ECN594: Math Bootcamp

A Crash Course in Real Analysis

Kun Zhang*

August 2022

*Department of Economics, Arizona State University. Email: kunzhang@asu.edu
Some parts of these notes base heavily on Section 1 of Vohra (2005) and Section A of Ok (2007). I have also benefited from Osborne (2016), as well as lecture notes written by (in ramdon order) Ahmet Altınok and Andreas Kleiner.

Contents

These notes are intended to provide a brief introduction to some basic real analysis concepts and results that are going to be useful in the rest of this class as well as many first-year courses. See footnotes for recommendations of some textbooks that can provide full details of the proofs that we will sometimes omit. Note that, however, results with a proof need not to be more important than results without a proof.

1 The space we work in

In the remainder of this class we are going to work in the space of all n-tuples of real numbers, $(x_1, x_2, ..., x_n)$, where n is a positive integer; we denote this space $\mathbb{R}^n$. An element of this space is called a **point** in $\mathbb{R}^n$; a point is equivalent to a vector whose tail is at the origin and the tip is at that point; so we sometimes use these two terms interchangeably. If $z \in \mathbb{R}^n$, the j–th component of z will be denoted z_j, where $j = 1, \dots, n$; we call the set of points all of whose components are nonnegative **the nonnegative orthant**, denote by $\mathbb{R}^n_+$. When $n = 1$, we get the real line $\mathbb{R}$.

When $n = 1$, the meaning of $\leq$, $\geq$, and $=$ should be clear. If $x, y \in \mathbb{R}^n$, $n \geq 2$, we define

- $x = y$ if $x_i = y_i$ for all i,
- $x \geq y$ if $x_i \geq y_i$ for all i,
- $x > y$ if $x_i \geq y_i$ for all i with strict inequality for at least one component, and
- $x \gg y$ iff $x_i > y_i$ for all i.

Next we define the notion of distance, and an important operation on $\mathbb{R}^n$.

Definition 1 (Distance)**.** Let x and y be any two vectors in $\mathbb{R}^n$. The (Euclidean) **distance**[1] between x and y, denote by $d(x, y)$, is

$$d(x, y) = \sqrt{\sum_{j=1}^{n} \left(x_j - y_j\right)^2}.$$

Observe that, for any $u, v \in \mathbb{R}$,

$$d(u, v) = |u - v|;$$

and for any $x \in \mathbb{R}^n$, we usually write $d(x, 0) = \|x\|$. If we take x as an vector, its length is $\|x\|$.

Definition 2 (Inner product)**.** The **inner product** of two vectors x and y is denoted $x \cdot y$ or xy and is defined as

$$x \cdot y = \sum_{j=1}^{n} x_j y_j.$$

[1] There are some other ways to define the distance between points, and they are all equivalent in $\mathbb{R}^n$.

For any set $S \subseteq \mathbb{R}$, an **upper bound** of this set is a real number u such that $u \geq x$ for all $x \in S$, and a **lower bound** of S is a real number ℓ such that $\ell \leq x$ for all $x \in S$. Next we formalize the idea of "least upper bound" and "greatest lower bound".

Definition 3 (Supremum and infimum)**.** Given a subset S of $\mathbb{R}$, $s^* \in \mathbb{R} \cup \{-\infty, \infty\}$ is the **supremum** of S, denoted $\sup S$, if

- s^* is an upper bound of S;
- $u \geq s^*$ for any upper bound u of S.

And $s_* \in \mathbb{R} \cup \{-\infty, \infty\}$ is the **infimum** of S, denoted $\inf S$, if

- s^* is an lower bound of S;
- $\ell \leq s^*$ for any lower bound ℓ of S.

Observation 1. *A nonempty set of real numbers always have a unique supremum and a unique infimum.*

Example 1. Let $S = \{x \in \mathbb{R} : 0 < x < 1\}$. Then $\sup S = 1$ and $\inf S = 0$.

Observation 2. *Neither* $\inf S$ *nor* $\sup S$ *need to be an element of* S*.*

2 Sequences

A **sequence** is a function whose domain is the set of natural numbers, $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$.[2] In these notes we require the codomain of a sequence to be $\mathbb{R}^n$; so a sequence is a specification of a point $x^k \in \mathbb{R}^n$ for each $k \in \mathbb{N}$. We denote a sequence by (x^k).

Definition 4 (Convergence of sequences)**.** A sequence (x^k) **converges** to $x \in \mathbb{R}^n$ if, for all $\varepsilon > 0$, there exists a $N \in \mathbb{N}$ such that $d(x^k, x) < \varepsilon$ for all $k \geq N$.

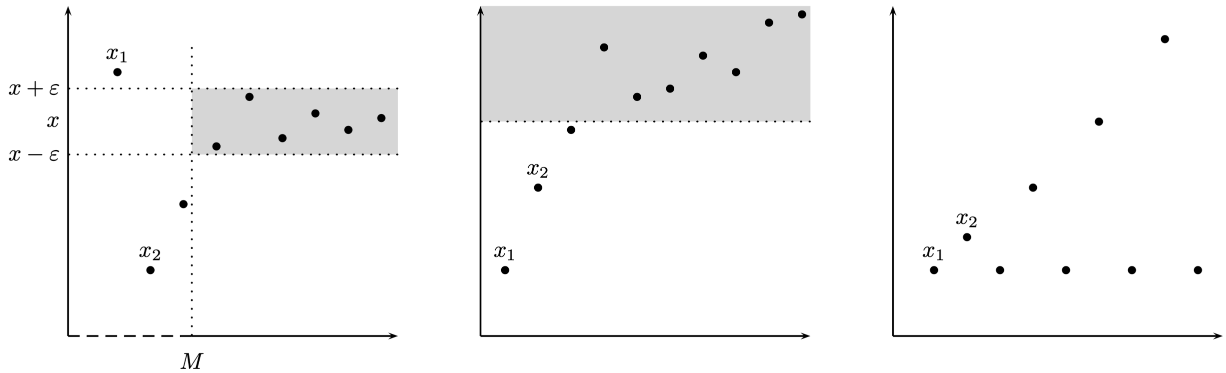
Remark 1. *Note that* N *in the definition above may depend on the chosen* ε*.*

If a sequence (x^k) converges to x, we say the sequence (x^k) is **convergent** and x is the limit of (x^k); we write $\lim_{k \to \infty} x^k = x$, or $x^k \longrightarrow x$. Intuitively, a sequence (x^k) is convergent if there is $x \in \mathbb{R}^n$ such that the later terms of the sequence get arbitrarily close to x. A sequence (x^k) is called **divergent** if it does not converge.

And we say that a *real* sequence[3] (x^k) **tends to infinity**, written as $\lim_{k \to \infty} x^k = \infty$, or $x^k \longrightarrow \infty$, if for every $Q \in \mathbb{R}$, there exists $M \in \mathbb{N}$ with $x^k \geq Q$ for each $k \geq M$.

[2] Some authors define the set of natural numbers as $\{0, 1, 2, 3, \ldots\}$; this difference does not really make any difference in our class.

[3] A real sequence is a sequence that specifies a real number x^k for each $k \in \mathbb{N}$.



(x_m) converges to x. (x_m) diverges to ∞. (x_m) is divergent but it does not diverge to ∞.

Next we solve some simple examples. We claim that $\lim_{k\to\infty}(1/k) = 0$. To show this, pick an arbitrary $\varepsilon > 0$, and ask if there is an $N \in \mathbb{N}$ large enough to guarantee that

$$\left|\frac{1}{k} - 0\right| < \varepsilon$$

for all $k \geq N$. It suffices to pick any N such that $N > 1/\varepsilon$. And using the definition, it is not difficult to check that, the sequence y^k defined as

$$y^k = k \quad \text{for all} \;\; k \in \mathbb{N}$$

tends to infinity.

The idea of convergence is that the tail of a sequence approximates its limit to any desired degree of accuracy. Some initial finitely many terms of the sequence may be quite apart from its limit, but eventually all terms of the sequence accumulates around this limit. To see this, let us consider a sequence (x^k) defined as

$$x^k = \begin{cases} k & \text{if } k \leq 100; \\ \frac{1}{k-100} & \text{if } k > 100. \end{cases}$$

To show that $\lim_{k\to\infty} x^k = 0$, we need to pick N large enough, larger than in the previous example; in fact, we need to pick $N > (1/\varepsilon) + 100$. So although real sequences $(1/k)$ and x^k appear to be quite different in their first 100 terms, they have the same long-run behavior: they both converge to 0. That is, the initial terms of the sequence have no say on the behavior of the tail of the sequence.

We summarize some basic properties of sequences below.

Claim 1. *A sequence converges to at most one limit.*

Proof. Suppose (x^k) is a sequence that converges to two limits x^0 and y^0 with $x^0 \neq y^0$. Let

$$\varepsilon = \frac{d(x^0, y^0)}{2},$$

we have $\varepsilon > 0$ since $x^0 \neq y^0$. Because $x^k \longrightarrow x^0$, there exists $N_1 \in \mathbb{N}$ such that $d(x^k, x^0) < \varepsilon$; and since $x^k \longrightarrow y^0$, there exists $N_1 \in \mathbb{N}$ such that $d(x^k, y^0) < \varepsilon$. Now take $N = \max\{N_1, N_2\}$, then for any $m > N$, we have

$$d(x^0, y^0) \leq d(x^0, x^m) + d(x^m, y^0) < \varepsilon + \varepsilon = d(x^0, y^0),$$

where the first inequality follows from the triangle inequality.[4] We have thus reached at a contradiction. ∎

Next we establish some properties of convergent sequences.

