PhD Math Camp Lecture Notes[1]

Nathan Yoder
University of Georgia

August 8, 2018

[1]These notes are based on notes by Samson Alva (UTSA) and Meghan Skira (UGA).

Contents

Chapter 1

Proofs, Sets, and Functions

This chapter serves two purposes.

First, it gives an overview of basic topics in higher mathematics which are useful in economic theory. It discusses what it means for a function to be continuous, and for a set to be open, closed, or compact. It then explains how these concepts are useful in optimization theory, and in particular, in showing that an economic model possesses an equilibrium. While deep understanding of all of these topics may not be possible in the few days we have, I hope that students will gain an intuitive understanding of which sets are open/closed/compact, which functions are continuous, and what sorts of assumptions in an economic model are important for agents' choices to be well-behaved.

Second, it introduces students to writing mathematical proofs. This not only helps build familiarity with the material, it also teaches students to rigorously develop a logical argument, and to present that argument clearly and succinctly. This skill is useful not only in writing proofs, but in all academic writing.

1.1 Set Theory

Notation

$\subseteq$ - subset (or set inclusion)

$\subset$ - strict subset (or strict set inclusion)

$\supseteq$ - superset

$\supset$ - strict superset

$\in$ - element of

$\ni$ - contains

$\mathbb{N}$ - the set of natural numbers

$\mathbb{R}$ - the set of real numbers

A **set** is a collection of objects. We'll see sets frequently in the first year economics PhD sequence.

- Consumer theory considers the choice of an agent from a **budget set**.
- Likewise, producer theory considers the choice of a firm from a **production set**.
- In game theory, players choose from a set of **actions**.
  - If they can randomize, they choose from their set of mixed strategies — the set of **probability distributions over their set of actions**.
- In principal-agent models, the principal chooses an element of the set of contracts that satisfy participation and incentive compatibility constraints.

Though most sets we'll see in the description of economic models are sets of things that agents can choose from, the notion of a set is far more general. Put simply, it is a device for keeping track of lots of objects at once. That way, we just have to think about the set, rather than all the elements in the set.

This is especially useful when we define a set by some rule. For instance, a budget set in a two-commodity world:

$$\{(x,y)|p_x x + p_y y \leq w\}$$

This means "consumption bundles (x,y) that a consumer with wealth w can afford at prices p."

Sometimes we'll want to consider sets that contain all of the objects in one set A, as well as all of the objects in another B. We call this set the **union** of A and B and denote it $A \cup B$. Unions are associated with "or": If A is the set of elements that satisfy one rule, and B is the set of elements that satisfy another, $A \cup B$ is the set of elements that satisfy at least one of the two rules, and perhaps both. For instance, if

$$A = \{(x,y)|p_x x + p_y y \leq w\},$$
$$B = \{(x,y)|p'_x x + p'_y y \leq w'\}$$

then $A \cup B = \{(x,y)|p_x x + p_y y \leq w \text{ or } p'_x x + p'_y y \leq w'\}$ — the set of consumption bundles that can be bought at prices p and wealth w OR at prices p' and wealth w'.

We also might want to consider a set that contains just the objects that are in both A and B. We call this the **intersection** of A and B and denote it $A \cap B$. Intersections are

associated with "and": If A is the set of elements that satisfy one rule, and B is the set of elements that satisfy another, $A \cap B$ is the set of elements that satisfy both rules. For instance, if

$$A = \{(x, y) | p_x x + p_y y \leq w\},$$
$$B = \{(x, y) | p'_x x + p'_y y \leq w'\}$$

then $A \cap B = \{(x, y) | p_x x + p_y y \leq w \text{ and } p'_x x + p'_y y \leq w'\}$ — the set of consumption bundles that can be bought at prices p and wealth w AND at prices p' and wealth w'.

Sometimes, A and B have no elements in common. Is $A \cap B$ still a set? Yes — it's a special set called the **empty set** $\emptyset$, the set without any elements. (The fact that a set can have no elements is useful in lots of places, but I address it here because it's the answer to a natural question that arises when we ponder the nature of intersections for a moment.)

There's one other way to build a new set out of A and B, and that's by considering the set of all elements that are in A but are NOT in B, denoted $A \setminus B$. (Well, two, really: we could also consider the set $B \setminus A$ of all elements that are in B but not A.) These are called **set differences**. Returning to our budget set example, $A \setminus B$ is the set of consumption bundles that our consumer can afford at prices p and wealth w, but NOT at prices p' and wealth w'.

Just like we sometimes make statements about how two numbers are related to one another by saying one is greater than another, we'll also want to make statements about how two sets are related to one another by saying one is a **subset** of the other. When we say that A is a subset of B — denoted $A \subseteq B$ — it means that every object in A is also in B. And just like we say two numbers are equal if they represent the same quantity, we'll write $A = B$ when A and B contain the same objects. (Similarly to the relationship between $<$ and $\leq$, we'll also write $A \subset B$ to mean that $A \subseteq B$ but $A \neq B$.)

Now, frequently when we talk about sets, we'll think of them as living in some larger set. For example, the budget sets we just discussed all live in the Euclidean plane $\mathbb{R}^2$. This brings us to the concept of the **universal set** X: the set containing all elements that are available in whatever context we're considering. All of the other sets of elements in that context are subsets of X.

This notion of a universal set allows us to define something else: the set of all elements that are not in a set A. This is called A's **complement**, and denoted $A^C \equiv X \setminus A$. As we can see, A's complement ("all objects not in A") depends crucially on what we picked as our universal set ("all objects"). Sometimes people will say things like "A's complement in X" to make it clear what universal set they are referring to.

Finally, we'll often want to take two disjoint sets or spaces and consider them side-by-side. For instance, I might have one set of colors, $C = \{r, g, b\}$, and another set of shapes, $S = \{\square, \triangle, \diamondsuit\}$, and want to form a set whose elements have both a color and shape, e.g. $\{\square, \triangle, \diamondsuit, \square, \triangle, \diamondsuit, \square, \triangle, \diamondsuit\}$. I'd do this by taking the **Cartesian product** of C and S, denoted $C \times S$. Similarly, I might want to take a set of quantities of good x (e.g., $\mathbb{R}$) and a set of quantities of good y (e.g., $\mathbb{R}$) and take their Cartesian product to get the set of bundles of x and y (e.g., $\mathbb{R}^2$).

1.2 Logic and Proofs

In mathematics, a **statement** is a sentence that is either true or false. A statement might be something like "$p_x \leq p'_x$ and $p_y \leq p'_y$" or "letting $S \equiv \{(x,y)|p_x x + p_y y \leq w\}$ and $T \equiv \{(x,y)|p'_x x + p'_y y \leq w'\}$, $S \cap T = \{(x,y)|p_x x + p_y y \leq w\}$".

For each statement P, we can form another statement $\neg P$ ("NOT P") which is the statement "P is false."

A **proposition** is a statement of the form "If P is true, then Q is true", where P and Q are statements. For instance, we can combine the two statements above to form the proposition

> "If $p_x \geq p'_x$, $p_y \geq p'_y$, $S \equiv \{(x,y)|p_x x + p_y y \leq w\}$, and $T \equiv \{(x,y)|p'_x x + p'_y y \leq w'\}$, then $S \cap T = \{(x,y)|p_x x + p_y y \leq w\}$."

Propositions are often written as "If P, then Q" or "P **implies** Q" or "$P \implies Q$".

Propositions are sometimes given other names that indicate their importance. Calling a proposition a **theorem** indicates that it is very important. Sometimes, because of the length of an argument for a theorem, the proof is broken into stages, with each linking proposition being proved as a **lemma**. So, lemmata (plural of lemma) are propositions whose subjective import derives not necessarily from its statement but from its role as a stage in the overarching construction of a proof of a theorem. However, occasionally a lemma has importance independent of the theorem for which it was constructed. Lastly, **corollaries** are propositions that follow almost immediately from a theorem; the proof of such a statement is usually trivial, but the subjective value of the knowledge of its truth is not.

Each proposition $P \implies Q$ has three related propositions that can be derived from it. First, the **converse** of $P \implies Q$ is the reverse implication $Q \implies P$. Importantly, whether $P \implies Q$ is true does not depend on whether its converse $Q \implies P$ is true: if P is "student a attended and participated in each ECON 8000 lecture and quiz" and

Q is "student a received the full 50% for their participation grade", then $P \implies Q$ is true but $Q \implies P$ is false (since you can miss up to three quizzes before it impacts your participation grade).

Second, P is **equivalent** to Q, or $P \Leftrightarrow Q$, if $P \implies Q$ and $Q \implies P$ are both true. We sometimes say "P if and only if Q".

Finally, the **contrapositive** of $P \implies Q$ is the statement $\neg Q \implies \neg P$. The contrapositive is important because it is essentially a different way of stating $P \implies Q$: the two statements are logically equivalent. However, it is sometimes easier to prove $\neg Q \implies \neg P$ than $P \implies Q$.

