Math Camp: Basic Concepts

Euclidean Space, Functions, and Sequences

Jacob Hukill

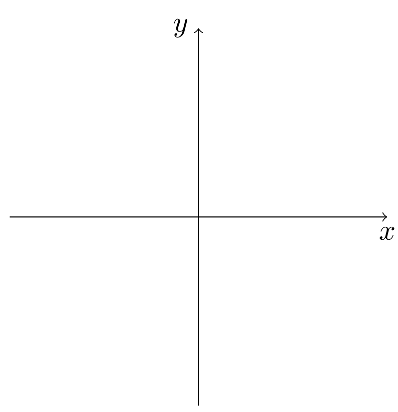
August 16, 2023

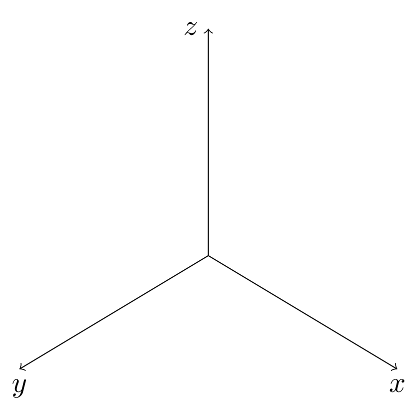
For the purpose of this math camp, the treatment of the material will be fairly informal - in some cases, providing background and intuition, while in others providing the 'cookbook' for how to approach different types of problems in economics. In general, much of the material assumes that you have some familiarity with the mathematics already. In these instances, this will act as a review.

Euclidean Space and Subsets of $\mathbb{R}^n$

The set of all real numbers, irrational and rational, is denoted by $\mathbb{R}$. You are likely already familiar with the real line and its properties. By applying the Cartesian product, we can get $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2 = \{(x, y) : x \in (-\infty, \infty), y \in (-\infty, \infty)\}$, which is now a two dimensional plane. Note that in contrast to the set $\{x, y\}$, where order does not matter, elements of $\mathbb{R}^2$ are defined as *ordered pairs*, where the order of the elements define distinct points in the two-dimensional plane.

We can apply the Cartesian product again to get $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3 = \{(x, y, z) : x \in (-\infty, \infty), y \in (-\infty, \infty), z \in (-\infty, \infty)\}$ which defines three dimensional space.





Similarly, for some positive integer n, we can define the n-fold Cartesian product of $\mathbb{R}$ as $\mathbb{R}^n$. $\mathbb{R}^n$ is known as *n-dimensional Euclidean space.* A point in $\mathbb{R}^n$ is a vector $x = (x,..., x_i, ...x_n)$, where x_i is the $i - th$ coordinate of x. Given any two n-vectors x and y, we write that:

$$
\begin{aligned}
&x = y,\ if\ x_i = y_i,\ i = 1, ..., n \\
&x \geq y,\ if\ x_i \geq y_i,\ i = 1, ..., n \\
&x \leq y,\ if\ x_i = \leq_i,\ i = 1, ..., n \\
&x,\ if\ x_i > y_i,\ i = 1, ..., n
\end{aligned}
$$

You will likely run into three special subsets of $\mathbb{R}^n$, the null vector and the *nonnegative* and *strictly* positive orthants[1] of $\mathbb{R}^n$, respectively defined as

$$0 = (0, ..., 0) \in \mathbb{R}^n$$

and

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n | x \geq 0\}$$

and

$$\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n | x \gg 0\}$$

I now describe three useful structures on $\mathbb{R}^n$: the Euclidean inner product, the Euclidean norm, and the Euclidean metric.

Definition 1 (Euclidean Inner Product)**.** Given $x, y \in \mathbb{R}^n$, the Euclidean inner product of the vectors x and y, denoted $x \cdot y$, is defined as:

$$x \cdot y = \sum_{i=1}^{n} x_i y_i$$

In $\mathbb{R}^1$, the Euclidean inner product is just the product of two real numbers, x and y.

Theorem 1. *The inner product has the following properties for any vectors $x, y, z \in \mathbb{R}^n$ and scalars $a, b \in \mathbb{R}$:*

1. *Symmetry: $x \cdot y = y \cdot x$*

2. *Bilinearity: $(ax + by) \cdot z = ax \cdot z + by \cdot z$ and $x \cdot (ay + bz) = x \cdot ay + x \cdot bz$*

[1] An orthant is the analogue in n-dimensional Euclidean space to the quadrant in a plane.

3. *Positivity:* $x \cdot x \geq 0$, *with equality holding if and only if* $x = 0$.

Theorem 2 (Cauchy-Schwartz Inequality)**.** *For any* $x, y \in \mathbb{R}^n$, *we have:*

$$|x \cdot y| \leq (x \cdot x)^{1/2}(y \cdot y)^{1/2}.$$

Definition 2 (Euclidean Norm)**.** The Euclidean norm of a vector $x \in \mathbb{R}^n$, denoted $\| x \|$ is defined as

$$\| x \| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

This gives us a notion of length or magnitude. The norm in one dimension is just the absolute value of x. In the second, $\| x \| = \sqrt{x_1^2 + x_2^2}$, and in $\mathbb{R}^3$ is $\| x \| = \sqrt{x_1^2 + x_2^2 + x_3^2}$. Effectively, it measures the the size of a vector x, or the distance from the point x to the origin. Notice that these these notions have tree important properties:

- Positive definiteness: $\| x \| \geq 0$ and $\| x \| = 0$ if and only if $x = 0$. This says that the length should be nonnegative and that only "zero" elements have length of zero.
- Positive homogeneity: $\| \alpha x \| = |\alpha| \, \| x \|$, where alpha is some scalar. This says that if we increase each element by a magnitude of α, then the length also increases by the same magnitude.
- Triangle inequality: $\| x \| + \| y \| \geq \| x + y \|$. In other words, if you are traveling from a point x to point z and along the way you deviate from the *straight* path, then you have to cover at least the same distance as was originally planned. In other words: the shortest distance between two points is a straight line.