Proposition 1 (Squeeze theorem)**.** *Let (x^k), (y^k) and (z^k) be real sequences such that $x^k \leq y^k \leq z^k$ for each $k \in \mathbb{N}$. If*

$$\lim_{k \longrightarrow \infty} x^k = \lim_{k \longrightarrow \infty} z^k = a,$$

then $\lim_{k \longrightarrow \infty} y^k = a$.

We omit the proof of Proposition 1; see Figure 1 for an illustration.

Proposition 2. *Let $\left(x^k\right)$ and $\left(y^k\right)$ be two real sequences such that $x^k \longrightarrow x \in \mathbb{R}$ and $y^k \longrightarrow y \in \mathbb{R}$. Then:*

(1) $\left|x^k\right| \longrightarrow |x|$;

(2) $x^k + y^k \longrightarrow x + y$;

(3) $x^k y^k \longrightarrow xy$; and

(4) if $x^k \neq 0$ for all $k \in \mathbb{N}$ and $x \neq 0$,

$$\frac{y^k}{x^k} \longrightarrow \frac{y}{x}.$$

[4] The triangle inequality states that, for any $x, y, z \in \mathbb{R}^n$,

$$d(x, y) \leq d(x, z) + d(z, y),$$

which essentially comes from the geometric intuition that for any triangle, the sum of the lengths of any two sides must be greater than or equal to the length of the remaining side. If x, y, z are real numbers, we get

$$|x - y| \leq |x - z| + |z - y|.$$

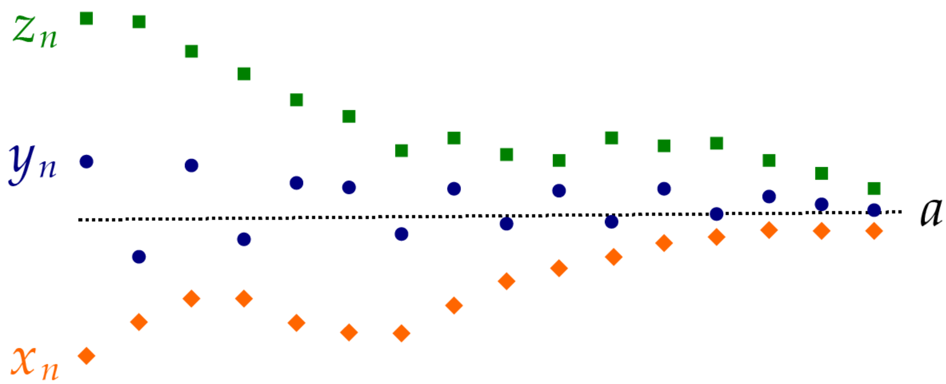


Figure 1: Illustration for the sandwich theorem (Proposition 1).

Proof (Sketch). For (1), use the inequality that[5]

$$\left||x^k| - |x|\right| \leq |x^k - x|.$$

For (2), use the triangle inequality to get that

$$|(x^k + y^k) - (x + y)| = |(x^k - x) + (y^k - y)| \leq |x^k - x| + |y^k - y|.$$

To prove (3), we use

$$\begin{aligned} |x^n y^n - xy| &= |x^n y^n - xy^n + xy^n - xy| \\ &\leq |x^n y^n - xy^n| + |xy^n - xy| \\ &= |x^n - x|\,|y^n| + |x||y^n - y| \end{aligned}$$

We omit the proof of (4); see, e.g., Rudin (1976), page 50, Theorem 3.3. ∎

Definition 5. (Bounded sequences) A real sequence (x^k) is **bounded** if there exists $K \in \mathbb{R}$ with $|x^k| \leq K$ for all $k \in \mathbb{N}$.

Claim 2. *Every convergent real sequence is bounded.*

Proof. Exercise. ∎

[5] For any $x, y \in \mathbb{R}$, we have

$$|x| = |x - y + y| \leq |x - y| + |y|;$$

where the inequality follows from the triangle inequality. Hence, $|x| - |y| \leq |x - y|$; similarly, we can show that $|y| - |x| \leq |x - y|$. Combining the two inequality above yields

$$\left||x| - |y|\right| \leq |x - y|.$$

3 Topological properties

Definition 6 (Open ball)**.** An **open ball** of radius $\varepsilon > 0$ centered at x, denote by $B(x, \varepsilon)$ is the set

$$\{y \in \mathbb{R}^n : d(x, y) < \varepsilon\}.$$

Definition 7 (Open and closed sets)**.** A set $S \subseteq \mathbb{R}^n$ is called **open** if for every $x \in S$ there is an $\varepsilon > 0$ such that any y within distance of ε of x, that is, $d(x, y) < \varepsilon$, is in S. A set $T \subseteq \mathbb{R}^n$ is **closed** if the complement of T, T^c, is open.

Equivalently, a set S is open if for any $x \in S$, there exists an open ball $B(x, \varepsilon)$ such that

$$B(x, \varepsilon) \subseteq S.$$

Observation 3. *The empty set $\emptyset$ and $\mathbb{R}^n$ are both open and closed.*

Claim 3. *An open ball is an open set.*

Proof. Exercise. ∎

The following characterization of closed sets is proved useful.

Proposition 3. *A set $S \subseteq \mathbb{R}^n$ is closed if and only if for any convergent sequence (x^k) in S, its limit $\lim_{k\to\infty} x^k$ is in S as well.*

Proof. For the "if" direction, let (x^k) be a convergent sequence in S with $\lim_{k\to\infty} x^k \in S$. To show that S is closed, it suffices to show that S^c is open. That is, we need to show that for all $x \in S^c$, there is an $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq S^c$. Suppose not, then for each $\varepsilon > 0$, there exists $y \in S^c$ such that

$$B(y, \varepsilon) \cap S \neq \emptyset.$$

In particular, for each positive integer k, there is an element y^k of S in $B(y, 1/k)$. Then the sequence (y^k) is in S and converges to y since

$$d(y^k, y) < 1/k.$$

But then we get $y \in S$, contradicts to the fact that $y \in S^c$. Thus, S is closed.

To prove the "only if" direction, let S be closed, and let (x^k) be a convergent sequence in S with $x^k \longrightarrow x$. We show that $x \in S$. Suppose not, so $x \in S^c$. But because S is closed, S^c is open; so there exists $\varepsilon > 0$ such that for all y with $d(x, y) < \varepsilon$, $y \in S$. But then since $x^k \longrightarrow x$, there exists $N \in \mathbb{N}$ such that $d(x, x^k) < \varepsilon$ whenever $k \geq N$. This implies that (x^k) is not completely contained in S, a contradiction. Therefore, whenever (x^k) is a convergent sequence completely contained in a closed set S, its limit is also contained in S. ∎

An important class of open and closed sets are called **intervals**. Given two real numbers a, b with $a < b$, the closed interval $[a, b]$ is the set $\{x \in \mathbb{R} : a \leq x \leq b\}$. The open interval (a, b) is the set $\{x \in \mathbb{R} : a < x < b\}$. The following two examples show that a closed interval is indeed closed, and an open interval is indeed open.

Example 2. $S = \{x \in \mathbb{R} : 0 < x < 1\} = (0, 1)$ is open. To see this, pick any $x \in S$. Define

$$\varepsilon = \min\{|x|, |x - 1|\},$$

it is not difficult to see that $(x - \varepsilon, x + \varepsilon) \subseteq (0, 1)$, hence $(0, 1)$ is open. And by definition,

$$S^c = (-\infty, 0] \cup [1, \infty)$$

is closed.

Example 3. $[0, 1]$ is closed. We use Proposition 3 to prove this assertion: let (x^k) be a convergent sequence in $[0, 1]$ with $\lim_{k\to\infty} x^k = x^0$. Suppose $x_0 \notin [0, 1]$; say, $x_0 > 1$. Pick $\varepsilon = (x_0 - 1)/2 > 0$. Since $\lim_{k\to\infty} x^k = x^0$, for all $\varepsilon > 0$ there exists an integer N such that $|x^k - x^0| \leq \varepsilon$ for all $k > N$. For our choice of ε, this implies $x^k > 1$ for all $k > N$, contradicting that $x^k \in [0, 1]$ for all k.

Example 4. $S = \{x \in \mathbb{R} : 0 < x \leq 1\} = (0, 1]$ is neither open nor closed. The sequence $(1/k)$ is in S, but it converges to 0 which is not in this set, so S is not closed. However, there is no $\varepsilon > 0$ sufficiently small such that every point within distance ε from 1 is in S. Thus, S is not open.

Observation 4. *If $S \subseteq \mathbb{R}$ is closed, the infimum and supremum of S are in S: they coincide with the smallest and largest member of S, respectively.*

The following proposition summarizes some useful facts about open and closed sets.

Proposition 4. *The union of any number of open sets is open, and the intersection of finitely many open sets is open. The intersection of any number of closed sets is closed, and the union of finitely many closed sets is closed.*

Proof. For the first part of the first statement, let $\{S_i\}_{i\in I}$ be a collection of open sets. Denote

$$S = \bigcup_{i\in I} S_i.$$

Suppose $x \in S$, then $x \in S_i$ for some $j \in I$. Since S_j is open, there is an open ball $B(x, \varepsilon)$ such that $B(x, \varepsilon) \subseteq S_j$. Then because $S_j \subseteq S$, we have $B(x, \varepsilon) \subseteq S$, so S is open.

For the second part of the first statement, let $\{T_i\}_{i=1}^m$ be a finite collection of open sets. Denote

$$T = \bigcap_{i=1}^{m} T_i.$$

Let $x \in T$, then $x \in T_i$ for all $i = 1, \dots, n$. Since each T_i is open, for each i there is an ε_i such that open ball $B(x, \varepsilon_i) \subseteq T_i$. Let $\varepsilon = \min_i \varepsilon_i$; then the open ball $B(x, \varepsilon)$ is contained in each $B(x, \varepsilon_i)$, and therefore is contained in each T_i. Thus, $B(x, \varepsilon) \subseteq T$, so T is an open set.