Example 1. We can see why the statements $\neg Q \implies \neg P$ and $P \implies Q$ are equivalent through the use of a **truth table**. A truth table tells us which truth values for the underlying statements P and Q are consistent with a given statement and which are not.

We start by making a table with columns for each of the statements whose truth we want to evaluate:

P	Q	$P \implies Q$	$\neg Q \implies \neg P$
T	T		
T	F		
F	T		
F	F		

We next determine whether each pair of truth values for P and Q is consistent with each statement (in which case we write T in the statement's column) or is not (in which case we write F:

P	Q	$P \implies Q$	$\neg Q \implies \neg P$
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

This lets us see that both statements are equivalent to "$\neg(P$ and $\neg Q)$".

1.2.1 Proving a Proposition Directly

A proof is an argument that a proposition follows from statements that are known to be true: e.g., definitions, axioms, lemmas that you have already proven, theorems that others have proven. Proofs of a proposition $P \Rightarrow Q$ can be broken up into three classes.

The first of these are **direct proofs**, which consist of a sequence of arguments, each following directly from the one before, connecting P and Q, so that we can write

$$P \Rightarrow P_1 \Rightarrow P_2 \Rightarrow \cdots \Rightarrow P_n \Rightarrow Q.$$

Example 2. Let's prove a couple propositions directly.

Proposition 1.1. *If $S \subseteq T$ and $T \subseteq S$, then $S = T$.*

Proof.

1.	$S \subseteq T$.	Hypothesis.
2.	$T \subseteq S$.	Hypothesis.
3.	$x \in S \Rightarrow x \in T$.	(1); definition of $\subseteq$.
4.	$x \in T \Rightarrow x \in S$.	(2); definition of $\subseteq$.
5.	$x \in S \Leftrightarrow x \in T$.	(3) and (4).
6.	$S = T$.	(5); definition of $=$.

□

Proposition 1.2. *If $S \subseteq T$, then $S \cap T = S$.*

Proof.

1.	$S \subseteq T$.	Hypothesis.
2.	$x \in S \Rightarrow x \in T$.	(1); definition of $\subseteq$.
3.	$x \in S \Rightarrow x \in S \cap T$.	(2); definition of $\cap$.
4.	$S \subseteq S \cap T$.	(3); definition of $\subseteq$.
5.	$x \in S \cap T \Rightarrow x \in S$.	Definition of $\cap$.
6.	$S \cap T \subseteq S$.	(5); definition of $\subseteq$.
7.	$S = S \cap T$.	(4); (6); Proposition 1.1.

□

We'll prove this next proposition through the use of a lemma. First, recall that multiplying by a positive number preserves inequalities. (We'll take this as an axiom — call it "A1.1".). Second, recall that inequalities can be added together. (Again, we won't prove this; instead, we'll take it as an axiom — call it "A1.2".)

Lemma 1.1. *If $p_x \geq p'_x$, $p_y \geq p'_y$, $S \equiv \{(x,y) \in \mathbb{R}^2_+ | p_x x + p_y y \leq w\}$ and $T \equiv \{(x,y) \in \mathbb{R}^2_+ | p'_x x + p'_y y \leq w'\}$, then $S \subseteq T$.*

Proof.

1.	$p_x \geq p'_x$ and $p_y \geq p'_y$.	Hypothesis.
2.	$(x,y) \in S \Rightarrow x \geq 0$.	Hypothesis.
3.	$(x,y) \in S \Rightarrow y \geq 0$.	Hypothesis.
4.	$(x,y) \in S \Rightarrow p_x x \geq p'_x x$.	(1); (2); A1.1.
5.	$(x,y) \in S \Rightarrow p_y y \geq p'_y y$.	(1); (3); A1.1.
6.	$(x,y) \in S \Rightarrow p_y y + p_x x \geq p'_y y + p'_x x$.	(4);(5); A1.2.
7.	$(x,y) \in S \Rightarrow w \geq p_y y + p_x x$.	Hypothesis.
8.	$(x,y) \in S \Rightarrow w \geq p'_y y + p'_x x$.	(6); (7); transitivity of $\geq$.
9.	$(x,y) \in S \Rightarrow (x,y) \in T$.	(8); hypothesis.
10.	$S \subseteq T$.	(9); definition of $\subseteq$.

□

Proposition 1.3. *If $p_x \geq p'_x$, $p_y \geq p'_y$, $S \equiv \{(x,y)|p_x x + p_y y \leq w\}$ and $T \equiv \{(x,y)|p'_x x + p'_y y \leq w'\}$, $S \cap T = \{(x,y)|p_x x + p_y y \leq w\}$.*

Proof.

1.	$S \subseteq T$.	Hypothesis; Lemma 1.1.
2.	$S \cap T = S$.	(1); Proposition 1.2.
3.	$S \cap T = \{(x,y)|p_x x + p_y y \leq w\}$.	(2); hypothesis.

□

You'll notice that — except for taking the properties of addition and multiplication of inequalities as axioms, because hopefully we all learned them in grade school — we're being super careful to leave out as few steps as possible. If I were writing a proof in a paper, I'd leave out a lot of these steps, because they are obvious enough that I don't really need to argue them — the reader should be able to follow my logic. Moreover, I wouldn't write them as a numbered list like this. But for now, when we're first learning how to write proofs, this extremely thorough way of writing them can help ensure that we don't make any unjustified logical leaps.

1.2.2 Conditional Statements and Quantifiers

A statement that is true or false conditional on the value of one or more variables is a **conditional statement**. E.g. $x^2 + 3y = 5$. Most statements one encounters are conditional statements. It is important to note that a conditional statement has a determinate truth value, conditional on the values of each of the variables upon which it depends.

Many statements combine a conditional statement with a **quantifier**, which delineates the scope or domain in which the truth of the conditional statement holds. There are two types of quantifiers: **existential** and **universal**. The existential quantifier can be recognized by the use of words such as "there exists" or "there is/are", and can be denoted

by $\exists$. When a statement includes such a quantifier, its truth is determined by the possibility of constructing or otherwise proving the existence of at least *one object satisfying* the conditional statement. E.g.,

"There exist x, y such that $x^2 + 3y = 5$"

is a statement, and furthermore is a true statement, since $x^2 + 3y = 5$ for $x = 1, y = \frac{4}{3}$ allows for the truth of the conditional statement. Therefore, the only way for a statement with a single, existential quantifier to be false is for there to be *no object* that satisfies the conditional statement.

Notice that there is an hidden assumption in the previous example. I argued that the statement is true by constructing an example (**proof by construction**). However, I assumed that $x, y \in \mathbb{R}$. This is neither allowed nor disallowed by the statement, which implies that the truth value of the statement is itself conditional on the domain of the variables x and y. Suppose the statement read "There exist x, y such $x^2 + 3y = 5$ and $x, y \in \mathbb{N}$". Then, the statement would be false, since there is no pair of values for the variables that would satisfy the expression and the domain restrictions.

The universal quantifier can be recognized by the use of words such as "for all/every/any" and can be denoted by $\forall$. "For some" is not a universal quantifier, but an existential quantifier, even though the word "for" appears. When a statement includes such a quantifier, its truth is determined by the possibility of constructing or otherwise proving the existence of at least *one object not satisfying* the conditions of the statement. Any such object would prove the statement false. E.g. (For all x, y, $x^2 + 3y = 5$) is false since $x = 1, y = 1$ yields a conditionally false statement.

Also, notice that some of the propositions used in the proofs from the previous example can be reformulated as conditional statements with universal quantifiers: e.g., $((x, y) \in S \Rightarrow p_x x \geq p'_x x)$ can also be written $(\forall (x, y) \in S, p_x x \geq p'_x x)$.

Sometimes, theorems or other statements involve the negation of quantifiers. Any statement of the form "$\neg(\forall x, P(x))$", where $P(x)$ is a conditional statement, can be written as "$\exists x$, such that $\neg P(x)$". Thus, when negated, the universal quantifier becomes an existential quantifier, with the attached conditional statement becoming negated. A similar algorithm allows for the negation of an existential quantifier: "$\neg(\exists x$, such that $P(x))$" is equivalent to "$\forall x, \neg P(x)$".

1.2.3 Indirect Proofs

There are two main methods of indirect proof. One is **proof by contrapositive**: since we already showed that $P \implies Q$ is equivalent to $\neg Q \implies \neg P$, one way of proving the

former is by formulating a direct proof of the latter. This is often really useful when P and Q both have existential quantifiers: instead of showing, e.g., $(\forall x, P(x)) \implies (\forall y, Q(y))$, we can instead show that $(\exists x, \neg Q(x)) \implies (\exists y, \neg P(y))$, which may be easier since we only need to find one such y.

