Math Boot Camp Notes
For Incoming SSE MSc Econ Students

Nils Lager

Stockholm School of Economics
nils.lager@hhs.se

Latest edit: May 5, 2026

Contents

Purpose of these notes

These notes are written for incoming SSE MSc Economics students as an introduction to some mathematical concepts and writing formal proofs during the summer. Before arriving at SSE, you are expected to have carefully read through these notes and worked with all the exercises, but not necessarily *master* this material.

The notes are mainly written as preparation for the course Math 5301, taught by Associate Professor Mark Voorneveld, but will also be useful for your other Term 1 courses. They are heavily inspired by excellent notes provided to me by my predecessors Johan Orrenius and Emil Bustos, along with Mark's course notes. Any remaining errors are my own.

Chapter 1 gives you an introduction to the course. Chapters 2 presents what constitutes a proof. Chapters 3–4 cover what constitutes a set and the canonical sets of numbers. While mastering formal proofs will take time, the underlying concepts should already be familiar to you. Chapters 5–11 introduce more advanced topics that you are not expected to master yet, but you should start to get a feel for.

Read the notes, try to make sense of the concepts, and attempt to solve the exercises. If you get confused, consult your favourite LLM for further help! None of the proofs in the suggested solutions are intended for memorisation – they are given to help you practice and understand how complete and rigorous arguments are built.

1 Introduction

Many incoming graduate students dread their first math course, expecting endless algebra or memorising formulas. But Math 5301 is different: it's about constructing clear, rigorous arguments from given definitions and results, just like in economic theory. The course mirrors foundations taught at top US PhD programs and teaches:

- **Mathematical Language**: To read and write advanced economic theory/models.
- **Logic of Economic Theory**: To see how definitions, theorems, and lemmas are built into arguments (of logical deduction) in a proof.
- **Mathematical Concepts**: To understand *why* results that underpin economic theory hold — not just *that* they hold — and how to prove them yourself.

Why do I need to know this?

If you lack interest in theory work or a PhD, you may wonder why this matters in an era dominated by AI and data. Even if you never write a formal proof again, clear thinking remains essential. Mathematical training is like lifting weights: the point of lifting weights at the gym is not to get good at weight-lifting, but to build strength for other parts of life. Likewise, math sharpens your ability to think, define problems, and follow arguments—whether in academic theory or real-world decisions.

What is an economic model?

Understanding economic models is one of the main reasons we study mathematical reasoning in this course. Purpose of economic models can roughly be divided as:

- **Pure theory models** — Built without data, focusing on internal logic and derivations from assumptions. They are the backbone of economic theory.
- **Structural models** — Used to estimate specific parameters in a fully specified economic model, often to test theories or simulate policy. Common in macro, IO, and the research-frontier of applied micro.
- **Stylized models** — Along with causal estimates using econometric techniques like regression discontinuity or difference-in-difference regressions, economists often include a *theoretical framework*. These are simple setups designed to illustrate mechanisms and are not used directly in the empirical specification. For example, Dal Bó et al. (ReStud, 2023) combine causal estimates from gender quotas with a model showing when quotas may improve political selection.

- **Econometric models** — Empirical models built to estimate relationships between economic variables using data. Technically, most econometrics takes place in *probability spaces* with random variables and expectations, which is beyond the scope of this course. Still, many mathematical concepts covered here are essential for understanding their structure.

On mastering mathematics

Solutions to the exercises are provided at the end of these notes, but I **strongly** encourage you to try each problem yourself first. Learning to write proofs is like learning to hammer a nail. A physics professor might understand the forces, angles, and materials involved—but still struggle to hit a nail straight. A carpenter, by contrast, becomes skilled through practice: missing, adjusting, and trying again until it becomes second nature.

Math works the same way. It's not just a science—it's an *art*. You learn by doing: struggling, failing, and improving. Reading suggested solutions can help *after* you've tried, but passive reading alone won't build skill.

2 What are Proofs?

2.1 Deduction

Imagine sitting by a pond, trying to predict the colour of the next swan to swim by.

Premise: All observed swans have been white.
Conclusion: The next swan will be white.

This is an example of **induction** — a pattern-based guess, not a logical guarantee.

Example 2.1 (Induction)**.** ***Premise****: In a large number of empirical settings, raising the marginal labour-income tax rate has caused reductions in hours worked.* ***Conclusion****: Raising the labour-income tax rate reduces labour supply in any setting.*

Now instead, consider the logical step:

Premise: All swans are white.
Conclusion: The next swan will be white.

This is **deduction** — if the premise is true, the conclusion must follow. In this course, we are concerned with deduction — understanding how to derive logically valid conclusions from clearly stated assumptions.

Example 2.2 (Deduction)**.** ***Premise****: A rational agent chooses leisure* $\ell \in [0, T]$ *to maximise quasi-linear utility* $u(c, \ell) = c + v(\ell)$*, with* v *strictly concave and increasing, subject to* $c = w(1 - \tau)(T - \ell)$*, where* $w > 0$ *is the wage and* $\tau \in [0, 1)$ *is the tax rate.* ***Conclusion****: Under above assumptions, optimal labour supply* $h^*(\tau) = T - \ell^*(\tau)$ *is strictly decreasing in* τ*, i.e. raising the labour-income tax rate reduces labour supply.*

2.2 Logical Connectives

One can build statements from simpler parts using **logical connectives**. In particular, let p and q be two arbitrary *propositions* (statements), regarding some number x, that can be either true or false. Define p as the proposition "$x > 0$".

- *Negation* ($\neg p$): This means "not p". The proposition $\neg p$ is true exactly when p is false. **Example**: $\neg p$ is the proposition "$x \leqslant 0$".
- *Conjunction* ($p \wedge q$): This means "p and q are both true". The conjunction is true only when *both* p and q are true. **Example**: Define q as "$x < 1$". Then $p \wedge q$ is the proposition "$0 < x < 1$".

- *Disjunction* ($p \vee q$): This means "p is true or q is true (or both)". It is false only when *both* p and q are false. **Example**: Define q as "$x < -1$". Then the proposition $p \vee q$ means "x is not in the interval $[-1, 0]$".

Unless explicitly stated otherwise, mathematicians use an "or". The statement "p or q is true" means that *at least* one of p and q is true, not that *exactly* one is true!

In this course, you will often see statements regarding *implications*: "if p then ", denoted $p \Rightarrow q$. This is declared to be false when p is true, but q is false, and is declared to be true in all other cases. For example, if p is the proposition "$x > 0$" and q is "$x^2 > 0$", then the implication $p \Rightarrow q$ is true. Note that this does not say anything about q if p is not true, or whether p is true or not.

In addition, you will encounter *equivalence* statements ($p \Leftrightarrow q$) like "p if and only if q". An equivalence statement is true when p and q have the same truth value (both true or both false). It can also be described as $(p \Rightarrow q) \wedge (q \Rightarrow p)$ being true. For example, $x^2 = 0 \Leftrightarrow x = 0$.

Finally, throughout this course, we will take as given for proofs by *contrapositive*:

Proposition 2.1.

$$p \Rightarrow q \quad \Leftrightarrow \quad \neg q \Rightarrow \neg p$$

Consider the previous example of $p : x > 0$ and $q : x^2 > 0$, so $p \Rightarrow q$. Then it follows that $\neg q \Rightarrow \neg p$, i.e. $x^2 \leqslant 0 \Rightarrow x \leqslant 0$. However, note that $x \leqslant 0 \not\Rightarrow x^2 \leqslant 0$.

Example 2.3 (Necessity and Sufficiency)**.** *Consider the statement "If it rains, then the street gets wet". Let R denote the proposition "it is raining" and W the proposition "the street is wet". Then the implication can be written as:*

$$R \Rightarrow W.$$

*That is, rain is a **sufficient** condition for the street to be wet. Using contrapositive logic, we can also write:*

$$\neg W \Rightarrow \neg R.$$

*In words: if the street is not wet, then it did not rain. However, rain is not a **necessary** condition for the street being wet:*

$$\neg R \not\Rightarrow \neg W.$$

In words: the street can be wet even if it didn't rain (e.g. a sprinkler was on).

2.3 How to prove something

Logical connectives guide how we combine assumptions and conclusions in proofs:

Direct Proof: To *prove* an implication ($p \Rightarrow q$), we assume p and show that q follows. To *disprove* an implication, we find a case where p is true but q is false, i.e. we find a counter-example. To prove an equivalence ($p \Leftrightarrow q$), we prove both $p \Rightarrow q$ and $q \Rightarrow p$.

As you will see throughout this course, proving that a statement holds is a lot more difficult than proving that a statement does not hold. To disprove a statement, you need to find a single counter-example where it does not hold. To prove a statement, you need to show that there does not exist a single counter-example. You can see this asymmetry clearly in **Exercise 9.1**.

Proof by Contradiction: If we want to show that p is true, it is the same as showing $\neg p$ is false. We assume p is false and then show that we reach a contradiction somewhere and then deduce that p must be true. For a classic example, see **Exercise 4.1**.

Proof by Induction: Assume we have a sequence of statements $p(n)$ with $n \in \mathbb{N}$. Then, if we show that (i) $p(1)$ is true, and (ii) $p(n+1)$ is true whenever $p(n)$ is true, then we have shown that $p(n)$ is true for all n.

Exercise 2.1. *Show that for all integers* $n \geqslant 1$,

$$2 + 4 + 6 + \cdots + 2n = n(n+1)$$

2.4 What to Prove?

Proving a proposition is much like making a dish, such as lasagna. If you had to run back and forth to the store each time you needed the next ingredient, the cooking would be chaotic and exhausting. Instead, you first gather all the necessary ingredients that are likely to be useful. Only then do you start assembling the dish, carefully layering the pieces together.

In our examples above, we have not necessarily gone into issues like "what exactly is a *number*?", "what is a *swan*?", "what is the colour *white*?". To make formal proofs, all such issues of definition must be resolved before starting to prove e.g. an implication.

The rest of these notes will familiarise you with mathematical definitions and theorems (the ingredients), and you will use the exercises to practice how to use clearly defined propositions and premises to draw conclusions (making the dish). Good luck!

3 What are Sets?

In this section, we learn about sets. This section draws on *Appendix A* in Mark's notes.

3.1 Collection of elements

Sets are (unordered) collections of *elements*, for example, $\{1, 2, 3\}$ or $\{a, b\}$. The elements can be all sorts of things, such as numbers, functions, names, or other sets. When we describe sets, we either list the elements directly, such as in the examples above, or define a condition for an element to be part of a set (see below).

3.2 Set Operations

Now that we have defined sets, we can start discussing relations between sets. Conceptually, we can look at parts of sets, join sets together and look at the parts sets have in common. First, we need some more notation.

- To say that element x belongs to the set X, we write $x \in X$. If an element x is not in X, we write $x \notin X$.
- We denote "for all elements in set X" as $\forall x \in X$.
- We denote "there exists some element in X" as $\exists x \in X$.

Now, consider two arbitrary sets A and B.

Definition 3.1 (Subset)**.** *If every element in A also belongs to B, then A is a **subset** of B. Formally, if $x \in A \Rightarrow x \in B$, then $A \subseteq B$.*

Definition 3.2 (Proper subset)**.** *If $A \subseteq B$ but $B \nsubseteq A$, then A is a **proper** subset of B. Formally, if $x \in A \Rightarrow x \in B$ and $\exists y \in B$ such that $y \notin A$, then $A \subset B$.*

Definition 3.3 (Identical sets)**.** *If each element in $A \subseteq B$ and $B \subseteq A$, then A and B are **identical**. Formally, if $x \in A \Rightarrow x \in B$ and $x \in B \Rightarrow x \in A$, then $A = B$.*

Example 3.1. *Denote the set of all students at SSE as X. Denote the set of students in Year 1 MSc Economics as E. In particular, we can define (using the definition operator $\equiv$) the subset as*

$$E \equiv \{x \in X : x \text{ is Year 1 MSc Econ}\} \subset X \tag{3.1}$$

*meaning "E is a set of every SSE student **such that** the student is a Year 1 MSc Econ student".*

The set of elements that are in A or B (or both!) is called the *union* of A and B, written as $A \cup B$. Likewise, if we have n sets $A_1, \ldots, A_n$, their union $A_1 \cup \ldots \cup A_n$ consists of all elements that belong to *any* of these sets. This may be written $\cup_{i=1}^n A_i$. Equivalently, we may define the set $I \equiv \{1, \ldots, n\}$ and write the union as $\cup_{i \in I} A_i$.

Example 3.2. *Denote the set of students at different programs at SSE as $D_1, \ldots, D_n$. Then $X = \cup_{i=1}^n D_i$.*

Example 3.3. *Denote some set of real numbers on the real interval as $D_i \equiv (-i, i)$, so e.g. $D_1 = (-1, 1)$. The union of all such intervals is $\cup_{i \in \mathbb{N}} D_i = (-\infty, \infty)$.*

The set of elements that are in A and in B is called the *intersection* of A and B, written as $A \cap B$. Likewise, if we have n sets $A_1, \ldots, A_n$, their intersection is $A_1 \cap \cdots \cap A_n$ or $\cap_{i=1}^n A_i$ and consists of all elements that belong to *each* of these sets. Equivalently, we may define the set $I \equiv \{1, \ldots, n\}$ and write the intersection as $\cap_{i \in I} A_i$

Example 3.4. *No students attend all programs at SSE. So $\cap_{i=1}^n D_i = \varnothing$, i.e. the set of all students at SSE that attends every program is the empty set $\varnothing \equiv \{\}$.*

Example 3.5. *The intersection of all intervals $D_i \equiv (-i, i)$ is $\cap_{i=1}^\infty D_i = (-1, 1)$.*

The set of elements in X that are not in A is written $X \backslash A$. If $A \subset X$, then we say that the *complement* of A is $X \backslash A$, denoted A^c.

Example 3.6. *The set of students at SSE that are not Year 1 MSc Econ students can be written as E^c.*

4 What are Numbers?

We are used to working with numbers to count and label things. For example, a set with two elements is said to have a *cardinality* of two. These basic counting numbers $1, 2, 3, \ldots$ are called the *natural numbers* $\mathbb{N}$:

$$\mathbb{N} \equiv \{1, 2, 3, \ldots\} \qquad \text{(Natural numbers)}$$

Sets are not generally "equipped" with an *order* "$>$". But here, we will assume sets of numbers "inherit" the natural order we are used to, so we can write, e.g., $3 > 2$.