The second statement follows from the first part and the De Morgan's law. (How?) ■

We also define the interior and closure of a set.

Definition 8. A point $x \in S$ is an **interior point** of S if there exists $\varepsilon > 0$ such that $\{y \in \mathbb{R}^n : d(x, y) < \varepsilon\} \subseteq S$. The **interior** of S, denoted $\text{int}(S)$, is the set of all interior points of S.

Definition 9. The **closure** of a set S, denote by $\text{cl}(S)$ is the set S combined with all points that are the limits of sequences of points in S.

Observation 5. *If $S \subseteq \mathbb{R}^n$ is open, $S = \text{int}(S)$; if $T \subseteq \mathbb{R}^n$ is closed, $T = \text{cl}(T)$.*

Example 5. The interior of $(0, 1)$ is $(0, 1)$, and the closure of $(0, 1)$ is $[0, 1]$; the interior of $[0, 1]$ is $(0, 1)$, and the closure of $[0, 1]$ is $[0, 1]$.

Definition 10 (Boundedness and compactness)**.** A set $S \subseteq \mathbb{R}^n$ is **bounded** if there exists a real number r such that $\|x\| \leq r$ for all $x \in S$. S is **compact** if it is closed and bounded.

Example 6. Let $a, b \in \mathbb{R}$. By taking $r = \max\{|a|, |b|\}$, we see that the closed interval $[a, b]$ is bounded, hence it is also compact. $\mathbb{R}$ is closed but not compact, so is

$$\{(x, y) \in \mathbb{R}^2 : x + y \leq 0\}.$$

4 Continuity

4.1 Limits of functions

Definition 11 (Limits of functions)**.** Let $S \subseteq \mathbb{R}^n$, and let $f : S \longrightarrow \mathbb{R}$. We say that $q \in \mathbb{R}$ is the **limit** of f at $x_0 \in S$, and write

$$\lim_{x \longrightarrow x_0} f(x) = q,$$

if q is such that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $d(x, x_0) < \delta$, we have $|f(x) - q| < \varepsilon$. If we cannot find such q, we say that the limit of f *does not exist* at x_0.

Roughly speaking, the limit of f at x_0 is q if whenever x is sufficiently close to x_0, $f(x)$ becomes arbitrarily close to q. Limits of functions can be equivalently defined in terms of limits of sequences.

Proposition 5. *Let $S \subseteq \mathbb{R}^n$, and let $f : S \longrightarrow \mathbb{R}$. Then $\lim_{x \longrightarrow x_0} f(x) = q$ if and only if for every sequence (x^k) in S such that $x^k \longrightarrow x_0$, we have $\lim_{n \longrightarrow \infty} f\left(x^k\right) = q$.*

Proof. Suppose $\lim_{x \longrightarrow x_0} f(x) = q$, and let (x^k) be a sequence in S for all k and $x^k \longrightarrow x_0$. Fix any $\varepsilon > 0$. Since $\lim_{x \longrightarrow x_0} f(x) = q$, there exists $\delta > 0$ such that $|f(x) - q| < \varepsilon$ whenever $d(x, x_0) < \delta$. But then since $x^k \longrightarrow x_0$, there exists $M \in \mathbb{N}$ such that whenever $k \geq M$ we have $d(x^k, x_0) < \delta$. Thus, $|f(x^k) - q| < \varepsilon$ for all $k \geq M$; consequently, $\lim_{n \longrightarrow \infty} f(p_n) = q$.

Conversely, suppose $\lim_{x \longrightarrow x_0} f(x) = q$ does not hold. Then by definition, there exists $\varepsilon > 0$ such that for every $\delta > 0$ there exists a point $x \in S$ with $d(x, x_0) < \delta$ but $|f(x) - q| \geq \varepsilon$. Letting (δ^k) be defined by $\delta^k = 1/k$, then for each $k \in \mathbb{N}$ we can find x^k with $d(x^k, x_0) < 1/k$ but $|f(x^k) - q| \geq \varepsilon$, and the resulting sequence (x^k) converges to x_0 but $(f(x^k))$ does not converge to q. ∎

Corollary 1 follows from Claim 1 and Proposition 5.

Corollary 1. *Let S be a subset of $\mathbb{R}^n$. If $f : S \longrightarrow \mathbb{R}$ has a limit at $x_0 \in \mathbb{R}$, then this limit is unique.*

Example 7. Let $f : [0, 2] \longrightarrow \mathbb{R}$ be defined by

$$f(x) = \begin{cases} (x-1)^2 + 1 & \text{if } 0 \leq x < 1 \text{ or } 1 < x \leq 2, \\ 2 & \text{if } x = 1; \end{cases} \tag{1}$$

see Figure 2 for its graph. By definition, the limit of f does not exist at $x = 1$.

The next result is an immediate consequence of Proposition 2 and Proposition 5.

Proposition 6. *Let $S \subseteq \mathbb{R}^n$ and $x_0 \in S$. Let f and g be real-valued functions on S and*

$$\lim_{x \longrightarrow x_0} f(x) = A, \quad \lim_{x \longrightarrow x_0} g(x) = B.$$

Then

(1) $\lim_{x \longrightarrow x_0} [f(x) + g(x)] = A + B$;

(2) $\lim_{x \longrightarrow x_0} f(x)g(x) = AB$; and

(3) if $B \neq 0$,

$$\lim_{x \longrightarrow x_0} \frac{f(x)}{g(x)} = \frac{A}{B}.$$

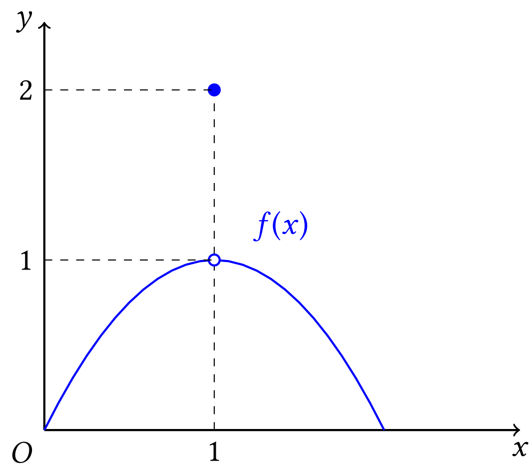


Figure 2: The graph of the function f defined by Equation (1).

4.2 Continuous functions

Definition 12 (Continuity)**.** Let S be a subset of $\mathbb{R}^n$. A real valued function $f : S \longrightarrow \mathbb{R}$ is **continuous at** x if for any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $y \in \mathbb{R}^n$ with $d(x, y) < \delta$, we have $|f(x) - f(y)| < \varepsilon$. And f is **continuous** if it is continuous at all x in S.

Observation 6. *Let S be a subset of $\mathbb{R}^n$. $f : S \longrightarrow \mathbb{R}$ is continuous at $x_0 \in S$ if and only if $\lim_{x \longrightarrow x_0} f(x)$ exists.*

Intuitively, if f is continuous at x, it maps points nearby x to points are close to $f(x)$. To make sure that you understand the definition, verify that the following functions are continuous:

- $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = x$;
- $g : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $g(x) = x^2$.

Example 8. $f(x) = |x|$ is continuous. Observe that we can write $f(x)$ as

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

so since both x and $-x$ are continuous, it suffices to check that f is continuous at $x = 0$. Fix any $\varepsilon > 0$, let $\delta = \varepsilon$; then whenever $|y - 0| < \delta = \varepsilon$, or $y \in (-\delta, \delta)$, we have

$$|f(y) - f(0)| = |y| = |y - 0| < \delta = \varepsilon,$$

which shows that f is continuous at $x = 0$.

Now consider $g : \mathbb{R} \longrightarrow \mathbb{R}$ defined as

$$g(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0, \end{cases}$$

we claim that this function is not continuous. At $x = 0$, so long as $\varepsilon < 1$, no matter how small δ we choose, we are not able to find $y \in (-\delta, 0) \subseteq (-\delta, \delta)$ such that $|f(y) - f(x)| < \varepsilon$.

Definition 13 (Sequential continuity). Let S be a subset of $\mathbb{R}^n$. A function $f : S \longrightarrow \mathbb{R}$ is **sequentially continuous at** x if for any sequence (x^k) in S that converges to x, it holds that

$$\lim_{k \longrightarrow \infty} f(x^k) = f(x).$$

And f is **sequentially continuous** if it is sequentially continuous at all x in S.

Observation 7. *Let S be a subset of $\mathbb{R}^n$. $f : S \longrightarrow \mathbb{R}$ is sequentially continuous at $x_0 \in S$ if and only if $\lim_{x \longrightarrow x_0} f(x)$ exists.*

The following important equivalence result follows from Observation 6 and Observation 7.

Theorem 1. *Let S be a subset of $\mathbb{R}^n$. A function $f : S \longrightarrow \mathbb{R}$ is continuous at x if and only if it is sequentially continuous at x.*

Intuitively, Theorem 1 says that continuity of f allows exchanging "f" and "lim":

$$\lim_{k \longrightarrow \infty} f(x^k) = f(x) = f\left(\lim_{k \longrightarrow \infty} x^k\right).$$

The equivalence of continuity and sequential continuity in $\mathbb{R}^n$ gives us a handy way to prove that a function is *not* continuous: in many cases, it is easier to check a function fails to be sequential continuous. Recall a function we discussed in Example 8:

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

It is easy to see that g is not sequentially continuous at $x = 0$: sequence (x^k) defined by $x^k = -1/k$ converges to 0, but

$$\lim_{k \longrightarrow \infty} g(-1/k) = 0 \neq 1 = g(0).$$

4.3 Properties of continuous functions

Theorem 2 shows that the sum, product, and quotient (provided well-defined) of continuous functions is continuous; and a continuous function of a continuous function is continuous.