You could construct an alternative norm to the Euclidean norm presented above (and construct an alternative distance function below), if it follows the above properties. More formally,:

Definition 3. Let V be a vector space and X be an underlying set of points. A norm of $x \in X$ is a function $\| \cdot \|$: $X \to \mathbb{R}$ if for all $x \in X$ it satisfied the following properties:

1. Positive definiteness: $\| x \| \geq 0, \| x \| = 0$ iff x=0,
2. Positive homogeneity: $\| \alpha x \| = |\alpha| \, \| x \|$,
3. Triangle inequality: $\| x \| + \| y \| \geq \| x + y \|$.

You may notice that we restrict attention to some space V, called a vector space. For our purposes, just think of a vector space as just $\mathbb{R}^n$ space with the usual notion of addition and scalar multiplication.

We can relate the inner product and the norm by the following identity:

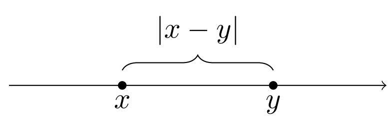
$$\| x \| \leq \| x \| \| y \|$$

Finally, we can construct some sort of distance $d(x, y)$ to describe how far apart two vectors x, y are.

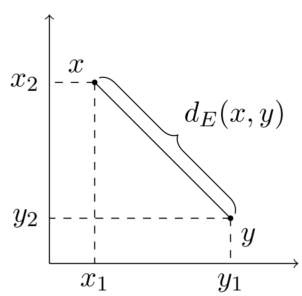
Definition 4. The Euclidean distance $d(x, y)$ between two vectors x and y in $\mathbb{R}^n$ is given by

$$d(x, y) = \| x - y \| = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}.$$

Notice how the distance function d is related to the norm. In one dimension, distance is simply given by $|x - y|$.



In two dimensions,



The distance function $d(\cdot)$ is called a *metric*. Notice that these natural notions of distance have the following properties:

- Non-negativity: For any points x, y we have that $d(x, y) \geq 0$ and equality holds iff $x = y$. That is, the distance always gives a non-negative number and distance is zero if and only if two points are the same.
- Symmetry: $d(x, y) = d(y, x)$. The distance between x and y is the same as the distance btween y and x.
- Triangle Equality: $d(x, y) + d(y, z) \geq d(x, z)$. This follows the same intuition as in the norm.

It turns out that these three properties are fundamental for the most useful and general results that one obtains. A formal definition of a metric is as follows:

Definition 5 (Metric, distance function)**.** A metric or distance function defined on a set X is a real-valued, non-negative fucntion $d : X \times X \to \mathbb{R}$ such that for every $x, y, z \in X$ we have:

1. $d(x, y) \geq 0$ and $d(x, y) = 0$ iff $x = y$,
2. $d(x, y) = d(y, x)$,
3. $d(x, y) + d(y, z) \geq d(x, z)$.

When we say "let (X, d) be a metric space, we generally mean that there is some set X and that the distance between elements from this set is given by $d(x, y)$. I.e., the "closeness" between elements are dictated by some function $d(x, y)$. For example, the $n-$dimensional Euclidean metric space is $(\mathbb{R}^n, d_E)$, where $d_E(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$.

Just like with norms, there are many other functions $d(x, y)$ other than Euclidean distance that have the above three properties and play fundamental roles in analysis. Below are some examples.

Example 1. In $\mathbb{R}^n$ the following define a distance:

- Euclidean distance: $d_E(x, y) = d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$.
- Manhattan or (Taxi-cab) distance $d_M(x, y) = \sum_{i=1}^{n} |x_i - y_i|$.
- Maximum distance: $d_{max}(x, y) = max\{|x_1 - y_1|, ..., |x_n - y_n|\}$
- l-metric: $d_l(x, y) = \left(\sum_{i=1}^{n}(x_i - y_i)^l\right)^{1/l},\ l > 1$

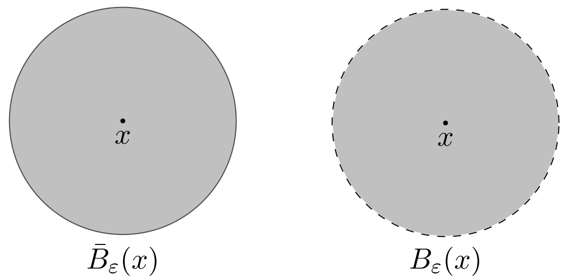
Why do we care about distance between x and y? We want to have some measure of how close two points in space that we work with are. If points are getting arbitrarily closer and closer to each other, then we can conclude that after a while, they are almost the same. If points are not getting closer to each other, than we might be interest in what exactly is the gap between them. With this notion of distance, we can define open and closed balls.