Next, we extend the natural numbers to include zero and the negatives of natural numbers. This gives the set of *integers* $\mathbb{Z}$. Formally, we define

$$\mathbb{Z} \equiv \{0\} \cup \mathbb{N} \cup \{-n : n \in \mathbb{N}\} \qquad \text{(Integers)}$$

The above might look tricky at first, but can be read as "the set of integers $\mathbb{Z}$ is the union of three sets: a singleton set with a zero, the set of natural numbers, and a set consisting of the negative of each element in the natural numbers (negative numbers)". We can extend the integers by allowing *ratios* of integers: the *rational numbers*.

Definition 4.1. *Denote the set of rational numbers as $\mathbb{Q}$. For any $x \in \mathbb{Q}$, we can write $x = \frac{m}{n}$ for some $m, n \in \mathbb{Z}$ with $n \neq 0$. Formally,*

$$\mathbb{Q} \equiv \left\{\frac{m}{n} : m \in \mathbb{Z},\ n \in \mathbb{Z}\backslash\{0\}\right\} \qquad \textit{(Rational numbers)}$$

Exercise 4.1. *Using the definition above: First (i) show that $x \in \mathbb{Z}$ is even if, and only if, x^2 is even. Then (ii) show that $\sqrt{2}$ is not rational.*

Now we get to the workhorse numbers of this course, the *real numbers*. Some numbers, such as π and $\sqrt{2}$, are not included in $\mathbb{Q}$, but are real numbers. Geometrically, the rationals sit as a dense scatter of points on the number line – but with countless gaps between them, like $\sqrt{2}$ and π. The real numbers are what you get when you fill those gaps in. For practical purposes, you can think of a real number as a number with a decimal expansion: one that terminates, like $\frac{3}{8} = 0.375$, or one that goes on forever, like $\frac{1}{9} = 0.1111\ldots$ or $\pi = 3.141592\ldots$. We denote the set of all real numbers by $\mathbb{R}$ and the non-negative real numbers as $\mathbb{R}_+ \equiv [0, \infty)$. For a more rigorous treatment, see: Wikipedia: Construction of the real numbers.

Thus, we may summarize the construction as

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

5 What are Spaces?

The next several chapters introduce different kinds of *spaces*. To help you keep track, here is a preview of how they relate. The key point is that each kind of space is just a set equipped with some *operators*; sometimes one kind of structure automatically gives you another for free, and sometimes the kinds of spaces are independent of each other.

Take a set X, simply a collection of elements. Then:

- **Vector space**: A (real) vector space V is a set X equipped with addition $(+)$ and real scalar multiplication $(\cdot)$ such that V is *closed* under both, i.e. vectors can be added together and rescaled without leaving the space. We can write a vector space as a *triple* $V \equiv (X, +, \cdot)$.
- **Inner product space**: An inner product space I is a vector space V equipped with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$. Notationally, read the previous as "an inner product is a *mapping* that takes two elements from V and spits out a number". In lower dimensions, the intuition is that an inner product $\langle x, y \rangle$ gives us the *angle* of two vectors x, y. We can write this inner product space as $I \equiv (V, \langle \cdot, \cdot \rangle)$.
- **Normed vector space**: A normed vector space N is a vector space V equipped with a norm $\| \cdot \| : V \to \mathbb{R}_+$, where $\|v\|$ measures the length of a vector v. We can write it as $N \equiv (V, \| \cdot \|)$. We can always derive a norm from an inner product by setting $\|v\| \equiv \sqrt{\langle v, v \rangle}$. The reverse is not always possible.
- **Metric space**: A metric space M is a set X equipped with a distance function $d : X \times X \to \mathbb{R}_+$, where the distance function $d(x, y)$ measures the distance between elements x, y. We can write this as $M \equiv (X, d)$. Note that, on its own, a metric space is structurally the simplest of the four – it requires no algebra, just a set and a way to measure distances. But whenever we already have a normed vector space in hand, we can derive a metric for free by setting $d(x, y) \equiv \|x - y\|$. The reverse is not always possible.

Schematically, the derivations only go one way:

$$\text{vector space} \xrightarrow{\text{add } \langle \cdot, \cdot \rangle} \text{inner product space} \xrightarrow{\|v\| \equiv \sqrt{\langle v, v \rangle}} \text{normed vector space} \xrightarrow{d(x,y) \equiv \|x - y\|} \text{metric space}$$

In each chapter ahead, we will pin down one of these structures. Once we have metric spaces in hand, we will then study *mappings* between metric spaces.

Example 5.1 (One set, four kinds of space). *To make the relationships above concrete, take the simplest non-trivial set: $X = \mathbb{R}^2$. Watch how each kind of structure can be put on X.*

- ***As a set.*** $X = \mathbb{R}^2 = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\}$. *No operations yet — just elements.*
- ***As a vector space.*** *Equip X with coordinate-wise addition and scalar multiplication: $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$, $\alpha(x_1, x_2) = (\alpha x_1, \alpha x_2)$. Both operations land back in $\mathbb{R}^2$, so X is closed under both. We can now add and rescale elements.*
- ***As an inner product space.*** *Equip X with the dot product $\langle x, y \rangle = x_1 y_1 + x_2 y_2$.*
- ***As a normed vector space.*** *The norm derived from this inner product is $\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + x_2^2}$ – the familiar Euclidean length (from origo).*
- ***As a metric space.*** *The metric derived from this norm is the Euclidean distance $d(x, y) = \|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$. We can now measure the distance between two elements.*

Same set $\mathbb{R}^2$ throughout, just different structures put on it. The same set $\mathbb{R}^2$ could also be turned into a different metric space directly, with no algebra involved at all — for example by using the discrete metric $d_1(x, y) = 1$ whenever $x \neq y$ and 0 otherwise. The standard choice (that we are used to working with) for $\mathbb{R}^2$ just happens to be the Euclidean one above.

Before we dig in, let me introduce someone you will see throughout these notes.

Econ Application: Meet Anna

Anna is a consumer choosing between two goods. Her *consumption bundle* is an ordered pair $x = (x_1, x_2) \in \mathbb{R}^2$, where x_1 is the quantity of good 1 and x_2 is the quantity of good 2 — exactly the kind of vector this chapter is about. The unit prices are $p = (p_1, p_2) = (1, 2)$, her income is $y = 10$, and her preferences are represented by a utility function $u : \mathbb{R}^2_+ \to \mathbb{R}$ such that

$$u(x_1, x_2) = \sqrt{x_1 x_2}.$$

Eventually we want to ask: which bundle $x^* = (x_1^*, x_2^*)$ maximises Anna's utility subject to her budget? We cannot answer that yet — we don't even have the language. But each chapter from here on will give us one tool that her problem needs, and by the end of these notes we'll watch her solve it.

6 What are Vectors?

When getting familiar with vector spaces, starting with the most simple notion of a vector is often useful. I find the YouTube course Essence of Linear Algebra by 3Blue1Brown illuminating to get an intuitive sense of (simple) vector spaces and how they connect to matrices. I strongly believe the video series will help in making sense of the first weeks of both mathematics and econometrics. In particular, begin with: 3Blue1Brown: Vectors.

6.1 Ordered List Vectors

A *set* is an unordered collection of elements: for example, $\{a, b\} = \{b, a\}$. But sometimes order matters — for instance, when working with coordinates. In such cases, we use *ordered pairs*, such as (a, b), where the order of elements is essential.

Vectors can be visualised as arrows in space. Define a collection of vectors $E \equiv \{e_1, e_2, e_3\}$, where $e_1 \equiv \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$ and so on. Probably, this is the *basis* you use when you interpret the below vector $\mathbf{v}_2 = (2, 1, 3)$ as having *coordinates*, see **Figure 6.1**.

$$\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{6.1}$$

Vectors can be added together. For instance, the vector $\mathbf{v}_1 = (1, 1, 1)$ points one unit in each of the three coordinate directions. Adding it to our vector $\mathbf{v}_2$ gives $\mathbf{v}_3 = \mathbf{v}_1 + \mathbf{v}_2 = (3, 2, 4)$, as depicted in **Figure 6.2**.

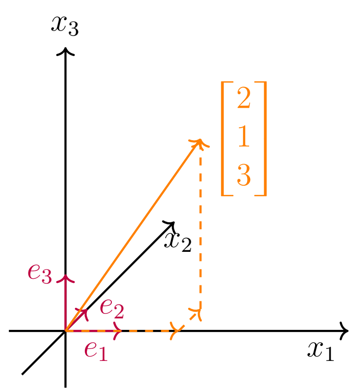


Figure 6.1: E as a basis

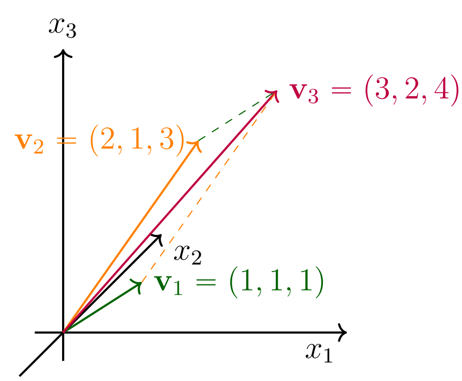


Figure 6.2: Vector $\mathbf{v}_3 = \mathbf{v}_1 + \mathbf{v}_2$

Finally, scaling by a positive number makes a vector longer or shorter while keeping its direction; a negative scalar flips its direction. For example, $2 \cdot (3, 2, 4) = (6, 4, 8)$.

6.2 Vector spaces

A *vector space* is a set V whose elements ("vectors") you can add together and multiply with numbers ("scalars") to produce new vectors in V. So far, we've discussed vectors as ordered lists of numbers. But a vector can also be a number, a matrix, or a function. What matters is that we can construct a well-functioning space of them. For intuition, see 3Blue1Brown: Abstract vector spaces.

Definition 6.1 (Vector spaces)**.** *A (real)* ***vector space*** V *is a set (of "vectors") on which two operations, addition* $+$ *and scalar multiplication* $\cdot$*, are defined such that*

- V *is closed under addition: for each pair of elements* $x, y \in V$ *there is a unique element* $x + y$*, the sum of* x *and* y*, in* V*,*
- V *is closed under scalar multiplication: for each* $x \in V$ *and each* $\alpha \in \mathbb{R}$*, there is a unique element* αx*, the (scalar) product of* α *and* x*, in* V*.*

Moreover, addition and scalar multiplication satisfy standard arithmetic properties.

Exactly what is meant by "standard arithmetic properties" is governed by the eight "vector axioms", which will be more closely defined later on in the course. Below are some simple examples of sets that can be turned into vector spaces using our basic notions of addition and scalar multiplication (and some that can't!).

Example 6.1 (The Real Line)**.** *The set* $\mathbb{R}$ *is a vector space. Vectors are real numbers and the zero element is* 0*. Addition is standard addition, and scalar multiplication is just multiplying by a real number: for* $x, y, \alpha \in \mathbb{R}$*, we have* $x + y \in \mathbb{R}$ *and* $\alpha x \in \mathbb{R}$*.*

Example 6.2 (The Interval (0,1))**.** *Consider the set* $(0, 1) = \{x \in \mathbb{R} : 0 < x < 1\}$ *with the same* $+$ *and* $\cdot$ *as above. This set is* ***not*** *a vector space. It is not closed under addition:* $0.8 + 0.5 = 1.3 \notin (0, 1)$*. Nor is it closed under scalar multiplication:* $2 \cdot 0.6 = 1.2 \notin (0, 1)$*.*

Example 6.3 (The Rational Numbers $\mathbb{Q}$)**.** *The set* $\mathbb{Q}$ *of rational numbers is not a* ***real*** *vector space. To be a real vector space, the set must be closed under scalar multiplication for all* ***real*** *scalars. But multiplying a rational number by an irrational scalar can yield an irrational number. For example,* $\sqrt{2} \cdot \frac{1}{2} = \frac{1}{\sqrt{2}} \notin \mathbb{Q}$*. Since* $\mathbb{Q}$ *is not closed under scalar multiplication by arbitrary real numbers, it is not a real vector space.*

Example 6.4 ($\mathbb{R}^2$)**.** *The set* $\mathbb{R}^2$ *consists of ordered pairs* (x_1, x_2) *of real numbers. The zero element is* $(0, 0)$*. Addition and scalar multiplication are defined coordinatewise: for* $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$ *and* $\alpha \in \mathbb{R}$*:*

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2) \quad \textit{and} \quad \alpha(x_1, x_2) = (\alpha x_1, \alpha x_2). \tag{6.2}$$

For example, $(1, 3) + (2, 5) = (3, 8)$ *and* $2 \cdot (1, 4) = (2, 8)$.

Example 6.5 (Quadratic Polynomials)**.** *The set* $P_2(\mathbb{R})$ *of real polynomials of degree at most 2 is a vector space. The zero element is the zero polynomial* $p(x) = 0$*. Each element is a function of the form* $p(x) = a_0 + a_1x + a_2x^2$ *with* $a_0, a_1, a_2 \in \mathbb{R}$*. Addition and scalar multiplication are defined coefficient-wise. For* $p(x), q(x) \in P_2(\mathbb{R})$*, where* $q(x) = b_0 + b_1x + b_2x^2$ *and* $\alpha \in \mathbb{R}$*,*

$$(p + q)(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2, \quad (\alpha p)(x) = \alpha a_0 + \alpha a_1x + \alpha a_2x^2.$$

If $q_1(x) \equiv 2x$ *and* $q_2(x) \equiv x^2 + x + 1$*, then* $p(x) = x^2 + 3x + 1$*, so* $p(x) \in P_2(\mathbb{R})$*.*

6.3 Subspaces

We often look at subsets of a vector space V with additional nice properties. If such a smaller set – with addition and scalar multiplication as in the larger set V – satisfies all properties of a vector space, it is called a (linear) *subspace*. Using the theorem below, one can verify that a given subset W is a subspace.