Another common method of proof that may be employable is **proof by induction**. Induction works to prove statements formed by combining a conditional statement — call it $P(n)$ — with a universal quantifier over the natural numbers $\mathbb{N}$. E.g., things like

"For all $n \in \mathbb{N}$, $n^2 \geq n$."

Proofs of $(\forall n \in \mathbb{N}, P(n))$ by induction consist of two steps. The **initial step** proves $P(1)$. Then, the **induction step** proves $(\forall n \in \mathbb{N}, P(n) \implies P(n+1))$. Together, these two steps prove $P(n)$ for each n: The initial step tells us that $P(1)$ is true, and the induction step forms a direct proof of $P(1) \implies P(n)$:

$$P(1) \implies P(2) \implies \cdots \implies P(n-1) \implies P(n).$$

Example 3. Let's use induction to prove the statement noted earlier:

Proposition 1.4. *For all $n \in \mathbb{N}$, $n^2 \geq n$.*

Proof. Initial step: $1 = 1^2$, by multiplicative identity $(\forall x, 1 \cdot x = x)$.
Induction step (proof of $n^2 \geq n \implies (n+1)^2 \geq n+1$):

1.	$n^2 \geq n$.	Induction hypothesis.
2.	$n \geq 0$.	Definition of $\mathbb{N}$.
3.	$2n \geq 0$.	(2); A1.1.
4.	$(n+1)^2 = n^2 + 2n + 1 \geq n^2 + 1$.	(4); A1.2.
5.	$n^2 + 1 \geq n + 1$.	(1); A1.2.
6.	$(n+1)^2 \geq n + 1$.	(4); (5); transitivity of $\geq$.

□

1.2.4 Set Theory/Proofs Homework

1. Simon & Blume Exercise A1.3 (p.857)

2. Simon & Blume Exercise A1.7 (p.858)

1.2.5 Set Theory/Proofs Homework Solutions

1. (a)

Proposition 1.5. $x \in (A \cup B)^C \Leftrightarrow x \in A^C \cap B^C$.

Proof. $x \in (A \cup B)^C \Rightarrow x \in A^C \cap B^C$:

1.	$x \in (A \cup B)^C$.	Hypothesis.
2.	$x \notin A \cup B$.	(1); definition of complement.
3.	$x \notin A$ and $x \notin B$.	(2); definition of union.
4.	$x \in A^C$ and $x \in B^C$.	(3); definition of complement.
5.	$x \in A^C \cap B^C$.	(4); definition of intersection.

$x \in A^C \cap B^C \Rightarrow x \in (A \cup B)^C$:

1.	$x \in A^C \cap B^C$.	Hypothesis.
2.	$x \in A^C$ and $x \in B^C$.	(1); definition of intersection.
3.	$x \notin A$ and $x \notin B$.	(2); definition of complement.
4.	$x \notin A \cup B$.	(3); definition of union.
5.	$x \in (A \cup B)^C$.	(4); definition of complement.

□

(b) follows similarly.

(c)

Proposition 1.6. $x \in A \cap (B \cup C) \Leftrightarrow x \in (A \cap B) \cup (A \cap C)$.

Proof. $x \in A \cap (B \cup C) \Rightarrow x \in (A \cap B) \cup (A \cap C)$:

1.	$x \in A \cap (B \cup C)$.	Hypothesis.
2.	$x \in A$ and $x \in B \cup C$.	(1); definition of intersection.
3.	$x \in A$ and either $x \in B$ or $x \in C$.	(2); definition of union.
4.	Either $x \in A$ and $x \in B$ or $x \in A$ and $x \in C$.	(3).
5.	Either $x \in A \cap B$ or $x \in A \cap C$.	(4); definition of intersection.
6.	$x \in (A \cap B) \cup (A \cap C)$.	(5); definition of union.

$x \in (A \cap B) \cup (A \cap C) \Rightarrow x \in A \cap (B \cup C)$:

1.	$x \in (A \cap B) \cup (A \cap C)$.	Hypothesis.
2.	Either $x \in A \cap B$ or $x \in A \cap C$.	(1); definition of union.
3.	Either $x \in A$ and $x \in B$ or $x \in A$ and $x \in C$.	(2); definition of intersection.
4.	$x \in A$ and either $x \in B$ or $x \in C$.	(3).
5.	$x \in A$ and $x \in B \cup C$.	(4); definition of union.
6.	$x \in A \cap (B \cup C)$.	(5); definition of intersection.

□

2.

Proposition 1.7. *If* $n \in \mathbb{N}$, *then* $2^n > n$.

Proof. Initial step: $2^1 = 2 > 1$.
Induction step (proof of $2^n > n \implies 2^{n+1} > n+1$):

1.	$2^n > n$.	Induction hypothesis.
2.	$2^{n+1} > 2n$.	(1); A1.1.
3.	$n \geq 1$.	Definition of $\mathbb{N}$.
4.	$2n \geq n+1$.	(3); A1.2.
5.	$2^{n+1} > n+1$.	(2);(4).

□

1.3 Functions

Frequently, we're going to want to map elements in one set X (for instance, levels of the inputs to a firm's production process) into elements in another set Y (e.g., output levels associated with those inputs). We'll call X the map's **domain** and Y its **codomain**. We can represent this map visually by making a list of each element in the domain on one side, and another list of each element in the codomain on the other, and drawing links between them. For instance:

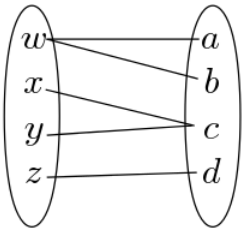

Figure 1.1: A map.

The most common kind of map we'll consider is one where each element in the domain maps to exactly one element in the codomain; i.e., no elements of the domain have more than one link. If the map satisfies this restriction, we call it a **function**. (We'll talk about other kinds of maps later.) On our diagram:

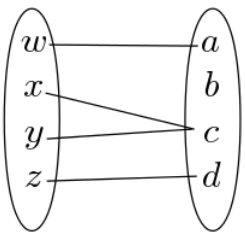

Figure 1.2: A function f.

We write a function f from domain X to codomain Y as $f : X \to Y$. If f maps an element $x \in X$ to $y \in Y$ we write $f(x) = y$.

While the function maps elements to elements, sometimes we want to take a subset of the domain $S \subseteq X$ and denote the set of elements in Y that elements $x \in S$ map to more easily than writing out $\{y \in Y | \exists x \in S, f(x) = y\}$. We call this the **image of** S **under** f, and write $f(S)$. For instance, suppose f is a production function, and S is a set of possible input levels: then $f(S)$ is the set of output levels that are feasible for the firm. If S is the entire domain X, we just call $f(X)$ the **image of** f. We can visualize this concept on our diagram:

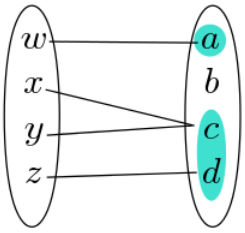

Figure 1.3: The image of f.

Another useful concept is the **preimage** or **inverse image** of some set $T \subseteq f(X)$ under f: the set of things in X that f maps to elements in T. We write this as $f^{-1}(T) \equiv \{x \in X | f(x) \in T\}$. This is a concept we use all the time in economics: if our function is a utility function, then the preimage of a number y under that function — the set of consumption bundles that give us utility of y — is an indifference curve. If f is a production function, then $f^{-1}(y)$ is an isoquant.

We can once again visualize this on our diagram:

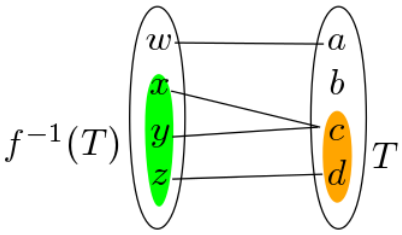

Figure 1.4: A set T and its preimage under f.

There are a few important properties that a function can have that I want to discuss now. First, if each element in the domain maps to a different element in the codomain, we say that the function is **one-to-one** or **injective**. On our diagram, this means that no element in the codomain list has more than one link attached to it:

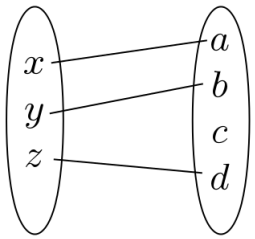

Figure 1.5: An injective function.

Formally, $f : X \to Y$ is injective if for all $x, z \in X$, $f(x) = f(z)$ implies $x = z$.

Injectivity means that the preimage of any one-element set $\{y\}$ under f, denoted $f^{-1}(y)$, is also a one-element set. So in other words, the preimage is itself a function from the image of f to the domain X — it maps each element in $f(X)$ to exactly one element in X — which we call the **inverse** of f and write $f^{-1} : f(X) \to X$.