Definition 6 (Open and closed ball)**.** Given a metric space (X, d), an open-ε ball centered around $x \in X$ with radius ε is the set given by

$$B_\varepsilon(x) = \{y \in X : d(x, y) < \varepsilon\}.$$

Similarly, a closed-ε ball is defined as

$$\overline{B}_\varepsilon(x) = \{y \in X : d(x, y) \leq \varepsilon\}.$$



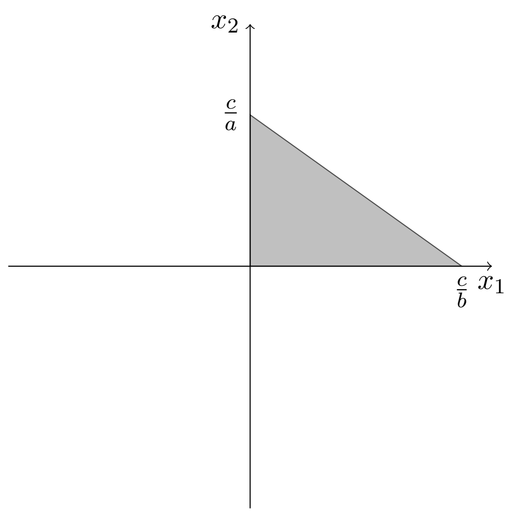
Properties of subsets in $\mathbb{R}^n$

Now that we have these notions of distance, we can describe additional properties of subsets in $\mathbb{R}^n$. For our purposes, the most useful "types" of sets are *open, closed, bounded, compact, and convex sets.* For example, in optimization theory, the most frequently used existence theorem relies on the assumption that the set is closed and bounded. We define these now.

Definition 7. A set $\subseteq \mathbb{R}^n$ is bounded if and only if, for any $x \in S$, there exists some $\varepsilon \in (0, \infty)$ such that $B_\varepsilon(x) \subset S$.

In other words, a set is bounded if it can be entirely contained within some $\varepsilon-$ ball. This boils down to the idea that all members of the set are less than some number. The example below illustrates how to show that a set is bounded.

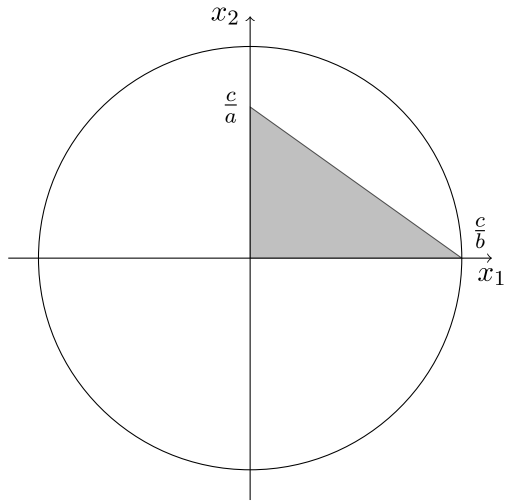
Example 2. Let $X = \{(x_1, x_2) \in \mathbb{R} : x_1 \geq 0, x_2 \geq 0, ax_1 + bx_2 \leq c\}$ where a, b, c are positive constants. The following figure illustrates this.



How do we show that this is bounded? Notice that as long as a, b, c are positive numbers, then we could define

$$\varepsilon = max\{\frac{c}{a}, \frac{c}{b}\} < \infty$$

and then the ball $B_\varepsilon(0)$ would contain X completely. Therefore, the set X is bounded.



Definition 8 (Open set)**.** A set $S \subseteq \mathbb{R}^n$ is open if and only if, for any x in S, there exists some $\varepsilon > 0$ such that $B_\varepsilon(x) \subset S$.

Intuitively, when the given set is open, then for any point in the set, any minor deviation leaves us within the set. Often we also define open sets using the notion of *interior points.*

Definition 9 (Interior point)**.** A point $x \in S$ is an interior point of S if there exists an open ball around x that is contained in S. That is,

$$\forall x \in S,\ \exists \varepsilon > 0 \ \ s.th.\ B_\varepsilon(x) \subseteq S.$$

Notice that the previous definition is obtained by saying that a set S is open if and only if every element of the set is an interior point. Below are some examples.

Example 3. The interval $(0, 1) = \{x \in \mathbb{R} | 0 < x < 1\}$ is an open set. Why? Intuitively, it is because it does not include endpoints 0 and 1, but includes any point that is arbitrarily close to the endpoints. To show that it is, we can just apply the definition of the open set. Take any $x \in (0, 1)$, then we have to show that there is some ε such that $B_\varepsilon(x) \subseteq (0, 1)$. Once we find this ε, we are done. In $\mathbb{R}$, an open ball around x with radius ε is just an interval $(x-\varepsilon, x+\varepsilon)$. Now let $\varepsilon = \frac{1-x}{2}$ if $x \geq 1/2$, and $\varepsilon = \frac{x}{2}$ if $x \leq 1/2$.

Example 4. A single point $a \in \mathbb{R}$ is not an open set. Why? The only ball that might give us $B_\varepsilon(x) \subseteq a$ is just the one which exactly is a (since it's a single element). But then ε must be zero, or the ball would include more elements than a. Therefore, $\nexists \varepsilon > 0$, and thus the set is not an open set.

Example 5. A line in $mathbbR^2$ is not an open set. Why? Suppose that it were open. Then there $\exists \varepsilon > 0$ such that $B_\varepsilon(x) \subset A$, where A is now a line in a plane. A similar argument to the previous example would imply that $\varepsilon = 0$, which shows that a line in $\mathbb{R}^2$ is not an open set.

Given the definition of an open set, we can define a closed set, which is its opposite.

Definition 10 (Closed set)**.** A set $S \subseteq \mathbb{R}^n$ is closed if and only if its complement $\mathbb{R}^n \diagdown S$ is open.