Definition 6.2 (Subspaces)**.** *A subset* W *of vector space* V *is a (linear)* ***subspace*** *of* V *if* W *itself (inheriting* $+$ *and* $\cdot$ *from* V*) is a vector space.*

Theorem 6.1. *A subset* W *of a vector space* V *is a subspace if and only if it satisfies the following three properties:*

i) W *contains the zero vector from* $V : 0 \in W$*,*

ii) W *is closed under addition:* $x + y \in W$ *whenever* $x \in W$ *and* $y \in W$*,*

iii) W *is closed under scalar multiplication:* $\alpha x \in W$ *whenever* $x \in W$ *and* $\alpha \in \mathbb{R}$*.*

Exercise 6.1. *Use the theorem above to show that*

a) The empty set $W_2 = \varnothing$ *is not a subspace of* $\mathbb{R}^2$*.*

b) The set $W_3 = \{x \in \mathbb{R}^2 : x_1 = 0$ *or* $x_2 = 0\}$ *is not a subspace of* $\mathbb{R}^2$*.*

c) The set $W_4 = \{x \in \mathbb{R}^2 : x_1, x_2 \in \mathbb{N}\}$ *is not a subspace of* $\mathbb{R}^2$*.*

d) The set $W_1 = \{x \in \mathbb{R}^2 : 3x_1 - 4x_2 = 0\}$ *is a subspace of* $\mathbb{R}^2$*.*

e) The set $W_5 = \{f \in P_2(\mathbb{R}) : f(x) = a + bx$ *for some* $a, b \in \mathbb{R}\}$ *is a subspace of the set* $P_2(\mathbb{R})$ *from* ***Example 6.5****.*

6.4 Angles and lengths

So far, vectors are objects we can add and scale. To talk about *angles* between vectors, or to relate the length of a vector to its components, we need a third operation: an **inner product**. A vector space equipped with an inner product is called an *inner product space.*

Example 6.6 (The inner product on $\mathbb{R}^n$)**.** *The standard inner product on $\mathbb{R}^n$ is the* ***dot product****: for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$,*

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i.$$

Other common notation includes $x \cdot y$ and $x^\top y$ – the latter because the inner product on $\mathbb{R}^n$ is the matrix product of the row vector $x^\top$ with the column vector y. We will use this connection in the next chapter when discussing matrix multiplication.

Econ Application: Anna's expenditure

Recall Anna (introduced above), with prices $p = (1, 2)$. The cost of a bundle $x = (x_1, x_2) = (3, 5)$ is the inner product

$$\langle p, x \rangle = p_1 x_1 + p_2 x_2 = 3 + 2 * 5 = 13.$$

More generally, with n goods and unit prices $p = (p_1, \ldots, p_n)$, the cost of a bundle $x = (x_1, \ldots, x_n)$ is $\langle p, x \rangle = \sum_i p_i x_i$.

Just as the absolute value $|x|$ measures the size of a number in one dimension, we can define the *length* (or *norm*) of a vector in $\mathbb{R}^n$ using the Euclidean norm:

$$\|\mathbf{v}\| = \sqrt{\langle v, v \rangle} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

If we equip a vector space with a norm, it becomes a *normed vector space.*

Example 6.7. *In $\mathbb{R}^2$ and $\mathbb{R}^3$, think of this as the Euclidean distance from the origin to the "point" of the vector. In $\mathbb{R}^2$, the vector $\mathbf{v}' = (3, 4)$ has length $\|\mathbf{v}'\| = \sqrt{3^2 + 4^2} = 5$. In $\mathbb{R}^3$, the length of $\mathbf{v}_3 = (2, 3, 4)$ is $\|\mathbf{v}_3\| = \sqrt{2^2 + 3^2 + 4^2} = \sqrt{29}$.*

As you will see later on, this is not the only possible norm we could use.

7 What are Matrices?

One particular type of vector space is a space of *matrices*. A matrix is a collection of rows and columns, organised in a rectangular fashion. We denote matrices by capital letters, A, and their element in row i and column j as a_{ij}.

7.1 Addition and Scalar Multiplication

First, we note that a vector space of matrices is simply a set of matrices (vectors), equipped with addition and scalar multiplication that obeys standard arithmetic rules.

Example 7.1 (2×3 Matrices)**.** *The set* $\mathbb{R}^{2\times 3}$ *of* 2×3 *real matrices is a vector space. If*

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix}, \quad N = \begin{bmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \end{bmatrix},$$

then addition and scalar multiplication are defined entrywise:

$$(M + N)_{ij} = m_{ij} + n_{ij}, \quad (\alpha M)_{ij} = \alpha m_{ij}.$$

The zero element is the 2×3 *matrix with all entries equal to* 0.

The addition operator for matrices is a commutative operation, so $A + B = B + A$.

Example 7.2 (Addition and scalar multiplication of matrices)**.** *We have:*

$$A \equiv \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B \equiv \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad A + B = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}, \quad 5 \cdot A = \begin{bmatrix} 5 & 10 \\ 15 & 20 \end{bmatrix}$$

7.2 Matrix multiplication

Recall the inner product on $\mathbb{R}^n$ from **Chapter 6**. *Matrix multiplication* extends this idea: each entry of a matrix product is itself an inner product of a row of the first matrix with a column of the second. To improve your intuition of these concepts, see 3Blue1Brown lessons on (i) linear transformations, (ii) matrix multiplication and (iii) dot products.

Matrix multiplication is not like number multiplication: it's not commutative, so we may have $A \cdot B \neq B \cdot A$. Furthermore, it only works when the "inner dimensions" match. If C is an $n \times m$ matrix and D is an $m \times p$ matrix, then the product $C \cdot D$ is defined and results in an $n \times p$ matrix. What matters is that the number of columns m of the first matrix must match the number of rows m of the second matrix.

Example 7.3 (Matrix multiplication)**.** *Define the matrices*

$$C = \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}}_{2\times 3}, \qquad D = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}}_{3\times 2}.$$

$$C \cdot D = \begin{bmatrix} 1\cdot 1 + 2\cdot 0 + 3\cdot 1 & 1\cdot 0 + 2\cdot 1 + 3\cdot 1 \\ 4\cdot 1 + 5\cdot 0 + 6\cdot 1 & 4\cdot 0 + 5\cdot 1 + 6\cdot 1 \end{bmatrix} = \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}.$$

We take the number of rows from C and the number of columns from D, so the result is a 2×2 matrix. If we instead would calculate $D \cdot C$, we would get a 3×3 matrix.

7.3 Transpose and Inverse

Two concepts that are relevant for solving matrix equations of models, both in economic theory and in econometrics, are those of the *transpose* and the *inverse* of a given matrix.

The *transpose* of a matrix is the matrix resulting from swapping columns and rows:

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad C^\top = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \tag{7.1}$$

Finally, the *inverse* of the matrix A above, denoted A^{-1}, is such that $AA^{-1} = I_2$, where I_2 is the 2×2 "identity matrix":

$$I_2 \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{7.2}$$

Note that for any 2×2 matrix, multiplying it with the identity matrix renders the same matrix – just like multiplying a scalar with the number 1. For example, for the matrix A defined above, we have $A \cdot I_2 = A$ (check yourself!).

Example 7.4 (Deriving the Inverse Formula for a 2×2 Matrix)**.** *Let*

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be a 2×2 matrix. We want to find a matrix

$$M^{-1} = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$$

such that $MM^{-1} = I_2$. We compute the product:

$$MM^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} ax + bz & ay + bw \\ cx + dz & cy + dw \end{bmatrix}.$$

We now require this product to equal the identity matrix:

$$\begin{bmatrix} ax + bz & ay + bw \\ cx + dz & cy + dw \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \iff \quad \begin{cases} ax + bz &= 1 \\ ay + bw &= 0 \\ cx + dz &= 0 \\ cy + dw &= 1 \end{cases}$$

Solving this system, we eventually find that the solution is:

$$M^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

provided that $ad - bc \neq 0$. This quantity $ad - bc$ is called the ***determinant*** *of A, and it must be nonzero for the inverse to exist. A matrix is* ***invertible*** *if $\det(M) \neq 0$ and the matrix is square (same number of rows and columns).*

Exercise 7.1. *Some basic exercises of matrix multiplication:*

a) Consider the matrix A above. What is $A^\top$? Get $A^\top A$. Is it the same as $AA^\top$?

b) Apply the inverse formula for A to get A^{-1}. Verify that $AA^{-1} = I_2$.

c) Consider a vector of residual values from an OLS regression:

$$\varepsilon \equiv \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{7.3}$$

Get $\varepsilon^\top \varepsilon$. Is it the same as $\varepsilon\varepsilon^\top$?

7.4 Systems of Linear Equations

In economics, matrices often describe *linear systems of equations* that arise from equilibrium conditions and/or budget constraints, e.g. from a firm choosing inputs, or a consumer choosing a bundle. Mathematically, we write such a system as

$$A\mathbf{x} = \mathbf{b}.$$

Econ Application: Anna's optimisation: A linear system

Take Anna (**Chapter 6**) with utility $u(x_1, x_2) = \sqrt{x_1 x_2}$, prices $p = (1, 2)$, and income $y = 10$. Later on, we will motivate that her optimal bundle will satisfy two conditions: (i) her budget binds: $x_1 + 2x_2 = 10$, (ii) a *tangency condition* $x_1 = 2x_2$. Together, these form a linear system of two equations in two unknowns:

$$\begin{cases} x_1 - 2x_2 = 0 \\ x_1 + 2x_2 = 10 \end{cases} \iff \underbrace{\begin{bmatrix} 1 & -2 \\ 1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 0 \\ 10 \end{bmatrix}}_{\mathbf{b}} \tag{7.4}$$

The inverse exists, so we get one unique solution:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \end{bmatrix} = \begin{bmatrix} 5 \\ 2.5 \end{bmatrix} \tag{7.5}$$

So Anna's optimal bundle is $x^* = (5, 2.5)$, giving utility $u(x^*) = \sqrt{12.5}$.

We often avoid computing A^{-1} directly. In practice, this inverse may not exist, or calculating it may be too cumbersome. Yet even without an inverse, a system may still have a solution. More generally, a system of linear equations can have exactly one, multiple or no solution.

Example 7.5 (One solution)**.** *The following system has one (unique) solution:*

$$\underbrace{\begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 1 \\ -5 \end{bmatrix}}_{\mathbf{b}} \iff \begin{cases} 3x - y = 1 \\ x - y = -5 \end{cases} \iff \begin{cases} y = 3x - 1 \\ y = x + 5 \end{cases}$$

Think of these as two lines in the plane (draw!). They have a single intersection (try solving it yourself by variable substitution), namely at $(x, y) = (3, 8)$.

Example 7.6 (No solution)**.** *The following system of parallell lines has no solution:*

$$\begin{bmatrix} 2 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \iff \begin{cases} x = 0.5y \\ y - 2x = 1 \end{cases} \iff \begin{cases} y = 2x \\ y = 2x + 1 \end{cases}$$

Example 7.7 (Infinity of Solutions)**.** *Finally, we can have an infinite number of solutions:*

$$\begin{bmatrix} 2 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff \begin{cases} x = 0.5y \\ y - 2x = 0 \end{cases} \iff \begin{cases} y = 2x \\ y = 2x \end{cases}$$

7.5 Gaussian Elimination

For large systems of equations, there is no general shortcut for computing A^{-1}, and solving by variable substitution quickly becomes infeasible beyond small examples. Fortunately, we have a powerful method for solving systems of equations efficiently — even when we don't compute A^{-1} explicitly.

Gaussian Elimination solves the system $A\mathbf{x} = \mathbf{b}$ by applying row operations to the augmented matrix $[A \mid \mathbf{b}]$, step by step, until the left side becomes the identity matrix:

$$[A \mid \mathbf{b}] \longrightarrow [I \mid \mathbf{x}].$$

The final column then gives the solution vector $\mathbf{x}$. In other words, Gaussian elimination produces the same result as computing $\mathbf{x} = A^{-1}\mathbf{b}$ — but without explicitly forming the inverse matrix. It is systematic, scalable, and works for systems with one, none, or infinitely many solutions. A quick Google search will give you the foundations and you can practice it yourself with the exercises below.

In practice, you'll often rely on a computer to do the heavy lifting of solving large systems of equations — but it's important to understand what the computer is doing. Manually applying Gaussian elimination gives you insight into how solutions are constructed step by step, and helps you build intuition for when solutions exist and when they don't.

Exercise 7.2. *Solve the following problems using Gaussian elimination.*

a) *Show that the system below has a unique solution:*

$$\begin{bmatrix} 1 & -1 & 1 \\ 2 & 3 & -1 \\ 3 & -2 & -9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 8 \\ -2 \\ 9 \end{bmatrix} \iff \begin{cases} x - y + z = 8 \\ 2x + 3y - z = -2 \\ 3x - 2y - 9z = 9 \end{cases} \tag{7.6}$$

b) *Show that the system below has no solution:*

$$\begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix} \iff \begin{cases} x - y = 3 \\ 2x - 2y = 7 \end{cases} \tag{7.7}$$

c) *Show that the system below has an infinite number of solutions.*

$$\begin{bmatrix} -1 & -2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \iff \begin{cases} -x - 2y + z = -1 \\ 2x + 3y = 2 \\ y - 2z = 0 \end{cases} \tag{7.8}$$

8 What are Metrics?

In a vector space V, we can add vectors and scale them by real numbers. But something is missing: there is no way to talk about how *close* two vectors are to each other. To handle distance, we need a different kind of structure: a *metric space.*

8.1 Metric spaces

To formally talk about distance, we need a different structure: a *metric space.* A metric space does not involve equipping a set with addition or scalar multiplication — just a set of elements X and a *metric* $d(x, x')$: a rule for measuring distances.

Definition 8.1 (Metric Space)**.** *Let X be a nonempty set. A function $d : X \times X \to \mathbb{R}$ is a* ***distance function*** *or* ***metric*** *if it is non-negative, symmetric, equals zero only when the two points are the same, and satisfies the triangle inequality. We call $d(x, y)$ the* ***distance*** *between x and y. The pair (X, d) is called a* ***metric space****.*

You will learn more about what generally constitutes a metric later on, but you already have a basic understanding of the Euclidean distance metric in $\mathbb{R}$ and $\mathbb{R}^2$.