Next, if f maps some element in X to every element in Y — that is, if $f(X) = Y$ — we say that it is **onto** or **surjective**. And if it's both injective and surjective, we say that it's **bijective**: each element of X maps to one element in Y, and each element in Y is mapped to by one element in X.

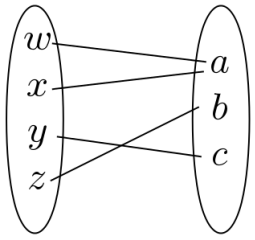

Figure 1.6: A surjective function.

This leads us to some conclusions that we will use later.

Proposition 1.8. *Given a function $f : X \to Y$:*

1. *For all $S \subseteq X$, $S \subseteq f^{-1}(f(S))$. If f is injective, $S = f^{-1}(f(S))$.*
2. *For all $T \subseteq Y$, $f(f^{-1}(T)) \subseteq T$. If f is surjective, $f(f^{-1}(T)) = T$.*

Proof. 1. We want to show that for each $x \in S$, $x \in f^{-1}(f(S))$. By definition of image, $f(x) \in f(S)$. By definition of preimage, since $f(x) \in f(S)$, $x \in f^{-1}(S)$. So $S \subseteq f^{-1}(f(S))$.

Now we want to show that if f is injective, for each $x \in f^{-1}(f(S))$, $x \in S$. By definition of preimage, $f(x) \in f(S)$. Since f is injective, the only element which f maps to $f(x)$ is x — and so we must have $x \in S$.

2. We want to show that for each $y \in f(f^{-1}(T))$, $y \in T$. By definition of image, $y = f(x)$ for some $x \in f^{-1}(T)$. By definition of preimage, f must map x to some element of T; i.e., $y \in T$.

   Now we want to show that if f is surjective, for each $y \in T$, $y \in f(f^{-1}(T))$. Since f is surjective, $y = f(x)$ for some $x \in X$. Then $x \in f^{-1}(T)$. By definition of image, since $f(x) = y$, $y \in f(f^{-1}(T))$.

□

1.3.1 Functions Homework

Given two functions $f : X \to Y$ and $g : Y \to Z$, their **composition** $g \circ f$ is given by the rule $g \circ f(x) = g(f(x))$.

1. Prove that if g and f are both injective, then $g \circ f$ is injective.

2. Prove that if g and f are both surjective, then $g \circ f$ is surjective.

1.3.2 Functions Homework Solutions

1. Since g is injective, for each $z \in Z$, there is at most one $y \in Y$ such that $g(y) = z$. And since f is injective, there is at most one x such that $f(x) = y$. Thus, there is at most one x such that $g(f(x)) = z$.

2. Since f is surjective, $f(X) = Y$. Since g is surjective, $g(f(X)) = g(Y) = Z$. Thus $g \circ f$ is surjective.

1.4 Sequences

A **sequence** is a function whose domain is $\mathbb{N}$. Instead of writing $s : \mathbb{N} \to X$, we'll usually identify this function using its image, whose elements we index by n and write $\{s_n\}$. For now, we'll consider the sequences in $\mathbb{R}$; later, we'll extend these ideas to more general spaces.

Some examples of sequences are in the textbook.

Usually, we are interested in sequences that approach a point as n increases. We say that a sequence $\{s_n\}$ in $\mathbb{R}$ **converges** to a real number s if for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that whenever $n \geq N$, $|s_n - s| < \epsilon$. That is, we can pick a distance ϵ as small as we want, and eventually, all further elements of the sequence will be within ϵ of s. When $\{s_n\}$ converges to s, we write $s_n \to s$ or $\lim s_n = s$.

A sequence that does not converge is said to **diverge**.

Example 4 (Examples of Sequences)**.** See Examples 12.1-12.2 in the book, p.254-255.

1.4.1 Properties of Sequences

A sequence can have at most one limit. There's a proof in the textbook, but it's needlessly complex, and it does not easily generalize to sequences in $\mathbb{R}^n$. We'll prove it using a fact called the **triangle inequality**:

$$\forall x, y \in \mathbb{R}, |x + y| \leq |x| + |y|.$$

The triangle inequality is, in fact, one of the fundamental properties of a **metric**, or measure of distance between elements of a set. Some manipulation of the triangle inequality yields its cousin, the **negative triangle inequality**:

$$\forall a, b \in \mathbb{R}, |a - b| \geq ||a| - |b|| .$$

Formally, we want to prove

Proposition 1.9. *If $\{s_n\}$ is a sequence in $\mathbb{R}$ with $s_n \to s$, and $s_n \to t$, then $s = t$.*

Note that this is one of those propositions that could be replaced with a conditional statement with a universal quantifier: we could write $(\forall \{s_n\} \subset \mathbb{R}$ such that $s_n \to s$ and $s_n \to t$, $s = t)$. These propositions are frequently most easily proven by contrapositive, which here can be written

Proposition 1.10. *If $s \neq t$, then there is no $\{s_n\}$ such that $s_n \to s$ and $s_n \to t$.*

Given the form of the contrapositive, this method is sometimes called **proof by contradiction**.

Proof.

1.	$s \neq t$.	Hypothesis.
2.	$\lvert s - t\rvert > 0$.	(1); definition of absolute value.
3.	$\lvert s_n - s\rvert < \lvert s - t\rvert/2 \implies -\lvert s_n - s\rvert > -\lvert s - t\rvert/2$.	A1.2.
4.	$\lvert s - t\rvert = \lvert t - s\rvert$.	Definition of absolute value.
5.	$\lvert s_n - s\rvert < \lvert s - t\rvert/2 \implies \lvert t - s\rvert - \lvert s_n - s\rvert > \lvert s - t\rvert/2$.	(3);(4); A1.2.
6.	$\lvert s_n - s\rvert < \lvert s - t\rvert/2 \implies \lvert t - s_n\rvert > \lvert s - t\rvert/2$.	(5); negative triangle inequality.
7.	$(\forall n \geq N, \lvert s_n - s\rvert < \lvert s - t\rvert/2) \implies (\forall n \geq N, \lvert t - s_n\rvert > \lvert s - t\rvert/2)$.	(6); universal quantifier.
8.	$(\forall n \geq N, \lvert s_n - s\rvert < \lvert s - t\rvert/2) \implies (\nexists n \geq N, \lvert t - s_n\rvert \leq \lvert s - t\rvert/2)$.	(7); $(\forall n, P(n)) \Leftrightarrow (\nexists n, \neg P(n))$.
9.	$s_n \to s \implies s_n \nrightarrow t$.	Defnition of convergence.

□

Sometimes, for one reason or another, we'll want to skip elements of a sequence. The result is what's called a **subsequence**. Formally, a subsequence of $\{s_n\}$ is given by the sequence $\{s_{n_m}\}$ for some increasing sequence $\{n_m\} \subseteq \mathbb{N}$.

Even if we can't say whether a sequence converges, we might still want to say that it at least doesn't get arbitrarily large or small. So we say that a sequence $\{s_n\}$ is **bounded** if $\exists M \in \mathbb{R}$ such that $|s_n| \leq M, \forall n$.

A sequence (a_n) is a **Cauchy sequence** if, $\forall \epsilon > 0, \exists N \in \mathbb{N}$ such that $\forall m, n \geq N, \|a_n - a_m\| < \epsilon$.

The following properties of sequences in $\mathbb{R}$ are frequently useful.

1. Every convergent sequence is bounded.

2. Let $\lim a_n = a$ and $\lim b_n = b$. Then,

   (a) $\lim ca_n = ca, \forall c \in \mathbb{R}$

   (b) $\lim(a_n + b_n) = a + b$

   - Proof: See Theorem 12.2 in the book, p.257. (We'll do this one in class)

   (c) $\lim(a_n b_n) = ab$

   - Proof: See Theorem 12.3 in the book, p.258.

   (d) $\lim(a_n / b_n) = a/b$, if $b \neq 0, b_n \neq 0 \forall n$

   (e) $(a_n \leq b_n, \forall n) \implies a \leq b$

   (f) $(\exists c \in \mathbb{R}, c \leq b_n, \forall n) \implies c \leq b$. A similar statement with the inequalities reversed also holds. See Theorem 12.4 in the book, p.259, for a proof.

3. Every monotone (i.e., increasing or decreasing) and bounded sequence converges.

4. Subsequences of a convergent sequence converge to the same limit as the original sequence. (Why?)

5. *(Bolzano-Weierstrass Theorem)* Every bounded sequence has a convergent subsequence.

6. *(Cauchy Criterion)* A sequence converges if and only if it is a Cauchy sequence.

