Contents

Pre-Requisite 4

1 Basic Set Theory 4
  1.1 Sets and Their Elements .......................................................... 4
  1.2 Operations on Sets ..................................................................... 4
  1.3 Mappings Between Sets ............................................................... 5
  1.4 Ordered Sets ............................................................................. 5
  1.5 Commonly Used Sets ................................................................... 6
  1.6 Cardinality ................................................................................ 6

2 Logic & Proof Writing 7
  2.1 Different types of propositions .................................................... 7
  2.2 Types of proofs .......................................................................... 9

3 Calculus & Analysis 13
  3.1 The Real Line: An Ordered Field . .............................................. 13
  3.2 Real Sequences ........................................................................ 14
    3.2.1 Basic Notions ..................................................................... 14
    3.2.2 Special Sequences ............................................................... 16
  3.3 Functions on $\mathbb{R}$ ................................................................. 17
    3.3.1 Basic Notions ..................................................................... 17
    3.3.2 Limits and Continuity ........................................................... 17
  3.4 Differentiation ........................................................................... 18
    3.4.1 Univariate and Multivariate Differentiability ....................... 18
    3.4.2 Differentiation Rules & Formulae ....................................... 20
    3.4.3 Continuous Differentiability & Higher-Order Derivatives ....... 20
  3.5 Integration .................................................................................. 21
    3.5.1 Riemann Integration ............................................................ 21
    3.5.2 Integration & Differentiation ............................................... 22
    3.5.3 Multivariate Integration ....................................................... 22
    3.5.4 Change of Variable ............................................................ 23

*Preliminary, please do not circulate.
†Watch for typos. Comments are welcome and should be sent to margaux.luflade@duke.edu.
A Few Words on this Handout.

The difficult thing with Math Camp (I mean for me) is to decide where to start. The level of heterogeneity among students is usually pretty high. Some of you have probably already mastered the content of the whole class; others will realize they remember very little from their college math classes. The goal of the class is to make sure everyone enters the first year equipped with the technical tools they need to make it through classes and problem sets.

This handout is supposed to set us up with a starting point for when we meet in class. I expect you to read it carefully and to make sure you understand and are familiar with everything in it by the first day of class. The material in it is usually covered in standard college calculus, analysis and linear algebra classes; so I have no doubt that if you have never seen this before, you should be able to catch up on everything by working through this handout. If you do not feel comfortable with everything though, whether it is because it is the first time you see this or because you forgot it, please email me. At the end of the handout (and in the syllabus that you all should have received by now), I provide references that you may want to check out if you feel you need more context to work through the material here.

This handout covers basic math concepts; it sets up the language we should all speak to be able to develop tools more directly applicable to study Econ questions. The first section presents basic notions of Set Theory; it itself sets up the language and concepts then used in the rest of the handout. The third section develops the bases of real analysis, starting with the definition of the real line as a field and then going through college calculus results (functions, differentiation, integration). This will serve as a basis for when we cover more advanced analysis results, topology and optimization in class. The last section recalls basics of matrix algebra.

Some parts of this handout may look to you pretty theoretical for a preparation to an Econ PhD, and doing math is not what you have signed for. However, there are some theoretical tools you need to have a working knowledge of to be able to navigate Econ research and literature. You need to be able to solve the optimization problem of the agent whose behavior you try to model; you need to be able to prove the existence of an equilibrium in your model before claiming it explains a phenomenon you are studying; etc. This requires technical tools but also knowledge of how to write proofs, an issue addressed in the second section of this handout. I believe the only way to become good at this is (1) to realize that, although there is no general method that will always give you a correct and elegant proof, there are typical ways to approach a proof; and (2) to practice. The second section of the handout is supposed to address the first point: it explains the questions you should ask yourself while constructing your proof, and the different ways you can proceed to do so. When it comes to the second point, you should see every proposition, result, lemma, theorem of the handouts of this class as an exercise: try to prove them! For many results in this handout, you do not need to know the proof to be successful in an Econ PhD. However, once again, you do need to know how to do proofs, I thus provide (some of) them to make sure you find examples that help you in your practice.

There is a problem set at the end of this handout. I’ll probably ask you to turn it in by the end of the first week of class, and you will have to work in groups. However, I strongly encourage you to use these problems as a way to test your knowledge of the material in this handout. Again, if you don’t feel comfortable with this material, email me.

In class, I will assume the content of this handout is common knowledge and build on it. Details on the content of the class are provided in the syllabus.

Finally, let me conclude this preamble by a few words on the notations used in this handout. As you will see, I make heavy use of symbolic notation. I introduce more formally and comment on these symbols in
section 2, but here are some hints to help those of you who are not familiar with them to read this handout.

- The right arrow $\Rightarrow$ reads ‘implies’. $A \Rightarrow B$ indicates that $B$ holds whenever $A$ does, i.e. that $A$ holding implies $B$ holding. For instance, definition 1.2:

  ‘$B$ is a subset of $A$ if $B$ is a set such that $b \in B \Rightarrow b \in A$.’

  can be understood as: ‘$B$ is a subset of $A$ if $B$ is a set such that whenever $b$ is in $B$, $b$ is also in $A$.’

- The double arrow $\iff$ reads ‘if and only if’ and indicates a double implication. $A$ holding implies $B$ holding, and conversely $B$ holding implies $A$ holding. For instance, definition 1.1:

  ‘Two sets $A$ and $B$ are equal, i.e. $A = B$, if $x \in A \iff x \in B$.’

  can be understood as: ‘Two sets $A$ and $B$ are equal if whenever an element is in $A$, it is also in $B$ and, conversely, whenever an element is in $B$, it is also in $A$.’

- The quantifier $\exists$ reads ‘there exists’, while $\forall$ reads ‘for all’. The vertical bar $|$ reads ‘such that’. The ‘such that’ is often implicit and it will come naturally when reading the statement out loud. I often use parentheses to organize my symbolic statements. I gather the quantifiers and the variables they are associated within parentheses, while the property satisfied is expressed after the sets of parentheses. For instance, in definition 1.18

  ‘$E$ is bounded above if $(\exists \, \beta \in S) \, (\forall \, x \in E) \, x \leq \beta$.’

  can be read as: ‘$E$ is bounded above if there exists a $\beta$ in $S$ such that for all $x$ in $E$, $x$ is smaller than $\beta$."

  Alternatively, reversing the two expressions between parentheses:

  ‘$(\forall \, x \in E) \, (\exists \, \beta \in S) \, x \leq \beta$.’

  would read: ‘for all $x$ in $E$, there exists a $\beta$ in $S$ such that this $x$ is smaller than $\beta$’ or ‘for all $x$ in $E$, there exists a $\beta$ in $S$ that is larger than $x$.’ Note that changing the order of the quantifiers is not innocuous: the two statements do not mean the same thing! (Are you clear with the difference?)

- Finally, a colon before an equal sign $:=\,$ denotes an equality that holds by definition. An equivalent symbol you can find in texts is $\equiv$. 


Part

Pre-Requisite

1 Basic Set Theory

1.1 Sets and Their Elements

Definition 1.1. • A set is a collection of elements. We denote $a \in A$ an element $a$ of a set $A$.

• The empty set $\emptyset$ is the set with no element.

• A singleton is a set with a unique element: $A = \{a\}$.

• Two sets $A$ and $B$ are equal, i.e. $A = B$, if $x \in A \iff x \in B$.

Definition 1.2. • $B$ is a subset of $A$ if $B$ is a set such that $b \in B \Rightarrow b \in A$. We denote $B \subseteq A$.

• $B$ is a proper subset of $A$ if $B$ is a subset of $A$ and there exists $a \in A$ such that $a \notin B$.

Remark. $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. This is the usual way to prove two sets are equal: first show that $A \subseteq B$, i.e. pick any element $a \in A$ and show that $a \in B$; then show that $B \subseteq A$, i.e. pick any element $b \in B$ and show that $b \in A$.

Definition 1.3. The power set of the set $A$, denoted $\mathcal{P}(A)$, is the set of all subsets of $A$: $\mathcal{P}(A) := \{B | B \subseteq A\}$. Note that $\emptyset, A \in \mathcal{P}(A)$.

Definition 1.4. An index set is a set whose elements are used to label the elements of another set.

Example 1.5. Consider the set $A = \{a_1, a_2, a_3\}$. Its elements are $a_1, a_2, a_3$. Its power set is $\mathcal{P}(A) = \{\emptyset, \{a_1\}, \{a_2\}, \{a_3\}, \{a_1, a_2\}, \{a_1, a_3\}, \{a_2, a_3\}, A\}$. Note that $a_1 \in A$ ($a_1$ is an element of $A$) while $\{a_1\} \subset A$ (the singleton $\{a_1\}$ is a subset of $A$). $\Lambda = \{1, 2, 3\}$ is an index set for $A$.

1.2 Operations on Sets

Definition 1.6. • Difference. Let $A, B$ be two sets. $A \setminus B := \{a \in A | a \notin B\}$.

• Complement. Let $A$ be a set and $U \in A$. $A^c := A \setminus U = \{u \in A | u \notin U\}$ is called the complement of $U$ (relative to $A$).

• Union. Let $A, B$ be two sets. $A \cup B := \{x | x \in A \text{ or } x \in B\}$ is called the union of $A$ and $B$. More generally, $\bigcup_{\lambda \in \Lambda} A_\lambda := \{a | a \in A_\lambda \text{ for some } \lambda \in \Lambda\}$ is the union of the family $\{A_\lambda\}_{\lambda \in \Lambda}$.

• Intersection. Let $A, B$ be two sets. $A \cap B := \{x | x \in A \text{ and } x \in B\}$ is called the intersection of $A$ and $B$. More generally, $\bigcap_{\lambda \in \Lambda} A_\lambda := \{a | a \in A_\lambda \text{ for all } \lambda \in \Lambda\}$ is the intersection of the family $\{A_\lambda\}_{\lambda \in \Lambda}$.

Proposition 1.7. – DeMorgan’s Laws. Let $X$ be a set and $A, B, (A_\lambda)_\Lambda \subseteq X$. All complements are to be understood relative to $X$.

• $(A \cup B)^c = A^c \cap B^c$ and more generally $(\bigcup_{\lambda \in \Lambda} A_\lambda)^c = \bigcap_{\lambda \in \Lambda} A_\lambda^c$.

• $(A \cap B)^c = A^c \cup B^c$ and more generally $(\bigcap_{\lambda \in \Lambda} A_\lambda)^c = \bigcup_{\lambda \in \Lambda} A_\lambda^c$.

Definition 1.8. Two sets $A$ and $B$ are called disjoint if $A \cap B = \emptyset$. A family $\{A_\lambda\}_{\lambda \in \Lambda}$ is called mutually disjoint if $A_\lambda \cap A_{\lambda'} = \emptyset$ for all $\lambda \neq \lambda'$.

Definition 1.9. The Cartesian product of two sets $A$ and $B$ is $A \times B := \{(a, b) | a \in A, b \in B\}$.

Example 1.10. Consider the set $E = \{w, x, y, z\}$. Let $A = \{x, y, z\}$ and $B = \{w, x, y\}$.

$A \setminus B = \{z\}$; $A \cap B = \{x, y\}$; $A \cup B = \{w, x, y, z\}$.

$A \times B = \{(x, w), (x, x), (x, y), (y, w), (y, x), (y, y), (z, w), (z, x), (z, y)\}$. 

4
1.3 Mappings Between Sets

Definition 1.11. A map (or rule of assignment) \( r \) from a set \( A \) to a set \( B \) is a subset of \( A \times B \) such that for each element \( a \in A \), there exists a unique element \( \tau \in r \) such that \( \tau = (a, b) \) for some \( b \in B \). The map is denoted: \( r : A \rightarrow B \).

Example 1.12. Let \( A = \{x, y, z\} \) and \( B = \{w, x, y\} \).
\[
\begin{align*}
\text{Example 1.12}: & \quad r_1 = \{(x, x), (y, x), (z, w)\} \text{ is a map.} \\
& \quad r_2 = \{(x, x), (y, x), (z, w)\} \text{ is a not map: } x \in A \text{ is paired to both } x \text{ and } y \in B. \\
& \quad r_3 = \{(y, x), (z, w)\} \text{ is a not map: } x \in A \text{ is not paired to any element of } B.
\end{align*}
\]

Definition 1.13. Let \( A, B \) be two sets. Consider a map \( r : A \rightarrow B \).

- For any \( a \in A \), the (unique) element \( b \in B \) such that \( (a, b) \in r \) is called the image of \( a \) by \( r \) and denoted \( b = r(a) \).
- For any subset \( A' \subset A \), the image of \( A' \) by \( r \) is the subset \( r(A') := \bigcup_{a \in A'} r(a) = \{b \in B | b = r(a), a \in A'\} \).
- \( A \) is called the domain of \( r \) and \( B \) is called its codomain. The image (or range) of \( r \) is the subset \( \text{Im}(r) := \{b \in B | \exists a \in A \text{ s.t. } b = r(a)\} \).
- For any \( b \in B \), the set \( r^{-1}(b) := \{a \in A | r(a) = b\} \) is called the pre-image of \( b \) (by \( r \)). Note that \( r^{-1}(b) \neq \emptyset \) if \( b \in \text{Im}(r) \). For any \( V \subseteq B \), the set \( r^{-1}(V) := \bigcup_{b \in V} r^{-1}(b) \) is called the pre-image of \( V \) (by \( r \)). Note that \( r^{-1}(V) \neq \emptyset \) if \( V \cap \text{Im}(r) \neq \emptyset \).

Definition 1.14. Let \( A, B \) be two sets. Consider a map \( r : A \rightarrow B \).

- \( r \) is injective (or one-to-one) if \( a \neq a' \in A \implies r(a) \neq r(a') \).
- \( r \) is surjective (or maps \( A \) onto \( B \)) if \( \text{Im}(r) = B \), i.e. \( (\forall b \in B) (\exists a \in A) b = r(a) \).
- \( r \) is bijective if it is both injective and surjective.

Proposition 1.15. If \( r : A \rightarrow B \) be a bijection between the sets \( A \) and \( B \), then the map denoted \( r^{-1} : B \rightarrow A \) and defined by: \( (\forall (a, b) \in A \times B) \)
\[
r^{-1}(b) = a \quad \text{where } b \text{ is such that } b = r(a). \tag{1}
\]
is well-defined. The map \( r^{-1} \) is called the inverse of \( r \).

1.4 Ordered Sets

Definition 1.16. Let \( S \) be a set. An order on \( S \) is a relation, denoted by \( < \), with the following two properties:

- (Completeness.) \( (\forall x, y \in S) \) one and only one of the following three statements is true: \( x < y \); \( x = y \); \( y < x \).
- (Transitivity.) \( (\forall x, y, z \in S) [x < y \text{ AND } y < z] \implies x < z \).

A set equipped with an order is called an ordered set.

Definition 1.17. Let \( S \) be an ordered set and \( E \subset S \). \( E \) is bounded above if \( (\exists \beta \in S) (\forall x \in E) x \leq \beta \). Such \( \beta \) is called an upper bound of \( E \).

\( E \) is bounded below if \( (\exists \beta \in S) (\forall x \in E) \beta \leq x \). Such \( \beta \) is called a lower bound of \( E \).