Example 8.1 ($\mathbb{R}$ with the Euclidean Distance)**.** *Let $X = \mathbb{R}$, and define the distance function $d_1(x, y) = |x - y|$. Then $(\mathbb{R}, d_1)$ is a metric space. The function d_1 measures the usual distance between two real numbers on the number line.*

Example 8.2 ($\mathbb{R}^2$ with the Euclidean Distance)**.** *Let $X = \mathbb{R}^2$, and define the distance function $d_2(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$, where $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$. Then $(\mathbb{R}^2, d_2)$ is a metric space. This is the standard distance in the plane, given by the Pythagorean theorem. For example, consider $x' \equiv (3, 2) \in \mathbb{R}^2$ and $y' \equiv (1, 1)$. Then $d_2(x', y') = \sqrt{(3-1)^2 + (2-1)^2} = \sqrt{5}$.*

However, we can also define a distance function for more exotic sets, such as the set of all continuous functions defined over some domain $[a, b]$.

Example 8.3. *Consider the set $C[a, b]$ of continuous functions $f : [a, b] \to \mathbb{R}$ with the* ***supremum metric*** *$d_\infty(f, g) = \max\{|f(x) - g(x)| : x \in [a, b]\}$ for each pair of functions $f, g \in C[a, b]$. Then $(C[a, b], d_\infty)$ is a metric space. The metric measures the maximum distance between function values over the domain.*

Spaces of functions show up throughout economics. In (micro) auction theory, agents choose an optimal bidding function $b^* \in U$. In (macro) dynamic growth models, agents choose optimal consumption-investment policy functions. The properties of these spaces and subsets determine whether an optimal solution can be guaranteed to exist.

8.2 Open Balls

You have been taught that $(0, 1)$ is "open" because it does not include its boundary points, and $[0, 1]$ is "closed" because it does. In $\mathbb{R}$, this intuition works well — but what does it even mean for a subset of functions or matrices to be open or closed? To answer these questions, we need a more general and precise notion of these concepts. First, we need a formal definition of what is meant by two elements being "close" to one another. We define:

Definition 8.2 (Open Ball)**.** *Let (X, d) be a metric space, and let $x \in X$ and $r > 0$. The **open ball** around x with radius r is the set*

$$B(x, r) = \{y \in X : d(x, y) < r\}$$

of all points whose distance from x is less than r.

Example 8.4. ***Figure 8.1*** *shows the open ball around $(1, 1)$ in $(\mathbb{R}^2, d_2)$ with radius $r = \frac{1}{2}$.*

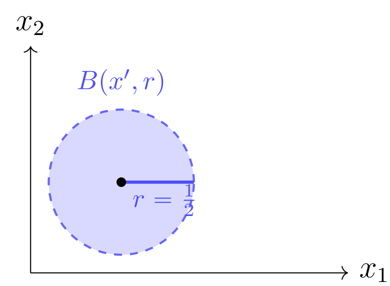


Figure 8.1: An open ball $B(x', r)$ around $x' = (1, 1)$ with radius $r = \frac{1}{2}$.

8.3 Openness and closedness

With the notion of open balls in place, we can now define what it means for a point to be *interior* (on the "inside" of the set) and for a set to be *open* or *closed.*

Definition 8.3 (Open and Closed Sets in a Metric Space)**.** *Let (X, d) be a metric space, and let $U \subseteq X$ be a subset in this space.*

- ☒ *A point $u \in U$ is an **interior point** of U if $\exists \varepsilon > 0$ such that $B(u, \varepsilon) \subseteq U$.*
- ☒ *A set $U \subseteq X$ is **open** if every point in U is an interior point.*
- ☒ *A set $U \subseteq X$ is **closed** if its complement $U^c = X \backslash U$ is open.*

Intuitively, for a given subset in a metric space, an *interior point* is an element that one can draw a sufficiently small ball around such that the entire ball is contained in the subset. For a given subset, we say that it is *open* if all elements of the set are interior. If not, then the subset is *not open*. Furthermore, we say that a subset is *closed* if all elements outside of the set are interior to the space outside of the set. If not, then the subset is *not closed*.

Note that sets don't behave like physical doors — a set can be open, closed, both, or neither. For example, the whole space X and the empty set $\varnothing$ are both open and closed, while the interval $[0, 1) \subset \mathbb{R}$ is neither open nor closed (see exercise below).

Example 8.5 (An interior point)**.** *Consider the (orange) interval* $(0, \infty) \subset (\mathbb{R}, d_1)$ *in **Figure 8.3**. The point* $x = 0.1$ *is an **interior** point, because it <u>exists</u> a small enough (purple)* ε*-ball for which all elements of the ball are included in* $(0, \infty)$*. This could be e.g.* $\hat{\varepsilon} = 0.05$*, so* $B(0.1, \hat{\varepsilon}) = (0.05, 0.15) \subset (0, \infty)$*. Note that* $B(0.1, \varepsilon) \subset (0, \infty)$ *need not hold for <u>any</u>* $\varepsilon > 0$ *for* $x = 0.1$ *to be an interior point (e.g. the green ones). It can be shown (see below) that any point in this interval is an interior point, so* $(0, \infty)$ *is an open set.*

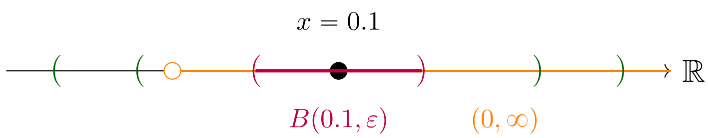


Figure 8.2: The point $x = 0.1$ is an interior point of the interval $(0, \infty)$.

Example 8.6 (A non-interior point)**.** *Consider the interval* $[0, \infty) \subset (\mathbb{R}, d_1)$*. The point* $x = 0$ *is **not** an interior point. For example, in any ball with radius* $\varepsilon > 0$*, there will be an element* $y \equiv -\frac{\varepsilon}{2}$ *such that* $y \in B(0, \varepsilon)$ *but* $y \notin [0, \infty)$*. Thus, the set is not open.*

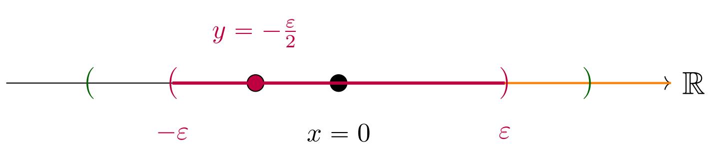


Figure 8.3: The point $x = 0$ is not an interior point of the interval $[0, \infty)$.

Exercise 8.1. *Consider some subsets of the metric space of the real line* $(\mathbb{R}, d_1)$*.*

a) Show that $A_1 = (0, \infty)$ *is open.*

b) Show that $A_2 = [0, 1)$ *is not open, nor closed.*

c) Take as given that open intervals on the real line are open and that unions of open sets are open. Show that $B_1 = [0, 1]$ *and* $B_2 = [0, \infty)$ *are closed.*

A property that will come in handy later on to show that more complicated sets are open or closed is the following:

Theorem 8.1. *Let (X, d) be a metric space. The union of $\cup_{i=1}^{n} A_i$ of any open sets $A_1, \ldots, A_n$ is itself an open set. Moreover, it is true for infinitely many sets as well.*

Econ Application: Anna's budget set is not open

Recall Anna from **Chapter 6**, with prices $p = (1, 2)$ and income $y = 10$. Her budget set is

$$A = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_1 + 2x_2 \leqslant 10\} \subset \mathbb{R}^2$$

The point $x = (3, 2)$ is an interior point, as we can draw a small ball around it that remains inside the set ($3 + 4 = 7 < 10$). The point $x' = (2, 4)$, however, lies on the boundary $x_1 + 2x_2 = 10$, so any ball around it will include points outside the budget set, e.g. $\left(2 + \frac{\varepsilon}{2}, 4\right)$. Thus, the budget set is not open.

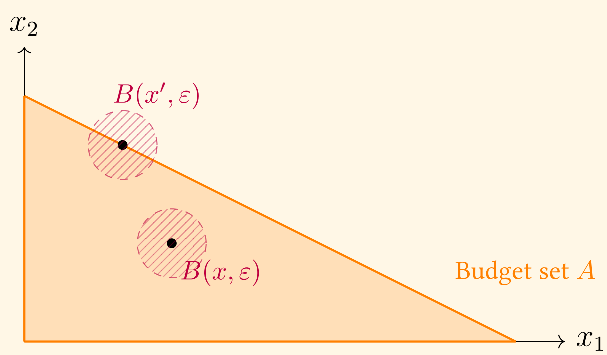


Figure 8.4: Two points in Anna's budget set: $x = (3, 2)$ is interior, $x' = (2, 4)$ is on the boundary.

8.4 Boundary points

We might also define a set's *boundary*. Although there is a trickier formal *definition*, boundary points can be *characterised* as follows: a point u is a boundary point of U if for *any* radius $\varepsilon > 0$, the open ball $B(u, \varepsilon)$ has at least one element in U and at least one element in the complement U^c.

Theorem 8.2. *Let (X, d) be a metric space and let $U \subseteq X$ be a subset in this space. A point $u \in U$ is a* ***boundary point*** *of U if $\forall \epsilon > 0$:*

$$B(x, \varepsilon) \cap U \neq \varnothing \quad \textit{and} \quad B(x, \varepsilon) \cap U^c \neq \varnothing \tag{8.1}$$

Here, note the difference in definition for interior points. A point $u \in U$ is an interior point if we can find *one* radius $\varepsilon > 0$ such that the entire ball lies inside of our set. Meanwhile, a point $u \in U$ is a boundary point if for *all* possible radius $\varepsilon > 0$, the ball lies both in U and outside. Furthermore, note that a boundary point need not lie in the set. For example, the point 0 is a boundary point to the open set $(0, 1)$, even though $0 \notin (0, 1)$.

Exercise 8.2. *Consider Anna's budget set A. Show that $x' = (2, 4)$ is a boundary point.*

8.5 Compactness

Finally, let's introduce two additional properties of a set. First, a set can be *bounded*:

Definition 8.4 (Bounded set)**.** *A subset Y of a metric space (X, d) is **bounded** if it is contained in some open ball, i.e. there exist $x \in X$ and $r > 0$ such that $Y \subseteq B(x, r)$.*

Second, a property of a set in $\mathbb{R}^n$ that will be crucial later on is *compactness.* A compact set behaves like a finite set, even when it has infinitely many elements. Continuous functions on a finite set are easy – they take finitely many values, so a maximum and a minimum always exist. Compact sets are an infinite version of this, where the same kind of guarantee survives. Later on, you will meet the formal definition of compactness, which works in any metric space and is stated in terms of *open coverings.* For now, we will use a much simpler characterisation that works in $\mathbb{R}^n$:

Theorem 8.3 (Heine-Borel)**.** *A subset of $\mathbb{R}^n$ is **compact** if and only if it is closed and bounded.*

So in $\mathbb{R}^n$, compactness is just the conjunction of two properties we already understand. A set with no holes (closed) that does not run off to infinity (bounded) is compact. At the end of the next chapter, you will see the most common application of compactness in economics: the *Extreme Value Theorem*, which states that a continuous function attains its maximum on any non-empty compact set. Thus, if an economic agent is maximising a continuous (utility) function over a non-empty compact (budget) set, the problem has a solution.

9 What are Functions?

Consider a *function*, or a *mapping*, $f : \mathbb{R} \to \mathbb{R}$, that takes elements from its *domain* $\mathbb{R}$ (a metric space) and maps them into the co-domain $\mathbb{R}$ (also a metric space). Here, we use $\mathbb{R}$ to represent the metric space $(\mathbb{R}, d_1)$, where d_1 is the usual distance in $\mathbb{R}$.

Exercise 9.1. *A function $f : \mathbb{R} \to \mathbb{R}$ is **strictly increasing** if $x' > x \Rightarrow f(x') > f(x)$. Show that $f(x) = x^2$ is **strictly increasing** on $[0, 1]$, but not on $\mathbb{R}$.*

9.1 Pre-images

We use the notation f^{-1} with two different meanings. Consider the subset $A = [0, 2] \subset \mathbb{R}$ in the co-domain. For a given element $x \in A$, the *inverse* $f^{-1}(x)$ is an *element*. For $g(x) = 2x$, an inverse exists: $g^{-1}(x) = x/2$. This is a function that "undoes" the action of f. For the given set A, the *pre-image* $f^{-1}(A)$ is a *set*, e.g. $g^{-1}(A) = [0, 1]$. This is the collection of all inputs that get mapped into A by the function.