This can be useful, for example, if we want to show that set S is closed and it suffices to show that S^c is open. Notice that the defintion does not say that whenever a set is not open, then it is closed - some sets are neither closed nor open. Consider for example, the interval $[0, 1)$ which is neither. A set can also be both closed and open, as is the case for and $\mathbb{R}^n$. Check the definitions carefully to ensure that this makes sense.

Theorem 3. *Let (X, d) be a metric space. Then,*

1. *Both and X are open in X,*
2. *The union of an arbitrary (possibly infinite) family of open sets is open,*
3. *The intersection of a finite collection of open sets is open.*

It is indeed true that for any point in we can find an open ball satisfying the above definition. In fact, we can do anything we want since there is no such point in . The same observation makes it clear that X is also open, since there is really nothing outside of X (it is a universal set).

To illustrate part 2 of the theorem, let the family of open intervals $\{A_n\}$ be given by $A_n = (-1/n, 1/n)$. Let's consider an infinite union of the family of these open sets. Define a partial union as $S_N = \bigcup_{n=1}^N A_n$. Then this clearly defines a recursion $S_N = S_{N-1} A_N$. Thus,

$$\begin{aligned} S_2 &= (-1, 1) \bigcup (-1/2, 1/2) \\ &= (-1, 1) \\ S_3 &= (-1, 1) \bigcup (-1/3, 1/3) \\ &= (-1, 1) \end{aligned}$$

and so forth, therefore as $N \to \infty$ we have

$$\bigcup_{n=1}^{N} A_n = (-1, 1)$$

which is an open set.

Whenever we have some useful properties for open sets, we can always translate it to closed sets as well since we have DeMorgan's Law[2].

Theorem 4. *Let (X, d) be a metric space. Then,*

1. *Both and X are closed in X,*
2. *The intersection of an arbitrary family of closed sets is closed,*
3. *The union of a finite collection of closed sets is closed.*

We also have special sets called *compact* sets, which play a crucial role in optimization theory. The "raw" definition of compact sets would require us to introduce a couple more definitions. However, since we are mostly working in $\mathbb{R}^n$, we can use the following theorem.

Theorem 5 (Heine-Borel)**.** *A set $S \subseteq \mathbb{R}$ is **compact** if and only if it is both closed and bounded.*

So whenever we are working in $\mathbb{R}$, compact sets are equivalent to closed and bounded sets. But what about $\mathbb{R}^n$? We have the result saying that if $A_1, ..., A_n$ are compact subsets of $\mathbb{R}$, then $A_1 \times ... \times A_n$ is also a compact subset of $\mathbb{R}^n$.

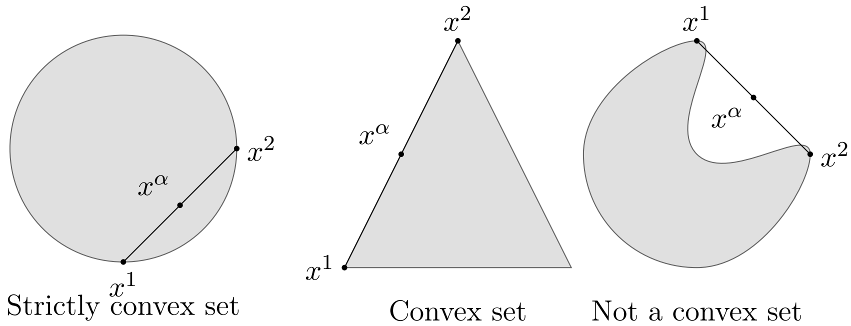
Finally, we will conclude this section by discussing convex sets.

Definition 11 (Convex sets)**.** A set $S \subseteq \mathbb{R}^N$ is **convex** if and only if, for any $x, y \in S$ and any $t \in [0, 1]$, the midpoint $tx + (1-t)y$ also belongs to S.
A set $S \subseteq \mathbb{R}^N$ is **strictly convex** if and only if, for any $x \neq y \in S$ and any $t \in (0, 1)$, there exists some $\varepsilon > 0$ such that the open ball around the midpoint $tx + (1-t)y$ lies entirely in S,

$$B_\varepsilon(tx + (1-t)y) \subseteq S$$

Since this property holds for all $t \in (0, 1)$, the entire straight line segment between x, y lies within the set. The midpoint is called a convex combination.

[2]Which relates the union and intersections between sets and their complements



There are various applications where convex sets play an important role. For example, in optimization theory, the uniqueness results rely on strict convexity of constraint sets.

Properties of Functions

We are often interested in defining relationships between sets. This is where functions and correspondences come in. A function f gives a recipe of how to associate elements of one set to another. We usually refer to this by saying that f maps X into Y. More formally,

Definition 12. Let X and Y be sets. A **function** $f : X \to Y$ is a rule that assigns to each element $x \in X$ a single element $y = f(x) \in Y$. The set X is called the Domain, and the set Y is called the Range, or Codomain. For any $A \subseteq X$, we call $f(A) = \{f(x)|x \in A\}$ the image of A.

You could think of a function as "arrows" that point from an element of the set X to one in the set Y. Each element in X always has an outgoing element, but an element in Y may have no, one, or many incoming arrows. In the previous definition, we emphasize *unique* for a reason. This is exactly the point that differentiates a function from a correspondence. You will run into correspondences throughout the upcoming semester. We can be a bit more precise on how the "arrows" describe a function.