Definition 1.18. Let \( S \) be an ordered set and \( E \subset S \). Suppose the set of all upper bounds of \( E \) is non-empty: \( \overline{E} := \{\beta \in S | (\forall x \in E) x \leq \beta\} \neq \emptyset \). The supremum (or least upper bound) of \( E \) is, if it exists, the element \( \alpha \in \overline{E} \) such that \( (\forall \beta \in \overline{E}) \alpha \leq \beta \). Such \( \alpha \) is denoted sup \( E \).

Suppose the set of all lower bounds of \( E \) is non-empty: \( \underline{E} := \{\beta \in S | (\forall x \in E) \beta \leq x\} \neq \emptyset \). The infimum (or greatest lower bound) of \( E \) is, if it exists, the element \( \alpha \in \underline{E} \) such that \( (\forall \beta \in \underline{E}) \beta \leq \alpha \). Such \( \alpha \) is denoted inf \( E \).

Example 1.19. We did not get to a formal definition of the real numbers as an ordered set but let’s use what we already know about them to get some more intuition about upper, lower bounds, suprema and infima.
Consider the interval \((0, 1) \subset \mathbb{R}\). The set of upper bounds of \((0, 1)\) is not empty: for instance, \(1, 2, 3\) are upper bounds of \((0, 1): (\forall x \in (0, 1)) x < 1 < 2 < 3. 1\) is the least upper bound, or supremum: any upper bound of \((0, 1)\) will be (weakly) larger than 1; conversely, any real number \(x < 1\) is not an upper bound of \((0, 1)\) since there will be a real number lying strictly between \(x\) and 1.

In the definition, I ‘suppose the set of all upper bounds of \(E\) is non-empty’. What could be a set in \(\mathbb{R}\) with no upper bound? Typically, unbounded sets such as \(I = [2, +\infty):\) pick any \(x \in \mathbb{R}\), if \(x < 2\), 2 is in \(I\) and larger than \(x\), so \(x\) is not an upper bound; if \(x \geq 2\), \(x + 1\) is in \(I\) and larger than \(x\) so \(x\) is not an upper bound of \(I\).

In the definition, I also hint at the fact that the supremum of a set may not exist, although this set does have upper bounds. Can we find such a set in \(\mathbb{R}\)? It turns out that no, we cannot. This is a property of \(\mathbb{R}\) (see definition 3.8, theorem 3.10). However, example 3.9 shows that such a set can be found in \(\mathbb{Q}\).

Remark. Note that the supremum or the infimum of a set, even if it exists, is not necessarily an element of the set itself. The previous example illustrates this point: 0 and 1 are respectively the infimum and supremum of \((0, 1)\).

1.5 Commonly Used Sets

I will not cover the construction of the real numbers here. The intuition we have about this set is sufficient for most of the economics problems you will be facing. Later in this handout, I define the field structure the real line is usually equipped with. Denote \(\mathbb{R}\) the set of real numbers.

Definition 1.20. • \(\mathbb{N}\) is the set of natural numbers or positive integers: \(\mathbb{N} = \{0, 1, 2, \ldots\}\)

• \(\mathbb{Z}\) is the set of all integers: \(\mathbb{Z} := \{\ldots, -2, -1, 0, 1, 2, \ldots\}\)

• \(\mathbb{Q}\) is the set of rational numbers, that is, all numbers that can be written \(\frac{m}{n}\) where \(m, n \in \mathbb{Z}\): for instance, \(\frac{1}{2}, \frac{1}{3}, 1, -\frac{3}{2}\) are rational.

• \(\mathbb{R}\setminus\mathbb{Q}\) is called the set of irrational numbers: for instance, \(\pi, \sqrt{2}, -e\) are irrational.

• \(\mathbb{C}\) is the set of complex numbers, that is, all numbers that can be written \(x + iy\) where \(x, y \in \mathbb{R}\), with \(i\) defined as the number satisfying \(i^2 = -1\): for instance, \(2, 2 - \frac{1}{2}i, 8\) are complex.

Remark. \(\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}\).

1.6 Cardinality

Definition 1.21. • The cardinality of a set \(A\), denoted \(\#A\), is its number of elements.

• A set \(A\) is finite if \(\#A < \infty\), that is, if there is \(n \in \mathbb{N}\) and a bijection between \(A\) and \(\{1, 2, \ldots, n\}\).

• A set \(A\) is infinite if it is not finite; then \(\#A = \infty\).

• Two sets are said to have the same cardinality if there exists a bijection between them.

• A set \(A\) is countable if there is a bijection between \(A\) and \(\mathbb{N}\).

Theorem 1.22. –Schroder-Bernstein Theorem. Let \(A, B\) be two sets. If there exist an injection \(A \rightarrow B\) and an injection \(B \rightarrow A\); then there exists an bijection between \(A\) and \(B\).

Example 1.23. \(\mathbb{Z}\) is countable. Indeed, consider the map

\[
\varphi : \mathbb{N} \rightarrow \mathbb{Z}, \quad n \mapsto \begin{cases} 
\frac{n}{2} & \text{if } n \text{ is even} \\
-\frac{n+1}{2} & \text{if } n \text{ is odd}
\end{cases}
\]

Let’s check that \(\varphi\) is a bijection. It is easy to see that \(\varphi\) is injective on \(\mathbb{N}\). Let \(m, n \in \mathbb{N}\) with \(m \neq n\). If \(m, n\) are both even, then \(n \neq m \Rightarrow \frac{n}{2} \neq \frac{m}{2} \Rightarrow \varphi(n) \neq \varphi(m)\). If \(m, n\) are both odd, then \(n \neq m \Rightarrow -\frac{n+1}{2} \neq -\frac{m+1}{2} \Rightarrow \varphi(n) \neq \varphi(m)\). Now, if \(m\) is odd while \(n\) is even, \(\varphi(m) = -\frac{m+1}{2} < 0\) while \(\varphi(n) = \frac{n}{2} \geq 0\) so \(\varphi(n) \neq \varphi(m)\).

To see that \(\varphi\) is surjective, pick any \(p \in \mathbb{Z}\). If \(p \geq 0\), note that \(2p \in \mathbb{N}\) and \(\varphi(2p) = p\). If \(p < 0\), note that \(-2p - 1 \in \mathbb{N}\) and \(\varphi(-2p - 1) = p\).

In other words, the inverse function of \(\varphi\) is well-defined as:

\[
\varphi^{-1} : \mathbb{Z} \rightarrow \mathbb{N}, \quad p \mapsto \begin{cases} 
2p & \text{if } p \geq 0 \\
-2p - 1 & \text{if } p < 0
\end{cases}
\]
Remark. This example illustrates the fact that the concept of cardinality can be a little tricky. When two sets are finite, then having the same cardinality is a straightforward idea: literally speaking they have the same number of elements. When the sets are not finite, the intuition may be harder to get. As we just saw, two infinite sets can have same cardinality, although one strictly contains the other: 

\[ \mathbb{Z} \supseteq \mathbb{N} \]

we just saw, two infinite sets can have same cardinality, although one strictly contains the other: \( \mathbb{Z} \supseteq \mathbb{N} \). The key trick is that \( \mathbb{N} \) and \( \mathbb{Z} \setminus \mathbb{N} = \{-1, -2, -3, \ldots \} \) are both countable. As shown in the problem set, the union of two countable sets is countable; so here \( \mathbb{Z} = \mathbb{N} \cup \mathbb{Z} \setminus \mathbb{N} \) is countable. However, not all infinite sets have same cardinality. By definition, an uncountable set \( B \) does not have the same cardinality as a countable set \( A \): there does not exist any bijection between \( B \) and \( A \). In the problem set, you will show that the infinite set \( \mathbb{R} \) is not countable.

Remark. Let \( A, B \) be two sets. Let \( f : A \to B \) be a map between them. Note that \( f \) being injective requires \( \#B \geq \#A \) while \( f \) being surjective requires \( \#B \leq \#A \).

## 2 Logic & Proof Writing

### 2.1 Different types of propositions

**Existential, Universal and Uniqueness statements**

An *existential* statement asserts the fact that there exists (at least) one element of the domain under consideration that satisfies some property \( P \). The *existential quantifier* \( \exists \) is used to denote existence: ‘(\( \exists x \in X \) \( P(x) \)’ reads ‘there exists an element \( x \) in \( X \) such that the property \( P \) holds for \( x \).’

A way to prove such a statement is to proceed *constructively*, i.e. to exhibit an element \( x \) of \( X \) that indeed satisfies \( P \). The term *constructively* comes from the fact that often this element \( x \) needs to be constructed. Often, such a proof proceeds in two steps: first a candidate \( x \) is found, and then it is shown to actually satisfy \( P \).

An *universal* statement asserts the fact that all elements of the domain under consideration satisfy some property \( P \). The *universal quantifier* \( \forall \) is used to denote universality: ‘(\( \forall x \in X \) \( P(x) \)’ reads ‘for all elements \( x \) in \( X \), the property \( P \) holds for \( x \).’

Exhibiting an example is not enough to prove such a statement; all what an example shows is that there exists one element for which the property holds. A general way to prove such a statement is pick an arbitrary element \( x \) of \( X \) and to show that the property \( P \) holds for that \( x \). By arbitrary, we mean an element whose only known characteristic is its belonging to \( X \) —nothing more, it can be any of the elements of \( X \). The logic is that if the property is proved for an arbitrary \( x \) in \( X \), then it holds for all such elements since, once again, the only thing assumed about \( x \) is that it belongs to \( X \).

A *uniqueness* statement asserts the fact that if some element of the domain under consideration that satisfies some property \( P \) then it is the *only one* to do so. Some existence statements also assert uniqueness, in which case we denote (\( \exists! x \in X \) \( P(x) \))’ the read ‘there exists a unique element \( x \) in \( X \) such that the property \( P \) holds for \( x \).’

A general way to prove a uniqueness statement is to assume there are two elements of \( X \) satisfying \( P \) and to show that they must be equal. To prove an existence-and-uniqueness statement, we usually proceed in two steps; one to prove existence, the other to prove uniqueness.

**Necessary vs. Sufficient conditions.**

\( A \) is a necessary condition for \( B \) if it must hold for \( B \) to obtain or, equivalently, if \( B \) holds *only if* \( A \) holds. \( A \) not holding prevents \( B \) from obtaining. \( A \) holding does not allow you to conclude that \( B \) holds, and \( B \) not holding does not tell you that \( A \) does not hold either, but \( B \) holding guarantees that \( A \) holds.
A is a **sufficient** condition for B if the fact that it holds guarantees B to obtain or, equivalently, if B holds if A holds. B may very well hold even if A does not. A not holding does not allow you to conclude that B does not hold, but B not holding tells you that A does not hold either.

**Example 2.1.** Let’s illustrate this by considering a grand set X and two subsets A and B such that \( B \subseteq A \subseteq X \). For instance, let \( X = \mathbb{R} \), \( A = \mathbb{Z} \) and \( B = \mathbb{N} \).

‘\( x \in A \)’ is a necessary condition for ‘\( x \in B \)’. Indeed, if ‘\( x \in A \)’ fails, then ‘\( x \in B \)’ must fail too. For instance: \( x = \frac{1}{2} \notin \mathbb{Z} \) tells you that for sure \( x = \frac{1}{2} \notin \mathbb{N} \).

However, ‘\( x \in A \)’ is not a sufficient condition for ‘\( x \in B \)’. Indeed, we can find instances where ‘\( x \in A \)’ holds, while ‘\( x \in B \)’ does not. For instance: \( x = -2 \in \mathbb{Z} \) while \( x = -2 \notin \mathbb{N} \).

A being a sufficient condition for B means that A *implies* B, thus the notation \( A \Rightarrow B \).

When defining a necessary condition, we noted that A being a necessary condition for B means that B cannot hold if A does not, i.e. that if B holds, it must be that A does as well. This is equivalent to B implies A, thus the notation \( A \Leftarrow B \).

Statements of the form \( A \iff B \) are called *if and only if* statements. They mean that A is a necessary *and* sufficient condition for B. In other words A and B are *equivalent*: A holds whenever B holds (necessity of A) and vice-versa (sufficiency of A).

The arrow symbols also hint at two other properties. First, rewriting \( A \Leftarrow B \) as \( B \Rightarrow A \) tells us that A is a necessary condition for B if and only if B is a sufficient condition for A. Second, considering \( A \Rightarrow B \) and \( B \Rightarrow C \), the notation suggests \( A \Rightarrow C \). The notion of *transitivity* actually holds for sufficient (and necessary) conditions: if A is sufficient (resp. necessary) for B and B is sufficient (resp. necessary) for C, then A is sufficient (resp. necessary) for C.

Finally, when defining a necessary condition, we noted that a way to see that A is a necessary condition for B is to see that the failure of A *implies* the failure of B. In other words, A is a necessary condition for B \((A \Leftarrow B)\) if and only if the failure of A is a sufficient condition for the failure of B \((\neg A \Rightarrow \neg B)\). This is the principle on which proofs by contradiction are built, as we will see later.

**Negation of statements.**

The previous paragraph hinted at the negation of statements \( \neg A \) and \( \neg B \). We thus need to be able to take the negation of A or B.

*Negation of an existence statement.* Consider the statement ‘there exists \( x \) in \( X \) such that \( x \) satisfies the property \( P \)’, i.e. \((\exists x \in X) \ P(x)\). Its negation writes: ‘there does not exist \( x \) in \( X \) such that \( x \) satisfies the property \( P \)’ or, equivalently ‘for all \( x \) in \( X \), \( x \) does not satisfy the property \( P \)’, i.e. \((\forall x \in X) \ \neg P(x)\). The negation of an existence statement is a universal statement.

This means that, although to show that an existence statement holds it is sufficient to find an example of an \( x \) satisfying the property, in order to show that an existence statement does not hold, it is not enough to find a counter-example, it is necessary to show that no \( x \) satisfies the property.

*Negation of a universal statement.* Consider the statement ‘for all \( x \) in \( X \), \( x \) satisfies the property \( P \)’, i.e. \((\forall x \in X) \ P(x)\). Its negation writes: ‘the property \( P \) is not satisfied by all \( x \) in \( X \)’ or, equivalently ‘there exists \( x \) in \( X \) such that \( x \) does not satisfy the property \( P \)’, i.e. \((\exists x \in X) \ \neg P(x)\). The negation of a universal statement is an existence statement.

This means that, although to show that a universal statement holds it is not enough to find an example of an \( x \) satisfying the property, in order to show that a universal statement does not hold, it is sufficient to find a counter-example.

These ideas will be useful below when discussing indirect methods to prove a statement.

**Vacuous truths.**

A *vacuously true* statement typically is a statement that asserts that some property holds for elements of an empty set: \((\forall x \in X) \ P(x)\) where \( X = \emptyset \). The idea of a vacuous truth is that the statement cannot be
proved to be wrong simply because since \(X\) is empty, no counter-example can be found. In other words, the negation \((\exists x \in X)\neg P(x)\) cannot be proved to hold.

**Example 2.2.** ‘\((\forall x \in \{y \in \mathbb{R} \mid y^2 < 0\}\) \(x\) is even’ is vacuously true; ‘\((\forall x \in \{y \in \mathbb{R} \mid y^2 < 0\}\) \(x\) is odd’ is also vacuously true. Actually, ‘\((\forall x \in \{y \in \mathbb{R} \mid y^2 < 0\}\) \(x\) is even and odd’ is also vacuously true. However, this does not imply that there exists an \(x\) that is even and odd, precisely because the \(x\)’s that are both even and odd are picked from an empty set. This last point should remind you that, when facing a set satisfying some property, before picking an element of this set, you must always check that the set is not empty.

