$$

finishing the induction proof.

Next, for $m \in \mathbb{N}$, we have $f\left(\frac{x}{m}\right) = f(x)^{\frac{1}{m}}$ because $f(x) = f\left(m\frac{x}{m}\right) = f\left(\frac{x}{m}\right)^m$ proves that $f\left(\frac{x}{m}\right) = \sqrt[m]{f(x)}$ as it is the unique positive solution to $z^m = f(x)$.

Combining we get

$$f(rx) = f\left(\frac{n}{m}x\right) = f\left(n\frac{x}{m}\right) = f\left(\frac{x}{m}\right)^n = \left(f(x)^{\frac{1}{m}}\right)^n = f(x)^{\frac{n}{m}}$$

QED.

Corollary
Let f be an exponential function. Then $f(x) = \lim_{n\to\infty} f(1)^{r_n}$ where $r_n \to x$ is any sequence of rational numbers converging to x. EOC.

Proof. Since f is continuous, $f(x) = \lim_{n\to\infty} f(r_n)$. By the lemma, $f(r_n) = f(1)^{r_n}$. QED.

Definition
Let f be an exponential function. Then $a \coloneqq f(1)$ is called the **base** of the exponential function. EOD.

Lemma
Let $f: \mathbb{R} \to \mathbb{R}$ be an exponential function with base $a > 1$. Then f is strictly monotone increasing with range $(0, \infty)$. EOL.

Proof. For any pair of rational numbers $r < s$, we know that $f(r) = a^r < a^s = f(s)$. This is a consequence of some UFOs.

Now let $x < y$ be real numbers and let $r_n \to x$ and $s_n \to y$ be sequence of rational numbers converging to x and y, respectively. As $\mathbb{R}$ is Archimedean we know that (x, y) contains a rational number, r, say. Likewise, the interval (r, y) contains an rational number s. We almost always have $r_n < r$ and $s_n > s$. Thus, we almost always have

$$r_n < r < s < s_n$$

and therefore, almost always

$$a^{r_n} < a^r < a^s < a^{s_n}$$

But then $f(x) = \lim a^{r_n} \le a^r < a^s \le \lim a^{s_n} = f(y)$, and so $f(x) < f(y)$.

It now follows that $\lim_{n\to\infty} f(n) = \infty$ and $\lim_{n\to\infty} f(-n) = \lim\left(\frac{1}{f(n)}\right) = 0$. The MMP now guarantees that the range of f is $(0, \infty)$: the range of f is an open interval, it cannot be bounded above, and is bounded below by 0. QED.

Corollary
If f is an exponential function with base $0 < a < 1$, then f is strictly monotone decreasing with range $(0, \infty)$. EOC.

Proof. $f(-x)$ is an exponential function with base $\frac{1}{a} > 1$. Now apply the lemma to $f(-x) = \frac{1}{f(x)}$. QED.

Lemma
The base uniquely determines an exponential function. In other words, if f, g are exponential functions with $f(1) = g(1)$, then $f = g$. EOL.

Proof. By the lemma, we have $f(r) = f(1)^r = g(1)^r = g(r)$ for all $r \in \mathbb{Q}$.

As $\mathbb{Q}$ is dense in $\mathbb{R}$ and since f, g are continuous we conclude that $f(x) = g(x)$ for all x and thus $f = g$: For any x there is a sequence of rational numbers $r_n \to x$. Then $f(x) = \lim f(r_n) = \lim g(r_n) = g(x)$. QED.

Theorem
For any $a > 0$ there exists a unique exponential function f with base a. EOT.

We often write a^x for $f(x)$, where f is the (unique) exponential function with base $a = f(1)$.

We will show later that $\exp x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ is an exponential function. For now, we will only observe that $\exp x$ satisfies properties 1. and 2. of an exponential function.

Note that $\exp(0) = 1 \neq 0$, so 1. is satisfied. We showed that $\exp x$ is continuous at 0 (see the last example in 4.5), so 2. is satisfied.

Note that the base $\exp(1) = 2 + \sum_{n=2}^{\infty} \frac{1}{n!} > 2 > 1$. Therefore $\exp x$ is strictly monotone increasing.