Definition 13. A function $f : X \to Y$ is **injective** or **one-to-one** if for every $x_1, x_2 \in X, x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$.
A function $f : X \to Y$ is **surjective** or **onto** if for all $y \in Y$, there exists an $x \in X$ such that $f(x) = y$.
A function $f : X \to Y$ is **bijective** or **invertible** if for every $y \in Y, x_1 \neq x_2$ there is a unique $x \in X$ for which $y = f(x)$.

A function is bijective if it is both injective and surjective. The reason why we are interested in bijective functions is that such functions are invertible, i.e. they have well defined inverse functions. The inverse function basically just reverses the process defined by f.

Theorem 6. *If $f : X \to Y$ and it is an invertible function then there exists $g : Y \to X$ such that $f(g(y)) = y$ and $g(f(x)) = x$. We denote g by f^{-1}.*

In the above notation, we used f^{-1}: don't confuse this with $1/f$. In order to avoid confusion, if we want to write the reciprocal, we write $(f(x))^{-1}$

It is useful at this point to define some common properties of functions.

Definition 14. Let $D \subset \mathbb{R}^n$. A function $f : D \to \mathbb{R}$ is **increasing** if and only if for all $x, y \in D$

$$x \geq y \implies f(x) \geq f(y).$$

A function with this property is also called **monotone**. If in addition, $x > y$ implies $f(x) > f(y)$, we say that the function is **strictly increasing**. Similarly, a function with this property is also called **strictly monotone.**

One of the very attractive properties of monotonic functions is given by the following theorem.

Theorem 7. *Any monotonic function is one-to-one.*

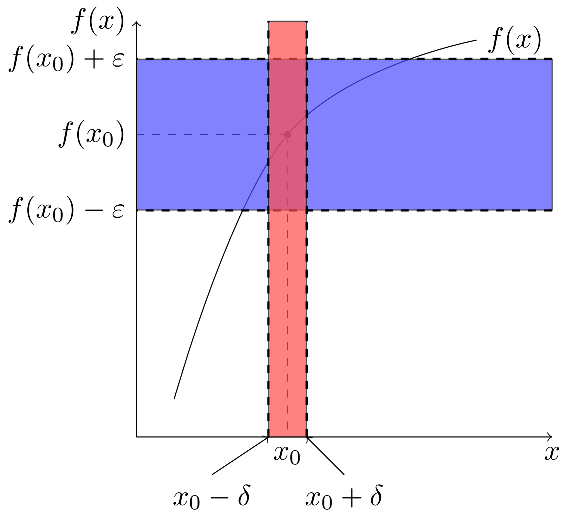
Another very convenient property of monotonic functions is that they either don't change the ordering of the elements in the domain, or they reverse it. For example, if function $f(x)$ is monotonically increasing, then $x_1 \geq x_2 \implies f(x_1) \geq f(x_2)$. So the ordering does not change.

Another property of functions that is extremely important is continuity. It plays a key role in ensuring that optimization problems have a solution. There are several equivalent definitions, and which one we use is usually context dependent - sometimes it is easier to apply one definition than another. We start with the classical definition, the so called "$\varepsilon - \delta$" definition.

Definition 15 (Continuity)**.** Let (X, d) be a metric space. A function $f : X \to \mathbb{R}$ is said to be continuous at x_0 if and only if for all $\varepsilon > 0$, there exists a δ so that

$$|f(x) - f(x_0)| < \varepsilon \quad \forall x \in B_\delta(x_0).$$

If f is continuous for all $x \in X$, then f is continuous.



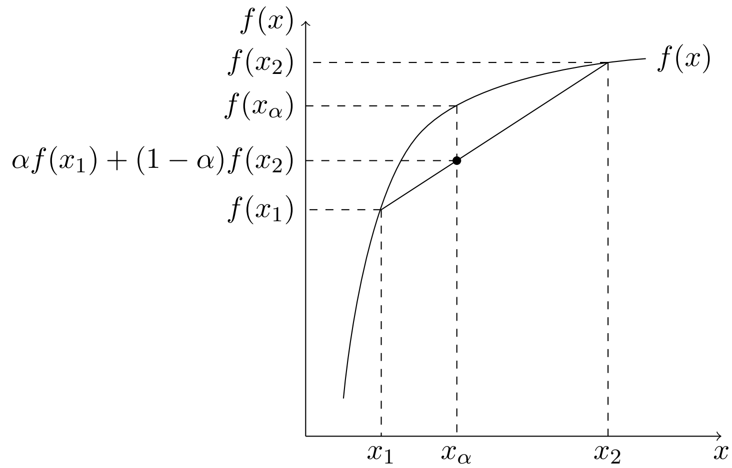
What the above definition says is that a function f is continuous if arbitrarily close points in the domain is mapped into arbitrarily close points in the range. The frgure illustrates this point. In the above figure, ε is given to us, and therefore we know the set of points in the range of f for which $|f(x)-f(x_0)|<\varepsilon$ (the blue shaded area). Then, we find a δ so for all points in the domain that are within an δ-ball of x_0, this implies that the distance between the function evaluated at this point and at x_0 is less than the ε that we were given. Intuitively, a function is continuous if you can draw its graph without lifting up your pen from the paper.

Definition 16. Let $D \subseteq \mathbb{R}^n$. A function $f: D \to \mathbb{R}$ is homogeneous of degree k if and only if for all $x \in D$ and $\alpha \geq 0, f(\alpha x) = \alpha^k f(x)$.

Definition 17. Let $D \subseteq \mathbb{R}^n$. A function $f: D \to \mathbb{R}$ is concave if and only if for all $x, y \in D$ and all $t \in [0,1]$,

$$f(tx+(1-t)y) \geq tf(x)+(1-t)f(y)$$

If $x \neq y$ and $t \in (0,1)$ implies strict inequality, then f is strictly concave.