### 2.2 Types of proofs

Suppose we are interested in proving a statement of the form \(A \Rightarrow B\).

**Direct proof.**

The principle of a direct proof simply builds on the transitivity of the implication \(\Rightarrow\). The idea is to assume that \(A\) holds and to deduce a sequence of intermediary results \(P_1, P_2, \ldots, P_k\) such that \(A \Rightarrow P_1 \Rightarrow P_2 \Rightarrow \cdots \Rightarrow P_k \Rightarrow B\).

**Example 2.3.** *Proposition.* Suppose \(x^2 - x > 0\) and \(x > 0\). Then \(x > 1\).

Let’s identify the structure of the statement first and make sure we understand what we need to prove. Writing \(A := “x^2 - x > 0 \text{ and } x > 0”\) and \(B := “x > 1”\), the statement rewrites \(A \Rightarrow B\).

**Proof.** Suppose \(x^2 - x > 0\) and \(x > 0\). Then \(x(x - 1) > 0\). Since \(x > 0\), we must have \(x - 1 > 0\) for this product to be positive. Therefore, \(x > 1\).

**Example 2.4.** *Proposition.* Let \(A, B\) be two sets. Let \(U \subseteq A, V \subseteq B\), Let \(f : A \rightarrow B\) be a map.

1. \(U \subseteq f^{-1}(f(U))\).
2. \(f(f^{-1}(V)) \subseteq V\).

Let’s identify the structure of the statement first and make sure we understand what we need to prove. Each of these statement claim that an inclusion holds. To reformulate the first statement in terms of our implication \(A \Rightarrow B\), let \(A := “x \in U”\) and \(B := “x \in f^{-1}(f(U))”\). Similarly for the second statement: let \(A := “x \in f^{-1}(U)”\) and \(B := “x \in V”\). This suggests a direct method to prove a first set is included in a second set: pick an arbitrary element of the first set and show it is an element of the second.

**Proof.** 1. Let \(x \in U\) (i.e. assume \(A\) holds); let’s prove that \(x \in f^{-1}(f(U))\) (i.e. that \(B\) holds). First, notice that \(f(U) := \{v \in A \mid \exists u \in U \text{ with } f(u) = v\} \subseteq B\). Since \(x \in U\), we know \(f(x) \in f(U)\). Now, notice that for any \(W \subseteq B\), \(f^{-1}(W) := \{u \in A \mid \exists w \in W \text{ with } f(u) = w\} \subseteq A\). Thus, choosing \(W = f(U)\), \(f^{-1}(f(U)) := \{u \in A \mid \exists w \in f(U) \text{ with } f(u) = w\} = \{u \in A \mid \exists u' \in U \text{ with } f(u) = f(u')\}\). It follows immediately that \(x \in f^{-1}(f(U))\): if \(x\) is our \(u\) in the requirement defining the set \(f^{-1}(f(U))\), setting \(u' = x\) is enough to see that \(x\) satisfies the requirement.

Because \(x\) was arbitrary in \(U\), proving \(x \in f^{-1}(f(U))\) proves that \(U \subseteq f^{-1}(f(U))\).

2. Let \(x \in f(f^{-1}(V))\); let’s prove that \(x \in V\). First, note that \(f^{-1}(V) := \{u \in A \mid \exists v \in V \text{ with } f(u) = v\} \subseteq U\). Next, see that for any \(W \subseteq A\), \(f(W) := \{v \in B \mid \exists w \in W \text{ with } f(w) = v\} \subseteq B\), thus choosing \(W = f^{-1}(V)\), \(f(f^{-1}(V)) := \{v \in B \mid \exists w \in f^{-1}(V) \text{ with } f(w) = v\} = \{v \in B \mid \exists v' \in V \text{ with } v' = v\}\). It follows immediately that \(x \in V\): if \(x \in f(f^{-1}(V))\), then choosing \(v' = x\) as well is enough to see that \(x\) satisfies the requirement.

Because \(x\) was arbitrary in \(f(f^{-1}(V))\), proving \(x \in V\) proves that \(f(f^{-1}(V)) \subseteq V\).
Proof by contradiction.

The idea of this method is (1) to assume \( \mathcal{A} \) holds but that \( \mathcal{B} \) does not, then (2) to develop on the consequences of \( \mathcal{B} \) not holding while \( \mathcal{A} \) does, and (3) to obtain a contradiction, that is, a conclusion that contradicts either \( \mathcal{A} \) or something else that we know holds.

Example 2.5. Proposition. \( \sqrt{2} \notin \mathbb{Q} \).

The structure of the statement is pretty clear here. Proving the statement by contradiction involves assuming the negated statement: \( \sqrt{2} \in \mathbb{Q} \), and deriving from there a contradiction.

Proof. By way of contradiction, suppose \( \sqrt{2} \) is rational: there are two integers \( n,m \) s.t. \( \sqrt{2} = \frac{n}{m} \). WLOG, assume \( m \) and \( n \) have no common factor. \( \sqrt{2} = \frac{n}{m} \iff n\sqrt{2} = m \Rightarrow 2n^2 = m^2 \), so \( 2 \) is a factor of \( m^2 \). But the factors of \( m^2 \) are simply the factors of \( m \), each taken twice\(^1\). So \( 2 \) is a factor of \( m \): \( m = 2k \) for some \( k \in \mathbb{N} \). Then \( m^2 = 4k^2 \) and \( 2n^2 = 4k^2 \iff n^2 = 2k^2 \). So \( 2 \) is a factor of \( n^2 \) and thus of \( n \). But then \( 2 \) is a common factor of \( n \) and \( m \), which contradicts our assumption. \( \square \)

Example 2.6. Proposition. Let \( A, B \) be two sets with \( \#A = \#B = n < \infty \). Let \( f : A \to B \) be a map. \( f \) is injective if and only if it is surjective.

Let’s identify the structure of the statement first and make sure we understand what we need to prove. The statement here is an ‘if and only if’ statement, i.e. \( \mathcal{A} \iff \mathcal{B} \) where \( \mathcal{A} = \{ \text{\( f \) is injective} \} \) and \( \mathcal{B} = \{ \text{\( f \) is surjective} \} \). It should be seen as a statement with a double implication: \( \mathcal{A} \iff \mathcal{B} \); or: \( \left[ \mathcal{A} \implies \mathcal{B} \right] \text{ AND } \left[ \mathcal{A} \leftarrow \mathcal{B} \right] \).

To prove an ‘if and only if’ statement, we usually proceed in two steps, one step to prove the ‘if’ part (i.e. \( \implies \)), and another to prove the ‘only if’ part (i.e. \( \iff \)).

Proof. \((\implies)\) By way of contradiction (‘BWOC’), suppose that \( f \) is surjective but not injective. \( f \) being surjective, \( \forall b \in B ) (\exists a \in A ) f(a) = b \). Therefore, \( f(A) = B \), thus \( \#f(A) = \#B \).

\( f \) not being injective, \( \exists a,a' \in A ) a \neq a' \Rightarrow f(a) = f(a') \). \( f \) being a map, each element of \( A \) is associated with a unique element of \( B \). Thus \( \#f(A) \leq \#A \) by \( f \) being a map, and in fact \( \#f(A) < \#A \) by \( f \) not being injective.

Combining both conclusions, we have \( n = \#B < \#A = n \): contradiction.

This proves that if \( f \) is surjective, \( f \) cannot be non-injective, i.e. \( f \) must be injective. In other words: \( \mathcal{B} \Rightarrow \mathcal{A} \).

\((\iff)\) Similarly, BWOC, suppose that \( f \) is injective but not surjective. \( f \) being injective, \( \forall a,a' \in A ) a \neq a' \Rightarrow f(a) \neq f(a') \). Thus \( \#f(A) \geq \#A \). \( f \) being a map, each element of \( A \) is associated with a unique element of \( B \). Thus \( \#f(A) \leq \#A \) and therefore \( \#f(A) = \#A \).

\( f \) not being surjective, \( \exists b \in B ) (\exists a \in A ) f(a) = b \). In other words, \( f(A) \subset B \setminus \{b\} \), therefore \( \#f(A) \leq \#B - 1 < \#B \).

Combining both conclusions, we have \( n = \#A < \#B = n \): contradiction.

This proves that if \( f \) is injective, \( f \) cannot be non-surjective, i.e. \( f \) must be surjective. In other words: \( \mathcal{A} \Rightarrow \mathcal{B} \). \( \square \)

Proof by the contrapositive.

As hinted in the previous subsection, a proof by the contrapositive builds on the fact that \( \mathcal{A} \Rightarrow \mathcal{B} \) is equivalent to \( \neg \mathcal{B} \Rightarrow \neg \mathcal{A} \). The idea then is to prove the equivalent statement instead of the original one: assume \( \mathcal{B} \) does not hold and prove that \( \mathcal{A} \) does not hold either.

Example 2.7. Proposition. Suppose \( x^2 - x > 0 \) and \( x > 0 \). Then \( x > 1 \).

Let’s identify the structure of the statement first and make sure we understand what we need to prove. This is the same statement as in the direct proof example; we still write \( \mathcal{A} := \{ x^2 - x > 0 \} \) and \( \mathcal{B} := \{ x > 1 \} \), but now we will prove \( \neg \mathcal{B} \Rightarrow \neg \mathcal{A} \), rather than the direct \( \mathcal{A} \Rightarrow \mathcal{B} \). Note that \( \neg \mathcal{A} = \{ x^2 - x \leq 0 \} \) OR \( x \leq 0 \) and \( \neg \mathcal{B} = \{ x \leq 1 \} \).

\(^1\)This follows from the fundamental theorem of arithmetics.
Proceed by the contrapositive, assume \( f(A) \Rightarrow B \) \( \Leftrightarrow \) \( \neg B \Rightarrow \neg A \).

Let’s identify the structure of the statement first and make sure we understand what we need to prove. As in the example for the direct proof, we are facing an inclusion, which rewrites: \( A \Rightarrow B \) with \( A = \{ x \in f^{-1}(V_0 \cup V_1) \} \) and \( B = \{ x \in f^{-1}(V_0) \cup f^{-1}(V_1) \} \).

**Example 2.8. Proposition.** Let \( A, B \) be two sets and \( V_0, V_1 \subseteq B \). Let \( f : A \rightarrow B \). \( f^{-1}(V_0 \cup V_1) \subseteq f^{-1}(V_0) \cup f^{-1}(V_1) \).

Proof. Proceeding by the contrapositive, assume \( \neg B \) and let’s prove \( \neg A \), or more explicitly for \( x \in A \), \( x \notin f^{-1}(V_0) \cup f^{-1}(V_1) \Rightarrow x \notin f^{-1}(V_0 \cup V_1) \). Consider an \( x \notin f^{-1}(V_0 \cup f^{-1}(V_1)) \), i.e. \( x \) such that \( x \notin f^{-1}(V_0) \) AND \( x \notin f^{-1}(V_1) \). Let’s make these properties of \( x \) more explicit: \( x \notin f^{-1}(V_0) \Leftrightarrow f(x) \notin V_0 \) and \( x \notin f^{-1}(V_1) \Leftrightarrow f(x) \notin V_1 \). But now, knowing that \( f(x) \notin V_0 \) and \( f(x) \notin V_1 \), we deduce that \( f(x) \notin (V_0 \cup V_1) \), which implies \( x \notin f^{-1}(V_0 \cup V_1) \), i.e. \( \neg A \).

**Example 2.9. Proposition.** Let \( A, B \) be two sets. Let \( U \subseteq A, V \subseteq B \), Let \( f : A \rightarrow B \).

1. \( U = f^{-1}(f(U)) \) if and only if \( f \) is injective.
2. \( f(f^{-1}(V)) = V \) if and only if \( f \) is surjective.

Let’s identify the structure of the statement first and make sure we understand what we need to prove. Each statement here is an ‘if and only if’ statement, so we will again in two steps, proving \((\Rightarrow)\) and \((\Leftarrow)\) separately. In the first statement, \( A = \{ U = f^{-1}(f(U)) \} \)” while \( B = “f \) is injective”. In the second statement, \( A = “f(f^{-1}(V)) = V \)” while \( B = “f \) is surjective”.

Proof. 1. \((\Leftarrow)\) Let’s start by the ‘only if’ statement: we want to prove that \( B \Rightarrow A \). Before anything, let’s look at our \( A \) statement. It is twofold: \( U \subseteq f^{-1}(f(U)) \) and \( U \supseteq f^{-1}(f(U)) \). In the direct proof above, we have already proved that \( U \subseteq f^{-1}(f(U)) \) holds for any map. Thus, a fortiori, \( U \subseteq f^{-1}(f(U)) \) holds for any injective map. Therefore, to prove that \( B \Rightarrow A \), only \( B \Rightarrow A' = “U \supseteq f^{-1}(f(U))” \) remains to be proved.

Proceed by the contrapositive: to prove that \( B \Rightarrow A' \), let’s actually prove that \( \neg A' \Rightarrow \neg B \). Assume that \( U \supseteq f^{-1}(f(U)) \) does not hold: \( \exists x \in f^{-1}(f(U)) \) such that \( x \notin U \). Note that \( x \in f^{-1}(f(U)) \Leftrightarrow f(x) \in f(U) \), i.e. \( \exists u \in U \) s.t. \( f(u) = f(x) \). Since be have assumed \( x \notin U \), there must exist \( u \neq x \) with \( f(u) = f(x) \), i.e. \( f \) is not injective.

\((\Rightarrow)\) We turn now to the ‘if’ statement: we want to prove that \( A \Rightarrow B \). Again, we proceed by the contrapositive, showing that \( \neg B \Rightarrow \neg A \). Note that to show that \( A \) fails, it is enough to show that either \( U \subseteq f^{-1}(f(U)) \) or \( U \supseteq f^{-1}(f(U)) \) fails. Assume \( f \) is not injective, i.e. \( \exists x, y \in A \) s.t. \( x \neq y \) and \( f(x) = f(y) \). Then, let’s find some \( U \subseteq A \) such that \( U \supseteq f^{-1}(f(U)) \) fails. Consider \( U = \{ x \} \). See that \( f(U) = \{ f(x) \} \).

Now, \( f^{-1}(f(U)) = \{ x \in A | f(z) = f(x) \} \) so at least \( x \) and \( y \) are in \( f^{-1}(f(U)) \): \( \{ x, y \} \subseteq f^{-1}(f(U)) \). Therefore \( f^{-1}(f(U)) \subseteq \{ x \} \) fails and so \( A \) fails.

2. \((\Leftarrow)\) Let’s start by the ‘only if’ statement: we want to prove that \( B \Rightarrow A \). Before anything, let’s look at our \( A \) statement. It is twofold: \( f(f^{-1}(V)) \subseteq V \) and \( f(f^{-1}(V)) \supseteq V \). In the direct proof above, we have already proved that \( f(f^{-1}(V)) \subseteq V \) holds for any map. Thus, a fortiori, \( f(f^{-1}(V)) \subseteq V \) holds for any surjective map. Therefore, to prove that \( B \Rightarrow A \), only \( B \Rightarrow A' = “f(f^{-1}(V)) \subseteq V” \) remains to be proved.