Definition 9.1 (Pre-image)**.** *Consider a function $f : X \to Y$. The **pre-image** $f^{-1}(V)$ of a set $V \subseteq Y$ consists of all points in the domain $x \in X$ with a function value $f(x)$ in V, i.e., all points that are mapped into V :*

$$f^{-1}(V) = \{x \in X : f(x) \in V\}$$

Example 9.1. *Consider $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = x^2$. The element $2 \in \mathbb{R}$ in the **domain** has the **image** $4 \subset \mathbb{R}$ in the **co-domain**. The function do not have a well-defined inverse. The **pre-image** in the **domain** of $\{4\} \subset \mathbb{R}$ in the **co-domain** is $\{-2, 2\} \subset \mathbb{R}$.*

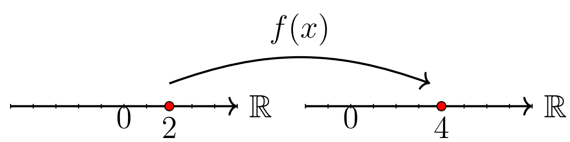


Figure 9.1: Image of 2 under x^2

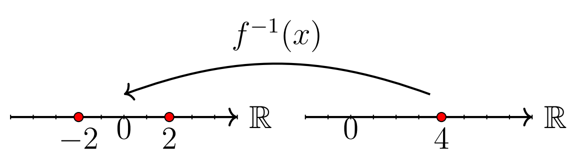


Figure 9.2: Pre-image of $\{4\}$ under x^2

Example 9.2. *Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, and consider the subset (in the co-domain) $A = [4, +\infty) \subseteq \mathbb{R}$. The pre-image of A is all points $x \in \mathbb{R}$ in the domain s.t. $f(x) \in A$:*

$$f^{-1}(A) = \{x \in \mathbb{R} : x^2 \geqslant 4\} = (-\infty, -2] \cup [2, +\infty).$$

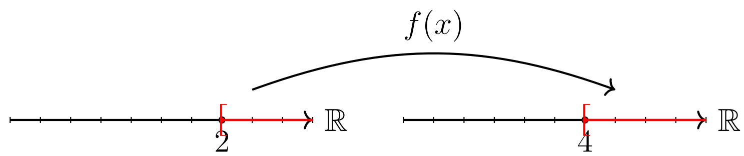


Figure 9.3: Image of $[2, +\infty)$ under x^2

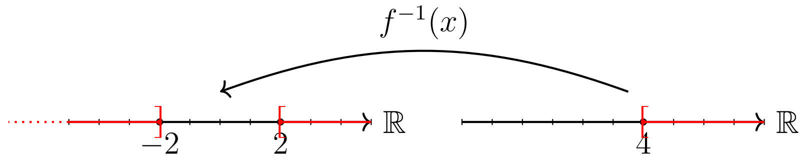


Figure 9.4: Pre-image of $[4, +\infty)$ under x^2

Example 9.3. *With a similar reasoning, the preimage of $A = (0, 4)$ under $f(x) = x^2$ is $f^{-1}(A) = (-2, 2)$. We could represent this as above, or as in* ***Figure 9.5*** *below.*

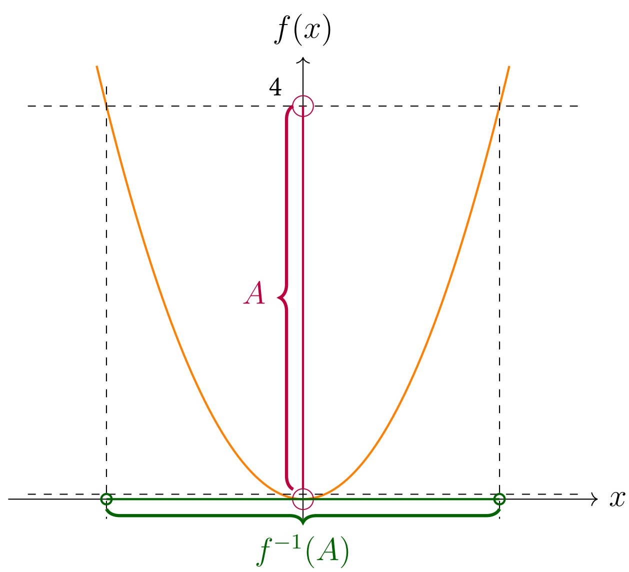


Figure 9.5: Pre-image of the set $A = (0, 4)$ under f is the set $f^{-1}(A) = (-2, 2)$.

Exercise 9.2. *Now, consider some functions $f : \mathbb{R}^2 \to \mathbb{R}$.*

a) Define $f_1(x) \equiv x_1$. Find the the image of $[0,1]^2$ and draw it in $\mathbb{R}$:

$$\{y \in \mathbb{R} : y = f_1(x) \text{ for some } x \in [0,1]^2\} \subset \mathbb{R} \tag{9.1}$$

b) Find the pre-image of $[0,+\infty)$ under $f_1(x)$ and draw it in $\mathbb{R}^2$:

$$\{x \in \mathbb{R}^2 : f_1(x) \in [0,+\infty)\} \subset \mathbb{R}^2 \tag{9.2}$$

Econ Application: Anna's indifference curves and upper contour sets

Take Anna with utility $u : \mathbb{R}^2_+ \to \mathbb{R}$ such that $u(x_1, x_2) = \sqrt{x_1 x_2}$. Two important pre-images come up everywhere in microeconomics:
The pre-image of the singleton $\{\bar{u}\}$ under $u(x)$ for some $\bar{u} > 0$ is the set of bundles that give Anna utility exactly $\bar{u}$ — her **indifference curve** at level $\bar{u}$:

$$u^{-1}(\{\bar{u}\}) = \{x \in \mathbb{R}^2_+ : \sqrt{x_1 x_2} = \bar{u}\}.$$

The pre-image of $[\bar{u}, \infty)$ under $u(x)$ is the set of bundles that give Anna utility at least $\bar{u}$ — bundles she weakly prefers to anything on her $\bar{u}$-indifference curve. This is her **upper contour set** at level $\bar{u}$:

$$u^{-1}([\bar{u}, \infty)) = \{x \in \mathbb{R}^2_+ : \sqrt{x_1 x_2} \geqslant \bar{u}\}.$$

Geometrically, slice Anna's utility surface horizontally at height $\bar{u}$. This cut traces out a curve, which can be projected down to the x_1–x_2 plane, giving the indifference curve. The bundles *above* the cut, projected down, give the upper contour set.

Econ Application (cont.)

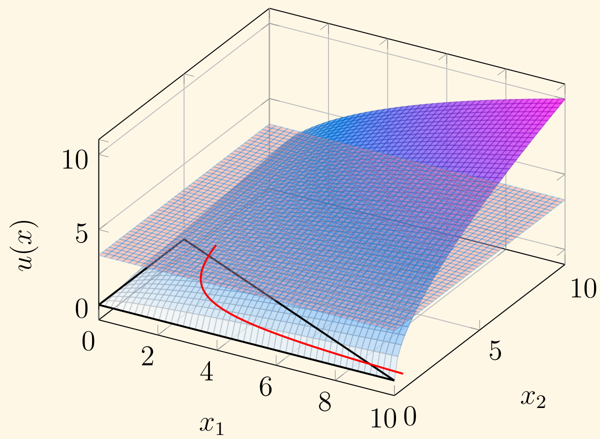


Figure 9.6: Anna's utility $u(x) = \sqrt{x_1 x_2}$ over the budget triangle, sliced by the horizontal plane $u(x) = \bar{u}$ (red). The intersection of the surface with the plane projects down to the indifference curve at level $\bar{u}$ in the x_1–x_2 plane.

For a flat 2D view of the same indifference curve, see **Figure 9.7** below. The intersection of the upper-contour set for $\bar{u}$, denoted $U_{\bar{u}}$, and Anna's budget set A is $\{x \in \mathbb{R}^2 : x \in A \cap U_{\bar{u}}\}$: these are all *feasible* bundles that give weakly greater utility than $\bar{u}$.

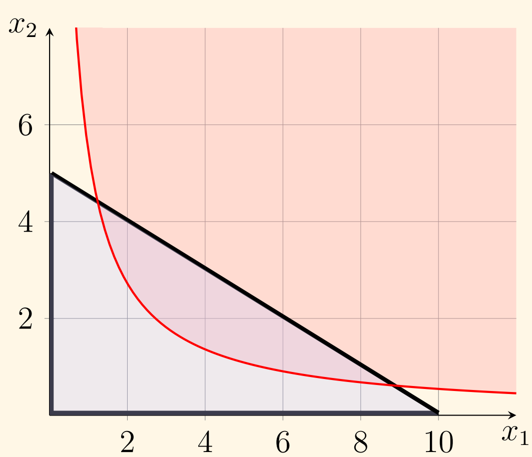


Figure 9.7: Anna's indifference curve at level $\bar{u}$ (red), her budget set (blue triangle), and the corresponding upper contour set $\{x : u(x) \geqslant \bar{u}\}$ (shaded red).

9.2 The (ϵ, δ)-definition of continuity

Although the entire 3Blue1Brown: Calculus series is great, I thoroughly recommend the chapters on limits and continuity. **After** reading these chapters, let's dive into how to rigorously *prove* properties like continuity and differentiability.

Continuous functions have no sudden jumps in their function values. Look at a point a in the domain with function value $f(a)$. For continuous functions, we have that if $a' \approx a$, then $f(a') \approx f(a)$. Formally, this is captured by the so-called (ε, δ)-definition of continuity:

Theorem 9.1 ((ε, δ)-definition of continuity)**.** *Let (X, d) and (Y, d') be metric spaces, $f : U \to Y$ a function on a subset U of X, and let $a \in U$. Then f is **continuous** at a if, and only if, for each $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$\forall x \in U : d(x, a) < \delta \Rightarrow d'(f(x), f(a)) < \varepsilon$$

Intuitively, suppose we are standing at the point x_0 on the real line. Someone hands us a desired level of precision $\varepsilon > 0$ for the function values. Our task is to construct a precision level $\delta > 0$ for the input such that staying within δ of x_0 guarantees that the function values stay within ε of $f(x_0)$. If this is always possible, then f is continuous.

As already mentioned, disproving a statement can be relatively easy to proving it. When proving that a function is not continuous, it is enough to find a single example of ε where for some x', we can never find a δ s.t. $d(x', x_0) < \delta \Rightarrow d(f(x'), f(x_0)) < \varepsilon$.

Exercise 9.3. *Suppose $x \in \mathbb{R}$ represents a gain/loss on a bet. Consider an agent with a "prospect theory"-inspired utility $u : \mathbb{R} \to \mathbb{R}$. The agent feels losses more keenly than she does wins, and the agent's utility takes a "jump" around zero: she cares about the "winning" and "losing" in itself:*

$$u(x) \equiv \begin{cases} \sqrt{x} + 0.5 & \textit{if } x \geqslant 0 \\ -2\sqrt{|x|} & \textit{if } x < 0 \end{cases} \tag{9.3}$$

Draw its graph. Then show $u(x)$ is not continuous at 0.

For the next proof, we will need to show that we can find an appropriate δ for any ε that may be thrown at us.

Exercise 9.4. *Show that the function $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = 2x + 1$ is continuous on $\mathbb{R}$ with the usual metric.*

9.3 Continuity and Openness

A useful result is the characterisation of continuous functions in terms of pre-images:

Theorem 9.2 (Continuity and openness/closedness)**.** *Let (X, d) and (Y, d') be metric spaces and $f : X \to Y$. The following three claims are equivalent:*

- *Function f is continuous;*
- *Pre-images of open sets are open sets: if $V \subseteq Y$ is open, then $f^{-1}(V)$ is open;*
- *Pre-images of closed sets are closed sets: if $V \subseteq Y$ is closed, then $f^{-1}(V)$ is closed.*

Note that the above theorem can be interpreted as an "if and only if" proposition:

$$f \text{ cont.} \quad \Longleftrightarrow \quad V \subseteq Y \text{ open} \Rightarrow f^{-1}(V) \text{ open} \quad \Longleftrightarrow \quad V \subseteq Y \text{ closed} \Rightarrow f^{-1}(V) \text{ closed}$$

<u>Openness of a set</u>: Instead of showing that an open ball can be drawn around every element in the set, one can prove openness using the theorem above:

Exercise 9.5. *Take as given that linear functions are continuous, e.g. $f(x_1, x_2) = x_1$. Use a pre-image argument to show that $C = \{x \in \mathbb{R}^2 : x_1 > 0\}$ is open.*

<u>Continuity of a function</u>: As we saw above, the epsilon-delta definition of continuity can be quite tricky to work with. At times, a more simpler proof is available to show properties of continuity of a function:

Exercise 9.6. *Show that the utility function u in **Exercise 9.3** is not continuous by finding an open set V whose preimage under u is not open.*

9.4 Existence of an Optimal Solution

We are now ready to answer a question we have been dancing around: when does an optimisation problem have a solution at all?

<u>*Why care about existence?*</u>

In an applied context this might feel pedantic – surely a maximum exists, why prove it? But it is easy to write down problems where no maximum exists. The function $f(x) = x$ on $[0, 1)$ has no maximum: any candidate $x \in [0, 1)$ admits a larger one. The function $f(x) = -x^2$ on $\mathbb{R}$ has a maximum at $x = 0$, but its analogue $g(x) = -x^2$ on the open set $(-1, 1)\backslash\{0\}$ does not – even though it is bounded above and continuous. The trouble in both cases is that the domain is not *compact.*

Theorem 9.3 (Extreme Value Theorem)**.** *Let B be a non-empty compact subset of a metric space, and let $f : B \to \mathbb{R}$ be continuous. Then f attains its maximum on B – i.e. there exist $x^* \in B$ with $f(x^*) \geqslant f(x)$ for all $x \in B$.*

The recipe for showing an economic optimisation problem has a solution is then short: verify (i) the feasible set is non-empty and compact, and (ii) the objective is continuous.

Econ Application: Anna's optimum exists

Take Anna with any continuous utility $u : \mathbb{R}^2_+ \to \mathbb{R}$ over her budget set

$$A = \{x \in \mathbb{R}^2 : x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_1 + 2x_2 \leqslant 10\}.$$

We prove that her optimisation problem has at least one solution.
A **is non-empty**: The bundle $0 = (0, 0)$ satisfies all three constraints, so $0 \in A$.
A **is bounded**: For any $x \in A$ we have $0 \leqslant x_1 \leqslant 10$ and $0 \leqslant x_2 \leqslant 5$, so $\|x\| \leqslant \sqrt{10^2 + 5^2} < 12$. This means that for any $x \in A$, the distance from origo is bounded $d(x, 0) < 12$. Hence $A \subseteq B(0, 12)$, and A is bounded.
A **is closed**: The complement of A is the union of three sets:

$$A^c = \underbrace{\{x \in \mathbb{R}^2 : x_1 < 0\}}_{A_1^c} \cup \underbrace{\{x \in \mathbb{R}^2 : x_2 < 0\}}_{A_2^c} \cup \underbrace{\{x \in \mathbb{R}^2 : x_1 + 2x_2 > 10\}}_{A_3^c}$$

With a similar reasoning as in **Exercise 9.5**, it can be shown that each A_1^c, A_2^c, A_3^c are open sets. By **Theorem 8.1**, A^c is open, so A is closed.
A **is compact**: Since A is closed and bounded in $\mathbb{R}^2$ with the usual metric, it is compact, by **Theorem 8.3**.
Conclusion. Since u is a continous function and A is a non-empty and compact set, u attains a maximum over A, by **Theorem 9.3**.
Anna's problem has at least one solution – and we have proven this without yet computing what the solution is. The gradient and convexity work in the chapters that follow will let us pin down exactly which bundle she chooses.