1.4.2 Sequences in Larger Spaces

Along with the triangle inequality, the properties of absolute value that we used in step 2 ($s = t \Leftrightarrow |s - t| = 0$) and in step 4 ($|s - t| = |t - s|$) of our proof of Proposition 1.10 are the fundamental properties of a metric. In another metric space (such as $\mathbb{R}^n$) with a distance measure that satisfies these properties (in $\mathbb{R}^n$, generally the Euclidean metric $||\mathbf{x} - \mathbf{y}|| = \sqrt{(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})}$) we generalize the notion of a convergent sequence correspondingly: a sequence $\{\mathbf{s}_n\}$ in $\mathbb{R}^n$ converges to $\mathbf{s}$ if $\forall \epsilon > 0$, $\exists N$ such that $||\mathbf{s}_n - s|| < \epsilon$ for all $n \geq N$. We generalize the ideas of boundedness and Cauchy sequence accordingly.

Lots of the properties we talked about regarding sequences in $\mathbb{R}$ carry over to $\mathbb{R}^n$. The uniqueness of limits is one of these: in fact, we can just replace $|$ with $||$ throughout our proof and it will go through for sequences in $\mathbb{R}^n$. Convergent sequences continue to be bounded; the Cauchy criterion can still be used to evaluate whether a sequence converges. In fact, on our list of properties, (1),(2a), (2b), (4), (5), and (6) all hold. (For the others, we'd have to generalize monotonicity, multiplication, or division to $\mathbb{R}^n$, and we're not there yet.)

1.4.3 Sequences Homework

1. Do Simon & Blume Exercise 12.7, and use the result to prove property 2(d). Hint: For all $x, y \in \mathbb{R}$, $|y||x| = |yx|$.

1.4.4 Sequences Homework Solutions

1. First, we prove a lemma: $\exists B > 0$ such that $|b_n| \geq B$ for all n. Since $b_n \to b$, $\exists M \geq 0$ such that $|b - b_n| < |b|/2 \forall n \geq M$. Then by A1.2,

$$|b - b_n| + |b_n| < |b|/2 + |b_n|,$$

and by the triangle inequality,

$$|b| < |b|/2 + |b_n| \Rightarrow |b|/2 < |b_n|.$$

Then for $B = \min\{|b|/2, |b_1|, |b_2|, \ldots, |b_M|\}$, $|b_n| \geq B$ for all n.

Next, we need another lemma: there exists M such that $b_n b > 0$ for all $n \geq M$. We know $\exists M \geq 0$ such that $|b - b_n| < |b|/2 \forall n \geq M$. Then by definition of absolute value, $b - b_n \leq |b - b_n| < |b|/2$ and $b_n - b \leq |b - b_n| < |b|/2 \forall n \geq M$. Then if $b > 0$

and so $|b| = b$, we have $b - b_n < b/2 \Rightarrow 0 < b/2 < b_n$; hence by A1.1, $bb_n > 0$. If $b < 0$ and so $|b| = -b$, we have $b_n - b < -b/2 \Rightarrow b_n < b/2 < 0$; hence $-b_n > 0$ and $-b > 0$, so $bb_n > 0$ by A1.1.

Next, we need to prove that $1/b_n \to 1/b$. Since $b_n \to b$, for all $\delta > 0$, $\exists N$ s.t. $|b_n - b| < \delta \forall n \geq N$. We want to show that given $\epsilon > 0$, we can choose δ and use the resulting inequality to show that $|\frac{1}{b_n} - \frac{1}{b}| < \epsilon$. Now in order to go from $|b_n - b|$ to $|\frac{1}{b_n} - \frac{1}{b}|$, we need to multiply both sides by $1/(b_n b)$, and to have this be positive so that it (a) doesn't change the direction of the inequality and (b) can pass through the absolute value brackets. We showed we can do this for $n \geq M$ in the second preceding lemma. Once we multiply by $1/(b_n b)$, we're going to have the inequality

$$|\frac{1}{b_n} - \frac{1}{b}| \leq \frac{\delta}{bb_n} = \frac{\delta}{|b||b_n|} \leq \frac{\delta}{|b|B}$$

So we want to pick $\delta = |b|B\epsilon$, and then we know that for all $n \geq \max\{M, N\}$, $|\frac{1}{b_n} - \frac{1}{b}| < \epsilon$.

So that's how we would figure out how to write the proof. Here's the proof itself:

Given $\epsilon > 0$, there exists N such that $|b_n - b| < B|b|\epsilon \forall n \geq N$. By the second preceding lemma, there exists M such that $\frac{1}{bb_n} > 0 \forall n \geq M$. Then for all $n \geq \max\{M, N\}$,

$$|\frac{1}{b_n} - \frac{1}{b}| \leq \frac{B|b|\epsilon}{bb_n} = \frac{B|b|\epsilon}{|b||b_n|} \leq \epsilon,$$

where the last inequality follows from the first preceding lemma.

1.5 Open, Closed, and Compact Sets

When we introduced the concept of convergent sequences, we adopted a definition of "close to" that depended on distance: we said that a sequence $\{s_n\}$ gets "close to" a point s if we could choose any distance ϵ and all but finitely many elements of the sequence would be less than ϵ away from s. But there's another, more abstract way of specifying "close to" that is not only more general, but (indirectly) gives us some tools that we'll need when we start talking about optimization.

1.5.1 Open Sets

Let's think back to $\mathbb{R}$. Another way of saying "within ϵ of s" is "in the interval $(s - \epsilon, s + \epsilon)$". You may recall that the parentheses denote the fact that such an interval is

open: it does not include its endpoints.

Now note that for *every* open interval I that contains s, there is some $\epsilon > 0$ such that $(s - \epsilon, s + \epsilon) \subseteq I$. So the statements

> "For all $\epsilon > 0$, there exists N such that whenever $n \geq N$, $|s_n - s| < \epsilon$."

and

> "For every open interval I containing s, there exists N such that whenever $n \geq N$, $s_n \in I$."

are equivalent. (Exercise: Write out a formal proof.)

In other words, we can represent the idea of "closeness" equally well with measures of distance, or by specifying which subsets of the appropriate universal set are **open**. This collection of sets is called a **topology**, and when given a topology, the universal set X is called a **topological space**. For a sequence $\{s_n\}$ in an arbitrary topological space, then, we say $s_n \to s$ if for every open set S containing s, there exists N such that whenever $n \geq N$, $s_n \in S$.

Our specification of the open sets has to satisfy some basic properties, or it won't work very well. These are as follows:

1. The empty and universal sets are open.
2. Arbitrary (i.e., either finite or infinite) unions of open sets are open.
3. Finite intersections of open sets are open.

It's usually impossible to directly specify all of the open sets in a space. For example, while we know that open intervals are open, the above axioms mean that *arbitrary unions* of open sets have to be open as well. So instead, we'll usually specify a **basis** for the topology (such as the open intervals in $\mathbb{R}$), and then say that the open sets are those that can be formed by taking unions of the basis elements. (Note that this only works if intersections of basis elements themselves contain basis elements — which is true for open intervals.)

The topology of a metric space (a space endowed with some notion of distance) can always be described through use of a basis consisting of the **ϵ-balls**: sets taking the form of all points within distance ϵ of some other point. In $\mathbb{R}^n$, for instance, an ϵ-ball about x takes the form $B_\epsilon(\mathbf{x}) = \{\mathbf{y} | \, ||\mathbf{x} - \mathbf{y}|| < \epsilon\}$.

Now, in your book, you'll note that Simon and Blume instead define open sets as sets S for which every point $\mathbf{x} \in S$ is contained in an ϵ-ball $B_\epsilon(\mathbf{x})$ about itself that is also contained in S. This definition is equivalent.

Proposition 1.11. *$S \subseteq \mathbb{R}^n$ is open $\Leftrightarrow \forall \boldsymbol{x} \in S, \exists B_\epsilon(\boldsymbol{x}) \subseteq S$.*

Proof. We first need a lemma:

Lemma 1.2. *If $\boldsymbol{y} \in B_\epsilon(\boldsymbol{x})$, then there exists $\gamma > 0, B_\gamma(\boldsymbol{y}) \subseteq B_\gamma(\boldsymbol{x})$.*

Proof. The proof of this lemma is identical to the proof of Theorem 12.7 in Simon and Blume. □

Let's start with the "if" or $\Leftarrow$ direction. By hypothesis, every point $\mathbf{x} \in S$ is contained in an open ball $V_\mathbf{x} \subseteq S$. Then the union of all the $V_\mathbf{x}$, $V \equiv \bigcup_{\mathbf{x} \in S} V_\mathbf{x}$, is open. Since each $\mathbf{x} \in S$ is contained in $V_\mathbf{x}$, V contains the union of all the $\mathbf{x}$ (which is just S): $S \subseteq V$. Since each $V_\mathbf{x}$ is contained in S, so is their union: $V \subseteq S$. Thus, $V = S$, so S is open.