Theorem 2. *Let f and g be continuous real-valued functions. Then*

(1) if f and g are defined on $T \subseteq \mathbb{R}^n$, $f + g$ and fg are continuous; f/g is also continuous provided that $g(x) \neq 0$ for all $x \in T$;

(2) let $S \subseteq \mathbb{R}^n$ and $Q \subseteq \mathbb{R}$; if f and g are defined on S and Q, respectively, with $f(S) \subseteq Q$, then $f \circ g$ is continuous.

Proof. Continuity of $f + g$, fg and f/g (assume it is well-defined) follow from Proposition 6 and Observation 6. Define $h = f \circ g$, it remains to show that h is continuous. The proof of this fact is left as an exercise. ■

An immediate consequence of Theorem 2 is that, if $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a continuous real-valued function and c is a real number, $g(x) = cf(x)$ is continuous. A polynomial is a function of a single variable x of the form $a_0 + a_1x + a_2x^2 + \ldots + a_kx^k$, where k is a nonnegative integer and $a_0, \ldots, a_k \in \mathbb{R}$. Because the function f defined by $f(x) = x$ for all $x \in \mathbb{R}$ is continuous, Theorem 2 implies that all polynomials are continuous. (Why?)

Next we state the celebrated intermediate value theorem, a simple but powerful result that are implied by continuity on an interval.[6]

Theorem 3 (Intermediate value theorem)**.** *Let $a, b \in \mathbb{R}$ with $a < b$. If $f : [a, b] \longrightarrow \mathbb{R}$ is a continuous function, then for any $s \in (f(a), f(b))$, there exists $x \in (a, b)$ such that*

$$f(x) = s.$$

For a proof, see, for example, de la Fuente (2000), page 76, Theorem 6.24; for an illustration, see Figure 3.

Another powerful result, which is implied by continuity and compactness of the domain, is the extreme value theorem. Due to its importance in optimization theory, we defer the discussion until the next module.

[6]For more general spaces, we need to require connectedness of the domain. We do not discuss connectedness in these notes; see, for example, Section D.2 of Ok (2007) for details.

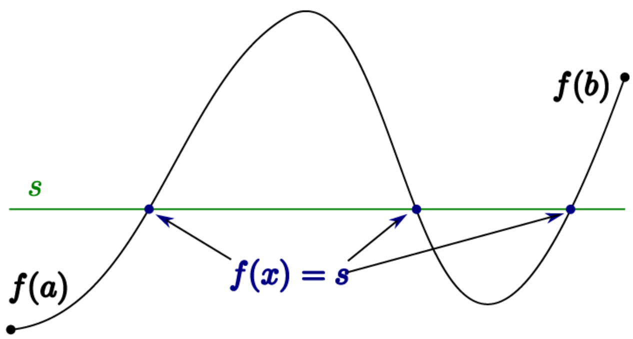


Figure 3: An illustration for the intermediate value theorem (Theorem 3).

5 Differentiation

In this section we mainly work on differentiation of univariate functions, that is, functions that map subsets of the real line to the reals. We also briefly present some definitions and a key result about differentiation of multivariate functions for future reference.

5.1 Definition, properties, and implications

Let $I = (a, b)$ be an open interval, and let $f : I \longrightarrow \mathbb{R}$. f is **differentiable** at $x_0 \in I$ if

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists. The **derivative** of f at x_0, denoted $f'(x_0)$, or $\frac{\mathrm{d}f}{\mathrm{d}x}(x_0)$, is this limit. The function f is said to be **differentiable** if it is differentiable at each point in I. In this case the **derivative** of f is defined as the function $f' : I \longrightarrow \mathbb{R}$ that maps each $x \in I$ to the derivative of f at x. If f' is also differentiable, then f is said to be **twice differentiable**, and the **second derivative** of f is defined as the function $f'' : I \longrightarrow \mathbb{R}$ that maps each $x \in I$ to the derivative of f' at x.

The derivative of a function of a single variable at a point x_0 in its domain, when it exists, is the slope of the tangent line to the graph of the function at x_0. For a small change in x, denote by $\Delta x = h$, the rate of change is defined as the ratio of the the change in a function of a variable to that of the variable; that is,

$$\text{rate of change } = \frac{\Delta y}{\Delta x} = \frac{f(x_0 + h) - f(x_0)}{h};$$

by definition, the derivative of f at x_0 is the limit of the rate of change as Δx approaches zero. For this reason, the derivative is often described as the "instantaneous rate of change", the ratio of the *instantaneous* change in $y = f(x)$ to that of x.

Proposition 7 shows that differentiable is a "stronger" requirement than continuous.

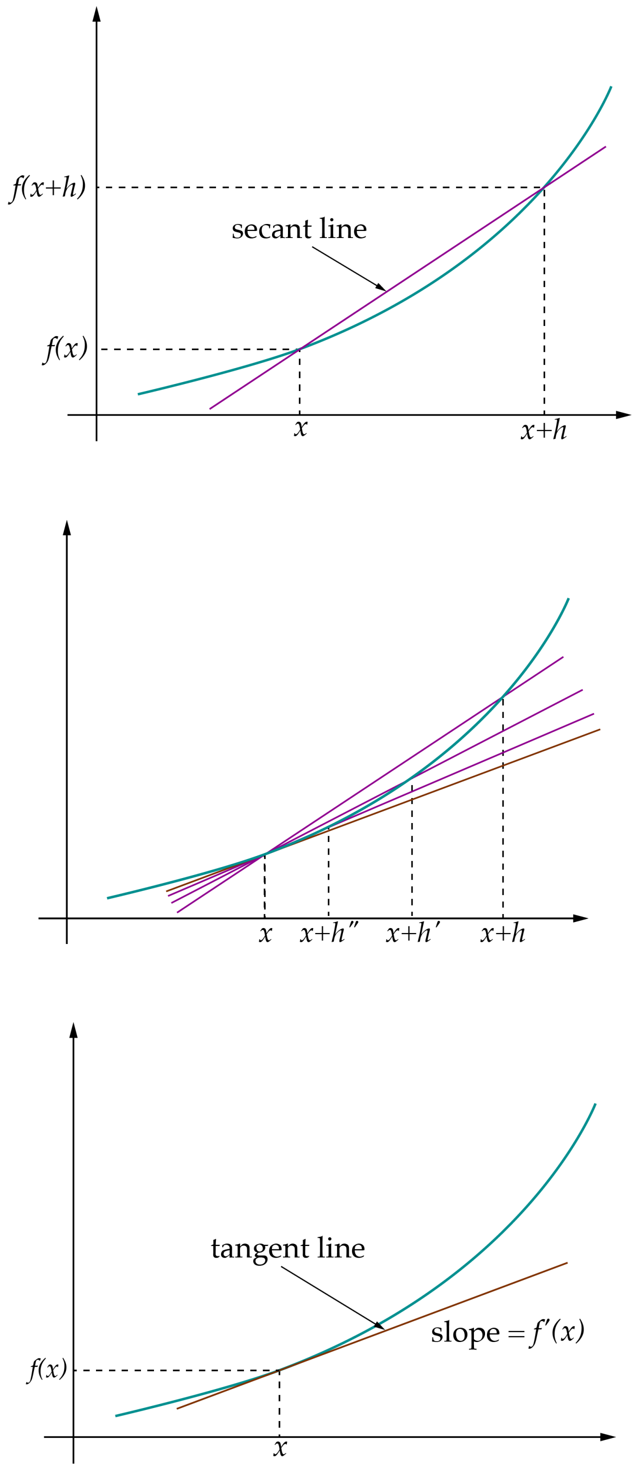


Figure 4: The derivative of a function at a point x is the limit (as $h \longrightarrow 0$) of secants to curve $y = f(x)$ determined by points $(x, f(x))$ and $(x + h, f(x + h))$; equivalently, it is the slope of the tangent line at $(x, f(x))$.

Proposition 7. *Let* $f : (a, b) \longrightarrow \mathbb{R}$ *If* f *is differentiable at* $x \in (a, b)$, *then it is continuous at* x.

Proof. Observe that

$$\lim_{h\to 0}[f(x+h) - f(x)] = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \cdot \lim_{h\to 0} h = f'(x) \cdot 0 = 0, \tag{2}$$

then Observation 6 implies that f is continuous at x. ∎

It is easy to find functions that are continuous at a point but not differentiable. An example is $f(x) = |x|$: it is continuous (recall Example 8) but not differentiable at 0. To see why, let

$$g(h) = \frac{f(x+h) - f(x)}{h};$$

to show that f is not differentiable at 0, it suffices to show that the limit of g at 0 does not exist. And observe that, by Proposition 5, to show that the limit of a function $g : \mathbb{R} \longrightarrow \mathbb{R}$ does not exist at a point x_0, it suffices to show that there exist two sequences y^n and z^n in $\mathbb{R}$, both converge to x_0, but $\lim_{n\to\infty} g(y^n) \neq \lim_{n\to\infty} g(z^n)$. Pick $(y^n) = (1/n)$ and $(z^n) = (-1/n)$, we see that both sequences converge to 0, $g(y^n) = 1$ for all $n \in \mathbb{N}$, and $g(z^n) = -1$ for all $n \in \mathbb{N}$, hence

$$\lim_{n\to\infty} g(y^n) = 1 \neq -1 = \lim_{n\to\infty} g(z^n).$$

Consequently, the limit of g does not exist at $h = 0$, so $f(x) = |x|$ is not differentiable at $x = 0$.