10 What are Derivatives?

Functions are mappings between metric spaces, and continuity indicates when small changes in the input result in small changes in the output. We now ask a sharper question: *how sensitive* is the output to the input? Suppose I consume x apples. What is the change in utility from consuming an infinitesimally small additional amount of apples? Does such an approximation exist?

10.1 Differentiability

In short, we say that a function is *differentiable* at a certain point if we can locally approximate it with a linear function.

Definition 10.1. *Let $f : \mathbb{R} \to \mathbb{R}$. We say that f is* ***differentiable*** *at a point x if the following limit exists:*

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

and we refer to $f'(x)$ as the ***derivative*** *of $f(x)$, also denoted as $\frac{df(x)}{dx}$.*

Exercise 10.1. *Using the above definition, show that $f'(x) = 2x$ for $f(x) = x^2$.*

We know that $f'(x) = 2x$, so at $x = 1$, the slope is $f'(1) = 2$. So $f(x) = x^2$ can be locally approximated near $x = 1$ by the linear function $g(x) = 2x - 1$, shown in **Figure 10.1**.

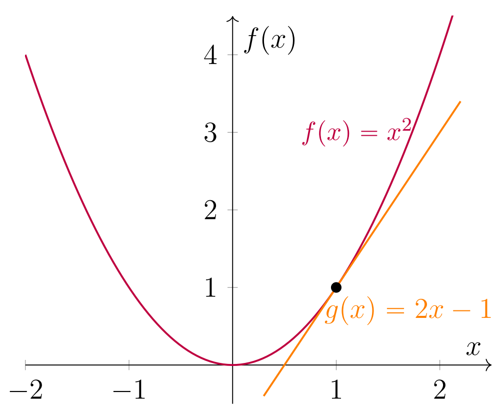


Figure 10.1: The function $f(x) = x^2$ and its linear approximation at $x = 1$.

Example 10.1. *The function $g : \mathbb{R} \to \mathbb{R}$ such that $g(x) = |x|$ is a continuous function (draw it!) has no derivative defined at $x = 0$: we can not approximate it by a linear function. We say that $g(x)$ is not differentiable at 0.*

10.2 Partial Derivatives

When a function depends on multiple variables, we can still talk about how it changes — but we examine one variable at a time, holding the others constant. For example, we call the change in function value of some function $f(x, y)$ from a infinitesmally small change in x the *partial derivative* of f with regards to x, denoted $\frac{\partial f(x,y)}{\partial x}$. Note that derivatives of single-variable functions are denoted $\frac{df}{dx}$, while functions of several variables use ∂ to indicate that only one variable (of many) is changing.

Definition 10.2 (Partial Derivative)**.** *Let $f(x_1, \ldots, x_n)$ be a real-valued function. The* ***partial derivative*** *of f w.r.t. x_i at the point $(a_1, \ldots, a_n)$ is defined as*

$$\frac{\partial f}{\partial x_i}(a_1, \ldots, a_n) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}.$$

Example 10.2 (Partial Derivatives)**.** *Consider $f(x, y) = -x^2 + 2xy - 2y^2$. To compute the partial derivative with respect to x, we fix y (treating it as a constant) and differentiate:*

$$\begin{aligned}\frac{\partial f}{\partial x} &= \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h} \\ &= \lim_{h \to 0} \frac{-(x + h)^2 + 2(x + h)y - 2y^2 - (-x^2 + 2xy - 2y^2)}{h} \\ &= \lim_{h \to 0}(-2x + 2y + h) = -2x + 2y.\end{aligned}$$

For the partial derivative with respect to y, we get $\frac{\partial f}{\partial y}(x, y) = 2x - 4y$.

10.3 Gradients

The *gradient* of a function is a vector that collects each partial derivative and thus points in the *direction of steepest ascent.*

Definition 10.3 (Gradient)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a point $a \in \mathbb{R}^n$. The* ***gradient*** *of f at a, denoted $\nabla f(a)$, is the vector of all partial derivatives. We write: $\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\right)$.*

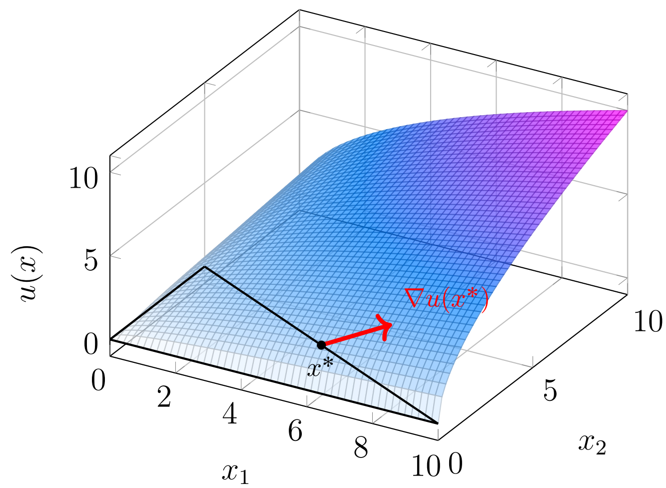
Suppose you are standing at a hill at position (x, y), where x is east–west coordinate, y is north–south coordinate and $f(x, y)$ is the elevation. The partial derivatives and tell you how steep the hill is if you take a tiny step east $\left(\frac{\partial f}{\partial x}\right)$ or north $\left(\frac{\partial f}{\partial y}\right)$. The gradient $\nabla f(x, y)$ points in the direction in which the hill rises most rapidly at (x, y).

Example 10.3 (Gradient). *For the function* $f(x, y) = 2xy - x^2 - 2y^2$*, the gradient is* $\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (-2x + 2y,\ 2x - 4y)$*. At* $(1, 1)$*, where* $f(1, 1) = -1$*, we get* $\nabla f(1, 1) = (0, -2)$*. This can be visualised as a vector in the tangent plane to the graph of* f*, anchored at the point* $(1, 1, -1)$*, pointing in the direction* $(0, -2, 0)$*.*

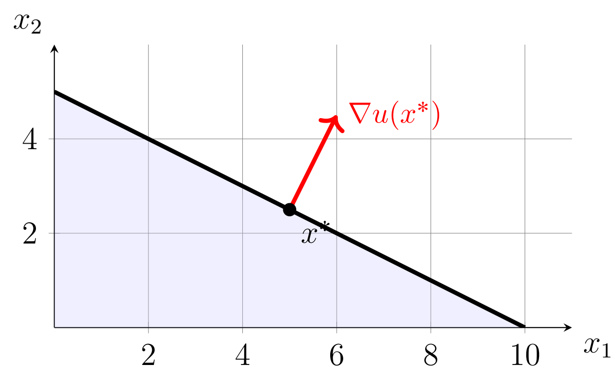
Exercise 10.2. *Consider Anna's utility* $u : \mathbb{R}^2_+ \to \mathbb{R}$ *of the more general form* $u(x) = x_1^\alpha x_2^{1-\alpha}$ *for some* $\alpha \in (0, 1)$*.*

a) Write the gradient $\nabla u(x)$ *as a function of* $u(x)$*. Interpret how the gradient changes with* x *and* α*.*

b) What is the value of the gradient at $x^* = (5, 2.5)$ *with* $\alpha = 0.5$*?*



(a) Surface $u(x) = \sqrt{x_1 x_2}$ with budget set and gradient at x^*.



(b) $\nabla u(x^*)$ at $x^* = (5, 2.5)$, perpendicular to the budget line.

Figure 10.2: Two views of Anna's gradient at her optimum.

11 What is Convexity?

Convexity is the property of sets and functions that, time and again, makes the difference between economic optimisation problems being well-behaved and being hopelessly tangled.

11.1 Convex sets

Intuitively, a set is convex if it has no "dents" or "holes" – pick any two points in the set, and the straight line between them stays inside the set. Formally:

Definition 11.1 (Convex set)**.** *A set $C \subseteq \mathbb{R}^n$ is* ***convex*** *if, for any two points $x, y \in C$ and any $\lambda \in [0, 1]$,*

$$\lambda x + (1 - \lambda)y \in C.$$

The expression $\lambda x + (1-\lambda)y$ is a point on the line segment between x and y: at $\lambda = 1$ we are at x, at $\lambda = 0$ we are at y, and as λ slides between the two, we trace out the segment. A rule of thumb to remember (shown below): circles are convex, bananas are not!

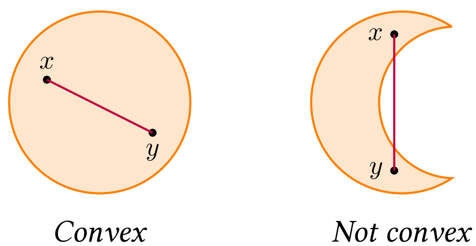


Figure 11.1: A convex set (left): the line segment between any two points stays inside. A non-convex set (right): some segments leave the set.

Econ Application: Anna's budget set is convex

Recall Anna's budget set $A = \{x \in \mathbb{R}^2 : x_1 + 2x_2 \leqslant 10\}$. For any two affordable bundles $x, y \in A$ and any $\lambda \in [0, 1]$, the mixed bundle $z = \lambda x + (1 - \lambda)y$ satisfies

$$z_1 + 2z_2 = \lambda(x_1 + 2x_2) + (1 - \lambda)(y_1 + 2y_2) \leqslant \lambda \cdot 10 + (1 - \lambda) \cdot 10 = 10,$$

and clearly $z_1, z_2 \geqslant 0$. So $z \in A$, and A is convex.

11.2 Convex and concave functions

Convexity for *functions* is closely related to convexity of sets. For convex functions any chord between two points on the graph lies above the curve. For concave functions, any chord between two points on the curve lies below it.

Definition 11.2 (Convex and concave functions)**.** *A function $f : \mathbb{R} \to \mathbb{R}$ is **convex** if, for any $x, y \in \mathbb{R}$ and any $\lambda \in [0, 1]$,*

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

*It is **concave** if $-f$ is convex.*

Example 11.1 (Convex and concave functions)**.** *The function $f(x) = x^2$ is convex. The function $g(x) = \ln x$ is concave.*

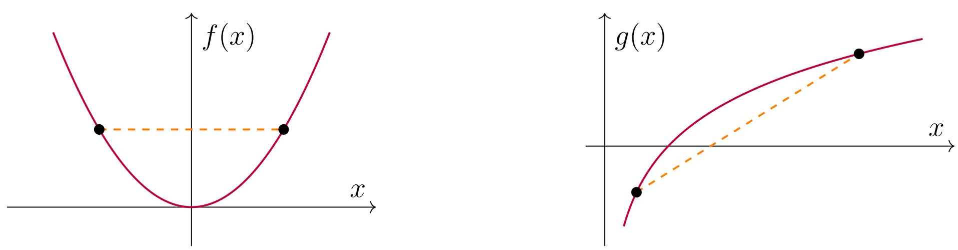


Figure 11.2: Left: $f(x) = x^2$ is convex. Right: $g(x) = \ln x$ is concave.

Think of the area above $f(x)$ and $g(x)$ in the above figures. We formally define this area as an *epigraph*:

Definition 11.3. *The epigraph of a function $f : \mathbb{R} \to \mathbb{R}$ is the set*

$$\text{epi}(f) = \{(x, t) \in \mathbb{R}^2 : t \geqslant f(x)\} \subseteq \mathbb{R}^2$$

11.3 Three ways to check the shape of a function

Suppose you stumble upon a function $f(x)$ in the wild. How to test if it is convex/concave?

(i) **Chord definition**: Show that the chord between any two points on the graph lies above (convex) or below (concave) the graph. Often quite tedious!

(ii) **Epigraph**: A function $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if epi(f) is a convex set.

(iii) **Derivative test.** For *differentiable* and *single-variable* functions: f is convex if $f''(x) \geqslant 0$ everywhere and concave if $f''(x) \leqslant 0$ everywhere.

The intuition for the derivative test is that f'' measures how the slope changes, and a convex function has a non-decreasing slope: the graph keeps curving upward, so the chord has to sit above it. But this test only works for a subset of functions you will encounter!

Example 11.2 (When the derivative test runs out)**.** *The function $g(x) = |x|$ is convex by route (i) — the chord clearly sits above the V-shape. But since g has no second derivative at 0, route (iii) is unavailable.*

Exercise 11.1. *Consider Anna's indifference curve for some utility level $\bar{u}$:*

$$\{x \in \mathbb{R}^2_+ : \sqrt{x_1 x_2} = \bar{u}\} \tag{11.1}$$

Along this indifference curve, we can describe the choice of x_2 as a function of x_1 (to keep the utility level constant) such that $f(x_1) = x_2$. In particular, this function is:

$$f(x_1) = \frac{\bar{u}^2}{x_1} \tag{11.2}$$

This is a differentiable single-variable function $f : (0, \infty) \to \mathbb{R}$. Show it is convex with each of the three routes explained above.

12 A First Look at Constrained Optimisation

Most of what economic theory do is constrained optimisation: an agent maximising an objective, subject to constraints. The bootcamp has equipped you with everything you need to set up and read such problems. We close by working through Anna's in full.

Your high-school maths gave you the rule "$f'(x) = 0$ at an optimum", which works perfectly when there are no constraints. With constraints, the situation is trickier: at Anna's optimum, $\nabla u(x^*) \neq 0$, so "set the gradient to zero" is not the right rule. The general tool is the *Lagrangian*, formalised through the *Karush–Kuhn–Tucker conditions*, introduced in great length in Mark's course. Intuitively, when a single constraint $h(x) \leqslant 0$ binds at the optimum, the gradient of the objective must be parallel to the gradient of the constraint:

$$\underbrace{\nabla u(x^*)}_{\text{direction of fastest utility increase}} = \underbrace{\lambda}_{\text{shadow price}} \, , \underbrace{\nabla h(x^*)}_{\text{direction of cost increase}} \, , \qquad \lambda \geqslant 0. \tag{12.1}$$

Geometrically, this forces $\nabla u(x^*)$ to lie perpendicular to Anna's budget line — and to point outward through it. The multiplier λ is the *shadow price* of the constraint: the marginal utility Anna would gain from a small relaxation of her budget. Below we apply the Lagrangian by hand, alongside a more simple method, to find Anna's optimum. Later on, you will learn how the convexity of the budget set, together with the concavity of the utility function, turns these first-order conditions from necessary into sufficient: guaranteeing that x^* *global optimum,* not just a candidate.