Definition
The function $\exp x : \mathbb{R} \to \mathbb{R}$ is invertible with range $(0, \infty)$. Its inverse function $\ln x : (0, \infty) \to \mathbb{R}$ is called the (natural) **logarithm**. EOD.

Exercise
Show that

1. $\ln x$ is strictly monotone increasing.
2. For all $x, y > 0$, $\ln(xy) = \ln x + \ln y$.
3. $\ln x$ is continuous everywhere.

EOE.

Note that because $\ln x$ is the inverse of $\exp x$ its range is the domain of $\exp x$, that is, $\mathbb{R}$.

Also, for any $a > 0$, consider the function f defined on $\mathbb{R}$ by $f(x) = \exp(x \ln a)$.

Then f is an exponential function: it is continuous (as a composition of continuous functions), it is positive everywhere, and $f(x + y) = \exp\big((x + y) \ln a\big) = \exp(x \ln a) \exp(y \ln a)$ for all x, y.

The base of f is $f(1) = \exp(\ln a) = a$.

This shows that once we have shown that $\exp(x)$ is an exponential function, then the theorem holds: for every $a > 0$ there is an exponential function with base a, which is unique by the lemma above.

4.8 Limits of functions

Definition
Let $f: D \to \mathbb{R}$ be a function and let x_0 be an accumulation point of D. We say the **limit** of f as x approaches x_0 is L if

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

for *all* sequences (x_n) in D with $x_n \neq x_0$ for all n and $\lim x_n = x_0$.
If such an L exists (proper or improper), we write

$$\lim_{x\to x_0} f(x) = L$$

(or sometimes also $\lim_{x\to x_0} f = L$). We extend the definition to the case of x_0 being an *improper accumulation point* (ie. $x_0 = \pm\infty$).
We say $f(x)$ **converges** to L as x approaches x_0 if $L = \lim_{x\to x_0} f(x)$ exists and is finite. In this case L is called a **proper limit**. If L exists but is $\pm\infty$ it is called an **improper limit**. EOD.

Because by assumption x_0 is an accumulation point of the domain, there is at least one sequence (x_n) with $x_n \in D$ and $x_n \neq x_0$ for all n.

Remark
In the definition of continuity, we asked for sequences (x_n) with $x_n \to x_0$ but did *not* require that $x_n \neq x_0$ for all n. This is a technical distinction. We could have required $x_n \neq x_0$ then as well but would have to make a special definition for isolated points: any function $f: \mathbb{N} \to \mathbb{R}$ is typically considered continuous, but $\mathbb{N}$ has no accumulation points.
Likewise, later it will be convenient to exclude x_0 as we then do not have to worry about whether x_0 is part of the domain or not. Some authors, however, do not require $x_n \neq x_0$ in this definition of a limit. EOR.

Lemma
Let I be an interval of positive size and $x_0 \in I$. Then f defined on I is continuous at x_0 if and only if $\lim_{x\to x_0} f(x) = f(x_0)$. EOL.

Proof. x_0 is an accumulation point of I (why?). If f is continuous at x_0 then certainly $\lim_{x\to x_0} f(x) = f(x_0)$, as the sequences considered for this definition form a subset of the sequences considered for continuity. Conversely suppose $\lim_{x\to x_0} f(x) = f(x_0)$. Let $x_n \to x_0$ be any sequence in D (in particular, $x_n = x_0$ for some or all n is allowed). We must show that $f(x_n) \to f(x_0)$. If $x_n = x_0$ for all except finitely many n, there is nothing to show. Otherwise let $z_k = x_{n_k}$ be the subsequence of (x_n) obtained by deleting all occurrences of x_0. Then still $z_k \to x_0$ and therefore $f(z_n) \to f(x_0)$ by assumption. Given $\varepsilon > 0$, there is k_0 such that if $k > n_0$ then $|f(z_k) - f(x_0)| < \varepsilon$. Let $n_0 = n_{k_0}$. Then for all $n > n_0$, we have $x_n = x_0$ or $n = n_k$ for some $k > k_0$. In both cases $|f(x_n) - f(x_0)| < \varepsilon$. Thus, $f(x_n) \to f(x_0)$ as well. QED.