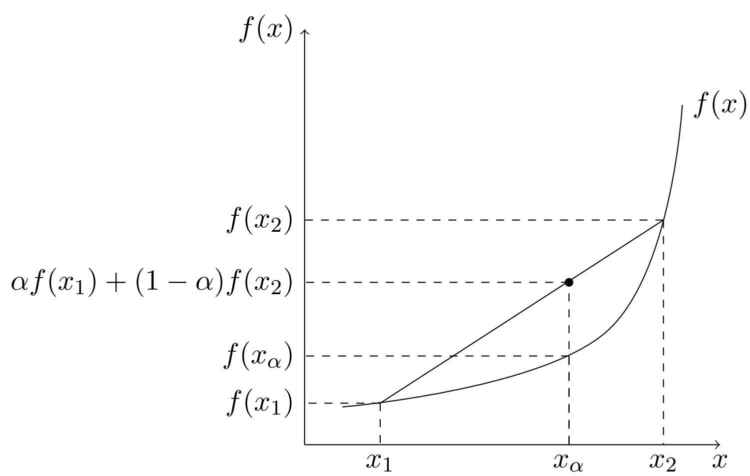


Concave function

Definition 18. Let $D \subseteq \mathbb{R}^n$. A function $f : D \to \mathbb{R}$ is convex if and only if for all $x, y \in D$ and all $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

If $x \neq y$ and $t \in (0, 1)$ implies strict inequality, then f is strictly concave.



Convex function

Sequences

A function from the natural numbers $f : \mathbb{N} \to S$ introduces a natural "ordering" of elements in S. We can now talk about $f(1)$ as the first element, $f(2)$ as the second, and $f(k)$ as the kth. This idea is the basis of the next defintion.

Definition 19. A sequence $\{x^k\}_{k\in\mathbb{N}} \subset S$ is an ordered list of numbers such that $x^k \in S$ for all $k \in \mathbb{N}$.

You could also define a sequence as a function $f : \mathbb{N} \to S$ with $f(k) = x^k$, but the above definition notation is more common. Some common sequences are:

- Let $x_n = n$. Then the sequence is $1, 2, 3, ...$
- Let $x_n = 1/n$. Then the sequence is $1, 1/2, 1/3, ...$
- Let $x_n = (-1)^{n-1}\frac{1}{n}$. Then the sequence is $1, -1/2, 1/3, ...$
- Let $x_1 = 1, x_2 = 1$. For some $n > 2, x_n = x_{n-1} + x_{n-2}$. Then the sequence is $1, 1, 2, 3, 5, 8, 13, ...$

Notice that the last is defined recursively, where the function f is given to us. If done this way, you would need the initial member of the sequence, x_0. This recursive formulation is a difference equation with initial condition x_0.

Sometimes we are interested in some parts of the sequence and not the whole sequence itself. Given a sequence $\{x_k\}$, we can extract another sequence that consists of members of the original sequence. For example, suppose a sequence is given by

$$x_k = \begin{cases} 1 & \text{if } k \text{ is odd} \\ -1 & \text{if } k \text{ is even} \end{cases}$$

so that $\{x_k\} = \{1, -1, 1, ...\}$. Then, if we extract the part of the sequence consisting of odd k, we will get another sequence $\{1, 1, ...\}$, a subsequence of $\{x_k\}$.

Definition 20. Consider any infinite subset $I \subset \mathbb{N}$. Then $\{x^{kn}\}_{k\in I}$ is a subsequence of $\{x^k\}_{k\in\mathbb{N}}$.

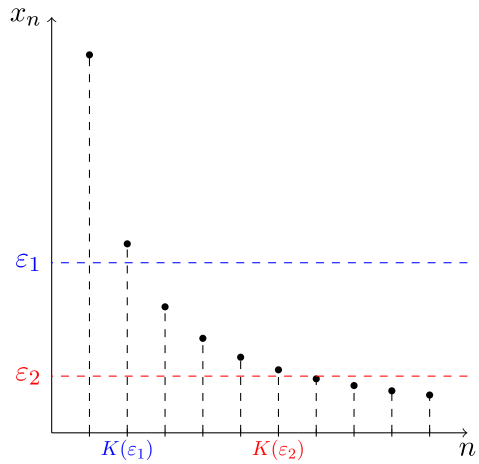
If we have a notion of distance, we can identify sequences that become more and more concentrated in the vicinity of a point as k grows.

Definition 21. A sequence $\{x^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ converges to $x_0 \in \mathbb{R}^n$ if and only if for every $\varepsilon > 0$, there exists a $K \in \mathbb{N}$ such that

$$\| x^k - x_0 \| < \varepsilon \quad \forall k \geq K$$

If there exists such an $x_0 \in \mathbb{R}^n$, we say that the sequence is convergent. We call x_0 the limit of the sequence, and write $x_0 = lim_{k\to\infty}$, or $x^k \to x_0$.

What the above definition really says is that if we manage to "trap" infinitely many members of the sequence in an arbitrary neighborhood of x_0 then it is a limit point. The following figure with sequence defined by $x_n = 1/n$ illustrates this idea.



Note that the limit x_0 does not have to be a member of the sequence itself - 0 is the limit of the sequence above, but is not actually part of it. It may be that all elements of a sequence are contained in some set S, such as the sequence above with $S = (0, 1]$, yet the limit is not contained within S. As it turns out, this only happens when the S is not closed. This allows for us to give another definition for what it means for a set to be closed.