Now for the "only if" or $\Rightarrow$ direction. Since the topology of $\mathbb{R}^n$ is defined using a basis, and S is open, it can be written as the union of basis elements $B_\epsilon(\mathbf{y})$. Thus, each $\mathbf{x} \in S$ is contained in such a $B_\epsilon(\mathbf{y}) \subseteq S$. By the preceding lemma, for each $\mathbf{x} \in S$, there is some $\gamma > 0$ such that

$$\mathbf{x} \in B_\gamma(\mathbf{x}) \subseteq B_\epsilon(\mathbf{y}) \subseteq S,$$

as desired. □

Example 5 (Examples of Open Sets)**.**

- Cartesian products of open intervals are open: Let $U = (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_k, b_k)$. For any point $\mathbf{x} \in U$, $B_\epsilon(\mathbf{x}) \subseteq U$ for $\epsilon = \min\{|x_1 - a_1|, |x_2 - a_2|, \ldots, |x_k - a_k|, |x_1 - b_1|, |x_2 - b_2|, \ldots, |x_k - b_k|\}$.
- More broadly, "things that don't contain their boundaries".

1.5.2 Closed Sets

So now we hopefully have some intuitive idea of what sets are open, and what "open" means. Another type of set that is going to be useful are the **closed** sets. The definition of a closed set T in a topological space X is simple: it's just a set whose complement T^C is open. (Remember that the complement is defined relative to the universal set X.)

Our axioms for open sets imply similar statements about closed sets.

Proposition 1.12. 1. *The empty and universal sets are closed.*

2. *Arbitrary (i.e., either finite or infinite) intersections of closed sets are closed.*

3. *Finite unions of closed sets are closed.*

Proof. (1) follows from the fact that $X^C = \emptyset$. (3) follows from the finite version of the generalized De Morgan's laws that we proved in class: $\bigcap_{i=1}^n T_i^C = \left(\bigcup_{i=1}^n T_i\right)^C$. When T_i are closed, we know that the intersection on the left-hand side is open, since it is the finite intersection of open sets. Thus, $\bigcup_{i=1}^n T_i$ is closed.

De Morgan's laws actually apply to arbitrary intersections/unions, however — which gives us (2). □

One of the most important facts about closed sets is that if a sequence of points in a closed set T converges, it converges to a point in T.

Proposition 1.13. *If T is closed, $\{s_n\} \subseteq T$, and $s_n \to s$, then $s \in T$.*

Proof. We prove by contradiction. Suppose $s \notin T$. Then $s \in T^C$, which is open. But if $\{s_n\} \subseteq T$, then there exists no N such that $s_n \in T^C$ for $n \geq N$. Then $s_n \nrightarrow s$. □

Example 6 (Examples of Closed Sets)**.**

- The nonnegative real numbers $[0, \infty)$.
- The integers are a closed subset of $\mathbb{R}$: Their complement is the union of open intervals $(n, n+1)$.
- The positive orthant (including the x and y axes) $\mathbb{R}^2_+$: For any $(x, y) \notin \mathbb{R}^2_+$, choose $\epsilon = \min\{|x|, |y|\}$. Then $B_\epsilon((x, y)) \subseteq (\mathbb{R}^2_+)^C$. So $(\mathbb{R}^2_+)^C$ is open.
- Closed rectangles in $\mathbb{R}^k$.
- **Half-spaces**: sets defined by linear inequalities, i.e., $\{\mathbf{x} \in \mathbb{R}^k | \mathbf{x} \cdot b \leq a\}$. (One can apply a similar technique as with rectangles and orthants to show this, but we would need to know some linear algebra... Alternatively, we can use continuity.)

1.5.3 Compact Sets

For optimization theory, there's a third important property that sets can have. We say that a set S is **compact** if for every collection of open sets that **cover** S (i.e., whose union contains S) there is a finite subcollection of those sets that *also* cover S.

This definition is a little complicated, but it has the advantage of being both very general and frequently useful. However, identifying compact sets is much easier in $\mathbb{R}^n$,

where compact sets are precisely those that are both closed and bounded. (A set $S \subset \mathbb{R}^n$ is bounded if there exists M such that $||\mathbf{x}|| < M$ for all $\mathbf{x} \in S$.) This is called the Heine-Borel Theorem.

Compact sets are useful because they behave as though they are finite sets (hence the word compact). In economics, we often assume that the sets we are working with are compact, particularly because of the Weierstrass Theorem, which we shall see later.

Example 7 (Examples of Compact Sets)**.**

- Budget sets in $\mathbb{R}^k$. (They are bounded, and the intersection of half-spaces.)
- The simplex $\{\mathbf{x} \in \mathbb{R}^k : \mathbf{x} \cdot \mathbf{1} = 1\}$. (Similarly.)

1.5.4 Open/Closed/Compact Sets Homework

1. Prove that if $S \subseteq \mathbb{R}^k$ is closed, and $T \subseteq \mathbb{R}^m$ is closed, then $S \times T \subseteq \mathbb{R}^{k+m}$ is closed.

   Hint: Recall that for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $||\mathbf{x} - \mathbf{y}|| = \sqrt{(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})}$. Suppose that we decompose two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{k+m}$ into its components $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{R}^k$ and $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{R}^m$. Then

$$\begin{aligned} ||\mathbf{x} - \mathbf{y}|| &= \sqrt{(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})} = \sqrt{(\mathbf{x}_1 - \mathbf{y}_1) \cdot (\mathbf{x}_1 - \mathbf{y}_1) + (\mathbf{x}_2 - \mathbf{y}_2) \cdot (\mathbf{x}_2 - \mathbf{y}_2)}, \\ &= \sqrt{||\mathbf{x}_1 - \mathbf{y}_1||^2 + ||\mathbf{x}_2 - \mathbf{y}_2||^2}. \end{aligned}$$

2. Prove that the topological and analytic definitions of convergence are equivalent for sequences in $\mathbb{R}^k$.

1.5.5 Open/Closed/Compact Sets Homework Solutions

1. For $\mathbf{x} \notin S \times T$, either $\mathbf{x}_1 \notin S$ or $\mathbf{x}_2 \notin T$. Since S^C and T^C are open, there either exists ϵ_1 such that $B_{\epsilon_1}(\mathbf{x}_1) \subseteq S^C$ **or** ϵ_2 such that $B_{\epsilon_2}(\mathbf{x}_2) \subseteq T^C$. Choose $\epsilon = \min\{\epsilon_1, \epsilon_2\}$. Then $||\mathbf{x} - \mathbf{y}|| < \epsilon \Leftrightarrow \sqrt{||\mathbf{x}_1 - \mathbf{y}_1||^2 + ||\mathbf{x}_2 - \mathbf{y}_2||^2} < \epsilon \Leftrightarrow ||\mathbf{x}_1 - \mathbf{y}_1||^2 + ||\mathbf{x}_2 - \mathbf{y}_2||^2 < \epsilon^2$. Then if $\mathbf{x}_1 \notin S$, $||\mathbf{x}_1 - \mathbf{y}_1|| < \epsilon \leq \epsilon_1$, and if $\mathbf{x}_2 \notin T$, $||\mathbf{x}_2 - \mathbf{y}_2|| < \epsilon \leq \epsilon_2$. Then either $\mathbf{y}_1 \notin S$ or $\mathbf{y}_2 \notin T$. Then $\mathbf{y} \notin S \times T$ for all $\mathbf{y} \in B_\epsilon(\mathbf{x})$. The proposition follows.

2. (Topological $\implies$ Analytic) Suppose $\mathbf{s}_n \to \mathbf{s}$ under the topological definition. For all $\epsilon > 0$, the ϵ-ball about s, $B_\epsilon(\mathbf{s})$, is open. Then there exists N such that $\mathbf{s}_n \in B_\epsilon(\mathbf{s}) \implies ||\mathbf{s}_n - \mathbf{s}|| < \epsilon$ for all $n \geq N$.

(Analytic $\implies$ Topological) Suppose $\mathbf{s}_n \to \mathbf{s}$ under the analytic definition. For each open set S such that $\mathbf{s} \in S$, by Proposition 1.11, there exists $\epsilon > 0$ such that $B_\epsilon(\mathbf{s}) \subseteq S$. Then there exists N such that $||\mathbf{s}_n - \mathbf{s}|| < \epsilon \Leftrightarrow \mathbf{s}_n \in B_\epsilon(\mathbf{s}) \subseteq S$.

1.6 Continuity

Now that we have defined the ideas of sequences and open sets, we can introduce another property of functions that is of extreme importance in economics. This property is called **continuity**, and it means, in a very formal way, that the function doesn't jump around: as Simon and Blume say, "you can draw its graph without lifting your pencil from the paper."

We can define continuity in four main ways, all of which are equivalent. The first definition follows naturally from the quotation above: a function $f : \mathbb{R} \to \mathbb{R}$ is continuous if its graph is a closed subset of $\mathbb{R}^2$.