Suppose “\( f(f^{-1}(V)) \subseteq V \)” fails and show that \( f \) cannot be surjective. Consider \( v \in V \) s.t. \( v \notin f(f^{-1}(V)) \). We’ll prove that \( \exists x \in A \) s.t. \( f(x) = v \). \( v \notin f(f^{-1}(V)) \Leftrightarrow (\exists x \in f^{-1}(V)) f(x) = v \). Now consider the complement of \( f^{-1}(V) \) in \( A \): \( A \backslash f^{-1}(V) = \{ x \in A | f(x) \notin f^{-1}(V) \} = \{ x \in A | f(x) \notin V \} \). Clearly, since \( v \in V \), \( \forall x \in A \backslash f^{-1}(V) f(x) \neq v \), i.e. \( \exists x \in f^{-1}(V)) f(x) = v \). Since \( A = f^{-1}(V) \cup A \backslash f^{-1}(V) \), we have proved that \( \exists x \in A \) s.t. \( f(x) = v \), i.e. \( f \) is not surjective.

\((\Rightarrow)\) We turn now to the ‘if’ statement: we want to prove that \( A \Rightarrow B \). Again, we proceed by the contrapositive, showing that \( \neg B \Rightarrow \neg A \). Note that to show that \( A \) fails, it is enough to show that either
Proposition.

Example 2.10. Proposition. \( (\forall n \in \mathbb{N}) \sum_{i=1}^{n} i = \frac{n(n+1)}{2}. \)

Let’s identify the structure of the statement first and make sure we understand what we need to prove. The proposition says that a statement that depends on \( n \), \( P_n = \sum_{i=1}^{n} i = \frac{n(n+1)}{2} \), holds for every single \( n \in \mathbb{N} \) that we could fix.

Proof. Base case. For \( n=1 \). On the one hand, \( \sum_{i=1}^{1} i = 1 \); on the other hand, \( \frac{n(n+1)}{2} = \frac{1 \times 2}{2} = 1 \), so the property holds for \( n = 1 \).

Induction hypothesis. Fix some arbitrary \( n \in \mathbb{N} \) and suppose the property \( P_n = \sum_{i=1}^{n} i = \frac{n(n+1)}{2} \) holds for that \( n \).

Inductive step. Let’s prove that, under the induction hypothesis, the property holds for \( n+1 \), i.e. \( P_{n+1} = \sum_{i=1}^{n+1} i = \frac{(n+1)(n+2)}{2} \).

\[
\sum_{i=1}^{n+1} i = \sum_{i=1}^{n} i + (n + 1) = \frac{n(n+1)}{2} + (n + 1) \text{ by the induction hypothesis} \\
= (n + 1) \left( \frac{n+2}{2} \right) = \frac{(n+1)(n+2)}{2}
\]

Final comments: tips on how to find your way in the problem and to construct a proof.

1. **What?** Identify the structure of the statement and make sure you understand what you need to prove.
   If you don’t know where you have to go, there is little chance you’ll actually make it there... Note that this is the first thing I did before starting my example-proofs. This requires you asking yourself questions such that: Am I facing an ‘if’ statement or an ‘if and only if’ statement? Is that an existence or a universal statement? What are my \( A \) and my \( B \)?

2. **Which way?** Identifying \( A \) and \( B \) is a necessary beginning, however sometimes it is not totally clear how to actually go from \( A \) to \( B \). Of course, there is no trick that will enable you to see the path, each time you face a new statement. The best I can do here is to give you questions you should ask yourself when you are stuck in the proof and you don’t know where to go next.

   • **What do I know?** Try to exploit all the information contained in the things you have. For instance, make explicit/clear to yourself what \( A \) and \( B \) mean. For instance, as in my direct proof above, if \( A \) or \( B \) is that \( x \) is some element of a set, write down the definition of the set, and then try to exploit the fact that \( x \) satisfies some requirement (the one defining the set), deducing all the consequences of the information you have. Use definitions of the notions involved and known theorems involving these notions. Maybe, from one of the conclusions you’ll reach that way, you’ll see an easy way to reach what you actually want.

   • **What would I like to hold?** Sometimes, trying to make your way to \( B \), you’ll tell yourself: “it would be convenient if I knew that \( C \) were true” because \( C \implies B \) seems much more straightforward
than $A \Rightarrow B$. Then proceed with $C$ as an intermediary step: try to see if you can prove $A \Rightarrow C$
and then prove $C \Rightarrow B$.
- **What if $B$ did not hold?** Assuming $B$ does not hold and then asking yourself *What do I know?*
may lead you to an obvious contradiction (in which case, you’ve made a proof by contradiction).

## 3 Calculus & Analysis

### 3.1 The Real Line: An Ordered Field

**Definition 3.1.** A bivariate operator $*$ is a function $* : A \times A \to B$.

**Definition 3.2.** A set $A$ is said to be closed under the bivariate operator $*$ if $* : A \times A \to A$.

**Definition 3.3.** A field $F$ is a set equipped with two bivariate operators (addition $+$ and multiplication $\cdot$) which associate each $(x, y) \in F \times F$ with unique elements $x + y \in F$ and $x \cdot y \in F$ and satisfy the following field axioms:

1. $[A]$ Axioms for addition $+$
   - (a) *Closedness under $+$* $(\forall a, b \in F) a + b \in F$
   - (b) *Commutativity* $(\forall a, b \in F) a + b = b + a$
   - (c) *Associativity* $(\forall a, b, c \in F) a + (b + c) = (a + b) + c$
   - (d) *Existence of an identity element* $\exists$ element $0 \in F$ such that $(\forall a \in F) a + 0 = a$
   - (e) *Existence of an (additive) inverse* $(\forall a \in F, \exists c \in F$ such that $a + c = 0$; We write $c$ as $-a$

2. $[M]$ Axioms for multiplication $\cdot$
   - (a) *Closedness under $\cdot$* $(\forall a, b \in F) a \cdot b \in F$
   - (b) *Commutativity* $(\forall a, b \in F) a \cdot b = b \cdot a$
   - (c) *Associativity* $(\forall a, b, c \in F) a \cdot (b \cdot c) = (a \cdot b) \cdot c$
   - (d) *Existence of an identity element* $\exists$ element $1 \in F$ such that $(\forall a \in F) a \cdot 1 = a$
   - (e) *Existence of a (multiplicative) inverse* $(\forall b \in F$ with $b \neq 0, \exists d \in F$ such that $b \cdot d = 1$; we write $d$ as $b^{-1}$

3. $[D]$ Distributive law: $(\forall a, b, c \in F) a \cdot (b + c) = a \cdot b + a \cdot c$

**Definition 3.4.** An ordered field is a field $F$ which is also an ordered set on which the following two properties hold:

1. $(\forall x, y, z \in F| x \leq y) \ x + z \leq y + z$.
2. $(\forall (x, y) \in F| x > 0, y > 0) \ x \cdot y > 0$.

**Example 3.5.** The set of real numbers $\mathbb{R}$ with the usual addition and multiplication operators is an ordered field.

*See problem set for a proof.*

**Theorem 3.6.** (i) Archimedean Property. $(\forall x, y \in \mathbb{R}$ with $0 < x) (\exists n \in \mathbb{N}) y < nx$.
(ii) Denseness of $\mathbb{Q}$ in $\mathbb{R}$. $(\forall x, y \in \mathbb{R}$ with $x < y) (\exists q \in \mathbb{Q}) x < q < y$.

*Proof.* See problem set.

**Definition 3.7.** Let $S$ be an ordered set, and $A \subseteq S$. An upper bound of $A$ in $S$ is an element $s \in S$ such that $(\forall a \in A) a \leq s$. The smallest of these upper bounds, if it exists, is called the supremum (or the least upper bound) of $A$ in $S$. If, in addition, the supremum of $A$ in $S$ is in $A$, then it is called the maximum of $A$. A lower bound of $A$ in $S$ is an element $s \in S$ such that $(\forall a \in A) a \geq s$. The largest of these lower bounds, if it exists, is called the infimum (or the greatest lower bound) of $A$ in $S$. If, in addition, the infimum of $A$ in $S$ is in $A$, then it is called the minimum of $A$. 

\[\text{\blacksquare}\]
Definition 3.8. A set $S$ is said to have the Least Upper Bound Property if every bounded subset $A$ of $S$ has a supremum in $S$.

Example 3.9. $\mathbb{Q}$ does not have the Least Upper Bound Property. To prove this, we only need a counterexample. Consider $A = \{ q \in \mathbb{Q} | q^2 < 2 \}$. Pick any $p \in \mathbb{Q}$ with $\sqrt{2} < p$, and let’s show that, although an upper bound of $A$ in $\mathbb{Q}$, $p$ is not the least upper bound of $A$ in $\mathbb{Q}$. This follows directly from the density of the rational numbers in the reals: $(\exists r \in \mathbb{Q}) \sqrt{2} < r < p$.

By contrast, $A = \{ q \in \mathbb{Q} | q^2 < 2 \} \subset \mathbb{R}$ does have a least upper bound in $\mathbb{R}$. Indeed, consider $\sqrt{2}$. Clearly, $(\forall x \in A) x < \sqrt{2}$. Now let’s show that any upper bound of $A$ in $\mathbb{R}$ is (weakly) larger than $\sqrt{2}$, or equivalently, $(\forall y < \sqrt{2}) y$ is not an upper bound of $A$. Let $y < \sqrt{2}$. Then by the density of $\mathbb{Q}$ in $\mathbb{R}$, $(\exists r \in \mathbb{Q} \subset \mathbb{R}) y < r < \sqrt{2}$, so, with this $r$, we have an element of $A$ that is strictly larger than $y$.

Note that showing that $A$ has a least upper bound in $\mathbb{R}$ is not enough to show that $\mathbb{R}$ has the Least Upper Bound Property: we would need to prove that any arbitrary subset $A$ has a least upper bound. The actual proof is long and tedious so we omit it here and just state the following theorem. Details of the proof can be found in Rudin [3].

Theorem 3.10. $\mathbb{R}$ has the Least Upper Bound Property.

Proposition 3.11. For any set $A \subset \mathbb{R}$, if $\bar{a} = \sup A$ exists in $\mathbb{R}$ then $(\forall \varepsilon > 0) (\exists a \in A) |\bar{a} - a| < \varepsilon$.

Proof. BWOC, suppose $(\exists \varepsilon > 0) (\forall a \in A) |\bar{a} - a| \geq \varepsilon$. Since $\bar{a} > a (\forall a \in A)$, this means $\bar{a} - a \geq \varepsilon (\forall a \in A)$, i.e. $a \leq \bar{a} - \varepsilon (\forall a \in A)$. But then, $a - \varepsilon$ is an upper bound of $A$ that is strictly smaller than $\bar{a}$, which contradicts the definition of a supremum. \hfill \Box

Remark. We also have that “For any set $A \subset \mathbb{R}$, if $\bar{a} = \sup A$ exists in $\mathbb{R}$ then $(\forall \varepsilon > 0) (\exists a \notin A) |\bar{a} - a| < \varepsilon$”, but this a trivial statement: one can just to pick $a = \bar{a} + \frac{1}{2}\varepsilon$.

Definition 3.12. The extended real numbers consist in the real ordered field $\mathbb{R}$ and two symbols $-\infty$, $+\infty$. Preserving the original order on $\mathbb{R}$, define: $-\infty < x < +\infty (\forall x \in \mathbb{R})$.

Remark. The extended real numbers does not form a field; however, the following conventions are made:

- $(\forall x \in \mathbb{R}) x + \infty = +\infty \ \ x - \infty = -\infty \ \ \frac{x}{\infty} = \frac{\infty}{-\infty} = 0$
- $(\forall x \in \mathbb{R}$ with $0 < x) x \cdot (+\infty) = +\infty \ \ x \cdot (-\infty) = -\infty$
- $(\forall x \in \mathbb{R}$ with $0 < x) x \cdot (+\infty) = -\infty \ \ x \cdot (-\infty) = +\infty$

3.2 Real Sequences

3.2.1 Basic Notions

Definition 3.13. • A real sequence is a map from $\mathbb{N}$ to $\mathbb{R}$. A sequence is denoted $(x_n)$, with the meaning that $(\forall n \in \mathbb{N}) x_n = x(n) \in \mathbb{R}$.

• Let $(x_n)$ be a real sequence. Let $(n_k)_k$ be a strictly increasing sequence of $\mathbb{N}$. $(x_{n_k})_k$ is a subsequence of $(x_n)$.

Remark. In my notation of a subsequence, $(x_{n_k})_k$, the subscript outside is designed to make clear that it is the $k$-indexing, rather than the $n$-indexing of the initial sequence, that runs over the natural numbers $\{1, 2, 3, \ldots\}$. Based on this logic, in the rest of the text, $(x_n)$ or $(x_n)_n$ will generally denote the main sequence; while $(x_{n_k})$ or $(x_{n_k})_k$ will denote a subsequence of this main sequence.

Definition 3.14. Let $(x_n)$ be a real sequence.

- $(x_n)$ is increasing (resp. decreasing) if $(\forall n \in \mathbb{N}) x_n \leq x_{n+1}$ (resp. $x_n \geq x_{n+1}$).
- $(x_n)$ is strictly increasing on $E$ (resp. strictly decreasing) if $(\forall n \in \mathbb{N}) x_n < x_{n+1}$ (resp. $x_n > x_{n+1}$).
- $(x_n)$ is (strictly) monotonic if it is either (strictly) increasing or (strictly) decreasing.

Definition 3.15. A real sequence $(x_n)$ is bounded if

$$(\exists M \in \mathbb{N}) (\exists N \in \mathbb{N}) (\forall n > N) |x_n| < M.$$
Definition 3.16. Let \((x_n)\) be a real sequence. \(x \in \mathbb{R}\) is the limit of \((x_n)\) if it satisfies:

\[
(\forall \varepsilon > 0) \, \exists N_\varepsilon \in \mathbb{N} \, (\forall n > N_\varepsilon) \, |x_n - x| < \varepsilon
\]

If such \(x\) exists, \((x_n)\) is said to be convergent, and we denote \(\lim x_n = x\) or \(x_n \to x\).

Proposition 3.17. If a limit exists, it is unique.

Proof. Let \((x_n)\) be a real sequence, and let \(x, x' \in \mathbb{R}\) such that \(x_n \to x\) and \(x_n \to x'\). BWOC, suppose \(x \neq x'\). Let \(\varepsilon_0 = \frac{1}{2}|x - x'| > 0\). Let \(N_{\varepsilon_0} = N'_{\varepsilon_0} \in \mathbb{N}\) such that, as in the definition of the limit:

\[
(\forall n > N_{\varepsilon_0}) \, |x_n - x| < \varepsilon_0 \text{ and } (\forall n > N'_{\varepsilon_0}) \, |x_n - x'| < \varepsilon_0.
\]

Let \(N = \max\{N_{\varepsilon_0}, N'_{\varepsilon_0}\}\) so that \((\forall n > N) \, |x_n - x| < \varepsilon_0\) and \(|x_n - x'| < \varepsilon_0\). Then pick any \(n > N\). \(|x - x'| \leq |x - x_n| + |x' - x_n| < \varepsilon_0 + \varepsilon_0 = 2\varepsilon_0\), which contradicts \(\varepsilon_0 = \frac{1}{2}|x - x'|\). \(\square\)

Remark. This is a typical way to prove uniqueness: BWOC assume two distinct elements satisfy the property and derive a contradiction.

Definition 3.18. Let \((x_n)\) be a real sequence.