Lemma 1 (Closed set). *A set $S \subset \mathbb{R}^n$ is closed if and only if the limit of any convergent sequence is contained in S,*

$$lim_{k\to\infty} x^k \in S \ \ \forall \textit{ convergent sequences } \ \{x^k\}_{k\in\mathbb{N}} \subset S.$$

Similarly, we can also define continuity using sequences

Lemma 2 (Continuity). *A function $f : \mathbb{R}^n \to \mathbb{R}$ is continuous at x if and only if, for all sequences $x_n \to x$, we have*

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

This results effectively says that we can exchange the limit operator and the function if and only if f is continuous at x. If we want to show that a function is discontinuous at a point x_0, then it suffices to find a sequence that converges to x such that $f(x_k)$ does not converge to $f(x)$, or $f(x)$ is not defined. The next two examples illustrate this point.

Example 6. Suppose that

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

and let us verify that at $x = 1$, the function is discontinuous. Since the function is not defined at $x = 1$, it is discontinuous at this point.

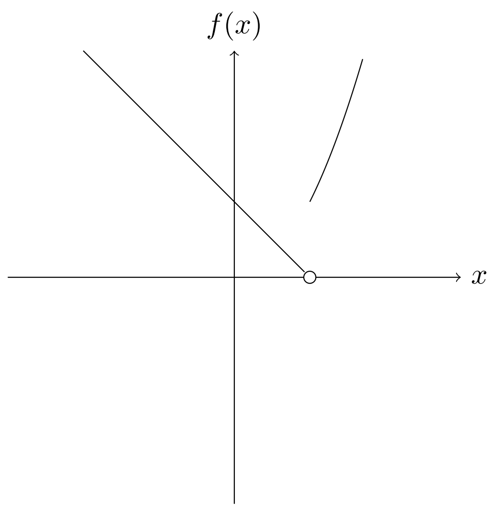
Example 7. Suppose that

$$f(x) = \begin{cases} x^2 \quad if \ x \geq 1 \\ 1 - x \quad if x < 1 \end{cases}$$

and let us verify that at x=1 the function is discontinuous. Suppose we take the sequence $x_n = 1 - 1/n$, so that $x_n \to 1$ from below. Then

$$f(x_n) = f(1 - 1/n) = 1 - (1 - 1/n) = 1/n.$$

which goes to 0 as $n \to \infty$. However, f(1)=1, and hence the function is discontinuous.



If a sequence does not grow "too far away" from zero, then we call it bounded.

Definition 22. A sequence $\{x^k\} \subset \mathbb{R}^n$ is bounded if and only if there exists $M \in \mathbb{R}$ such that $\| x^k - 0 \| < M$ for all $k \in \mathbb{N}$.

Theorem 8. *Any convergent sequence in* $\mathbb{R}^{\ltimes}$ *is bounded.*

Note that the opposite is not true - consider the sequence $x_n = (-1)^n$, which oscillates between -1 and 1. If we take every other element, however, we are looking at a constant subsequence, which is convergent. This is true more generally, as the following theorem states.

Theorem 9 (Bolzano-Weierstrass)**.** *Any bounded sequence* $\{x^k\}_{k\in\mathbb{N}}$ *in* $\mathbb{R}^n$ *has a convergent subsequence.*

This effectively says that no matter how crazy of a sequence we take, as long as it is bounded, then some part of it must converge. Thus is used often in various proofs, and becomes very important in the characterization of compact sets.

The Derivative and Differentiability of a Function

This, is the bread and butter for optimization theory. Differential calculus makes our lives really easy in terms of characterizing maxima and minima of functions. Derivatives are well studied objects, very tractable, and well suited for optimization problems. We start with derivatives of univariate functions, and later on when dealing with optimization problems in $\mathbb{R}^n$, we will study how these objects generalize in higher dimensions.

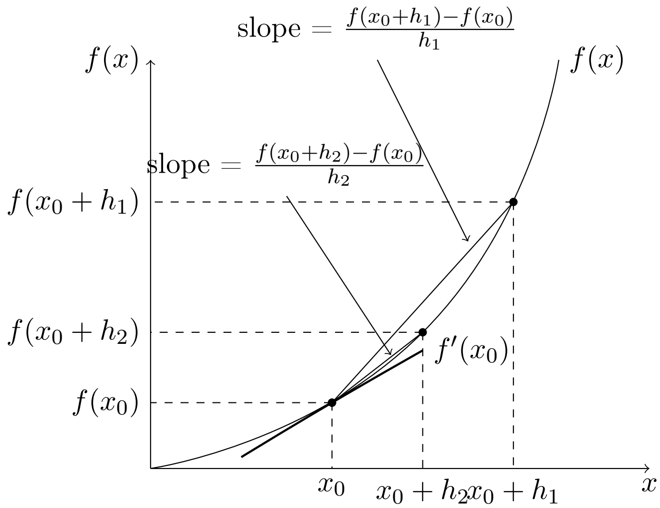
Definition 23. Let $f : D \to \mathbb{R}$ be given, where D is an open subset of $\mathbb{R}$. Then the derivative of f at x, denoted by $f'(x)$ is defined as

$$f'(x) = \lim_{y\to x} \frac{f(y) - f(x)}{y - x} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

where $h \in \mathbb{R}$.

Whenever the above limit exists, we interpret it as an instantaneous rate of change of a function f. Whenever the above derivative exists, we say that the function is differentiable at x; if this holds for all x, then we say that a function is differentiable. There are various notations for derivatives: $df/dx, df(x)/dx, Df(x)$, *etc.*

You will likely be familiar with the geometric interpretation of the derivative that is illustrated below.