For some complicated functions, computing derivatives using definition could be a tricky and daunting task. The following few results provide some "shortcuts" for it. Using the definition, we can calculate the following formulas for the derivative of specific functions, where a, n, and k are constants.[7]

$f(x)$	$f'(x)$
k	0
kx^n	knx^{n-1}
$\log x$	$1/x$
e^x	e^x
a^x	$a^x \log a$
$\cos x$	$-\sin x$
$\sin x$	$\cos x$

We do not attempt to prove all of the derivatives in the table above; instead, we find the

[7] Whenever I use "log" I mean natural log–this is the only kind of logarithms that is relevant in economics and business.

derivative of $f = x^2$ as an example to fix ideas. By definition, fix any $y \in \mathbb{R}$, we have

$$\lim_{h \longrightarrow 0} \frac{(y+h)^2 - y^2}{h} = \lim_{h \longrightarrow 0} \frac{y^2 + 2hy + h^2 - y^2}{h} = \lim_{h \longrightarrow 0} \frac{2hy + h^2}{h} = \lim_{h \longrightarrow 0} (2y + h) = 2y.$$

Because y is arbitrary, we have $f'(x) = 2x$ for all $x \in \mathbb{R}$. Derive the formula of $f'(x)$ for $f(x) = kx^n$ for general n is a very good exercise. I do not require you to do the same for the last five rows because some of you might know little about the Euler's number e, exponential functions, logarithm, and trigonometric functions; for those who know these well, now you have five more good exercises.[8]

Next we introduce some more general rules of finding derivatives.

Theorem 4 (Rules of differentiation)**.** *Let $f : [a, b] \longrightarrow \mathbb{R}$ and $g : [a, b] \longrightarrow \mathbb{R}$ be differentiable at $x \in [a, b]$. Then $f + g$, fg, and f/g are differentiable at x, and*

(1) $[f(x) + g(x)]' = f'(x) + g'(x)$;

(2) $[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$; *and*

(3) if $g(x) \neq 0$,

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

Proof (Sketch). (1) is a direct consequence of Proposition 6 (1). For (2), observe that

$$\begin{aligned} f(x+h)g(x+h) - f(x)g(x) &= f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x) \\ &= f(x+h)[g(x+h) - g(x)] + [f(x+h) - f(x)]g(x); \end{aligned}$$

divide this by h and let $h \longrightarrow 0$, since by (2) we have $\lim_{h \longrightarrow 0} f(x+h) = f(x)$, (2) follows. To prove (3), let $z = f/g$. Then

$$\frac{z(x+h) - z(x)}{h} = \frac{1}{g(x+h)g(x)} \left[g(x)\frac{f(x+h) - f(x)}{h} - f(x)\frac{g(x+h) - g(x)}{h} \right].$$

Letting $h \longrightarrow 0$ yields (3). ■

Theorem 5 (Chain rule)**.** *Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous, and differentiable at $x \in (a, b)$. Let $g : I \longrightarrow \mathbb{R}$, where $f([a, b]) \subseteq I$, be differentiable at $f(x)$. Define $h : [a, b] \longrightarrow \mathbb{R}$ by*

$$h(t) = g(f(t))$$

[8] At least for students major in Economics and Finance, you really need to be comfortable working with logarithm functions, the Euler's number e and exponential functions. For those who would like to learn, Chapter 5 of Simon and Blume (1994) is an excellent introduction.

for all $t \in [a, b]$; then h is differentiable at x, and

$$h'(x) = g'(f(x))f'(x).$$

The chain rule is often summarized as "the derivative of the outside times the derivative of the inside," but one must remember that the derivative of the outside function is evaluated at the inside function's value at x. See, for example, Rudin (1976), page 105, Theorem 5.5, for a proof of Theorem 5.

We state two very famous and useful results below without proof;[9] we focus on geometric intuition instead.

Theorem 6 (Rolle)**.** *Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous, and differentiable on (a, b). If $f(a) = f(b) = 0$, then there exists $\theta \in (a, b)$ such that $f'(\theta) = 0$.*

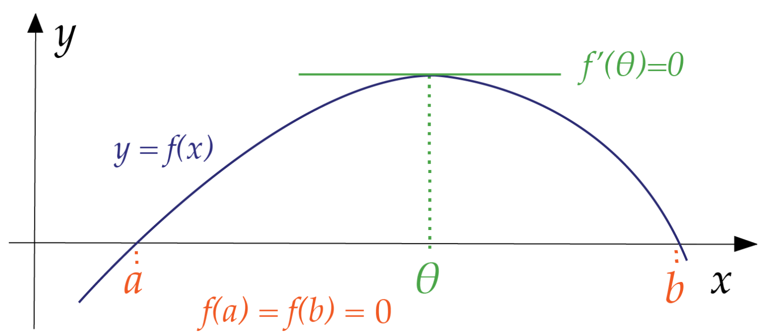


Figure 5: Illustration to Rolle's theorem

Figure 5 provides a very neat geometric explanation for Rolle's theorem. It says that a differentiable function that attains equal values at two distinct points must have at least one point, somewhere strictly between them, where the slope of the tangent line is zero. Rolle's theorem can be generalized to the celebrated mean value theorem.

Theorem 7 (Mean value theorem)**.** *Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous, and differentiable on (a, b). Then there exists $\xi \in (a, b)$ such that*

$$f(b) - f(a) = f'(\xi)(b - a). \tag{3}$$

The mean value theorem can be proved using Rolle's theorem and a carefully constructed

[9]Interested readers may consult, for example, Ok (2007), page 69. The proof of Rolle's theorem is a bit intricate but not difficult; the reason I choose to omit it is that we have not introduced the celebrated *first-order (necessary) condition*. Everyone who knows it is encouraged to try.

auxiliary function. Observe that (3) can be written as

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}. \tag{4}$$

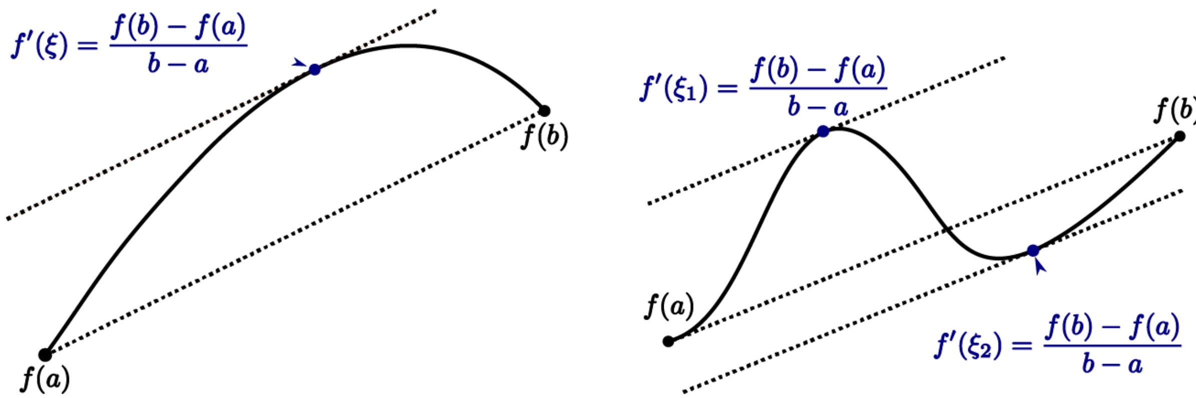


Figure 6: Illustration to the mean value theorem

The geometric intuition of the mean value theorem is illustrated in Figure 6. As shown in the left panel, (4) says that, For any function that is continuous on $[a, b]$ and differentiable on (a, b), there exists some ξ in the interval (a, b) such that the secant joining the endpoints of the interval $[a, b]$ is parallel to the tangent at ξ; in other words, the function f attains the slope of the secant between a and b as the derivative at the point ξ. And the right panel shows that ξ need not to be unique; that is, it is also possible that there are multiple tangents parallel to the secant.

5.2 Differentiable monotone functions

Definition 14 (Monotonicity). Let T be a nonempty subset of $\mathbb{R}$. A function $f : T \longrightarrow \mathbb{R}$ is said to be

- **increasing** if for any $x, y \in T$, $x \geq y$ implies $f(x) \geq f(y)$;
- **strictly increasing** if for any $x, y \in T$, $x > y$ implies $f(x) > f(y)$;
- **decreasing** if for any $x, y \in T$, $x \geq y$ implies $f(x) \leq f(y)$; and
- **strictly decreasing** if for any $x, y \in T$, $x > y$ implies $f(x) < f(y)$.

We call a function that is either increasing or decreasing **monotone**. It is not difficult to see that f is (strictly) decreasing if and only if $-f$ is (strictly) increasing. The next result characterizes differentiable monotone functions.

Theorem 8. *Let I be an open interval and $f(x) : I \longrightarrow \mathbb{R}$ be differentiable, then*

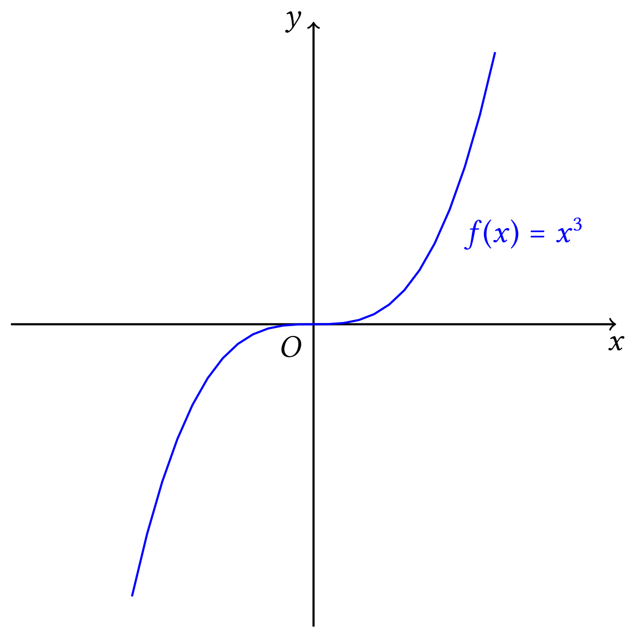


Figure 7: The graph of $f(x) = x^3$.