Next is the sequential definition of continuity: a function $f : X \to Y$ is continuous if for every sequence $\{s_n\} \subseteq X$ such that $s_n \to s$ for some s, the sequence $\{f(s_n)\} \subseteq Y$ converges to $f(s)$. (This is the definition given in the book.)

The third is the classical analytic definition: a function $f : \mathbb{R}^k \to \mathbb{R}^m$ is continuous if for each $\mathbf{x}_0 \in \mathbb{R}^k$ and every $\epsilon > 0$, there exists a $\delta > 0$ such that $||\mathbf{x} - \mathbf{x}_0|| < \delta$ implies $||f(\mathbf{x}) - f(\mathbf{x}_0)|| < \epsilon$.

Finally, the topological definition, which makes up what it lacks in intuitiveness by simplicity and ease of use: a function $f : X \to Y$ is continuous if for every open set $S \subseteq Y$, its preimage $f^{-1}(S) \subseteq X$ is open.

(Exercise: prove that the analytic and topological definitions are equivalent.)

Example 8 (Sums, Differences, Products, Cartesian products of Continuous Functions)**.** See Theorems 13.4 and 13.5 (p.294-5) in the book.

Example 9 (Continuous and Discontinuous Functions)**.**

- Example 13.8 (p.294) in the book.
- The identity is continuous (trivially).
- Linear combinations of continuous functions are continuous: Recall that $ax_n \to ax$ for $x_n \to x$, so if $f_i : \mathbb{R}^k \to \mathbb{R}$ is continuous, $\mathbf{x}_n \to \mathbf{x} \implies f_i(\mathbf{x}_n) \to f_i(\mathbf{x}) \implies af_i(\mathbf{x}_n) \to af_i(\mathbf{x})$, so $af_i : \mathbb{R}^k \to \mathbb{R}$ is continuous. By Theorem 13.5 in the book, $a\mathbf{f} : \mathbb{R}^k \to \mathbb{R}^m$ is continuous. Then by Theorem 13.4 in the book, $a_1\mathbf{f}_1 + a_2\mathbf{f}_2 + \cdots + a_n\mathbf{f}_n$ is continuous.

- Polynomials are continuous: they are linear combinations of products of the identity.
- The absolute value function is continuous.

1.6.1 Continuity and Compactness

Continuity and compactness interact in a way that is very useful for optimization. Namely, the image of a compact set under a continuous function is also compact.

Proposition 1.14. *If $f : X \to Y$ is continuous, and $T \subseteq X$ is compact, then $f(T) \subseteq Y$ is compact.*

Proof. Consider a collection of open sets $\{S_i\}_{i \in I}$ in Y that cover $f(T)$. Their preimages $f^{-1}(S_i)$ are each open (since f is continuous) and cover T (since by definition, each $f^{-1}(S_i)$ must contain all elements of X that f maps into S_i, and so their union must contain all elements of X that f maps into *some* S_i — which includes all the elements of T). Since T is compact, finitely many of the $f^{-1}(S_i)$ also cover T; relabel their indices $i = 1, 2, \ldots, N$.

Thus for all $x \in T$, $x \in f^{-1}(S_i)$ for $i \in \{1, 2, \ldots, N\}$; by definition of the preimage, then, $f(x) \in S_i$. Since for all $y \in f(T)$, there exists $x \in T$ such that $y = f(x)$, it follows that $\forall y \in T$, $y \in S_i$ for some $i \in \{1, 2, \ldots, N\}$. Thus $\{S_1, S_2, \ldots, S_N\}$ cover $f(T)$.

□

This fact leads to **Weierstrass' Theorem**.

1.6.2 Continuity Homework

1. (not assigned) Prove that the topological and analytic definitions of continuity are equivalent.

1.6.3 Continuity Homework Solutions

1. (Topological $\implies$ Analytic) Suppose $f : \mathbb{R}^k \to \mathbb{R}^m$ is continuous under the topological definition. Then since ϵ-balls are open, for each $\mathbf{x}_0 \in \mathbb{R}^k$ and each $\epsilon > 0$, $f^{-1}(B_\epsilon(f(\mathbf{x}_0)))$ is open. By Proposition 1.11, there is some $\delta > 0$ such that $B_\delta(\mathbf{x}_0) \subseteq f^{-1}(B_\epsilon(f(\mathbf{x}_0)))$, and thus $||\mathbf{x} - \mathbf{x}_0|| < \delta$ implies $||f(\mathbf{x}) - f(\mathbf{x}_0)|| < \epsilon$.

   (Analytic $\implies$ Topological) Suppose $f : \mathbb{R}^k \to \mathbb{R}^m$ is continuous under the analytic definition. Let $T \subseteq \mathbb{R}^m$ be open. For each $\mathbf{t}_0 \in T$, and each $\mathbf{x}_0 \in f^{-1}(\mathbf{t}_0)$, and

each $\epsilon > 0$, there exists $\delta > 0$ such that $||\mathbf{x} - \mathbf{x}_0|| < \delta$ implies $||f(\mathbf{x}) - \mathbf{t}_0|| < \epsilon$. Equivalently, there exists $\delta > 0$ such that $\mathbf{x} \in B_\delta(\mathbf{x}_0)$ implies $f(\mathbf{x}) \in B_\epsilon(\mathbf{t}_0)$. Or more usefully, $B_\delta(\mathbf{x}_0) \subseteq f^{-1}(B_\epsilon(\mathbf{t}_0))$.

By Proposition 1.11, for all $\mathbf{t}_0 \in T$, $\exists \epsilon$, $B_\epsilon(\mathbf{t}_0) \subseteq T$. Take the union of all these ϵ-balls and call it U; clearly $U \subseteq T$. Since $\mathbf{t}_0 \in U$ for all $\mathbf{t}_0 \in T$, $T \subseteq U$. So $U = T$. Now take the union of the corresponding δ-balls in $\mathbb{R}^k$ and call it V. For each of these δ-balls $B_\delta(\mathbf{x}_0)$, we have from the above that $B_\delta(\mathbf{x}_0) \subseteq f^{-1}(B_\epsilon(\mathbf{t}_0)) \subseteq f^{-1}(T)$, so $V \subseteq T$. By construction, V contains each $\mathbf{x}_0 \in f^{-1}(T)$, so $f^{-1}(T) \subseteq V$. So $V = f^{-1}(T)$. V is a union of basis elements, so it is open — and thus so is $f^{-1}(T)$.

1.7 Optimization: Theory

Economics is fundamentally about agents making decisions, and usually economic models assume they do so to maximize some objective function over a set of possible choices. I.e., economic agents frequently solve variations on the problem

$$\max_{x \in T} f(x)$$

where f maps some set X into $\mathbb{R}$, and C is some subset of X.

The theory of optimization tells us when this problem is guaranteed to have a solution, and how that solution changes with continuous changes in C or f.

1.7.1 Existence of a Maximum: Weierstrass' Theorem

Weierstrass' Theorem tells us that an optimization problem has a solution when the choice set is compact and the objective function is continuous. To prove it, we need to introduce a fundamental property of the real numbers: Any bounded set $S \subseteq \mathbb{R}$ has a **least upper bound**, or **supremum** — the smallest b such that $b \geq x$ for all $x \in S$. This number is denoted $\sup S$.

Proposition 1.15 (Weierstrass' Theorem)**.** *If $f : X \to \mathbb{R}$ is continuous, and $T \subseteq X$ is compact, there is some $x^* \in T$ such that for all $x \in T$, $f(x) \leq f(x^*)$.*

Proof. $f(T)$ is compact; therefore it is closed and bounded. Then $\sup T$ exists, and T^C is open.

Suppose $\sup T \in T^C$. Then by Proposition 1.11, for some $\epsilon > 0$, $\sup T \in B_\epsilon(\sup T) = (\sup T - \epsilon, \sup T + \epsilon) \subseteq T^C$. Then $\sup T - \epsilon \geq x$ for all $x \in f(T)$, a contradiction since

$\sup T$ is the least upper bound of T. So $\sup T \in T$. The proposition follows by letting $x^* \in f^{-1}(\sup T)$. □

1.7.2 Continuity of Maxima and Maximizers: Berge's Maximum Theorem

Frequently economists are interested in *comparative statics*: changes in behavior when a parameter changes, or when other agents' behavior changes. That is, we now have $f : X \times \Theta \to \mathbb{R}$, where Θ is a space of parameters. Our problem is thus

$$\max_{x \in T} f(x, \theta) \tag{1.1}$$

and we want to know how the solution changes as θ does. θ might represent price (in producer theory), or actions by other agents (in game theory), or a preference parameter (in principal-agent theory). First, some notation: For each $\theta \in \Theta$, let $x^*(\theta)$ be the solution $\arg\max_{x \in T} f(x, \theta)$ to (1.1), and $f^*(\theta)$ be its maximized value $\max_{x \in T} f(x, \theta)$.