Lemma ($\boldsymbol{\varepsilon - \delta}$-definiton for proper limits)
Let $f: D \to \mathbb{R}$ be a function and x_0 a (proper) accumulation point of D. Then $\lim_{x\to x_0} f(x) = L$ with $L \in \mathbb{R}$ if and only if for every $\varepsilon > 0$ there is $\delta > 0$ such that for all $x \in D$ with $0 < |x - x_0| < \delta$ we have $|f(x) - L| < \varepsilon$. EOL.

Proof. Suppose $\lim_{x\to x_0} f(x) = L$, and let $\varepsilon > 0$. Suppose there is no $\delta > 0$ such that $|f(x) - L| < \varepsilon$ if $0 < |x - x_0| < \delta$. Then for $\delta = \frac{1}{n}$ there exists $x_n \neq x_0 \in D$ with $|x_n - x_0| < \frac{1}{n}$ but $|f(x_n) - L| \geq \varepsilon$. As $x_n \to x_0$ this contradicts the fact that $\lim_{x\to x_0} f(x) = L$.

If on the other hand such a $\delta > 0$ exists for each $\varepsilon > 0$, and if $x_n \to x_0$ is a sequence in D with $x_n \neq x_0$ then there exists n_0 such that $|x_n - x_0| < \delta$ for all $n > n_0$. For each such n we then have $|f(x_n) - L| < \varepsilon$ and $f(x_n) \to L$. QED.

Example

1. Let f be defined on all of $\mathbb{R}$ as $f(x) = x$.
   Then $\lim_{x\to x_0} f(x) = x_0$ for all x_0 (including $\pm\infty$).
2. Let f be defined on all of $\mathbb{R}$ as $f(x) = x^n$ (for some $n \in \mathbb{N}$).
   Then $\lim_{x\to x_0} f(x) = x_0^n$ (even if $x_0 = \pm\infty$, if we put $(-\infty)^n = (-1)^n\infty$).
3. For any polynomial function f and any sequence $x_n \to x_0$, we know that $\lim_{x\to x_0} f(x) = f(x_0)$.

Remark

Let $f: D \to \mathbb{R}$ be a function, and x_0 an accumulation point of D. Let x_0 be an accumulation point of D and define $E = \{h \in \mathbb{R} \mid x_0 + h \in D\}$.
Then 0 is an accumulation point of E (why?), and $\lim_{x\to x_0} f(x) = \lim_{h\to 0} f(x_0 + h)$ (meaning if one exists, the other does, and they are equal). Here $f(x_0 + h)$ should be viewed as the function defined on E by $h \mapsto f(x_0 + h)$. EOR.

Remark

As with continuity, the domain of a function is important:
If $f: D \to \mathbb{R}$ is a function and x_0 is an accumulation point of both D and some subset $E \subset D$, then if $\lim_{x\to x_0} f(x) = L$ exists, it follows that $\lim_{x\to x_0} f|_E(x)$ exists and is equal to L. The converse is false. If the restriction $g \coloneqq f|_E$ has a limit for $x \to x_0$, it does not follow that f has. If f has a limit, though, it must be equal to the limit of g by the above.
As an example, consider (as usual) the Heaviside function H defined on $D = \mathbb{R}$ as $H(x) = 0$ for $x < 0$, and $H(x) = 1$ for $x \geq 0$. For $E_1 = (-\infty, 0)$ and $E_2 = [0, \infty)$ we have that $g_1 \coloneqq H|_{E_1}$ and $g_2 = H|_{E_2}$ both have limits for $x \to 0$, but H does not. EOR.

Exercise

Show that if x_0 is an accumulation point of D and $f: D \to \mathbb{R}$ is a function. Then $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x)$ where $g = f|_{D'}$ with $D' = D \setminus \{x_0\}$. (Note if $x_0 \notin D$, then $f = g$.) The point here is that if either limit exists then so does the other and then they are equal. EOE.

4.8.1 Properties of limits

Since we based our definition of limits for functions on limits of sequences it should not be surprising that they behave quite similarly.