- The limit superior (or \(\limsup\)) of \((x_n)\) is \(\limsup_{n \to \infty} x_n := \lim_{n \to \infty} \sup_{m \geq n} x_m = \inf_{n \geq 0} \sup_{m \geq n} x_m\).
- The limit inferior (or \(\liminf\)) of \((x_n)\) is \(\liminf_{n \to \infty} x_n := \lim_{n \to \infty} \inf_{m \geq n} x_m = \sup_{n \geq 0} \inf_{m \geq n} x_m\).

Proposition 3.19. For any real sequence \((x_n)\), \(\liminf x_n \leq \limsup x_n\). The limit \(\lim x_n\) of \((x_n)\) exists if and only if \(\liminf x_n = \limsup x_n\) in which case \(\lim x_n = \limsup x_n = \lim x_n\).

Proposition 3.20. - Operations on limits. Suppose the sequences \((u_n)\) and \((v_n)\) have limits; let \(\lambda \in \mathbb{R}\).

- \(\lim (\lambda u_n) = \lambda \lim u_n\)
- \(\lim (u_n + v_n) = \lim u_n + \lim v_n\)
- \(\lim (u_n v_n) = \lim u_n \lim v_n\)
- \(\lim \frac{u_n}{v_n} = \frac{\lim u_n}{\lim v_n}\) if \(\lim v_n \neq 0\)

Proposition 3.21. A real sequence \((x_n)\) converges to \(x\) if and only if every subsequence of \((x_n)\) converges to \(x\).

Proof. \((\Rightarrow)\) Suppose \((x_n)\) converges to \(x\). Pick an arbitrary subsequence \((x_{n_k})\) of \((x_n)\) and let’s show it converges to \(x\). Fix \(\varepsilon > 0\). By convergence of \((x_n)\), \((\exists N \in \mathbb{N}) \, (\forall n > N) \, |x_n - x| < \varepsilon\). By definition of a subsequence, \(\{x_{n_k} \mid n_k > N\} \subset \{x_n \mid n > N\}\) so in particular \((\forall n_k > N) \, |x_{n_k} - x| < \varepsilon\), i.e. \((x_{n_k})\) converges to \(x\).

\((\Leftarrow)\) By the contrapositive, suppose \((x_n)\) does not converge to \(x\) and let’s construct a subsequence that does not converge to \(x\) either. Fix \(\varepsilon > 0\), say \(\varepsilon = 1\). Since \((x_n)\) does not converge to \(x\), \((\exists n_0 \in \mathbb{N}) \, |x_{n_0} - x| \geq \varepsilon\). Now, since \((x_n)\) does not converge to \(x\), \((\exists n > n_0 \in \mathbb{N}) \, |x_{n_1} - x| \geq \varepsilon\). This procedure can be repeated infinitely: after \(k\) steps, such that \(\{x_{n_0}, x_{n_1}, \ldots, x_{n_{k-1}}\}\) has been constructed and since \((x_n)\) does not converge to \(x\), \((\exists n_k > n_{k-1} \in \mathbb{N}) \, |x_{n_k} - x| \geq \varepsilon\). Let’s check that the following holds for the resulting subsequence \((x_{n_k})\): \((\exists \varepsilon > 0) \, (\forall N \in \mathbb{N}) \, (\exists n_k > N) \, |x - x_{n_k}| > \varepsilon\). Consider our \(\varepsilon = 1\) and pick any arbitrary \(N \in \mathbb{N}\). Notice that by definition of a subsequence, indexes \((n_k)\) eventually get over our fixed \(N\). Let \(n_{k*} := \min\{n_k \mid n_k > N\}\), the first indice of the subsequence over \(N\). By construction, \(|x_{n_{k*}} - x| > \varepsilon\). So our subsequence does not converge to \(x\). \(\square\)

Proposition 3.22. - Bolzano-Weierstrass Theorem (Basic Case). Every bounded sequence of \(\mathbb{R}\) has a convergent subsequence.

Proof. The result immediately follows from the two lemmas:

Lemma 3.23. Every sequence of \(\mathbb{R}\) has a monotone subsequence.
Proof of the lemma. Consider a sequence \((x_n)_n\). Define a peak of the sequence any \(N \in \mathbb{N}\) s.t \((\forall n > N) x_n < x_N\). We distinguish two cases now: \((x_n)_n\) has either infinitely many peaks or only finitely many of them. First consider the case of infinitely many peaks: \(\{n_1, n_2, \ldots, n_k, \ldots\}\). Consider the induced subsequence \((x_{n_k})_k\). By construction, it is monotonically decreasing. Now consider the second case of finitely many peaks. The set of peaks is a finite subset of \(\mathbb{N}\) and therefore as a max, say \(N\). Let \(n_1 = N + 1\). \(n_1\) not being a peak, \((\exists n_2 > n_1) (x_{n_2} \geq x_{n_1})\). Similarly, \(n_2 > N\) so \(n_2\) is not a peak and thus \((\exists n_3 > n_2) (x_{n_3} \geq x_{n_2})\). Proceeding iteratively like this, we construct a monotonically increasing subsequence \((x_{n_k})_k\).

**Lemma 3.24.** A monotone sequence that is bounded converges.

Proof of the lemma. WLOG consider an increasing sequence \((x_n)_n\). By boundedness of the sequence and the Least Upper Bound Property of \(\mathbb{R}\), \(x := \sup_n \{x_n | n \in \mathbb{N}\}\) exists. We show now \(x\) is the limit of \((x_n)_n\). Fix \(\varepsilon > 0\). Recall a lemma proved earlier: For any set \(A \subset \mathbb{R}\), if \(\bar{a} = \sup A\) exists in \(\mathbb{R}\) then \((\forall \varepsilon > 0) (\exists a \in A) |\bar{a} - a| < \varepsilon\). Applied to our set \(A = \{x_n\} : (\exists x_N \in A) |x - x_N| < \varepsilon\). Now the sequence being increasing and \(x\) being the sup of \(A\), this means means that all subsequent terms of the sequence must lie between \(x_N\) and \(x\), so that: \((\forall n > N) |x_n - x| < \varepsilon\). Thus \((x_n)_n\) converges to \(x\).}

**Definition 3.25.** A real sequence \((x_n)_n\) is Cauchy if:

\[(\forall \varepsilon > 0) (\exists N_\varepsilon \in \mathbb{N}) (\forall n, m > N_\varepsilon) \ |x_n - x_m| < \varepsilon\]

**Proposition 3.26.** A sequence \((x_n)_n\) is Cauchy if and only if it is convergent.

Proof. See problem set.

### 3.2.2 Special Sequences

**Geometric Sequences.**

**Definition 3.27.** A sequence \((x_n)_n\) is geometric if \((\exists q \in \mathbb{R}) (\forall n \in \mathbb{N}) x_{n+1} = q \cdot x_n\). \(q\) is called the parameter of the sequence.

Remark. If \((\forall n \in \mathbb{N}) x_{n+1} = q \cdot x_n\), then \((\forall n \in \mathbb{N}) x_n = x_0 \cdot q^n\).

**Proposition 3.28.** Let \((x_n)_n\) be a geometric sequence of parameter \(q\). \((x_n)_n\) converges to 0 if and only if \(|q| < 1\), it goes to \(\infty\) if and only if \(|q| > 1\). It is constant (equal to \(x_0\)) if and only if \(q = 1\); if \(q = -1\), it alternatively takes values \(x_0\) and \(-x_0\).

**Proposition 3.29.** Let \((x_n)_n\) be a geometric sequence of parameter \(q\) such that \(|q| < 1\). For any \(p < m \in \mathbb{N}\) the sum of the terms \(\{x_p, x_{p+1}, \ldots, n_{m-1}, x_m\}\) is \(\sum_{n=p}^{m} x_n = \frac{x_p + q^m x_m}{1-q}\). In particular \(\sum_{n=0}^{\infty} x_n = \frac{x_0}{1-q}\).

**Arithmetic Sequences.**

**Definition 3.30.** A sequence \((x_n)_n\) is arithmetic if \((\exists q \in \mathbb{R}) (\forall n \in \mathbb{N}) x_{n+1} = x_n + q\). \(q\) is called the parameter of the sequence.

Remark. If \((\forall n \in \mathbb{N}) x_{n+1} = q + x_n\), then \((\forall n \in \mathbb{N}) x_n = x_0 + nq\).

**Proposition 3.31.** Let \((x_n)_n\) be an arithmetic sequence of parameter \(q\). \((x_n)_n\) converges to \(x_0\) if and only if \(q = 0\), it goes to \(+\infty\) if and only if \(q > 0\) and to \(-\infty\) if \(q < 0\).

**Proposition 3.32.** Let \((x_n)_n\) be an arithmetic sequence of parameter \(q\). For any \(p < m \in \mathbb{N}\) the sum of the terms \(\{x_p, x_{p+1}, \ldots, n_{m-1}, x_m\}\) is \(\sum_{n=p}^{m} x_n = (m - p + 1) \frac{x_p + x_m}{2}\).
3.3 Functions on \( \mathbb{R} \)

3.3.1 Basic Notions

Definition 3.33. A function on \( \mathbb{R} \) is a map \( f : \mathbb{R} \to \mathbb{R} \).

Definition 3.34. – Operations on functions. Let \( X \subseteq \mathbb{R} \); let \( f_1 : X \to \mathbb{R} \), \( f_2 : X \to \mathbb{R} \). Let \( Y \subseteq \text{Im}(f_1) \subseteq \mathbb{R} \); let \( g : Y \to \mathbb{R} \). Let \( \lambda \in \mathbb{R} \).

- **Multiplication by a scalar.** \((\lambda f_1)\) is the function \( X \to \mathbb{R} \) such that \((\forall x \in X)\) \((\lambda f_1)(x) := \lambda \cdot f_1(x)\).
- **Sum.** \((f_1 + f_2)\) is the function \( X \to \mathbb{R} \) such that \((\forall x \in X)\) \((f_1 + f_2)(x) := f_1(x) + f_2(x)\).
- **Multiplication.** \((f_1 f_2)\) is the function \( X \to \mathbb{R} \) such that \((\forall x \in X)\) \((f_1 f_2)(x) := f_1(x) f_2(x)\).
- **Quotient.** \( \left( \frac{f_1}{f_2} \right) \) is the function \( X \setminus f_2^{-1}\{0\} \to \mathbb{R} \) such that \((\forall x \in X \setminus f_2^{-1}\{0\})\) \(\left( \frac{f_1}{f_2} \right)(x) := \frac{f_1(x)}{f_2(x)}\).
- **Composition.** \((g \circ f_1)\) is the function \( X \to \mathbb{R} \) such that \((\forall x \in X)\) \((g \circ f_1)(x) := g(f_1(x))\).

Definition 3.35. Let \( X \subseteq \mathbb{R} \) and \( f_1 : X \to \mathbb{R} \), \( f_2 : X \to \mathbb{R} \).

- \( f_1 \geq f_2 \) (resp. \( f_1 \leq f_2 \)) if \((\forall x \in X)\) \( f_1(x) \geq f_2(x) \) (resp. \((\forall x \in X)\) \( f_1(x) \leq f_2(x) \)).

- Similarly, \( f_1 > f_2 \) (resp. \( f_1 < f_2 \)) if \((\forall x \in X)\) \( f_1(x) > f_2(x) \) (resp. \((\forall x \in X)\) \( f_1(x) < f_2(x) \)).

Definition 3.36. Let \( X \subseteq \mathbb{R} \) and \( f : X \to \mathbb{R} \). Let \( E \subseteq X \).

- \( f \) is increasing on \( E \) (resp. decreasing on \( E \)) if \((\forall x_1, x_2 \in E)\) \( x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2) \) (resp. \((\forall x_1, x_2 \in E)\) \( x_1 < x_2 \Rightarrow f(x_1) \geq f(x_2) \)).
- \( f \) is strictly increasing on \( E \) (resp. strictly decreasing on \( E \)) if \((\forall x_1, x_2 \in E)\) \( x_1 < x_2 \Rightarrow f(x_1) < f(x_2) \) (resp. \((\forall x_1, x_2 \in E)\) \( x_1 < x_2 \Rightarrow f(x_1) > f(x_2) \)).
- \( f \) is (strictly) monotonic on \( E \) if it is either (strictly) increasing on \( E \) or (strictly) decreasing on \( E \).

3.3.2 Limits and Continuity

Definition 3.37. \( l \in \mathbb{R} \cup \{-\infty, +\infty\} \) is the limit at \( c \) of a function \( f : X \to \mathbb{R} \) if

\[
(\forall \varepsilon > 0) \quad \exists \, \delta > 0 \quad \left( c - \frac{\delta}{2} < c + \frac{\delta}{2} \right) \quad \forall \varepsilon \in \left( c - \frac{\delta}{2}, c + \frac{\delta}{2} \right) \quad |f(c) - l| < \varepsilon
\]

Proposition 3.38. – Operations on limits. Properties here can be deduced straightforwardly from the properties of limits of sequences.

Definition 3.39. Let \( X \subseteq \mathbb{R} \) and \( f : X \to \mathbb{R} \).

Let \( x \in X \). \( f \) is continuous at \( x \) if:

\[
(\forall \varepsilon > 0) \quad \exists \, \delta > 0 \quad (\forall x' \in X) \quad |x - x'| < \delta \Rightarrow |f(x) - f(x')| < \varepsilon
\]

Note that if \( f \) is continuous at \( x \in X \), \( \lim_{x' \to x} f(x') = f(x) \).

Let \( E \subseteq X \). \( f \) is continuous on \( E \) if it is continuous at every \( x \in E \).

Definition 3.40. Let \( X \subseteq \mathbb{R} \) and \( f : X \to \mathbb{R} \).

Let \( E \subseteq X \). \( f \) is uniformly continuous on \( E \) if:

\[
(\forall \varepsilon > 0) \quad \exists \, \delta > 0 \quad (\forall x, x' \in X) \quad |x - x'| < \delta \Rightarrow |f(x) - f(x')| < \varepsilon
\]

Remark. Clearly, any uniformly continuous function is continuous. The definition of uniform continuity says that in the definition of continuity, for any \( \varepsilon > 0 \), the same \( \delta > 0 \) will work for all \( x \)'s.

Proposition 3.41. If \( f : [a,b] \to \mathbb{R} \) is continuous then for every convergent sequence \((x_n)_n\) of \([a,b]\), \( \lim f(x_n) = f(\lim x_n) \).

Proof: Suppose \( f \) is continuous on \([a,b]\). Let \((x_n)_n\) be a convergent sequence of \([a,b]\), denote \( x := \lim x_n \).

Consider the sequence \((f(x_n))_n\). Fix an arbitrary \( \varepsilon > 0 \). We want to show \((\exists N \in \mathbb{N}) \quad (\forall n > N) \quad |f(x_n) - f(x)| < \varepsilon \).

By continuity of \( f \), \((\exists \delta > 0) \quad |x - y| < \delta \Rightarrow |f(x_n) - f(x)| < \varepsilon \).

Now, by convergence of \((x_n)_n\), \((\exists N_\delta \in \mathbb{N}) \quad (\forall n > N_\delta) \quad |x_n - x| < \delta \).

Thus for \( N = N_\delta \): \((\forall n > N) \quad |x_n - x| < \delta \) which implies \(|f(x_n) - f(x)| < \varepsilon \), as desired. \( \square \)
Proposition 3.42. – Heine-Cantor Theorem (Basic Case). A function \( f : [a, b] \to \mathbb{R} \) that is continuous on a closed and bounded interval of \( \mathbb{R} \) is uniformly continuous.