(1) f is increasing if and only if $f' \geq 0$ on I;

(2) if $f' > 0$ on I, then f is strictly increasing;

(3) f is decreasing if and only if $f' \leq 0$; and

(4) if $f' < 0$ on I, then f is strictly decreasing.

Proof. Exercise. ■

Example 9. Observe that the converse of (2) and (4) in Theorem 8 need not hold. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be given by $f(x) = x^3$; see Figure 7 for its graph. Then $f'(x) = 3x^2$, so $f'(0) = 0$, but f is a strictly increasing function: for any $x, y \in \mathbb{R}$ with $x > y$,

$$f(x) - f(y) = x^3 - y^3 = (x - y)(x^2 + xy + y^2) = (x - y)x(x + y) + (x - y)y^2 > 0,$$

where the inequality follows from the assumption that $x - y > 0$, which in turn implies that $x(x + y) > 0$ no matter the signs of x and y are.

Lemma 1. *Let I be an open interval; if $f : I \longrightarrow \mathbb{R}$ is strictly monotone, then it is injective.*

Proof. We prove the case that f is strictly increasing; the proof for strictly decreasing f is completely analogous. Let $x, y \in I$ with $x \neq y$, then either $x < y$ or $x > y$. If $x < y$, then since f is strictly increasing on I, we have $f(x) < f(y)$. Similarly, if $x > y$, then since f is strictly increasing on I, we have $f(x) > f(y)$. In both cases, $f(x) \neq f(y)$; hence f is injective on I. ■

Now we are ready to state the celebrated inverse function theorem.

Theorem 9 (Inverse function theorem)**.** *Let $f : (a, b) \longrightarrow \mathbb{R}$ be strictly monotone and continuous, and f is differentiable at $x_0 \in (a, b)$ with $f'(x_0) \neq 0$. Then f is invertible, and its inverse $g : f((a, b)) \longrightarrow \mathbb{R}$ is differentiable at $y_0 = f(x_0)$ and*

$$g'(y_0) = \frac{1}{f'(x_0)} = \frac{1}{f'(g(y_0))}.$$

Proof. Again, we prove the case that f is strictly increasing. Because f is strictly increasing and continuous, $f((a, b)) = (f(a), f(b))$; in other words, the range of f is also an interval. Denote $J = (f(a), f(b))$. Then by Lemma 1, the function $f : (a, b) \longrightarrow J$ is injective,[10] and since $f((a, b)) = J$, it is also bijective. Thus, f is invertible; call its inverse g.

Since $y_0 = f(x_0)$, we have $g(y_0) = x_0$, so

$$\frac{g(y) - g(y_0)}{y - y_0} = \frac{g(y) - g(y_0)}{f(g(y)) - f(g(y_0))} = \frac{x - x_0}{f(x) - f(x_0)} = \frac{1}{\frac{f(x)-f(x_0)}{x-x_0}}, \tag{5}$$

where the first and second equality follows since g is the inverse of f. Because f is differentiable at x_0,

$$\lim_{y \longrightarrow y_0} \frac{g(y) - g(y_0)}{y - y_0} = \lim_{x \longrightarrow x_0} \frac{1}{\frac{f(x)-f(x_0)}{x-x_0}} = 1/f'(x_0),$$

where the first equality follows from (5),[11] and the second follows from the fact that $f'(x_0) \neq 0$ and Proposition 6. ■

5.3 Partial derivatives

Definition 15 (Partial derivative)**.** Let S be a nonempty subset of $\mathbb{R}^n$. If $f : S \longrightarrow \mathbb{R}$, the **partial derivative of** f with respect to its j-th argument, x_j, at $x^0 = (x_j^0, x_{-j}^0)$, is

$$\lim_{h \longrightarrow 0} \frac{f\left(x_j^0 + h, x_{-j}^0\right) - f\left(x^0\right)}{h},$$

provided that the above limit exists.

When the partial derivative of f with respect to x_j exists, it is the derivative of f with respect to x_j *holding all other variables fixed*; it is denoted $\frac{\partial f}{\partial x_j}$, or sometimes f_j. That said, to calculate

[10] Formally this function is no longer f, since we changed its codomain. The original function is only injective, and the new one is bijective. To keep our argument simple, we abuse notation and still call it f.

[11] In fact, to show that $y \longrightarrow y_0$ implies $x \longrightarrow x_0$, we need to prove that g is a continuous function. It is not too difficult to prove but we skip it here; interested readers might try to prove the following statement: if a real-valued function defined on an interval I is continuous and injective, then its inverse is also continuous.

partial derivatives, we just need to take all other variables as constants, and apply the rules we introduced in Section 5.1 to x_j.

For functions of many variables, the definition of "differentiability" is beyond the scope of this class.[12] When we say a function of many variables is differentiable, it implies that all partial derivatives exist.

Example 10. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined as

$$f(x, y) = x^2 + xy + y + 1,$$

then

$$\frac{\partial f}{\partial x} = 2x + y, \quad \text{and} \quad \frac{\partial f}{\partial y} = x + 1.$$

Let $g : \mathbb{R}^2_{++} \longrightarrow \mathbb{R}$ be defined as

$$g(x, y) = e^{xy} + y \log x + 3,$$

then

$$\frac{\partial g}{\partial x} = ye^{xy} + \frac{y}{x}, \quad \text{and} \quad \frac{\partial g}{\partial y} = xe^{xy} + \log x.$$

For a differentiable function $f : S \longrightarrow \mathbb{R}$ and $x^0 \in S$, we use $Df(x^0)$ or $\nabla f(x^0)$ to denote the vector of all partial derivatives at x^0; that is,

$$\left(\frac{\partial f}{\partial x_1}(x^0), \frac{\partial f}{\partial x_2}(x^0), \dots, \frac{\partial f}{\partial x_n}(x^0) \right).$$

We call this vector the **gradient** of f at x^0.

Having defined partial derivatives, we are ready to generalize the chain rule beyond Theorem 5.

Theorem 10 (The chain rule generalized)**.**

(1) Suppose that for each $j = 1, 2, \dots, n$, $g_j : \mathbb{R} \longrightarrow \mathbb{R}$ is differentiable, and f is a diffentiable function of n-variables. Then the function $F : \mathbb{R} \longrightarrow \mathbb{R}$ given by

$$F(x) = f(g_1(x), \dots, g_n(x))$$

is differentiable, and

$$F'(x) = f_1(g_1(x), \dots, g_n(x))g_1'(x) + \dots + f_n(g_1(x), \dots, g_n(x))g_n'(x),$$

[12] For a formal definition, see Section 1.6 of Osborne (2016).

where f_j is the partial derivative of f with respect to its j-th argument.

(2) *Suppose that for each $j = 1, 2, \dots, n$, $h_j : \mathbb{R}^m \longrightarrow \mathbb{R}$ is differentiable, and f is a diffentiable function of n-variables. Then the function $G : \mathbb{R}^m \longrightarrow \mathbb{R}$ given by*

$$G(x_1, \dots, x_m) = f(h_1(x_1, \dots, x_m), \dots, h_n(x_1, \dots, x_m))$$

is differentiable, and for each $k = 1, \dots, m$,

$$\begin{aligned}\frac{\partial G}{\partial x_k}(x_1, \dots, x_m) = f_1(h_1(x_1, \dots, x_m), \dots, h_n(x_1, \dots, x_m))\frac{\partial h_1}{\partial x_k}(x_1, \dots, x_m)\\ + \dots + f_n(h_1(x_1, \dots, x_m), \dots, h_n(x_1, \dots, x_m))\frac{\partial h_n}{\partial x_k}(x_1, \dots, x_m)\\ = \sum_{j=1}^{n} f_j(h_1(x_1, \dots, x_m), \dots, h_n(x_1, \dots, x_m))\frac{\partial h_j}{\partial x_k};\end{aligned}$$

The first assertion is a direct analog of the chain rule formula in Theorem 5; loosely speaking, we take derivatives with respect to each of the g_i's and then use the linearity of differentiation. Some knowledge required in proving (2) is beyond the scope of this course; interested readers are directed to Rudin (1976), Theorem 9.15, page 214.

6 Convexity and concavity

Definition 16 (Convex sets)**.** A subset T of $\mathbb{R}^n$ is said to be **convex** if the line segment connecting any two elements of T lies entirely within T; that is,

$$\lambda x + (1 - \lambda)y \in T$$

for all $x, y \in T$ and $\lambda \in [0, 1]$.

See Figure 8 for an illustration. On the real line, convex sets coincide with intervals. Convex sets play a central role in modern economics because of many nice properties, many of which you will discover in your first year courses.

6.1 Convex and concave functions

Definition 17 (Convex and concave functions)**.** Let T be a nonempty *convex* subset of $\mathbb{R}^n$.

- A function $f : T \longrightarrow \mathbb{R}$ is called **convex** if

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y) \quad \text{for any } x, y \in T \text{ and } 0 \le \lambda \le 1,$$

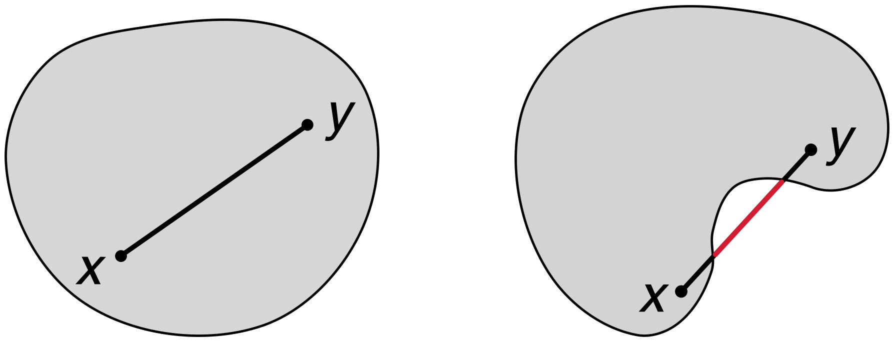


Figure 8: Convex (left) and non-convex (right) sets

and **strictly convex** if the above inequality holds strictly for any distinct $x, y \in T$ and $0 < \lambda < 1$.