Theorem (Useful Limit Theorem for Functions; ULTF)

Let $f, g: D \to \mathbb{R}$ be functions and suppose x_0 is a proper or improper accumulation point. Suppose $L = \lim_{x\to x_0} f(x)$, $M = \lim_{x\to x_0} g(x)$ (so both limits exist).

1. If L, M are finite, or if $L \neq -M$, then $\lim_{x\to x_0} (f(x) + g(x)) = L + M$.
2. If L, M are finite, or both are nonzero, then $f(x)g(x) \to LM$ as $x \to x_0$.
3. If $c, d \in \mathbb{R}$ such hat $cL + dM$ is defined, $cf(x) + dg(x) \to cL + dM$ as $x \to x_0$.

4. If $M \neq 0$, and L or M is proper, then $\frac{f(x)}{g(x)} \to \frac{L}{M}$ for $x \to x_0$.

EOT.

In 3. it is meant that if L is improper, then $c \neq 0$, if M is improper then $d \neq 0$, and if one of them is improper then $cL \neq -dM$.

Proof. This is an immediate consequence of the definition of limit and the Useful Limit Thereom.

For 4. we observe that as $\lim_{x \to x_0} g(x) = M \neq 0$, there must be $\delta > 0$ such that $g(x) \neq 0$ for all $x \in D$ with $0 < |x - x_0| < \delta$ (if x_0 is proper), and if x_0 is improper, then there is B such that $g(x) \neq 0$ for all $x > B$ (if $x_0 = \infty$) or $x < B$ (if $x_0 = -\infty$). QED.

Exercise
Restate the theorem for each of the cases: L, M proper; L proper, but M not, Mproper, but L not. Both L, M improper. EOE.

Example

1. If $f(x) = ax^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is a polynomial of degree n (so $a \neq 0$), then $\lim_{x \to \infty} f(x) = a \cdot \infty$ (= ∞ if $a > 0$ and = $-\infty$ if $a < 0$).
   Similarly, $\lim_{x \to -\infty} f(x) = a \cdot (-\infty)$
   We will discuss the first case:
   Note that for $x > 0$, $\frac{f(x)}{x^n} = a + \frac{a_{n-1}}{x} + \frac{a_{n-2}}{x^2} + \cdots + \frac{a_0}{x^n} \to a \neq 0$ for $x \to \infty$. By UTALOF1, we find $f(x) \to a \cdot \lim_{x \to \infty} x^n = a \cdot \infty$ as claimed.

The Squeeze Principle and the Comparison Lemma also have analogues for limits of functions.

Lemma (Comparison Lemma for Functions; CLF)
Let $f, g: D \to \mathbb{R}$ be functions and let x_0 be a proper or improper accumulation point of D. Suppose $L = \lim_{x \to x_0} f(x)$ and $M = \lim_{x \to x_0} g(x)$ exist.
If $f(x) \leq g(x)$ for all $x \in D$, then $L \leq M$. EOL.

Proof. Again, this is an immediate consequence of the definition of limits and the Comparison Lemma (CL). QED.

Lemma (Squeeze Principle for Functions; SPF)
Let $f, g, h: D \to \mathbb{R}$ be functions and let x_0 be a proper or improper accumulation point of D. If $L = \lim_{x \to x_0} f(x) = \lim_{x \to x_0} h(x)$ exists and if $f(x) \leq g(x) \leq h(x)$ for all $x \in D$, then $\lim_{x \to x_0} g(x) = L$. EOL.

Proof. This is a consequence of SP for sequences. QED.

4.8.2 One sided limits

Occasionally, we have a situation where we want to discuss a limit of a function as x approaches x_0, but only want to consider cases $x > x_0$ or cases $x < x_0$. This is not really a new concept. It is rather a convenient notation for a somewhat often-encountered situation.

Definition
Let x_0 be an accumulation point of the domain D of a function f.

If x_0 is still an accumulation point of $D \cap (x_0, \infty)$, we write $\lim_{x \to x_0^+} f(x)$ for $\lim_{x \to x_0} \bar{f}(x)$ where $\bar{f}$ is the restriction of f to $D \cap (x_0, \infty)$. This is called the **right sided** or **right hand** limit of $f(x)$ as x approaches x_0.
Similarly, we write $\lim_{x \to x_0^-} f(x)$ for $\lim_{x \to x_0} \bar{\bar{f}}(x)$ where $\bar{\bar{f}}$ is the restriction of f to $D \cap (-\infty, x_0)$, provided x_0 is still an accumulation point of this set. This is called the **left sided** or **left hand** limit of $f(x)$ as x approaches x_0. EOD.