Proof. BWOC suppose \( f \) is continuous but not uniformly continuous: \( \exists \varepsilon > 0 \) \( \forall \delta > 0 \) \( \exists x, y \in [a, b] \) \( |x - y| < \delta \) and \( |f(x) - f(y)| > \varepsilon \). In particular, consider the sequence \( (\delta_n)_n \) with \( \delta_n = \frac{1}{n} \); then we have an associated sequence of points of \([a, b] : (x_n, y_n)_n \) s.t

\[
(\forall n \in \mathbb{N}) |x_n - y_n| < \delta_n \text{ and } |f(x_n) - f(y_n)| > \varepsilon
\]  

(2)

Note that the sequences \((x_n)_n\) and \((y_n)_n\) are bounded (their terms lie in \([a, b]\)). By the Bolzano-Weierstrass Theorem, \((x_n)_n\) has a convergent subsequence, say \((x_{n_k})_k\). Let \( x := \lim x_{n_k} \). Based on the lemma below, we note that \( x \in [a, b] \), so \( f \) is defined at \( x \).

Now consider the subsequence \((y_{n_k})_k\) of \((y_n)_n\). Since \( (\forall n \in \mathbb{N}) |x_n - y_n| < \delta_n \), \( \lim y_{n_k} = x \). \( f \) being continuous, \( \lim f(x_{n_k}) = f(\lim x_{n_k}) = f(x) \) and \( \lim f(y_{n_k}) = f(\lim y_{n_k}) = f(x) \). So \( \lim f(x_{n_k}) = \lim f(y_{n_k}) \). This induces: \( (\exists M \in \mathbb{N}) (\forall k > M) |f(x_{n_k}) - f(y_{n_k})| < \varepsilon \), which then contradicts (2).

Let’s see how to derive that \( (\exists M \in \mathbb{N}) (\forall k > M) |f(x_{n_k}) - f(y_{n_k})| < \varepsilon \). \( \lim f(x_{n_k}) = f(x) \) implies that \( (\exists M_y \in \mathbb{N}) (\forall k > M_y) |f(x_{n_k}) - f(x)| < \frac{1}{2} \varepsilon \). Similarly, \( \lim f(y_{n_k}) = f(x) \) implies that \( (\exists M_y \in \mathbb{N}) (\forall k > M_y) |f(y_{n_k}) - f(x)| < \frac{1}{2} \varepsilon \). Let \( M = \max(M_x, M_y) \). Then,

\[
(\forall k > M) |f(x_{n_k}) - f(y_{n_k})| = |f(x_{n_k}) - f(x) + f(x) - f(y_{n_k})| \\
\leq |f(x_{n_k}) - f(x)| + |f(x) - f(y_{n_k})| \\
< \frac{1}{2} \varepsilon + \frac{1}{2} \varepsilon = \varepsilon
\]

where the first inequality comes from the triangle inequality, a property of the absolute value that you are asked to check in the problem set.

Lemma 3.43. Let \((x_n)_n\) be a sequence of the closed interval \([a, b] \subset \mathbb{R}\). If converges \((x_n)_n\) to \( x \in \mathbb{R} \) then \( x \in [a, b] \).

Proof. First, suppose \( x < a \) and let \( \varepsilon := a - x > 0 \). By definition of convergence to \( x \), \( (\exists M \in \mathbb{N}) (\forall n > M) |x_n - x| < \frac{1}{2} \varepsilon \). But then \( x_n < a \), which contradicts the fact that all terms of the sequence fall within \([a, b]\). Thus, \( x < a \) cannot hold, i.e \( x \geq a \)

Supposing \( b < x \) and proceeding just as above leads us to the fact that \( b < x \) cannot hold either, so that \( x \leq b \).

So we have that \( x \geq a \) and \( x \leq y \), i.e. \( x \in [a, b] \).

Remark. See that the proof of the lemma would not work if the interval was not closed. Consider the sequence \((x_n)_n\) with \( x_n = \frac{1}{n} \) \( (\forall n \in \mathbb{N}) \). All terms of the sequence are in \((0, 1]\) and the limit of the sequence is 0, which is not in \((0, 1]\).

We will come back to the relationship between limits of sequences and closed sets in the topology section of the class.

3.4 Differentiation

3.4.1 Univariate and Multivariate Differentiability

Differentiability of univariate functions

Definition 3.44. Let \( U \) be an open interval of \( \mathbb{R} \). Let \( x \in U \). A function \( f : U \to \mathbb{R} \) is differentiable at \( x \) if the limit

\[
\lim_{h \to 0} \frac{f(x + h) - f(x)}{h}
\]

exists and is finite. In such case, the limit is denoted \( f'(x) \) and is called the derivative of \( f \) at \( x \).

The function \( f : U \to \mathbb{R} \) is differentiable if it is differentiable at every \( x \in U \).

When \( f : U \to \mathbb{R} \) is differentiable, the (well-defined) function \( U \to \mathbb{R}, x \mapsto f'(x) \) is denoted \( f' \) and called the derivative of \( f \).
Remark. Writing the limit (3) requires us to evaluate the function \( f \) at points \( x' = x + h \) getting closer and closer to \( x \) as \( h \) goes to 0. How do we know that we have the right to do so?, meaning how do we know that \( f \) is defined at such points \( x' \)?, i.e. how do we know \( x' \) lies in the domain \( U \) of \( f \)?. This is a property of open intervals of \( \mathbb{R} \) that, if it is not already, will become clear when we get to the Topology section of the class: for any open interval \( U \subseteq \mathbb{R} \) and for any point \( x \in U \), there is some \( \varepsilon > 0 \) such that \((\forall x' \in \mathbb{R}) |x - x'| < \varepsilon \Rightarrow x' \in U\).

For some this reason, in the next paragraph, I will also make an openness requirement on the domain of functions of \( \mathbb{R}^n \) to define their differentiability. You may have less intuition about what it means for a subset of \( \mathbb{R}^n \) to be open. The definition of the concept will become clear later; for now, you can see an open subset \( U \) of \( \mathbb{R}^n \) as the Cartesian product of \( n \) open intervals of \( \mathbb{R} \): \( U = \prod_{i=1}^{n} U_i \) where \( U_i \subseteq \mathbb{R} \) is an open interval.

Differentiability of multivariate functions

Definition 3.45. Let \( U \) be a open subset of \( \mathbb{R}^n \). Let \( x \in U \), let \( j \in \{1, \ldots, n\} \). Let \( e_j \) denote the vector of \( \mathbb{R}^n \) whose elements are all 0 but the \( j \)th one which is equal to 1. A function \( f : U \to \mathbb{R} \) is differentiable at \( x \) with respect to its \( j \)th component if the limit

\[
\lim_{h \to 0} \frac{f(x + he_j) - f(x)}{h} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_j-1, x_j + h, x_{j+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}
\]

exists and is finite. In such case, the limit is denoted \( \frac{\partial}{\partial x_j} f(x) \) (or \( \partial_j f(x) \)) and is called the \( j \)th partial derivative of \( f \) at \( x \).

When all of them exist, the \( n \)-vector of partial derivatives of \( f \) at \( x \) is called the gradient of \( f \) at \( x \) and denoted \( \nabla f(x) = Df(x) := (\partial_1 f(x), \ldots, \partial_n f(x)) \).

When \( f : U \to \mathbb{R} \) is differentiable at all \( x \in U \), the (well-defined) operator \( U \to \mathbb{R}^n, x \mapsto Df(x) \) is denoted \( Df(\cdot) \) and called the derivative of \( f \).

Example 3.46. – Marginal utility. Consider a consumer who derives utility from consuming two (perfectly divisible) goods \( a \) and \( b \). Suppose her level of utility from consuming \( x_a \) units of \( a \) and \( x_b \) units of \( b \) is given by \( u(x_a, x_b) = x_a^2 + x_b^2 \). Note that \( u : \mathbb{R}_+^2 \to \mathbb{R} \) is differentiable with respect to both of its arguments and the partial derivatives are given by:

\[
\begin{align*}
\partial_{x_a} u(x_a, x_b) &= 2x_a \\
\partial_{x_b} u(x_a, x_b) &= 2x_b
\end{align*}
\]

The quantity \( \partial_{x_a} u(x_a, x_b) \) (resp. \( \partial_{x_b} u(x_a, x_b) \)) is called the marginal utility good \( a \) (resp. good \( b \)). It measures the change in the level of utility induced by a marginal increase in the quantity of good \( a \) (resp. \( b \)) consumed, given that the agent already consumes \( (x_a, x_b) \). This can be understood from the limit (4) defining partial derivatives. In the case of good \( b \) here, the marginal utility is equal to a constant \((\frac{1}{2})\); this means that no matter how much of each good is already consumed, an increase in the consumption of \( b \) by \( 1 \) unit will induce an increase in the level of utility of \( \frac{1}{2} \). In the case of good \( a \), the marginal utility is a non-constant function of the quantity already consumed \( x_a \).

Differentiability of maps \( \mathbb{R}^n \to \mathbb{R}^p \)

Definition 3.47. Let \( U \) be an open subset of \( \mathbb{R}^n \). Consider a map \( f : U \to \mathbb{R}^p \),

\[
f : x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longmapsto \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix}
\]

where \((\forall j = 1, \ldots, p) f_j(\cdot)\) is a function \( U \to \mathbb{R} \). If \((\forall j = 1, \ldots, p) f_j(\cdot)\) is differentiable at \( x \), then the \((n \times p)\)-array of all partial derivatives of \( f \) is called the Jacobian matrix of \( f \) and denoted

\[
Jf(x) = Df(x) := \begin{pmatrix} \nabla f_1(x) & \partial_{x_1} f_1(x) & \ldots & \partial_{x_n} f_1(x) \\
\vdots & \vdots & \ddots & \vdots \\
\nabla f_p(x) & \partial_{x_1} f_p(x) & \ldots & \partial_{x_n} f_p(x) \end{pmatrix}
\]
3.4.2 Differentiation Rules & Formulae

### Proposition 3.48.
- For $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^n$, $f'(x) = nx^{n-1}$ ($\forall x \in \mathbb{R}$) ($\forall n \in \mathbb{Q}$)
- For $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto e^x$, $f'(x) = e^x$ ($\forall x \in \mathbb{R}$)
- For $f : \mathbb{R}_{x+} \to \mathbb{R}$, $x \mapsto \ln(x)$, $f'(x) = \frac{1}{x}$ ($\forall x \in \mathbb{R}_{x+}$)

### Proposition 3.49.
- Let $f : U \to \mathbb{R}$ be differentiable and $\lambda \in \mathbb{R}$. Then $\lambda f : U \to \mathbb{R}$ is differentiable and $(\lambda f)'(x) = \lambda f(x)$ ($\forall x \in U$).
- Let $f, g : U \to \mathbb{R}$ be differentiable. Then $f + g : U \to \mathbb{R}$ is differentiable and $(f + g)'(x) = f'(x) + g'(x)$ ($\forall x \in U$).
- Let $f, g : U \to \mathbb{R}$ be differentiable. Then $fg : U \to \mathbb{R}$ is differentiable and $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$ ($\forall x \in U$).
- Let $f, g : U \to \mathbb{R}$ be differentiable and such that ($\forall x \in U$) $g(x) \neq 0$. Then $\frac{1}{g} : U \to \mathbb{R}$ is differentiable and $(\frac{1}{g})'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$ ($\forall x \in U$).
- (Chain Rule.) Let $f : U \to \mathbb{R}$ and $g : f(U) \to \mathbb{R}$ be differentiable. Then $g \circ f : U \to \mathbb{R}$ is differentiable and $(g \circ f)'(x) = g'(f(x))f'(x)$ ($\forall x \in U$).
- Let $f : U \to \mathbb{R}$ be differentiable and strictly monotonic so that $f^{-1} : f(U) \to \mathbb{R}$ is well-defined. Then $f^{-1} : f(U) \to \mathbb{R}$ is differentiable and $f^{-1}'(y) = \frac{1}{f'(f^{-1}(y))}$ where $x$ is such that $f(x) = y$.

### Proposition 3.50. – L'Hospital Rule.
Let $f, g$ be two differentiable functions on $(a, b)$, with $-\infty < a < b < +\infty$, such that $g'(x) \neq 0$ ($\forall x \in (a, b)$). Suppose:

1. $\frac{f'(x)}{g'(x)} \to A \in \mathbb{R}$ as $x \to a$
2. and one of the following
   - $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$
   - $g(x) \to \infty$ as $x \to a$

Then $\frac{f(x)}{g(x)} \to A \in \mathbb{R}$ as $x \to a$.

3.4.3 Continuous Differentiability & Higher-Order Derivatives

### Definition 3.51.
If $f : U \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable on the open set $U$, and if its derivative $f' : U \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable itself, then the derivative of $f'$ is called the second derivative of $f$ and is denoted $f''$. Similarly, taking successively the derivative of each differentiable derivative, we obtain functions $f'' , f''' , f^{(3)}, \ldots , f^{(k)} , f^{(k+1)}, \ldots$. The function $f^{(k)}$ is called the $k^{th}$ derivative of $f$.

### Definition 3.52.
If $f : U \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable on the open set $U$, and if its derivative $f' : U \subseteq \mathbb{R} \to \mathbb{R}$ is continuous, $f$ is said to be continuously differentiable, which is denoted $C^1$.

If $f : U \subseteq \mathbb{R} \to \mathbb{R}$ is $k$ times differentiable on the open set $U$, and if its $k^{th}$ derivative $f^k : U \subseteq \mathbb{R} \to \mathbb{R}$ is continuous, $f$ is said to be $k$ times continuously differentiable, which is denoted $C^k$.

If for all $k$, $f^k$ is continuous and differentiable, $f$ is said to be infinitely continuously differentiable, which is denoted $C^\infty$.

We denote by $C^k(U)$ (resp. $C^\infty(U)$) the set of all real functions that are $C^k$ (resp. $C^\infty$) on $U$.

### Definition 3.53.
Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be twice differentiable with respect to both its arguments. For $(x, y) \in U$, the array

$$
H(f)(x, y) = \begin{pmatrix}
\partial_{x_1}\partial_{x_1}f(x, y) & \partial_{x_2}\partial_{x_1}f(x, y) & \ldots & \partial_{x_n}\partial_{x_1}f(x, y) \\
\partial_{x_1}\partial_{x_2}f(x, y) & \partial_{x_2}\partial_{x_2}f(x, y) & \ldots & \partial_{x_n}\partial_{x_2}f(x, y) \\
\vdots & \vdots & \ddots & \vdots \\
\partial_{x_1}\partial_{x_n}f(x, y) & \partial_{x_2}\partial_{x_n}f(x, y) & \ldots & \partial_{x_n}\partial_{x_n}f(x, y)
\end{pmatrix}
$$

is called the Hessian matrix of $f$ at $(x, y)$.

### Theorem 3.54. – Young’s Theorem.
If $f : U \subseteq \mathbb{R}^2 \to \mathbb{R}$ is $C^2$ on the open set $U$, then $\partial_y \partial_x f(x, y) = \partial_x \partial_y f(x, y)$ $\forall (x, y) \in U$. 


3.5 Integration

3.5.1 Riemann Integration

Definition of the Riemann Integral

Definition 3.55. Consider a non-empty interval $[a, b] \subseteq \mathbb{R}$.