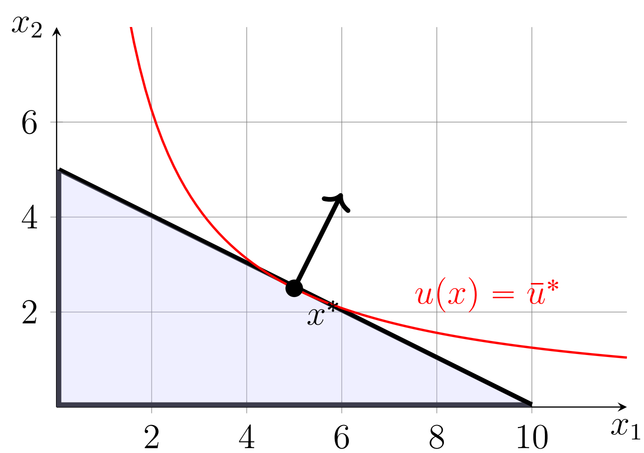


Figure 12.1: Anna's optimum $x^* = (5, 2.5)$ at the tangency between her indifference curve $u(x) = \bar{u}^*$ (red) and her budget line. The black arrow is $\nabla u(x^*)$.

Econ Application: Anna's optimal bundle

Anna's bundle is

$$A \equiv \{x \in \mathbb{R}^2 : x_1, x_2 \geqslant 0,\ x_1 + 2x_2 \leqslant 10\}$$

Anna's problem is

$$\max_{x \in A} \sqrt{x_1 x_2}.$$

From **Chapter 9**, we know that an optimum exists. Since $u(x_1, 0) = u(0, x_2) = 0$, while there exist a better feasible bundle $u(1,1) = 1$, the non-negativity constraints will not bind in optimum.

Method 1: Substitution. Her utility is strictly increasing in each good, so the budget constraint will bind in optimum: $x_1 + 2x_2 = 10$. Since the budget binds,

$$\begin{aligned} & x_1 = 10 - 2x_2 \\ \Rightarrow \quad & u(x_2) = \sqrt{(10 - 2x_2)x_2} \end{aligned}$$

Under some conditions, that will be stated more rigorously later on in the course, optimum can be found at:

$$u'(x_2) = \frac{10 - 4x_2}{2\sqrt{(10 - 2x_2)x_2}} = 0 \quad \Rightarrow \quad \begin{cases} x_2 = 2.5, \\ x_1 = 5. \end{cases} \tag{12.2}$$

Method 2: Lagrangian. The Lagrangian is $\mathcal{L} = \sqrt{x_1 x_2} - \lambda(x_1 + 2x_2 - 10)$. The first-order condition $\nabla \mathcal{L} = 0$ translates to:

$$\begin{cases} \frac{1}{2}\sqrt{x_2/x_1} = \lambda \\ \frac{1}{2}\sqrt{x_1/x_2} = 2\lambda \end{cases} \quad \Rightarrow \quad x_1 = 2x_2 \tag{12.3}$$

Substituting back into either FOC gives $\lambda = \frac{1}{2\sqrt{2}} > 0$. A positive multiplier means the budget constraint must be active (this is a *complementary slackness* condition, which will be stated precisely later in the course), so $x_1 + 2x_2 = 10$. Combined with $x_1 = 2x_2$, we get $x^* = (5, 2.5)$.

The Lagrangian is the standard tool of constrained optimisation in economics. Substitution is its more elementary cousin, useful when the active constraints are obvious and the problem collapses to a single variable. By the end of Math 5301, you will see both made rigorous and extended to multi-constraint settings via the Karush–Kuhn–Tucker conditions. Anna is the first of many agents you will solve for.

13 Concluding Words: The Game is Set

On these pages, we've seen some of the building blocks of mathematics. While it might be difficult to piece everything together, hopefully, you will soon realize that it's all about taking small pieces and putting them together in more and more complex ways. Just like chess games can be very complicated, it all stems from the rules of how to move the different pieces.

Look back, for instance, at Anna. By the time you reached the convexity chapter you had already met every tool her problem requires: bundles as vectors (**Chapter 6**), prices as inner products (**Chapter 6**), her problem as a system of linear equations (**Chapter 7**), the budget set as a closed and bounded subset of $\mathbb{R}^2$ (**Chapter 8**), utility as a continuous function (**Chapter 9**), the gradient as her direction of steepest ascent (**Chapter 10**), and the convexity of her indifference curve (**Chapter 11**). Mark's first lectures will revisit each piece more rigorously and give you a framework for understanding the environments of economic agents.

I hope you enjoyed this "boot camp" and that you feel more prepared for the Master's program in Economics. I'll see you after the summer, where we will begin our mathematics adventure by reviewing this material together and addressing any remaining questions that may have come up during your work with these notes.

14 Suggested Solutions

Exercise 2.1. The proof goes:
Base case: For $n = 1$, we have $2 = 1(1 + 1) = 2$, so true.
Inductive step: Assume that for $n = k$, we have $2 + 4 + \cdots + 2k = k(k + 1)$.
Induction proof: Using the inductive hypothesis:

$$2 + 4 + \cdots + 2k + 2(k + 2) = k(k + 1) + 2(k + 1) = (k + 1)(k + 2)$$

By induction, the formula holds for all $n \geqslant 1$. $\square$

Exercise 4.1:
($\Rightarrow$) If x is even, then $x = 2k$ for some $k \in \mathbb{Z}$. Squaring both sides,

$$x^2 = (2k)^2 = 4k^2 = 2(2k^2),$$

Since $2k^2 \in \mathbb{Z}$, we have that x^2 is divisible by 2, so x^2 is even.
($\Leftarrow$) If x^2 is even, then $x^2 = 2m$ for some $m \in \mathbb{Z}$. Suppose, for contradiction, that x is odd. Then $x = 2k + 1$ for some $k \in \mathbb{Z}$, so

$$x^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1,$$

Since $2k^2 + 2k \in \mathbb{Z}$, then x^2 is odd. But this contradicts that x^2 is even.
We conclude that x is even if, and only if, x^2 is even. $\square$
b.ii): Suppose, for contradiction, that $\sqrt{2}$ is rational. Then for some $p, q \in \mathbb{Z}$, not both even, we have

$$\sqrt{2} = \frac{p}{q} \quad \Rightarrow \quad 2 = \frac{p^2}{q^2} \quad \Rightarrow \quad 2q^2 = p^2 \tag{14.1}$$

Clearly, p^2 is even (as it is divisible by 2), so p is even. Since p is even, we can write $p = 2m$ for some $m \in \mathbb{Z}$. We have:

$$2q^2 = (2m)^2 \quad \Rightarrow \quad q^2 = 2m^2 \tag{14.2}$$

But then q is also even. This means that both p and q are even. But this is a contradiction, as we assumed that not both p and q were not even. We have arrived at a contradiction. $\square$

Exercise 6.1: Proofs for each sub-questions are below.
a): Since $0 \notin W_2$, it does not satisfy all the properties of a subspace. Since a set is a subspace if and only if it satisfies all three properties, we conclude that W_2 is not a subspace of $\mathbb{R}^2$. □
b): Consider $(0,1) \in W_2$ and $(1,0) \in W_2$. Their sum $(1,0) + (0,1) = (1,1) \notin W_2$. Since the set is not closed under addition, we conclude by **Theorem 6.1** that it is not a subspace of $\mathbb{R}^2$. □
c): Consider $\alpha' = 0.5 \in \mathbb{R}$ and $x' = (1,1) \in W_4$. Then $\alpha' x' = (0.5, 0.5) \notin W_4$. Since the subset is not closed under scalar multiplication, we conclude by **Theorem 6.1** that W_4 is not a subspace. □
d): Let us show that the set satisfies all three properties in **Theorem 6.1**.

(i) Since $3 \cdot 0 - 4 \cdot 0 = 0$, we have $(0,0) \in W_1$.

(ii) For any $x \in W_1$, we have $x_2 = \frac{3}{4}x_1$. Consider some $z, z' \in W_1$. Denote their sum $v \equiv z + z'$. Then

$$v \equiv z + z' = \begin{bmatrix} z_1 + z_1' \\ z_2 + z_2' \end{bmatrix} = \begin{bmatrix} z_1 + z_1' \\ \frac{3}{4}z_1 + \frac{3}{4}z_1' \end{bmatrix} = \begin{bmatrix} z_1 + z_1' \\ \frac{3}{4}\left[z_1 + z_1'\right] \end{bmatrix} \tag{14.3}$$

Thus $v_2 = \frac{3}{4}v_1$, so $v \in W_1$. We conclude that W_1 is closed under addition.

(iii) For any $x \in W_1$, we have $x_2 = \frac{3}{4}x_1$. Then for $v \equiv \alpha x$, we have $v_1 = \alpha x_1$ and $v_2 = \alpha x_2$. We can write $v_2 = \alpha \frac{3}{4} x_1$. So $v_2 = \frac{3}{4}v_1$. Thus, $v \in W_1$, so the set is closed under scalar multiplication.

We conclude that W_1 is a subspace of $\mathbb{R}^2$. □**e)**: Let us show that the set satisfies all three properties in **Theorem 6.1**.
(i) We can write the zero function $f_0(x) = 0$ as $f_0(x) = 0 + 0 \cdot x$, so $f_0 \in W_5$.
(ii) Consider any two vectors in W_5 and denote them $f(x) = a + bx \in W_5$ and $g(x) = c + dx \in W_5$, where $a, b, c, d \in \mathbb{R}$. Their sum is

$$(f+g)(x) = (a+c) + (b+d)x,$$

where $a + c \in \mathbb{R}$ and $b + d \in \mathbb{R}$. Hence $f + g \in W_5$, so W_5 is closed under addition.
(iii) Let $\alpha \in \mathbb{R}$ and $f(x) = a + bx \in W_5$, where $a, b \in \mathbb{R}$. Then

$$(\alpha f)(x) = \alpha a + \alpha b x,$$

where $\alpha a \in \mathbb{R}$ and $\alpha b \in \mathbb{R}$, so $\alpha f \in W_5$. Thus, W_5 is closed under scalar multiplication.
We conclude that W_5 is a subspace of $P_2(\mathbb{R})$. □

Exercise 7.1 a) We compute the transpose:

$$A^\top = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \quad \Rightarrow \quad A^\top A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 10 & 14 \\ 14 & 20 \end{bmatrix},$$

while

$$AA^\top = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 11 \\ 11 & 25 \end{bmatrix}.$$

So $A^\top A \neq AA^\top$.

b) Applying the inverse formula for A gives:

$$A^{-1} = \frac{1}{-2} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix}.$$

We verify that $AA^{-1} = I_2$:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix} = \begin{bmatrix} 1 \cdot (-2) + 2 \cdot 1.5 & 1 \cdot 1 + 2 \cdot (-0.5) \\ 3 \cdot (-2) + 4 \cdot 1.5 & 3 \cdot 1 + 4 \cdot (-0.5) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

c) We have that $\varepsilon'\varepsilon \in \mathbb{R}$ is a scalar:

$$\varepsilon'\varepsilon = \sum_{i=1}^{n} \varepsilon_i^2.$$

This product is the *residual sum of squares* — it summarises how far the predicted values are from the observed values. In contrast, $\varepsilon\varepsilon' \in \mathbb{R}^{n \times n}$ is an $n \times n$ matrix with elements:

$$(\varepsilon\varepsilon')_{ij} = \varepsilon_i \varepsilon_j.$$

This product is the *variance-covariance matrix.*

Exercise 7.2 a): Write in matrix form:

$$\begin{pmatrix} 1 & -1 & 1 & | & 8 \\ 2 & 3 & -1 & | & -2 \\ 3 & -2 & -9 & | & 9 \end{pmatrix} \tag{14.4}$$

Take $-2R_1$ in second row and $-3R_1$ in third row:

$$\begin{pmatrix} 1 & -1 & 1 & | & 8 \\ 0 & 5 & -3 & | & -18 \\ 0 & 1 & -12 & | & -15 \end{pmatrix} \tag{14.5}$$

Take $-5R_3$ in second row:

$$\begin{pmatrix} 1 & -1 & 1 & | & 8 \\ 0 & 0 & 57 & | & 57 \\ 0 & 1 & -12 & | & -15 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & -1 & 1 & | & 8 \\ 0 & 0 & 1 & | & 1 \\ 0 & 1 & -12 & | & -15 \end{pmatrix} \tag{14.6}$$

Take $+R_3$ to first row:

$$\begin{pmatrix} 1 & 0 & -11 & | & -7 \\ 0 & 0 & 1 & | & 1 \\ 0 & 1 & -12 & | & -15 \end{pmatrix} \tag{14.7}$$

Take $+11R_2$ to first row and $+12R_2$ to third row:

$$\begin{pmatrix} 1 & 0 & 0 & | & 4 \\ 0 & 0 & 1 & | & 1 \\ 0 & 1 & 0 & | & -3 \end{pmatrix} \tag{14.8}$$

We get:

$$\begin{cases} x = 4 \\ y = -3 \\ z = 1 \end{cases} \tag{14.9}$$

Exercise 7.2 b): Suppose, for contradiction, that the system has a solution. Write in matrix form:

$$\begin{pmatrix} 1 & -1 & | & 3 \\ 2 & -2 & | & 7 \end{pmatrix} \tag{14.10}$$

Now, let's eliminate the first column of the second row by subtracting $2R_1$ from the second row:

$$\begin{pmatrix} 1 & -1 & | & 3 \\ 0 & 0 & | & 1 \end{pmatrix} \tag{14.11}$$

This results in the equation $0 = 1$, which is a contradiction. Therefore, the system of equations has no solution.
c): Write in matrix form:

$$\begin{pmatrix} -1 & -2 & 1 & | & -1 \\ 2 & 3 & 0 & | & 2 \\ 0 & 1 & -2 & | & 0 \end{pmatrix} \tag{14.12}$$