Thus, to define a right sided limit, we need at least one sequence $x_n \to x_0$ with $x_n > x_0$, and for a left handed limit we need at least one with $x_n < x_0$.

Remark
Some authors simply write $\lim_{\substack{x \to x_0 \\ x < x_0}} f(x)$ instead of $\lim_{x \to x_0^-} f(x)$. This has the (limited) advantage that one could easily adapt it to other requirements on x (e.g. $x \in \mathbb{Q}$, or $x \neq 0$).
Some authors also write $\lim_{x \to a^+} f(x)$ instead of $\lim_{x \to a} f(x)$ in cases where $D = [a, b]$, for example. We usually will not do that, as there is nothing ambiguous about $\lim_{x \to a} f(x)$ if f is only defined on $[a, b]$. EOR.

Example
Suppose f is defined on some set D, and suppose x_0 is an accumulation point of both $D \cap (-\infty, x_0)$ and $D \cap (x_0, \infty)$. Then $\lim_{x \to x_0} f(x)$ exists and is equal to L if and only if both $\lim_{x \to x_0^+} f(x) = L$ and $\lim_{x \to x_0^-} f(x) = L$. EOE.

Exercise
Verify that the example is true. EOE.

Example
Combining the previous example with the first lemma in 4.8 above, we obtain the following result: Let f be defined on some set D and let x_0 be an interior point of D. Then f is continuous at x_0 if and only if $\lim_{x \to x_0^-} f(x) = f(x_0) = \lim_{x \to x_0^+} f(x)$. EOE.

Definition (One-sided continuity)
Let f be a function and x_0 an interior point of its domain. We say f is **right-continuous** or **right-sided continuous** if $\lim_{x \to x_0^+} f(x) = f(x_0)$. A similar definition applies to **left-continuous** or **left-sided continuous**, and then $\lim_{x \to x_0^-} f(x) = f(x_0)$. EOD.

Note that a function is continuous at an interior point x_0 of its domain if and only if it is both left- and right-continuous.

Example
A typical example of one-sided continuous functions are certain types of step functions. The prototype of such a function is the Heaviside[8] step function, here defined at

$$H(x) := \vdash \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

[8] Oliver Heaviside (1850 – 1925)

It is right-continuous at $x_0 = 0$ but not continuous. Sometimes $H(0)$ is defined as $\frac{1}{2}$, in which case the function is neither right- nor left-continuous at 0. EOE.

The following example is occasionally useful:

Lemma
Let $f: [a, b] \to \mathbb{R}$ be a monotone function. Then for all $x_0 \in (a, b)$ both one-sided limits exist. For a the right-hand and for b the left-hand limit exists. EOL.

Proof. Indeed, let $x_0 \in [a, b)$, and let (x_n) be any monotone decreasing sequence converging to x_0 with $x_n \neq x_0$. Then $f(x_n)$ is bounded (by $f(a)$ below and $f(b)$ above), monotone, and hence $f(x_n)$ converges. Let (y_n) be another such sequence. Then so is $z_n = x_1, y_1, x_2, y_2, x_3, y_3, \dots$. It follows $(x_n), (y_n), (z_n)$ all monotonely converge to x_0, and $f(x_n), f(y_n)$ are convergent subsequences of $f(z_n)$. Thus, $\lim\limits_{x \to x_0^+} f(x)$ exists and is proper. The argument for left-handed limits with $x_0 \in (a, b]$ is similar. EOE.

Of course, the one-sided limits at interior points need not coincide (and f therefore need not be continuous). But this cannot happen too often, as the following nice observation (taken from Heuser, *Lehrbuch der Analysis, Teil 1*, 1993, p239) shows:

Corollary
A monotone function $f: [a, b] \to \mathbb{R}$ is continuous except at possibly a countable number of points. EOC.