- A partition of $[a, b]$ is a finite collection $\pi := \{x_0, x_1, \ldots, x_n\}$ such that $a = x_0 < x_1 < \cdots < x_n = b$.
- Let $\pi = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$. A refinement of $\pi$ is a partition $\pi' := \{y_0, y_1, \ldots, y_m\}$ of $[a, b]$ such that $\{x_0, x_1, \ldots, x_n\} \subseteq \{y_0, y_1, \ldots, y_m\}$. We also say that $\pi'$ is finer than $\pi$.
- Let $\pi = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$. A selection associated with $\pi$ is a finite collection $\sigma(\pi) = \{\xi_1, \ldots, \xi_n\}$ such that $(\forall k = 1, \ldots, n) \xi_k \in [x_{k-1}, x_k)$.

Definition 3.56. Let $[a, b] \subseteq \mathbb{R}$ and $f : [a, b] \to \mathbb{R}$. The $\pi$-upper Riemann sum of $f$ is

$$S_\pi(f) := \sum_{k=1}^{n} u_{f, \pi}(k)(x_k - x_{k-1})$$

where $u_{f, \pi}(k) := \sup\{f(\xi_k)|\xi_k \in [x_{k-1}, x_k]\}$.

The $\pi$-lower Riemann sum of $f$ is

$$s_\pi(f) := \sum_{k=1}^{n} l_{f, \pi}(k)(x_k - x_{k-1})$$

where $l_{f, \pi}(k) := \inf\{f(\xi_k)|\xi_k \in [x_{k-1}, x_k]\}$.

Remark. For any partition $\pi$, $S_\pi(f) \geq \pi(f)$. Also, as the chosen partition becomes finer, $S_\pi(f)$ decreases (the sups are taken over smaller intervals) while $s_\pi(f)$ increases (the infs are taken over smaller intervals).

Definition 3.57. Let $[a, b] \subseteq \mathbb{R}$ and $f : [a, b] \to \mathbb{R}$. The upper Riemann integral of $f$ is

$$R(f) := \inf\{S_\pi(f)|\pi \text{ is a partition of } [a, b]\}.$$

The lower Riemann integral of $f$ is

$$r(f) := \sup\{s_\pi(f)|\pi \text{ is a partition of } [a, b]\}.$$

Remark. For any bounded function $f : [a, b] \to \mathbb{R}$, $R(f) \geq r(f)$ (see the problem set for a proof).

Definition 3.58. A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if $R(f) = r(f)$. In such case, $I := R(f) = r(f)$ is called the (Riemann) integral of $f$ over $[a, b]$ and denoted $I := \int_{a}^{b} f(x)dx$.

Properties of the Riemann Integral

Proposition 3.59. Any $f \in C[a, b]$ is Riemann integrable.

Proof. If $a = b$, the claim is obvious, so suppose $a < b$ and take $f \in C[a, b]$. To show that $R(f) = r(f)$, we will show that $R(f) = r(f) = 0$, that is, $R(f) - r(f) = 0$. Fix some arbitrary $\varepsilon > 0$. Since $-\infty < a < b < +\infty$, $f$ is uniformly continuous on $[a, b]$. Thus, $(\exists \delta > 0)$ $(\forall t, t' \in [a, b]) |t - t'| < \delta \Rightarrow |f(t) - f(t')| < \frac{\varepsilon}{b-a}$.

Let’s choose any partition $\pi = \{a_0, a_1, \ldots, a_{n}\}$ of $[a, b]$ s.t $|a_k - a_{k-1}| < \delta$ $(\forall k = 1, \ldots, n)$. Then:

$$R_\pi(f) - r_\pi(f) = \sum_{k=1}^{n} (u_{f, \pi}(k) - l_{f, \pi}(k))(a_k - a_{k-1})$$

$$< \sum_{k=1}^{n} \frac{\varepsilon}{b-a}(a_k - a_{k-1})$$

$$= \frac{\varepsilon}{b-a} \left( (a_1 - a_0) + (a_2 - a_1) + \cdots + (a_n - a_{n-1}) \right)$$

$$= \frac{\varepsilon}{b-a} (a_n - a_0) = \frac{\varepsilon}{b-a} (b - a) = \varepsilon$$

Now note that by definition, $R(f) \leq R_\pi(f)$ and $r(f) \geq r_\pi(f)$ $(\forall \pi')$. Therefore, for our partition $\pi$, $R(f) - r(f) \leq R_\pi(f) - r_\pi(f) < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, the claim is proven. □
Proposition 3.60. Let \( f, g, h : [a, b] \to \mathbb{R} \) such that \( f, g \) and \( |h| \) are integrable.

- Linearity. \((\forall \alpha, \beta \in \mathbb{R}) \int_a^b (\alpha f + \beta g) = \alpha \int_a^b f + \beta \int_a^b g \)
- Monotonicity. \( f \leq g \Rightarrow \int_a^b f \leq \int_a^b g \)
- Chasles’ Identity. \((\forall c \in [a, b]) \int_a^b f = \int_a^c f + \int_c^b f \)
- \( \int_a^b f g \leq \left( \int_a^b f \right)^{\frac{1}{2}} \left( \int_a^b g \right)^{\frac{1}{2}} \)
- \(|\int h| \leq \int |h| \)

### 3.5.2 Integration & Differentiation

**Theorem 3.61.** – **Fundamental Theorem of Calculus.** Let \( f \in C[a, b] \). Let \( F : [a, b] \to \mathbb{R} \). Then

\[
F' = f \iff (\forall x \in [a, b]) F(x) = F(a) + \int_a^x f(t) dt
\]

**Definition 3.62.** An antiderivative of a function \( f : [a, b] \to \mathbb{R} \) is a function \( F : [a, b] \to \mathbb{R} \) such that \((\forall x \in (a, b)) F'(x) = f(x) \).

**Corollary 3.63.** – **Integration by Parts.** For \( f, g \in C^1[a, b] \),

\[
\int_a^b f(x)g'(x)dx = \left[ f(x)g(x) \right]_a^b - \int_a^b f'(x)g(x)dx
\]

where \([f(x)g(x)]_a^b := f(b)g(b) - f(a)g(a)\).

\[\text{Proof.} \quad \text{Note first that } f, g, f', g' \in C[a, b] \text{ are all integrable on } [a, b], \text{ and so is any product of them. Seeing that } (fg)'(x) = f'(x)g(x) + f(x)g'(x) \text{ the claim follows immediately from the Fundamental Theorem of Calculus.} \]

**Proposition 3.64.** – **Leibniz’s Rule.** Let \( f : [a, b] \times [c, d] \to \mathbb{R} \) be continuous and such that \( \partial_x f(u; x) : [a, b] \times [c, d] \to \mathbb{R} \) exists and is continuous. Let \( \alpha, \beta : [c, d] \to [a, b] \) be two differentiable functions.

\[
\frac{d}{dx} \int_{\alpha(x)}^{\beta(x)} f(u; x) du = \beta'(x)f(\beta(x); x) - \alpha'(x)f(\alpha(x); x) + \int_{\alpha(x)}^{\beta(x)} \partial_x f(u; x) du
\]

### 3.5.3 Multivariate Integration

Above, we have defined integration for real functions of one variable. Defining integration over sets of higher dimensions proceed in a comparable way. Here we formally define integrals over sets of dimension 2; based on this, the definition of integrals over sets of dimension \( \geq 2 \) would be straightforward.

**Defining a double integral over a rectangle.**

**Theorem 3.65.** A (closed) rectangle of \( \mathbb{R}^n \) is the Cartesian product of \( n \) (closed) bounded intervals of \( \mathbb{R} \):

\( R := \prod_{i=1}^n [a_i, b_i] \).

**Definition 3.66.** Let \( R \) be a rectangle of \( \mathbb{R}^2 \). A partition \( \mathcal{A} \) of \( R \) is a finite collection \( \{A_1, A_2, \ldots, A_n\} \) of rectangles of \( \mathbb{R}^2 \) such that \( \bigcup_{i=1}^n A_i = R \) and \((\forall i, j = 1, \ldots, n) i \neq j \) \( A_i \cap A_j = \emptyset \).

Given a partition \( \mathcal{A} = \{A_1, A_2, \ldots, A_n\} \) of \( R \), a selection associated with \( \mathcal{A} \) is \( \sigma(\mathcal{A}) = \{(u_1, v_1), \ldots, (u_n, v_n)\} \) such that \((\forall k = 1, \ldots, n) (u_i, v_i) \in A_k \).

**Definition 3.67.** Let \( f : R \to \mathbb{R} \) be a function defined on a rectangle \( R \) of \( \mathbb{R}^2 \). Given a partition \( \mathcal{A} = \{A_1, A_2, \ldots, A_n\} \) of \( R \) and a selection \( \sigma \) associated with \( \mathcal{A} \), the Riemann sum of \( f \) is \( \sum_{i=1}^n f(u_i, v_i) \Delta(A_i) \) where \( \Delta(A_i) = (b_1^i - a_1^i)(b_2^i - a_2^i) \) is the area of the rectangle \( A_i = [a_1^i, b_1^i] \times [a_2^i, b_2^i] \).
**Definition 3.68.** A function \( f : R \to \mathbb{R} \) defined on a rectangle \( R \) of \( \mathbb{R}^2 \) is integrable over \( R \) if the following limit exists:

\[
\lim_{n \to \infty} \text{ and } \lim_{(v_i)\to 0} \sum_{i=1}^{n} f(u_i,v_i)\Delta(A_i)
\]

In such case, the limit is denoted \( \iint_{R} f(x,y)d(x,y) \) and called the integral of \( f \) over \( R \).

**Defining a double integral over a general subset of \( \mathbb{R}^2 \).**

**Definition 3.69.** Let \( f : D \subset \mathbb{R}^2 \to \mathbb{R} \). \( f \) is said to be integrable over \( D \) if there exists some rectangle \( R \supset D \) such that the function

\[
F : R \to \mathbb{R}, (x,y) \mapsto \begin{cases} f(x,y) & \text{if } (x,y) \in D \\ 0 & \text{otherwise} \end{cases}
\]

is integrable over \( R \). In such case, the integral of \( f \) over \( D \) is denoted \( \iint_{D} f(x,y)d(x,y) \) and defined as \( \iint_{D} f(x,y)d(x,y) := \iint_{R} F(x,y)d(x,y) \).

**Computing multiple integrals in practice.**

**Theorem 3.70. – Fubini’s Theorem.** (i) Rectangular domain. Let \( D := [a,b] \times [c,d] \subset \mathbb{R}^2 \), \( f : D \to \mathbb{R} \) be a continuous function. Then

\[
\iint_{D} f(x,y)d(x,y) = \int_{a}^{b} \int_{c}^{d} f(x,y)dy dx = \int_{c}^{d} \int_{a}^{b} f(x,y)dx dy
\]

(ii) Non-rectangular domain. Let \( D \) be the domain of a continuous function \( f : D \to \mathbb{R} \). Suppose there are continuous functions \( g_1, g_2 : [a,b] \to \mathbb{R} \) and \( h_1, h_2 : [c,d] \to \mathbb{R} \) such that \( D = \{(x,y)|x \in [a,b], y \in [g_1(x),g_2(x)]\} \cup \{(x,y)|y \in [c,d], x \in [h_1(y),h_2(y)]\} \). Then

\[
\iint_{D} f(x,y)d(x,y) = \int_{a}^{b} \int_{g_1(x)}^{g_2(x)} f(x,y)dy dx = \int_{c}^{d} \int_{h_1(y)}^{h_2(y)} f(x,y)dx dy
\]

**Remark.** Fubini’s theorem says that double integrals of continuous bivariate functions can be computed as two successive univariate integrals, and that the order of integration does not matter. It can be generalized to higher dimensions: the \( n \)-dimensional integral of a continuous function \( f : D \subset \mathbb{R}^n \to \mathbb{R} \) can be computed as \( n \) successive univariate integrals, as above, and the order of integration does not matter. Remember to be careful with the bounds of integration when changing the order of integration, though.

**3.5.4 Change of Variable**

**Theorem 3.71. – The Inverse Function Theorem.** (i) Let \( \phi : \mathbb{R} \to \mathbb{R} \) be \( C^1 \) on some open set of \( \mathbb{R} \) containing \( a \). Suppose \( \phi'(a) \neq 0 \). There exists an open set \( V \) containing \( a \) and an open set \( W \) containing \( \phi(a) \) such that \( \phi : V \to W \) has a continuous inverse \( \phi^{-1} : W \to V \) which is differentiable on \( W \). Also, \( \forall y \in W \) \( \phi^{-1}(y) = \frac{1}{\phi'(\phi^{-1}(y))} \).

(ii) Let \( \phi : \mathbb{R}^n \to \mathbb{R}^n \) be \( C^1 \) on some open set of \( \mathbb{R}^n \) containing \( a \). Suppose \( \det[D\phi(a)] \neq 0 \). There exists an open set \( V \) containing \( a \) and an open set \( W \) containing \( \phi(a) \) such that \( \phi : V \to W \) has a continuous inverse \( \phi^{-1} : W \to V \) which is differentiable on \( W \). Also, \( \forall y \in W \) \( D(\phi^{-1})(y) = \frac{1}{\det[D\phi(\phi^{-1}(y))]}. \)

**Corollary 3.72. – Change of variable.** (i) Let \( f : U \subset \mathbb{R} \to \mathbb{R} \) be a continuous function. Let \( \phi : U \to \mathbb{R}, x \mapsto u \) be a continuous transformation so that \( \phi^{-1} : \phi(U) \subset \mathbb{R} \to U, u \mapsto x \) exists and is differentiable. Then,

\[
\int_{U} f(x)dx = \int_{\phi^{-1}(U)} f \circ \phi^{-1}(u)|\phi^{-1}(u)|du
\]
(ii) Let \( f : U \subseteq \mathbb{R}^n \to \mathbb{R} \) be a continuous function. Let \( \phi : U \to \mathbb{R}^n \), \( x = (x_i)_i \mapsto u = (u_i)_i \) be a continuous transformation so that \( \phi^{-1} : \phi(U) \subseteq \mathbb{R}^n \to U \), \( u \mapsto x \) exists and is differentiable. Then,
\[
\int_U f(x)dx = \int_{\phi^{-1}(U)} f \circ \phi^{-1}(u) \det[J\phi^{-1}(u)]du
\]

4 Matrix Algebra

4.1 Basic Definitions & Operations

Definition 4.1. A matrix is a rectangular array of elements (or entries). The size of a matrix \( A \) is a pair \( (m, n) \in \mathbb{N}^* \times \mathbb{N}^* \) where \( m \) is the number of rows of \( A \) and \( n \) the number of columns of \( A \).

We will focus on matrices whose elements are real numbers. Given the field structure of \( \mathbb{R} \), we can define operations on real matrices as follows.

Definition 4.2. – Matrix Operations.

- **Transpose.** Let \( A \) be \( m \times n \). \( A' \) is the \( (n \times m) \)-matrix such that \((A')_{ij} := A_{ji}\).
- **Multiplication by a scalar.** Let \( A \) be \( m \times n \), and \( \lambda \) a scalar. \((\lambda A)\) is the \( (m \times n) \)-matrix such that \((\lambda A)_{ij} := \lambda \cdot A_{ij}\).
- **Addition.** Let \( A, B \) be \( (m \times n) \)-matrices. \( A + B \) is the \( (m \times n) \)-matrix such that \((A + B)_{ij} := A_{ij} + B_{ij}\).
- **Multiplication.** Let \( A \) be \( m \times n \) and \( B \) be \( n \times p \). \( AB \) is the \( (m \times p) \)-matrix such that \((AB)_{ij} := \sum_{k=1}^n A_{ik}B_{kj} \), i.e. the scalar product of the \( n \)-vectors \( A_i \) and \( B_j \).