- A function $f : T \longrightarrow \mathbb{R}$ is called **concave** if

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y) \quad \text{for any } x, y \in T \text{ and } 0 \leq \lambda \leq 1,$$

and **strictly concave** if the above inequality holds strictly for any distinct $x, y \in T$ and $0 < \lambda < 1$.

- A function $f : T \longrightarrow \mathbb{R}$ is called **affine** if it is both convex and concave; equivalently, f is affine if

$$f(\lambda x + (1 - \lambda)y) = \lambda f(x) + (1 - \lambda)f(y) \quad \text{for any } x, y \in T \text{ and } 0 \leq \lambda \leq 1.$$

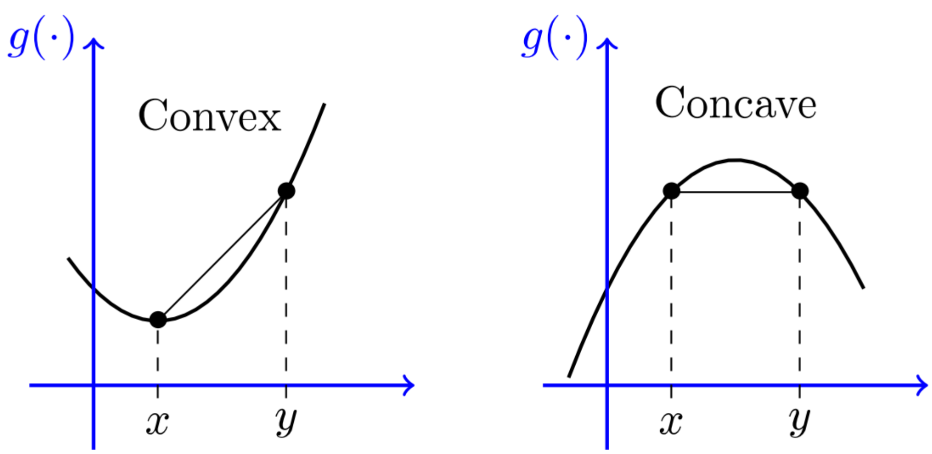


Figure 9: (Strictly) convex and concave functions of a single variable

Example 11. $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = x^2$ is a strictly convex function, and $g : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $g(x) = \sqrt{x}$ is a strictly convex function.

Example 12. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by

$$f(x) = 2x + 1.$$

Then for all $\lambda \in [0, 1]$ and all $x', x'' \in \mathbb{R}$,

$$f(\lambda x' + (1 - \lambda)x'') = 2[\lambda x' + (1 - \lambda)x''] + 1 = \lambda(2x' + 1) + (1 - \lambda)(2x'' + 1) = \lambda f(x') + (1 - \lambda)f(x''),$$

so f is both convex and concave, and hence affine. Is $g : \{0, 1\} \longrightarrow \{1, 3\}$ defined by $g(x) = 2x + 1$ convex? Concave? Why?

The following useful results are direct consequence of the above definitions.

Observation 8. *A function f is (strictly) convex if and only if $-f$ is (strictly) concave.*

Observation 9. *If f and g are (strictly) convex functions, and $\alpha, \beta \geq 0$, then $\alpha f + \beta g$ is a (strictly) convex function; if f and g are (strictly) concave functions, and $\alpha, \beta \geq 0$, then $\alpha f + \beta g$ is a (strictly) concave function.*

The next result shows that a increasing convex/concave transformation of a concave function is convex/concave. It is not difficult to prove: just apply the definition twice and use the fact that f is increasing to "pass" the inequality.

Proposition 8. *If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is an increasing convex (concave) function and $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a convex (concave) function, $f \circ g$ is also convex (concave).*

Proof. Exercise. ∎

Observe that, if we drop the requirement that f is convex (concave), then Proposition 8 may fail. To see this, let $g(x) = x$ and $f(x) = x^2$. It is easy to see that g is a concave function, and f is increasing. Let $h = f \circ g$, we see that $h(x) = x^2$ is a strictly convex function, hence not concave.

The remaining discussion on convex and concave functions are confined to differentiable functions that are defined on intervals. See Appendix A for a result on convexity and concavity of functions of two variables. For more general results, Chapter 3 of Osborne (2016) is an excellent reference.

Theorem 11. *Let $f : (a, b) \longrightarrow \mathbb{R}$ be differentiable. Then*

- *f is (strictly) convex if and only if f' is (strictly) increasing;*

- *f is (strictly) concave if and only if f' is (strictly) decreasing.*

A nice proof of Theorem 11 can be found at Rockafellar (1970), page 26, Theorem 4.4. Combining Theorem 11 and Theorem 8, we arrive at the following corollary.

Corollary 2. *Let $f : (a, b) \longrightarrow \mathbb{R}$ be twice differentiable. Then*

- *f is convex if and only if $f'' \geq 0$, and if $f'' > 0$, f is strictly convex;*
- *f is concave if and only if $f'' \leq 0$, and if $f'' < 0$, f is strictly concave.*

Example 11 Revisited. Note that both f and g are twice differentiable, with $f''(x) = 2 > 0$ and

$$g''(x) = -\frac{1}{4x^{\frac{3}{2}}} < 0,$$

hence f is strictly convex and g is strictly concave.

We conclude our analysis by showing that the graph of a differentiable concave (convex) function lies on or below (on or above) every tangent to the function. We illustrate it in Figure 10.

Proposition 9. *Let $f : (a, b) \longrightarrow \mathbb{R}$ be differentiable, then*

- *f is convex if and only if*

$$f(x) - f(y) \geq f'(y)(x - y) \text{ for all } x, y \in (a, b),$$

  and strictly convex if and only if the above inequality is strict;

- *f is concave if and only if*

$$f(x) - f(y) \leq f'(y)(x - y) \text{ for all } x, y \in (a, b),$$

  and strictly concave if and only if the above inequality is strict.

Proof. We only prove the first bullet point; the proof for the second is completely analogous. Suppose f is convex first. Then for all $x, y \in (a, b)$ and all $\lambda \in [0, 1]$, we have

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y),$$

straightforward algebra yields

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(y + \lambda(x - y)).$$

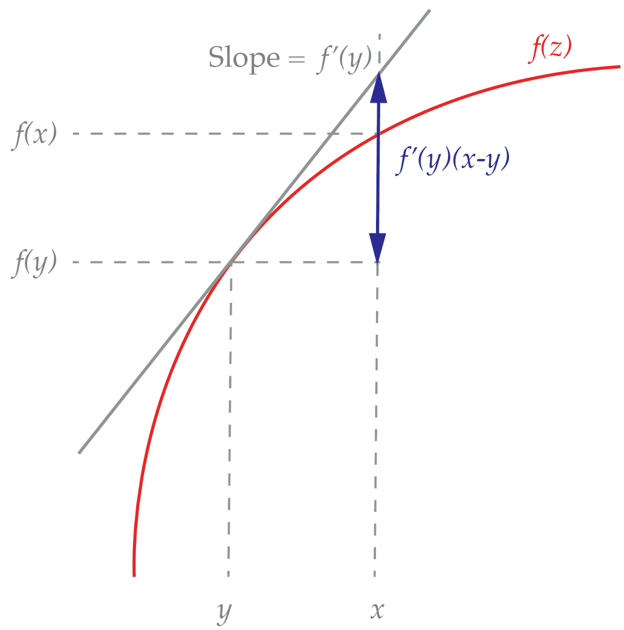


Figure 10: Illustration of Proposition 9 for (strictly) concave functions; draw a picture for convex functions!

Divide both sides by λ, we obtain

$$f(x) + \frac{f(y)}{\lambda} - f(y) \geq \frac{f(y + \lambda(x - y))}{\lambda};$$

rearrange, we get

$$f(x) - f(y) \geq \frac{f(y + \lambda(x - y)) - f(y)}{\lambda(x - y)}(x - y).$$

Letting $\lambda \longrightarrow 0$ yields the desired inequality.

Now suppose

$$f(u) - f(v) \geq f'(v)(u - v) \text{ for all } u, v \in (a, b). \tag{6}$$

Take any $x, y \in (a, b)$ with $x \neq y$, and any $\lambda \in [0, 1]$. Let $z = \lambda x + (1 - \lambda)y$; applying Equation (6) on x, z and y, z respectively, we get

$$f(x) - f(z) \geq f'(z)(x - z), \quad f(y) - f(z) \geq f'(z)(y - z).$$

Multiplying the inequality on the left by λ, and on the right by $1 - \lambda$, and adding them up yields

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(z);$$

hence f is convex.

Following basically the same proof as above and replacing inequalities by strict inequalities, the statement for strictly convex functions can be established. ■

6.2 Quasiconvex and quasiconcave functions

As discussed above, convex and concave functions have many appealing properties; but sometimes some weaker assumptions are sufficient for our purpose. In economics, the most used weakening of convexity (concavity) is called quasiconvexity (quasiconcavity).

Definition 18 (Quasiconvex and quasiconcave functions)**.** Let T be a nonempty *convex* subset of $\mathbb{R}^n$.

- A function $f : T \longrightarrow \mathbb{R}$ is called **quasiconvex** if

  $$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\} \quad \text{for any } x, y \in T \text{ and } 0 \leq \lambda \leq 1,$$

  and **strictly quasiconvex** if the above inequality holds strictly for any distinct $x, y \in T$ and $0 < \lambda < 1$.