Proof. Let $x_0 \in (a, b)$. Then $L_-(x_0) = \lim\limits_{x \to x_0^-} f(x)$ and $L_+(x_0) = \lim\limits_{x \to x_0^+} f(x)$ both exist. By UTALOF 4, we must have $L_-(x_0) \leq f(x_0) \leq L_+(x_0)$, with equality if and only if f is continuous at x_0 (see also the first example in 4.8). f is not continuous at x_0 if $L_-(x_0) \neq L_+(x_0)$. If this is the case there exists a rational number $R(x_0) \in (L_-(x_0), L_+(x_0))$. Note that if $x_1 < x_2$ then $L_+(x_1) < L_-(x_2)$.

Let N be the set of all $x \in [a, b]$, where f is not continuous. Then for each $x \in N$ we obtain a rational number $R(x)$ as described. Also, if $x \neq y \in N$, then $R(x) \neq R(y)$ (if $x < y$, say, then $R(x) \leq L_+(x) < L_-(y) \leq R(y)$).

It follows that R defines an injective mapping $R: N \to \mathbb{Q}$. As $\mathbb{Q}$ is countably infinite, this means N is at most countably infinite as well. QED.

Remark
It is known that any monotone function on a closed interval $[a, b]$ is integrable, and therefore by the Lebesgue Criterion continuous almost everywhere. But this is a much more involved argument, and actually yields a weaker result: almost everywhere means outside a set of measure zero. Not every set of measure zero is countable though. EOR.

4.8.3 Continuous extension

Suppose $f: D \to \mathbb{R}$ is a continuous function, and $x_0 \notin D$ is an accumulation point of D. The most relevant cases for us is $D = I \setminus x_0$ where $x_0 \in I$ and I is an interval.

It is a natural question to ask whether f can be **extended** to $D \cup \{x_0\}$ in a *continuous* fashion, that is, is there $\tilde{f}$ defined and continuous on $D \cup \{x_0\}$ such that $f = \tilde{f}\,|_D$ is the restriction of $\tilde{f}$ to D (of course, f can always be extended, by e.g. putting $f(x_0) = 0$).

Theorem (Extension Theorem)
Let $f: D \to \mathbb{R}$ be continuous and x_0 an accumulation point of D. Then there exists a unique continuous extension $\tilde{f}: D \cup \{x_0\} \to \mathbb{R}$ (with $\tilde{f}\mid_D = f$) if and only if $\lim_{x \to x_0} f(x)$ exists and is proper. If that is the case $\tilde{f}(x_0) = \lim_{x \to x_0} f(x)$. EOT.

Proof. There is nothing much to do: if a continuous extension $\tilde{f}$ exists, then $\tilde{f}(x_0) = \lim_{x \to x_0} \tilde{f}(x) = \lim_{x \to x_0} f(x)$ (because $f = \tilde{f}$ on D). This also shows that $\tilde{f}$ is unique in this case (as $\tilde{f}(x_0)$ is determined by f). Conversely if $L = \lim_{x \to x_0} f(x)$ is exists and is proper, then $\tilde{f}(x) \coloneqq L$ if $x = x_0$ and $\tilde{f}(x) \coloneqq f(x)$ if $x \in D$ defines a continuous function on $D \cup \{x_0\}$ (see the lemma above in 4.8). QED.

***Remark**
In this context the following question becomes relevant: in some courses you might be asked to "find the domain of the function" $f(x) = \frac{x^2-1}{x+1}$. Usually this means to find the "maximal" domain where the function is defined. What is often meant is that you should observe that $x^2 - 1 = (x-1)(x+1)$ and therefore the $x+1$ in the denominator cancels. This works for rational functions on the real line, as multiplication (and hence division) can be defined independently of pointwise evaluation. However, to make this precise one should use the notion of a polynomial ring.
An alternate interpretation of the question could be "find the maximal domain where this function can be defined continuously" (which is a slightly imprecise question).
But as a word of warning, in general, if f, g are functions (even continuous ones) defined on an interval I, say, it makes no sense to define $\frac{fg}{g} = f$. It could be that $fg = 0$ even if none of f, g are the zero function. If that is the case, then if we put $\frac{fg}{g} = f$, we also have $\frac{fg}{g} = \frac{0}{g} = 0$, which is a contradiction if $f \neq 0$. (Of course, one could say, "OK, $\frac{fg}{g} = f$ wherever $g(x) \neq 0$." This is essentially the approach we are taking.)
For this reason we stick to a formal approach: a function defined by a formula such as $\frac{x^2-1}{x+1}$ is defined exactly where the formula can be evaluated "as is", that is in this case, where $x \neq -1$. Whether or not this function could then be extended (continuously or otherwise) to a larger domain is a secondary question. In that sense the "domain" of the function $\frac{x}{x}$ is $\mathbb{R} \setminus \{0\}$. While this rarely causes confusion, one should distinguish between a defining formula and a function. EOR.