Remark. Note that any two matrices cannot be added or multiplied: they must be conformable. To be added, two matrices must have same size, i.e. same number of row and same number of columns. To be multiplied, the number of columns of the left matrix must equal the number of rows of the right matrix.

Proposition 4.3. – Properties of matrix operations. Assuming all left-hand side operations are well-defined:

1. \((A')' = A\)
2. \(B + A = A + B\)
3. \((A + B)' = A' + B'\)
4. \((AB)' = B'A'\)
5. \(A(B + C) = AB + AC\)
6. \((B + C)A = BA + CA\)
7. \(\lambda(B + C) = \lambda B + \lambda C = (B + C)\lambda\)
8. \(\lambda(AB) = \lambda(A)B = A(\lambda B) = (AB)\lambda\)

Proof. Assuming all left-hand side operations above are well-defined, let \((i, j) \in \{1, \ldots, m\} \times \{1, \ldots, n\} \)

1. \((A')'_{ij} = A'_{ji} = A_{ij}\)
2. \((B + A)_{ij} = B_{ij} + B_{ij} = B_{ij} + B_{ij} = (A + B)_{ij}\) where the second equality holds because of the commutativity of + of the underlying field of scalars.
3. \((A + B)'_{ij} = (A + B)_{ji} = (A_{ij} + B_{ij}) = A'_{ij} + B'_{ij}\)
4. \((AB)'_{ij} = (AB)_{ji} = \sum_{k=1}^n A_{jk}B_{ki} = \sum_{k=1}^n A_{ik}B'_{kj} = \sum_{k=1}^n B'_{ik}A'_{kj} = B'A'\) where the third equality holds because of the commutativity of \( \cdot \) on the underlying field of scalars.
5. \(A(B + C) = \sum_{k=1}^n A_{ik}(B_{kj} + C_{kj}) = \sum_{k=1}^n (A_{ik}B_{kj} + A_{ik}C_{kj}) = AB + AC\) where the second equality holds because of the distributivity of \( \cdot \) with respect to + on the underlying field of scalars, and the fourth equality by the commutativity and associativity of +.
6. \((B + C)A = \sum_{k=1}^n (B_{ik} + C_{ik})A_{kj} = \sum_{k=1}^n (B_{ik}A_{kj} + C_{ik}A_{kj}) = BA + CA\) where the second equality holds because of the distributivity of \( \cdot \) with respect to + on the underlying field of scalars, and the fourth equality by the commutativity and associativity of +.
7. \((\lambda(B+C))_{ij} = \lambda(B_{ij} + C_{ij}) = \lambda B_{ij} + \lambda C_{ij} = (\lambda B)_{ij} + (\lambda C)_{ij} = B_{ij} \lambda + C_{ij} \lambda = (B_{ij} + C_{ij}) \lambda = (B+C)_{ij} \lambda\)
where the second and fifth equalities hold because of the distributivity of \(\cdot\) with respect to + on the underlying field of scalars and the fourth by the commutativity of \(\cdot\) on that field.

8. \(\lambda(AB))_{ij} = \lambda((AB)_{ij}) = \lambda \sum_{k=1}^{n} A_{ik} B_{kj} = \sum_{k=1}^{n} (\lambda A_{ik}) B_{kj} = \sum_{k=1}^{n} (\lambda A_{ik}) \lambda B_{kj}\) where the third equality holds by the distributivity of \(\cdot\) with respect to + and the associativity of \(\cdot\) on the underlying field of scalars. This proved the first equality of the result. Similarly, for the second: \(\lambda(AB))_{ij} = \lambda((AB)_{ij}) = \lambda \sum_{k=1}^{n} A_{ik} B_{kj} = \sum_{k=1}^{n} A_{ik} (\lambda B_{kj}) = \sum_{k=1}^{n} A_{ik} (\lambda B_{kj}).\) Finally for the third: \(\lambda(AB))_{ij} = \lambda((AB)_{ij}) = \lambda \sum_{k=1}^{n} A_{ik} B_{kj} = (\sum_{k=1}^{n} A_{ik} B_{kj}) \lambda = ((AB) \lambda)_{ij}.

\(\square\)

Remark. Note that matrix multiplication is not commutative.

First, note that \(AB\) being well-defined does not guarantee that \(BA\) will be well-defined. Indeed, \(AB\) being well-defined requires \(A\) to be \(m \times n\) and \(B\) to be \(n \times p\). If \(m \neq p\) then \(BA\) is not well-defined. Suppose that \(m = p\), so that both products make sense. Note then that \(AB\) and \(BA\) may not have the same size, in which case they cannot be equal. Indeed, \(AB\) is \(m \times m\) \((m \times p\) with \(p = m\) while \(BA\) is \(n \times n\). For the two products to have same size, we thus need \(m = n\), that is \(A, B\) are both square matrices \(n \times n\). Suppose that is the case. \(AB\) and \(BA\) are equal if \((\forall (i, j) \in \{1, \ldots, n\}^2\) \((AB)_{ij} = (BA)_{ij}\). Note that \((AB)_{ij} := \sum_{k=1}^{n} A_{ik} B_{kj}\) while \((BA)_{ij} := \sum_{k=1}^{n} B_{ik} A_{kj}\). In general, these two things are not equal, simple because \((A_{i1}, B_{1j})\) and \((B_{i1}, A_{1j})\) are not the same pair of vectors. The best way to prove that the statement “\(AB\) and \(BA\) are equal when \(A, B\) are both \(n \times n\)” does not hold is to provide a counter-example: with \(n = 2\), consider:

\[
A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}, \quad \text{then } AB = \begin{pmatrix} 10 & 13 \\ 22 & 29 \end{pmatrix}, \quad \text{while } BA = \begin{pmatrix} 11 & 12 \\ 19 & 28 \end{pmatrix}
\]

Remark. Subtraction of matrices is straightforwardly defined from addition and multiplication by a scalar: for any two \((m \times n)\)-matrices \(A, B\), \(A - B := A + (-1) \cdot B\).

Things are not that straightforward when it comes to division. First, just as in the real numbers where dividing by 0 is not defined, we cannot divide by any matrix. Seeing division as a multiplication by the inverse, we can divide by a given matrix if that matrix has an inverse. In the Linear Algebra section of the class, we will come back to the properties a matrix must satisfy to be invertible. As a second note, since multiplication is not commutative, we should expect something similar to happen with division: taking two matrices \(A, B\) such that \(A^{-1}B\) and \(BA^{-1}\) both exist, we should not expect them to be equal in general.

Definition 4.4. The trace of an \((n \times n)\)-matrix is the sum of its diagonal terms: \(tr(A) := \sum_{i=1}^{n} A_{ii}\).

Proposition 4.5. Let \(A, B, C\) be matrices of sizes \((m \times n)\), \((n \times p)\) and \((p \times m)\), respectively. \(tr(ABC) = tr(BCA) = tr(CAB)\).

Definition 4.6. • An \((n \times n)\)-matrix \(A\) is symmetric if \((\forall i, j \in \{1, \ldots, n\}) A_{ij} = A_{ji}.

• An \((n \times n)\)-matrix \(A\) is diagonal if \((\forall i, j \in \{1, \ldots, n\}) i \neq j \Rightarrow A_{ij} = 0.

• An \((n \times n)\)-matrix \(A\) is lower triangular (resp. upper triangular) if \((\forall i, j \in \{1, \ldots, n\}) i > j \Rightarrow A_{ij} = 0 \) (resp. \(i < j \Rightarrow A_{ij} = 0\).

• An \((n \times n)\)-matrix \(A\) is idempotent if \(AA = A\).

Remark. These definitions may sound a little abstract for now. In the Linear Algebra section of the class, we’ll go back to the properties and uses of such matrices.

4.2 Matrix Form of a Linear System

Given the definition of addition and multiplication of matrices, and seeing (column-)vectors as matrices with a single column, we can rewrite a linear system of \(m\) equations and \(n\) unknowns as:

\[
\begin{pmatrix}
    a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n &=& b_1 \\
    a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n &=& b_2 \\
    \vdots & \vdots & \vdots \\
    a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n &=& b_m
\end{pmatrix}
\]
Given a \((m \times n)\) matrix \(A\) of coefficients for the left-hand side and a \(m\)-vector \(b\) of coefficients for the right-hand side, such system is said to have a solution if and only if there exists some \(n\)-vector \(x\) that satisfies the \(m\) equations of the system. We know that:

- If \(m > n\), i.e. if there are more equations than unknowns (and if no equation is redundant given the others), then the system has no solution.
- If \(m < n\), i.e. if there are more unknowns than equations, then the system has multiple solutions.
- If \(m = n\), i.e. if there are exactly as many equations as unknowns (and if no equation is redundant given the others), then the system has a unique solution.

The conditions on \(m\) and \(n\) translates into (necessary) size conditions on the matrix \(A\) for the matrix-form system \(Ax = b\) to have a solution. As we will see in the Linear Algebra section of the class, the requirement on non-redundancy of the equations of the system will translate into a requirement on the relationship existing between the columns of \(A\). In the case where \(m = n\), the non-redundancy requirement will in turn translate into a condition for the square matrix \(A\) to be invertible so that the expression \(x = A^{-1}b\) makes sense and actually gives the unique solution \(x\) to the system.

References

5 Problem Set

Problem 1. Show that \( \mathbb{R} \) equipped with the usual addition and multiplication is a field.

Problem 2. (i) Let \( \mathbb{Q} \) be the set of all rational numbers. Show that \( \mathbb{Q} \) is countable.

(ii) Let \( \mathbb{R} \setminus \mathbb{Q} \) be the set of all irrational numbers. Show that \( \mathbb{R} \setminus \mathbb{Q} \) is not countable.

(iii) Let \( \mathbb{N} \) be the set of all natural numbers. Show that \( \mathbb{N} \) is countable.

Problem 3. Let \( x > -1 \) and \( n \in \mathbb{N} \). Prove Bernoulli’s inequality: \( (1 + x)^n \geq 1 + nx \).

Problem 4. We can define a distance (or metric) on a space \( \mathcal{F} \), a function \( d : \mathcal{F} \times \mathcal{F} \to \mathbb{R} \) that satisfies:

1. \((\forall x,y \in \mathcal{F}) \ d(x,y) = 0 \iff x = y \)
2. \((\forall x,y \in \mathcal{F}) \ d(x,y) = d(y,x) \)
3. \((\forall x,y,z \in \mathcal{F}) \ d(x,y) \leq d(x,z) + d(z,y) \)

Prove that

(i) \( d_1 : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \), \((x,y) \mapsto |x-y|\) is a distance on \( \mathbb{R} \).

(ii) \( d_2 : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R} \), \((x,y) \mapsto \left( \sum_{i=1}^{k} (x_i - y_i)^2 \right)^{1/2} \) is a distance on \( \mathbb{R}^k \).

Remark. Property 3. defining a distance is called the triangle inequality. It is used very often.

Problem 5. Prove the Archimedean property of the real numbers and the denseness of \( \mathbb{Q} \) in \( \mathbb{R} \).

Problem 6. Note that a sequence of \( \mathbb{R}^k \) is \((x_n)_n\) where \((\forall n \in \mathbb{N}) \ x_n = (x_{n,1}, \ldots, x_{n,k})' \) with \( x_{n,j} \in \mathbb{R} \) (\( \forall j = 1, \ldots, k \)). We say that a sequence \((x_n)_n\) of \( \mathbb{R}^k \), converges if

\[
(\exists x \in \mathbb{R}^k)(\forall \varepsilon > 0) (\exists N_x \in \mathbb{N})(\forall n > N_x) \ ||x_n - x|| < \varepsilon
\]

where \( ||x - y|| := \left( \sum_{i=1}^{k} (x_i - y_i)^2 \right)^{1/2} \).

Show that a sequence of \( \mathbb{R}^k \), \((x_n)_n\) converges if and only if \((\forall j = 1, \ldots, k)\), the sequence of \( \mathbb{R} \), \((x_{n,j})_n\) converges.

Problem 7. Sometimes we are interested in sequences, not of real numbers, but of real-valued functions. Consider the sequence \( \{f_n\} \), with \( f_n : X \to \mathbb{R} \)

- The sequence \( \{f_n\} \) is said to converge pointwise to the function \( f : X \to \mathbb{R} \) if \( (\forall x \in X) \ \lim f_n(x) = f(x) \).

- The sequence \( \{f_n\} \) is said to converge uniformly to the function \( f \) if \( (\forall \varepsilon > 0) (\exists N \in \mathbb{N})(\forall n > N) (\forall x \in X) ||f(x) - f_n(x)|| < \varepsilon \).

Let \( I \) be an interval, and let \( f_n : I \to \mathbb{R} \) (\( \forall n \in \mathbb{N} \)). Let \( x \in I \). Suppose that \((\forall n \in \mathbb{N}) f_n \) is continuous at \( x \), and that \( \{f_n\} \) converges uniformly on \( I \) to \( f \). Show that if \( \{x_n\} \) is a sequence in \( I \) that converges to \( x \), then \( f_n(x_n) \to f(x) \).

Problem 8. Prove a sequence in \( \mathbb{R} \) is Cauchy if and only if it is convergent.

Problem 9. (i) Let \( U \) be an open interval of \( \mathbb{R} \). Show that \((\forall x \in U)(\exists \varepsilon > 0)(\forall x' \in \mathbb{R}) ||x' - x|| < \varepsilon \Rightarrow x' \in U \).

(ii) Let \( U = \prod_{i=1}^{n} U_i \) where \((\forall i = \{1, \ldots, n\}) U_i \) is an open interval of \( \mathbb{R} \). Show that \((\forall x \in U)(\exists \varepsilon > 0)(\forall x' \in \mathbb{R}^n) ||x' - x|| < \varepsilon \Rightarrow x' \in U \).

Problem 10. Consider \( f : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R} \), \((\mu, \sigma^2) \mapsto -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \). Justify that \( f \) is twice differentiable and derive its Hessian matrix.

\(^2\text{Hint. Check out Minkowski’s inequality.}\)
Problem 11. Going back to the definition of the Riemann integral, prove that for any function $f : [a, b] \to \mathbb{R}$, $R(f) \geq r(f)$.

Problem 12. Evaluate the following:

1. $\int_{-\infty}^{+\infty} x^2 e^{3x} \, dx$

2. $\int_D xy^2 \, d(x,y)$ where $D := \{(x,y) \in \mathbb{R} \mid x \in [0, 2], \ y \in [0, \frac{4}{7}]\}$. Proceed twice, using both orders of integration.

3. $\int_0^3 2x(x^2 + 1)^3 \, dx$. Use the change of variable $u = x^2 + 1$.

4. $\partial_r V(s,r)$ and $\partial_s V(s,r)$ where $(\forall (s,r) \in \mathbb{R}_+ \times \mathbb{R}_+) \ V(s,r) = \int_s^T f(t)e^{-(t-s)r} \, dt$, with $f : \mathbb{R}_+ \to \mathbb{R}$ and $T \in \mathbb{R}_+$