To eliminate the first column of the second row, add $2R_1$ to R_2:

$$\begin{pmatrix} -1 & -2 & 1 & | & -1 \\ 0 & -1 & 2 & | & 0 \\ 0 & 1 & -2 & | & 0 \end{pmatrix} \tag{14.13}$$

Next, add R_2 to R_3:

$$\begin{pmatrix} -1 & -2 & 1 & | & -1 \\ 0 & -1 & 2 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \tag{14.14}$$

From the second row, we solve for y in terms of z. Substitute $y = 2z$ into the first equation to express z in terms of x:

$$-x - 2(2z) + z = -1 \quad \Rightarrow \quad x = 1 - 3z \tag{14.15}$$

Thus, the solution for y and z as functions of x is:

$$\begin{cases} y = \frac{2(1-x)}{3} \\ z = \frac{1-x}{3} \end{cases} \tag{14.16}$$

Exercise 8.1: Answers for each subquestion is presented below.
a): Wish to show that each element of A_1 is an interior point: for any $x \in A_1$, there exists some $\varepsilon > 0$ s.t. $B(x, \varepsilon) \subseteq A_1$. Let $x \in (0, \infty)$ and consider the open ball $B(x, \varepsilon) = \{y \in \mathbb{R} : |x - y| < \varepsilon\}$. Then for every point $y \in B(x, \varepsilon)$, we have:

$$|x - y| < \varepsilon \Rightarrow -\varepsilon < x - y < \varepsilon \Rightarrow x - \varepsilon < y < x + \varepsilon.$$

Define $\varepsilon = \frac{x}{2} > 0$. We get $y \in (x - \varepsilon, x + \varepsilon) = \left(\frac{x}{2}, \frac{3x}{2}\right) \subset (0, \infty)$. Hence, $B(x, \varepsilon) \subset A_1$. Since each element of A_1 is an interior point, it is open.
b): Consider $0 \in A_2$. For any $\varepsilon > 0$, we have $-\frac{\varepsilon}{2} \in B(0, \varepsilon)$ but $-\frac{\varepsilon}{2} \notin A_2$, so $B(0, \varepsilon) \nsubseteq A_2$. So 0 is not an interior point. Thus, A_2 is not open.
Suppose A_2 was closed. Then A_2^c is open. But 1 is not an interior point to A_2^c, as $1 - \frac{\varepsilon}{2} \in B(1, \varepsilon)$ but $1 - \frac{\varepsilon}{2} \notin A_2^c$. We conclude that A_2 is not closed.
c): $B_1 = [0, 1]$ is closed since its complement is open, as it is the union of two open intervals $B_1^c = (-\infty, 0) \cup (1, \infty)$. Similarly, B_2 is closed, since its complement $B_2^c = (-\infty, 0)$ is open.

Exercise 8.2: Any ball around $x' = (2, 4)$ will include points outside the budget set, e.g. $\left(2 + \frac{\varepsilon}{2}, 4\right) \in B(x', \varepsilon) \cap A^c$. In addition, $x' \in B(x', \varepsilon) \cap A$. Since $B(x', \varepsilon) \cap A^c \neq \varnothing$ and $B(x', \varepsilon) \cap A \neq \varnothing$ for any $\varepsilon > 0$, we conclude by **Theorem 8.2** that x' is a boundary point.

Exercise 9.1: Consider points $z \equiv 0$ and $z' \equiv -1$. We have $z > z'$. However, we have $f(z') > f(z)$, since $f(-1) = 1$ and $f(0) = 0$. Thus, f is not strictly increasing.
Next, suppose f was not strictly increasing on $[0, 1]$. That means that there exist some pair $x', x \in [0, 1]$ such that $x' = x + k$ for some $k > 0$ but

$$(x + k)^2 \leqslant x^2 \quad \Rightarrow \quad 2xk + k^2 \leqslant 0 \tag{14.17}$$

which is a contradiction, since $2xk + k^2 > 0$ when $x \in [0, 1]$.

Exercise 9.2 a): If $x_1 \in [0,1]$ and $x_2 \in [0,1]$, it follows that $f(x) = x_1 \in [0,1]$.

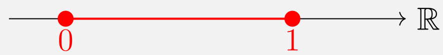


Figure 14.1: Image of $[0,1]^2$ under $f_1(x) = x_1$.

b): The pre-image is the “half-space” $\{x \in \mathbb{R}^2 : x_1 \geqslant 0\}$.

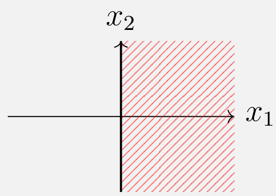


Figure 14.2: Pre-image of $[0,+\infty)$ under $f_1(x) = x_1$.

Exercise 9.3: Graph is shown below.

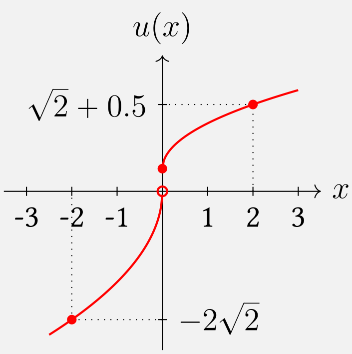


Figure 14.3: Prospect Theory Utility

If $u()$ is continuous at 0, then for each $\varepsilon > 0$, there is a $\delta > 0$ such that each $x \in X$ with $d(x,0) < \delta$ has $d'(u(x), u(0)) < \varepsilon$. Since this must hold for any $\varepsilon > 0$ if $u()$ is continuous, we simply need to find *one* counter-example. Consider $\varepsilon = \frac{1}{2}$ and let us pick the point $x' = -\frac{\delta}{2}$, so $d(x', x_0) < \delta$. Then $d(u(x'), u(0)) = d\left(-\sqrt{2\delta}, \frac{1}{2}\right) = \sqrt{2\delta} + \frac{1}{2} > \varepsilon$. We conclude that $u()$ is not continuous.

Exercise 9.4: Fix a point $x_0 \in \mathbb{R}$. For a given $\varepsilon > 0$, we wish to find a characterisation $\delta(\varepsilon)$, i.e. a δ as a function of ε, s.t. for any point $x \in \mathbb{R}$, we have

$$d(x, x_0) < \delta \quad \Rightarrow \quad d(f(x), f(x_0)) < \varepsilon. \tag{14.18}$$

First, note that our δ may not be contingent on the point x. Second, note that we wish to prove the implication (14.18): ***if*** $|x - x_0| < \delta$, ***then*** $|f(x) - f(x_0)| < \varepsilon$. So we are only concerned with what happens when $|x - x_0| < \delta$. For the case $|x - x_0| < \delta$, written with our usual metric, we wish to find a δ s.t.

$$|x - x_0| < \delta \quad \Rightarrow \quad |f(x) - f(x_0)| < \varepsilon. \tag{14.19}$$

Note that $|f(x) - f(x_0)| = |(2x + 1) - (2x_0 + 1)| = 2|x - x_0|$. It follows that $|f(x) - f(x_0)| < 2\delta$. To ensure that $|f(x) - f(x_0)| < \varepsilon$, we set

$$\delta = \frac{\varepsilon}{2}.$$

Under this choice of δ, the implication (14.18) follows. Since our choice of x_0 was arbitrary, this holds for any point on $\mathbb{R}$. This finalises the proof.

Exercise 9.5: The function $f : \mathbb{R}^2 \to \mathbb{R}$ s.t. $f(x_1, x_2) \equiv x_1$ is linear, thus continuous.
Argument I: Note that we can write $C = \{x \in \mathbb{R}^2 : f(x) \in (0, \infty)\}$, where $f(x) = x_1$. C is the pre-image of the open set $(0, \infty)$ under a continuous function f, thus open.
Argument II: The complement C^c is the pre-image of an closed set $(-\infty, 0]$ under a continuous function f, thus closed. Since the complement of C is closed, we conclude that C is open.

Exercise 9.6: Consider the open set $A = (0, 2.5)$. The pre-image of A under $u()$ is $u^{-1}(A) = \{x \in \mathbb{R} : u(x) \in (0, 2.5)\} = [0, 4)$, which is not an open set as the point 0 is not interior. See **Figure 14.4** for a visual interpretation of this argument.

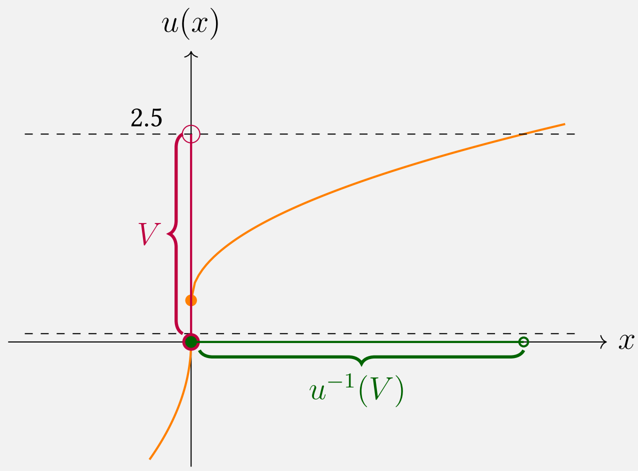


Figure 14.4: Pre-image $u^{-1}(A) = [0, 4)$ of open set $A = (0, 2.5)$ is not open.

Exercise 10.1: Using the definition of a derivative:

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \frac{2xh + h^2}{h} = \lim_{h \to 0} 2x + h = 2x \tag{14.20}$$

Exercise 10.2:
a): The partial derivatives are:

$$\frac{\partial u}{\partial x_1} = \alpha \cdot \frac{u(x_1, x_2)}{x_1} \quad \text{and} \quad \frac{\partial u}{\partial x_2} = (1-\alpha) \cdot \frac{u(x_1, x_2)}{x_2}.$$

So the gradient at $(x_1', x_2') > 0$ is:

$$\nabla u(x_1', x_2') = \left(\alpha \cdot \frac{u(x_1', x_2')}{x_1'}, (1-\alpha) \cdot \frac{u(x_1', x_2')}{x_2'}\right).$$

The gradient increases in α for the x_1 component and decreases for the x_2 component, reflecting that α is the output elasticity of x_1. For a given level of x_2', the x_1 component decreases in x_1', due to diminishing marginal returns.
b: At $\alpha = 0.5$, the utility level is $u(5, 2.5) = \sqrt{12.5}$. The gradient takes the value

$$\nabla u(5, 2.5) = \left(0.5 * \frac{\sqrt{12.5}}{5}, 0.5 * \frac{\sqrt{12.5}}{2.5}\right) = \sqrt{2} * (1, 2) \tag{14.21}$$

Exercise 11.1: We verify each function via each route in turn.
Chord definition: Using **Definition 11.2**, we wish to show that for any $x, x' \in \mathbb{R}_+$ and $\lambda \in [0, 1]$:

$$f(\lambda x + (1-\lambda)x') \leqslant \lambda f(x) + (1-\lambda) f(x') \tag{14.22}$$

$$\frac{\bar{u}^2}{\lambda x + (1-\lambda)x'} \leqslant \lambda \frac{\bar{u}^2}{x} + (1-\lambda)\frac{\bar{u}^2}{x'} \tag{14.23}$$

Let's multiply both sides with $\frac{\lambda x + (1-\lambda)x'}{\bar{u}^2}$:

$$1 \leqslant \lambda \frac{\lambda x + (1-\lambda)x'}{x} + (1-\lambda)\frac{\lambda x + (1-\lambda)x'}{x'} \tag{14.24}$$

$$xx' \leqslant (\lambda x' + (1-\lambda)x)(\lambda x + (1-\lambda)x') \tag{14.25}$$

$$xx' \leqslant xx' + \underbrace{\lambda(1-\lambda)\,(x - x')^2}_{\geqslant 0} \tag{14.26}$$

Derivative test: The function has a positive second derivative $f''(x) = 2/x^2 > 0$, thus convex.

Epigraph: The epigraph of $f(x) = \frac{\bar{u}^2}{x}$ is

$$E \equiv \left\{(x, t) \in \mathbb{R}^2_+ : t \geqslant \frac{\bar{u}^2}{x}\right\} \tag{14.27}$$

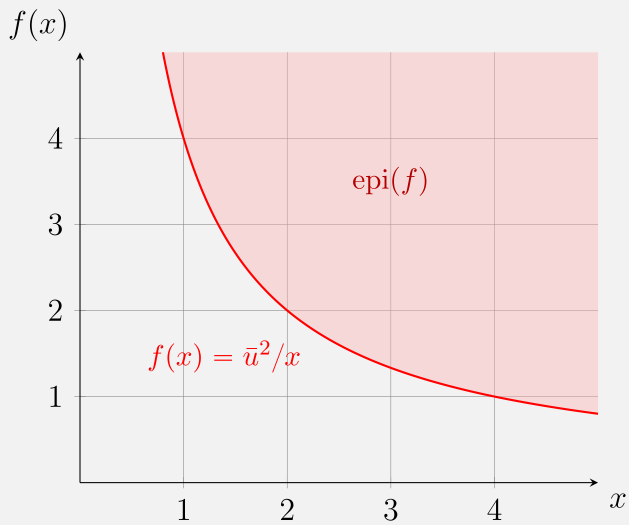


Figure 14.5: The epigraph of $f(x) = \bar{u}^2/x$ on $\mathbb{R}_+$ (shaded red): the set of all points lying weakly above the graph of f.

Using **Definition 11.3**, we wish to show that for any $y, y' \in E$ and $\lambda \in [0, 1]$:

$$\frac{\bar{u}^2}{\lambda y + (1-\lambda)y'} \leqslant \lambda\frac{\bar{u}^2}{y} + (1-\lambda)\frac{\bar{u}^2}{y'} \tag{14.28}$$

$$\frac{yy'}{\lambda y + (1-\lambda)y'} \leqslant \lambda y' + (1-\lambda)y \tag{14.29}$$

$$yy' \leqslant (\lambda y + (1-\lambda)y')(\lambda y' + (1-\lambda)y) \tag{14.30}$$

and the same logic as above for the chord definition gives the result. Since the epigraph is convex, f is convex.
Finally, note that the epigraph coincides with an upper-contour set in the (x_1, x_2) domain. Later in the course, you will learn the connection between the convexity of upper contour sets and a property of the utility function called *quasiconcavity*.