- A function $f : T \longrightarrow \mathbb{R}$ is called **quasiconcave** if

  $$f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\} \quad \text{for any } x, y \in T \text{ and } 0 \leq \lambda \leq 1,$$

  and **strictly quasiconcave** if the above inequality holds strictly for any distinct $x, y \in T$ and $0 < \lambda < 1$.

The "symmetry" of convex and concave functions stated in Observation 8 holds for quasiconvex and quasiconcave functions.

Observation 10. *A function f is (strictly) quasiconvex if and only if $-f$ is (strictly) quasiconcave.*

By definition, it is not difficult to show that quasiconvexity (quasiconcavity) is indeed a weakening of convexity (concavity).

Observation 11. *If a function f is (strictly) convex, it is (strictly) quasiconvex; if a function f is (strictly) concave, it is (strictly) quasiconcave.*

Observe that the converse of Observation 11 is not true. For instance, the function $f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ defined by $f(x) = x^2$ is quasiconcave: take any $x', x'' \in \mathbb{R}_+$; without loss of generality, assume $x' \leq x''$, then we have

$$\lambda x' + (1-\lambda)x'' \geq x'';$$

and since f is increasing on $\mathbb{R}_+$, we have

$$f(\lambda x' + (1-\lambda)x'') \geq f(x'') = \min\{f(x'), f(x'')\}.$$

However, f is not concave; in fact, it is convex. Then Observation 11 implies that f is also quasiconvex; so a function could be both quasiconvex and quasiconcave.

And for quasiconvex and quasiconcave functions, we have a "stronger" result than Proposition 8 for convex and concave functions.

Proposition 10. *If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is an increasing function and $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a quasiconvex (quasiconcave) function, $f \circ g$ is also quasiconvex (quasiconcave).*

Proof. Suppose g is quasiconcave,[13] so for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$,

$$g(\lambda x + (1-\lambda)y) \geq \min\{g(x), g(y)\}.$$

Then because f is increasing,

$$f\left(g(\lambda x + (1-\lambda)y)\right) \geq f\left(\min\{g(x), g(y)\}\right) = \min\{f(g(x)), f(g(y))\}.$$

Draw a picture to make sure that you understand why does the equality above hold. ■

Definition 19. Let $T \subseteq \mathbb{R}^n$ and let $f : T \longrightarrow \mathbb{R}$ be a function. For any $y \in \mathbb{R}$, the set

$$U^f(y) = \{x \in T : f(x) \geq y\} = f^{-1}([y, \infty))$$

is called the **upper level set** (or **superlevel set**) of f for y; and the set

$$L_f(y) = \{x \in T : f(x) \leq y\} = f^{-1}((-\infty, y])$$

is called the **lower level set** (or **sublevel set**) of f for y.

Lemma 2. *Let $f : C \longrightarrow \mathbb{R}$ where C is a convex subset of $\mathbb{R}^n$. Then*

(1) f is quasiconcave if and only if $U^f(y)$ is convex for all $y \in \mathbb{R}$;

(2) f is quasiconvex if and only if $L_f(y)$ is convex for all $y \in \mathbb{R}$.

Proof. We only prove Part (1); Part (2) is completely analogous. Suppose f is quasiconcave, and fix any $y \in \mathbb{R}$. To show that $U^f(y)$ is convex, take $x', x'' \in U^f(y)$. By definition of $U^f(y)$, $f(x') \geq y$

[13] The proof for quasiconvex g is completely analogous.

and $f(x'') \geq y$. But then

$$f(\lambda x' + (1-\lambda)x'') \geq \min\{f(x'), f(x'')\} \geq y;$$

this implies that $\lambda x' + (1-\lambda)x'' \in U^f(y)$, and hence $U^f(y)$ is convex.

Now suppose $U^f(y)$ is convex for all y. Take $x', x'' \in C$ and any $\lambda \in [0,1]$, and let $y = \min\{f(x'), f(x'')\}$. By construction, $x', x'' \in U^f(y)$. Then because $U^f(y)$ is convex, $\lambda x' + (1-\lambda)x'' \in U^f(y)$. Consequently, $f(\lambda x' + (1-\lambda)x'') \geq y = \min\{f(x'), f(x'')\}$, which implies that f is quasiconcave by definition. ■

With the help of Lemma 2, we are ready to establish the following characterization of quasiconvex and quasiconcave functions defined on subsets of the real line.

Theorem 12. *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f : [a,b] \longrightarrow \mathbb{R}$ be a function.*

(1) f is quasiconcave if and only if there exists $x^* \in [a,b]$ such that f is increasing on $[a, x^*]$ and decreasing on $[x^*, b]$.

(2) f is quasiconvex if and only if there exists $x^* \in [a,b]$ such that f is decreasing on $[a, x^*]$ and increasing on $[x^*, b]$.

Proof. Again, we only prove Part (1); Part (2) is completely analogous. We proceed by contraposition. If f does not satisfy any of the conditions, then we can find $x_1, x_2, x_3 \in [a,b]$ such that $x_1 < x_2 < x_3$ and

$$f(x_2) < \min\{f(x_1), f(x_3)\}.$$

Define

$$\lambda = \frac{x_3 - x_2}{x_3 - x_1},$$

then $\lambda \in (0,1)$, and $x_2 = \lambda x_1 + (1-\lambda)x_3$. Let $y = \min\{f(x_1), f(x_3)\}$, then the upper level set $U^f(y)$ contains x_1 and x_3, but not x_2, and hence is not convex. By Lemma 2, f is not quasiconcave. ■

We sometimes call a function satisfying the condition described in (1) "single-peaked", and "single-dipped" if it satisfies the condition described in (2). Taking $x^* \in \{a, b\}$, we immediately get Corollary 3.

Corollary 3. *Let $f : [a,b] \longrightarrow \mathbb{R}$ be a function. If f is (strictly) monotone, it is both (strictly) quasiconvex and quasiconcave.*

Example 13. $f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ defined by $f(x) = x^2$ is both quasiconvex and quasiconcave.

7 Further reading

For those who know some analysis (for example, those who have taken at least one semester of calculus and one semester of advanced calculus or undergraduate real analysis), Chapter A of Ok (2007) is an excellent warm-up read. Section A.1 could be a bit too abstract, so you might want to focus on the other three sections. It is not necessarily an easy read but it would be very helpful if you read carefully.

Appendix

A Concave and convex function of two variables

To determine whether a twice-differentiable function of many variables is concave or convex is much more difficult than a function defined on the real line: we need to examine all its second partial derivatives. In this short section we focus on the case that f is defined on a subset of $\mathbb{R}^2$, since we rarely need to check concavity or convexity of a function of more variables.

Definition 20. Let $S \subseteq \mathbb{R}^2$ and let $f : S \longrightarrow \mathbb{R}$ be a twice-differentiable function. The **Hessian** of f at $s \in S$ is

$$H(s) = \begin{pmatrix} f_{11}(s) & f_{12}(s) \\ f_{21}(s) & f_{22}(s) \end{pmatrix},$$

where

$$f_{11} = \frac{\partial^2 f}{\partial x_1^2}, f_{12} = \frac{\partial^2 f}{\partial x_1 \partial x_2}, f_{21} = \frac{\partial^2 f}{\partial x_2 \partial x_1}, \text{ and } f_{22} = \frac{\partial^2 f}{\partial x_2^2}.$$

The **determinant** of a 2×2-matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

denoted $|A|$, is defined as

$$|A| = ad - bc.$$

Then the determinant of the Hessian at $s \in S$ is

$$|H(s)| = f_{11}(s) f_{22}(s) - f_{12}(s) f_{21}(s).$$

Theorem 13. *Let $S \subseteq \mathbb{R}^2$ be convex and open, and let $f : S \longrightarrow \mathbb{R}$ be twice-differentiable. Then* *a) f is convex if and only if*

$$f_{11}(s) \geq 0, f_{22}(s) \geq 0, \text{ and } |H(s)| \geq 0$$

for all $s \in S$;

b) if

$$f_{11}(s) > 0 \text{ and } |H(s)| > 0$$

for all $s \in S$ then f is strictly convex;

a) f is concave if and only if

$$f_{11}(s) \leq 0, f_{22}(s) \leq 0, \text{ and } |H(s)| \geq 0$$

for all $s \in S$; and

d) if

$$f_{11}(s) < 0 \text{ and } |H(s)| > 0$$

for all $s \in S$ then f is strictly concave.

Note that Theorem 13 does not claim that if f is strictly concave then $H(s)$ is negative definite for all $s \in S$; it is indeed false.

Example 14. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined by

$$f(x, y) = 2x - y - x^2 + 2xy - y^2.$$

Then we have

$$H(x, y) = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$$

for all $(x, y) \in \mathbb{R}^2$. $H(x, y)$ is negative semidefinite, so f is concave. (Note that $H(x, y)$ is not negative definite, so f *may or may not be* strictly concave.)

In Example 14, the Hessian does not depend on (x, y); in general it does.

References

de la Fuente, A. (2000): *Mathematical Methods and Models for Economists*, Cambridge: Cambridge University Press.

Ok, E. A. (2007): *Real Analysis with Economic Applications*, Princeton: Princeton University Press.

Osborne, M. J. (2016): "Mathematical Methods for Economic Theory," https://mjo.osborne.economics.utoronto.ca/index.php/tutorial/index/1/int/i.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton: Princeton University Press.

Rudin, W. (1976): *Principles of Mathematical Analysis*, New York: McGraw Hill.

Simon, C. P. and L. Blume (1994): *Mathematics for Economists*, New York: W. W. Norton & Company.

Vohra, R. V. (2005): *Advanced Mathematical Economics*, New York: Routledge.