***Exercise**
Find examples of (continuous) functions f, g defined on $[0,1]$ such that $f, g \neq 0$ but $fg = 0$. EOE.

4.8.4 The oscillation of a function

Let f be defined on some set D and bounded above or below. For any interval I we may define

$$\Omega_f(I) \coloneqq \sup f(I \cap D) - \inf f(I \cap D) \geq 0$$

If f is bounded, then this is always a real number. It is called the **oscillation** of f on I.

Now let $x_0 \in D$. For any $\delta > 0$ we may define $\Omega_{f,x_0}(\delta) \coloneqq \Omega_f\big(D \cap (x_0 - \delta, x_0 + \delta)\big)$. This defines a function on $(0, \infty)$, and in fact it is a monotone increasing function.

By the previous section, we conclude that

$$\omega_f(x_0) \coloneqq \lim_{\delta \to 0} \Omega_{f,x_0}(\delta)$$

exists, and $\omega_f(x_0) \geq 0$. We call this the **oscillation** of f at x_0.

Exercise
Show that
$\Omega_f(I) = \sup_{x,y \in D \cap I} \big(f(x) - f(y)\big) = \sup_{x,y \in D \cap I} |f(x) - f(y)|$. EOE.

Proposition
f is continuous at x_0 if and only if $\omega_f(x_0) = 0$. EOP.

Proof. Suppose f is defined on D and continuous at x_0. Let $\varepsilon > 0$. There exists $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0 + \delta) \cap D$, we have $|f(x) - f(x_0)| < \frac{\varepsilon}{3}$. It follows that $\sup f\big((x_0 - \delta, x_0 + \delta) \cap D\big) \leq f(x_0) + \frac{\varepsilon}{3}$. Likewise, $\inf f\big((x_0 - \delta, x_0 + \delta) \cap D\big) \geq f(x_0) - \frac{\varepsilon}{3}$. Together we get

$$\Omega_{f,x_0}(\delta) \leq \frac{2\varepsilon}{3} < \varepsilon$$

This shows that $\omega_f(x_0) = \lim_{\delta \to 0} \Omega_{f,x_0}(\delta) = 0$.

For the converse, suppose $\omega_f(x_0) = 0$, and let $\varepsilon > 0$. Then there exists δ such that $\Omega_{f,x_0}(\delta') < \varepsilon$ for all $0 \leq \delta' < \delta$. By the previous exercise this means that $|f(x) - f(y)| < \varepsilon$ for all $x, y \in (x_0 - \delta, x_0 + \delta) \cap D$. This applies in particular if $y = x_0$, and then $|f(x) - f(x_0)| < \varepsilon$ for all $x \in (x_0 - \delta, x_0 + \delta) \cap D$. This shows f is continuous at x_0. QED.

This concept will be of great use when we discuss the conditions on a function to be integrable.

4.8.5 The Cauchy criterion for proper limits

We have seen that a sequence is convergent if and only if it is a Cauchy sequence. There are similar results for proper limits of functions.

Lemma (Cauchy Criterion for limits of functions)
Let f be a function and x_0 an accumulation point of its domain. Then f has a proper limit as x approaches x_0, ie. there is $L \in \mathbb{R}$ such that $\lim_{x \to x_0} f(x) = L$ if and only if for every $\varepsilon > 0$ there is $\delta > 0$ such that whenever $0 < |x - x_0|, |y - x_0| < \delta$ (and $x, y \in D$), $|f(x) - f(y)| < \varepsilon$. EOL.