Below I list derivatives that are more or less frequently used in applications:

Proposition 1. *Frequently used rules for differentiation:*

- $\frac{d(c)}{dx} = 0$, *where c is a constant*
- $\frac{dx^n}{dx} = nx^{n-1}$
- $\frac{de^x}{dx} = e^x$
- $\frac{dln(x)}{dx} = \frac{1}{x}$
- $\frac{da^x}{dx} = a^x ln(a), \quad a > 0$
- $\frac{dsin(x)}{dx} = cosx$
- $\frac{dcosx}{dx} = -sinx$

Below are some useful properites of differentiation for general functions that you should remember.

Proposition 2. *Suppose that f, g are differentiable functions. Then,*

1. *Additive rule:* $(f(x) + g(x))' = f(x)' + g(x)'$
2. *Product rule:* $(f(x)g(x))' = f'(x)g(x) + g'(x)f(x)$
3. *Chain rule:* $(f(g(x)))' = f'(g(x))g'(x)$

You might also recall there being a "quotient rule" that is not lists in the above rules. The reason is that it is actually a combination of (2) and (3), since:

$$\frac{f(x)}{g(x)} = f(x)[g(x)]^{-1}$$

and given that $g(x) \neq 0$, we can simply apply the rules to get that

$$\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

Differentiability of a function puts some implicit necessary structure on the function.

Proposition 3. *Every differentiable function on an open domain is continuous.*

This is very useful whenever we want to show that a function is continuous.

So far we have discussed first order derivatives for functions. The information that the derivative of a function contains is useful, since it tells whether the function at a given point is increasing, decreasing, or constant. We might also want to know more about the shape of the function - is it $\bigcup -shaped$ like a valley, or $\bigcap -shaped$ like a hill? In the context of optimization, this will tell us whether we have a max or a min. The second order derivative captures exactly this kind of information. We denote it by $f''(x), d^2f/dx^2$, or $D^2f(x)$, and define it as

$$f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h}$$

or the derivative of the derivative. Higher order derivatives are useful, as they let us more accurately approximate a function.

The Taylor approximation is used extensively in proofs of various theorems. Besides, whenever the system that we have to solve is a highly complicated, non-linear system, we often linearize it to study its local behavior. If we want to solve for a root of a function $f(x)$ numerically, the basis for widely used methods lie in $Newton's$ method that uses successive first order Taylor series.

What do Taylor series do? They approximate the function (using polynomials) around some point using the information on the derivatives. Given a function f(x) and a point x_0, the

nth order Taylor series are defined as follows:

$$\begin{aligned} f(x) &\approx f(x_0) + \frac{df(x_0)}{dx}(x-x_0) + \frac{1}{2!}\frac{d^2 f(x_0)}{dx^2}(x-x_0)^2 + ... + \frac{1}{n!}\frac{d^n f(x_0)}{dx^n}(x-x_0)^n \\ &= f(x_0) + \sum_{k=1}^{n} \frac{1}{k!}\frac{d^k f(x_0)}{dx^k}(x-x_0)^k \end{aligned}$$

The more terms (higher n) that we include in the Taylor series, the more accurate the approximation becomes.

Example 8. Let $f(x) = 1/(1-x)$ and consider the nth order Taylor approximation around $x = 0$. We have a sequence of derivatives

$$\begin{aligned} f^{(1)}(x) &= (1-x)^{-2} \\ f^{(2)}(x) &= 2(1-x)^{-3} \\ f^{(3)}(x) &= 2 \cdot 3(1-x)^{-4} \end{aligned}$$

and so on. From the patter above, the general nth term derivative will be

$$f^{(n)} = n!(1-x)^{-(n+1}$$

so that $f^{(n)}(0) = n!$. Putting these together in a Taylor series give us

$$\begin{aligned} f(x) &\approx 1 + \sum_{k=1}^{n} \frac{1}{k} k! x^k \\ &= 1 + x + x^2 + x^3 + ... + x^n \end{aligned}$$

If we kept adding more terms, as $n \to \infty$, we would get $1/(1-x) = 1 + x + x^2 + x^3 + ...$, a formula for the sum of a geometric series.

Finally, we return to the properties of concavity and convexity of a function. Concavity and convexity put some implicit structure on a function. It turns out, whenever f is concave (convex), the f should also be continuous.

Theorem 10. *Let f be a concave or convex function on an open domain X, then f is continuous on X.*

When we are dealing with differentiable functions, we have a theorem that characterizes concavity.

Theorem 11. *Let $f : X \to \mathbb{R}$ be a differentiable function and $X \subseteq \mathbb{R}$ be an open and convex set, then f is concave if and only if*

$$f(x_1) \leq f(x_2) + f'(x_2)(x_1 - x_2)$$

for all $x_1, x_2 \in X$. If the inequality is flipped, then concave changes to convex.

The simplest way to see what the above theorem says is to notice that the right hand side is just a tangent line to $f(x_1)$ (think of a first order Taylor series). Thus, the theorem says that the function is concave if and only if for any two points, a tangent line lies above the graph.

The more frequently used characterization is given below

Theorem 12. *Let function $f : D \to \mathbb{R}$, where D is a convex subset of $\mathbb{R}$, be twice continuously differentiable. Then, f is concave (convex) if and only if $f''(x) \leq 0 \ \ (if \ f''(x) \geq 0)$ for all $x \in D$. Moreover, if the inequality is strict, then f is strictly concave (convex).*