Proof. Let $\varepsilon > 0$. If $\lim_{x \to x_0} f(x) = L$ with $L \in \mathbb{R}$, then by the $\varepsilon - \delta$-definition of proper limits above, there is $\delta > 0$ such that for all $x \in D$ with $0 < |x - x_0| < \delta$ we have $|f(x) - f(x_0)| < \frac{\varepsilon}{2}$. If x, y both are in the domain of f and $|x - x_0|, |y - x_0| < \delta$, then $|f(x) - f(y)| \leq |f(x) - L| + |L - f(y)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$ as required.

For the converse suppose that for every $\varepsilon > 0$ a $\delta > 0$ as claimed exists. Let x_n be any sequence in the domain of f with $x_n \neq x_0$ for all n and $x_n \to x_0$. Let $\varepsilon > 0$. Then there is $\delta > 0$ such that whenever $x, y \neq x_0$ are in the domain and in $(x_0 - \delta, x_0 + \delta)$, then $|f(x) - f(y)| < \varepsilon$. There is n_0 such that for all $n > n_0$ we have $|x_n - x_0| < \delta$. Consequently, $|f(x_m) - f(x_n)| < \varepsilon$ for all $m, n > n_0$. In other words, $f(x_n)$ is a Cauchy sequence, and therefore convergent.

Now let x_n, y_n be two sequences in the domain of f converging to x_0, with $x_n, y_n \neq x_0 \forall n$. Then the combined sequence z_n defined as $x_1, y_1, x_2, y_2, \ldots$ also converges to x_0. Since by the above all three $f(x_n), f(y_n), f(z_n)$ are convergent sequences, and since x_n, y_n are subsequences of z_n, it follows that all three have the same (proper) limit. It follows there is a well defined number L such that $\lim_{n\to\infty} f(x_n) = L$ for all such sequences x_n and $\lim_{x\to x_0} f(x) = L$ exists and is proper. QED.

Exercise
Show that $\lim_{x\to\infty} f(x)$ exists and is finite if and only if for every $\varepsilon > 0$ there exists $B > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all $x, y > B$. Show an analogous result for $\lim_{x\to-\infty} f(x)$. EOE.

We have observed before that for sequences, the Cauchy criterion is usually not helpful for finding the actual limit. Not surprisingly, the same applies to functions.

4.9 Trigonometric functions revisited

We have defined $\sin x$ and $\cos x$ geometrically. As mentioned, these definitions were not rigorous. We have now the tools to provide an axiomatic definition of these functions.

Theorem (Trigonometric Function Theorem; TFT)
There are two functions $\sin x$ and $\cos x$, defined on $\mathbb{R}$ and continuous at 0 such that

1. $\sin x$ is odd.
2. $\cos x$ is even.
3. $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$.
4. $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$.
5. $\lim_{x\to 0} \frac{\sin x}{x} = 1$.
6. $\cos(0) = 1$.

EOT.

Most of these six properties should be intuitively clear, given the geometric origin of the definition. We will not waste time proving these properties geometrically (in particular, as indicated, as this would have no real bearing on our analytic discussion of functions).

It turns our that these six properties characterize the two functions uniquely, and we will assume that there are precisely these functions for the time being.

A rigorous proof of these properties (or, rather of the existence of a unique pair of functions with these properties) will be provided later.

Lemma
$\sin x$ and $\cos x$ are continuous everywhere. EOL.

Proof. We prove the lemma in case of $\sin x$. The case of $\cos x$ is similar.

Let $x_0 \in \mathbb{R}$. We must show that $\lim_{x \to x_0} \sin(x) = \sin(x_0)$. Equivalently, we must show that $\lim_{h \to 0} \sin\,(x_0 + h) = \sin(x_0)$. Now

$$\sin(x_0 + h) = \sin(x_0)\cos(h) + \cos(x_0)\sin(h)$$

Taking the limit for $h \to 0$ on the right-hand side and observing that $\sin x$ and $\cos x$ are continuous at 0, we get

$$\lim_{h \to 0} \sin(x_0 + h) = \sin(x_0) \cdot 1 + \cos(x_0)\sin(0) = \sin(x_0)$$

QED.