Theorem 1.1 (Maximum Theorem, Single-Valued Case)**.** *If $f : X \times \Theta \to \mathbb{R}$ is continuous, T is compact, and the solution $x^*(\theta)$ to problem (1.1) is unique for each θ, then $x^* : \Theta \to X$ and $f^* : \Theta \to \mathbb{R}$ are both continuous.*

The maximum theorem tells us that so long as f is continuous (in **both** its arguments) the solution $x^*(\theta)$ is continuous too. This is frequently important for things like existence of equilibrium: In a simple partial equilibrium model with a representative consumer and firm, it tells us that demand can't jump over supply, or vice versa. It can also be important for justifying an identification strategy in empirical work: for instance, if you are identifying a treatment effect on choice behavior using a regression discontinuity design, you need that choice behavior to be continuous in the "running variable".

1.7.3 Detour: Correspondences and Hemicontinuity

What about when $x^*(\theta)$ *isn't* unique for each θ? Well, first of all, the maximizer is no longer a function of θ, since functions take only one value in the codomain (here, X) for each element of the domain (here, Θ). Instead, we have to generalize to the idea of a **correspondence**.

Correspondences

A **correspondence** maps a given element of its domain to a *set* of elements in its codomain. To note this, we add an extra arrow when we're describing correspondences: e.g., $x^* : \Theta \rightrightarrows X$.

Hemicontinuity

For functions, we said that a function $f : X \to Y$ is continuous if for all sequences $x_n \to x$, $f(x_n) \to f(x)$. Or equivalently, f has a closed graph. We can carry this notion over to correspondences: we say a correspondence $G : X \rightrightarrows Y$ is **upper hemicontinuous (UHC)** if for all pairs of sequences $x_n \to x$ in X and $y_n \to y$ in y, if $y_n \in G(x_n)$ for all n then $y \in G(x)$. (Or equivalently, G has a closed graph.) That is, the correspondence might get discontinuously **larger**, but it will never get discontinuously **smaller**.

If a correspondence is UHC and single-valued (i.e., a function) that's equivalent to continuity.

We can now give a more general case of the maximum theorem.

Theorem 1.2 (Maximum Theorem, Multi-Valued Case)**.** *If $f : X \times \Theta \to \mathbb{R}$ is continuous and T is compact, then $x^* : \Theta \rightrightarrows X$ is upper hemicontinuous and nonempty, and $f^* : \Theta \to \mathbb{R}$ is continuous.*

(There's an even more general version of the Maximum Theorem that allows the constraint set T to vary with θ, but we'd need to introduce lower hemicontinuity, which is not as intuitive as upper hemicontinuity.)

1.8 Fixed Points and Convexity

Most equilibrium concepts in economics require that each consumer/player/agent/etc. behaves optimally, given the behavior of everyone else. For instance, in a two-player game where each player has strategies S and payoff function $u : S \times S \to \mathbb{R}$, a Nash equilibrium (s_1^*, s_2^*) requires that

- s_1^* solves

$$\max_{x \in S} u(x, s_2^*) \tag{1.2}$$

- s_2^* solves

$$\max_{x \in S} u(x, s_1^*) \tag{1.3}$$

Essentially, Nash equilibria are points $(s_1, s_2) \in S \times S$ such that $(s_1, s_2) \in x^*(s_1, s_2)$, where $x^* : S \times S \rightrightarrows S \times S$ is the correspondence defined by $x^*(s_1, s_2) = x^*(s_2), x^*(s_1)$.

This is called a **fixed point** of x^*: a point that is an element of its own image under x^*.

This section discusses three of the four main fixed point theorems used in economics. But in order to state these theorems, we first need to define the notion of a **convex set**.

1.8.1 Convex Sets and Functions

In $\mathbb{R}^n$, a set A is **convex** if the line between any two points in A lies entirely within A.

More formally, $A \subseteq \mathbb{R}^n$ is convex if for all $\mathbf{x}, \mathbf{y} \in A$ and all $\lambda \in [0, 1]$, $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in A$.

Example 10. (Draw some convex sets.)

We can also call **functions** $f : \mathbb{R}^n \to \mathbb{R}$ convex, if the set of points that lie above their graph is convex.

(Draw picture.)

More formally, $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and all $\lambda \in [0, 1]$,

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}).$$

If this inequality is always strict, we say that f is **strictly convex**. If $-f$ is convex, we say that f is **concave**; likewise, if $-f$ is strictly convex, f is strictly concave.

Example 11. Examples of convex functions:

- $f(x) = x^2$
- $f(x) = |x|$
- $f(x) = e^x$

You may recall from calculus class that for a twice continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$, f is concave if and only if $f''(x) \leq 0$ for all x. (There is an analogue to this criterion for functions $f : \mathbb{R}^n \to \mathbb{R}$ which we will discuss a few lectures from now.)

1.8.2 Fixed Point Theorems

The most commonly used fixed point theorem for functions on $\mathbb{R}^n$ is **Brouwer's fixed point theorem**.

Theorem 1.3 (Brouwer)**.** *If $A \subseteq \mathbb{R}^n$ is nonempty, convex, and compact, and $f : A \to A$ is continuous, then f has a fixed point in A.*

The intuition when $n = 1$ is clear. Since A is convex, it must be an interval in $\mathbb{R}$; since it is compact, it must be bounded. So we can represent $A \times A$ as a square. Any point in the graph of f that lies on the 45° line through that square is a fixed point. If f is continuous, its graph has to go through that line, since it must form an unbroken line between the left and right edges of the square.

For correspondences, the most commonly used fixed point theorem is **Kakutani's fixed point theorem**.

Theorem 1.4 (Kakutani)**.** *If $A \subseteq \mathbb{R}^n$ is nonempty, convex, and compact, and $G : A \rightrightarrows A$ is upper hemicontinuous, nonempty, and convex-valued, then G has a fixed point in A.*

(When we say G is nonempty and convex valued, we mean that for each $\mathbf{x}$, $G(\mathbf{x}) \neq \emptyset$ and $G(\mathbf{x})$ is a convex set.)

The intuition for Kakutani's fixed point theorem — which is responsible for the existence of Nash equilibria in normal form games — is similar. However, UHC correspondences need to be convex-valued for it to work, because otherwise they can jump over the 45° line.

When will choice correspondences be convex-valued? The answer lies in an extension of the Maximum Theorem:

Theorem 1.5 (Maximum Theorem, Concave Case)**.** *If $f : X \times \Theta \to \mathbb{R}$ is concave and continuous and T is compact, then $x^* : \Theta \rightrightarrows X$ is upper hemicontinuous, nonempty, and convex-valued and $f^* : \Theta \to \mathbb{R}$ is continuous and concave.*

The third main fixed point theorem is called the **contraction mapping theorem**. This theorem is used less frequently to show that equilibria exist (usually the Brouwer or Kakutani theorems work better for this) but more often in **dynamic programming** — which you'll do in the Macro sequence.

We say that a function $f : A \to A$ is a **contraction** if there is some $\beta < 1$ such that for all $\mathbf{x}, \mathbf{y} \in A$, $||f(\mathbf{x}) - f(\mathbf{y})|| \leq \beta||\mathbf{x} - \mathbf{y}||$. Contractions are continuous: use the analytic definition and choose $\delta = \epsilon$.

Theorem 1.6 (Contraction Mapping Theorem, $\mathbb{R}^n$ Version)**.** *If $A \subseteq \mathbb{R}^n$ is closed, and $f : A \to A$ is a contraction, then f has a unique fixed point $\boldsymbol{x}^*$. Moreover, for any initial $\boldsymbol{x}_0 \in A$, the sequence defined recursively by $\boldsymbol{x}_{n+1} = f(\boldsymbol{x}_n)$ converges to $\boldsymbol{x}^*$.*

The beauty of the contraction mapping theorem is that we don't need to assume much about the domain. In fact, A doesn't even need to be a subset of $\mathbf{R}^n$; it just needs to be a complete metric space — i.e., any metric space where every Cauchy sequence converges. This makes it useful in lots of situations where the Brouwer or Kakutani theorems aren't applicable: for instance, when you do dynamic programming, you'll use the contraction mapping theorem on a space of **functions**.

Further, the contraction mapping theorem is surprisingly easy to prove. What follows is based on the original proof by Banach, which you can find on Wikipedia.

Proof. First, a lemma:

Lemma 1.3. *If $f : A \to A$ is a contraction with constant β, then $||\boldsymbol{x}_n - \boldsymbol{x}_{n+1}|| \leq \beta^n ||\boldsymbol{x}_0, \boldsymbol{x}_1||$.*

Proof. Use induction. Initial step:

$$||\mathbf{x}_1 - \mathbf{x}_2|| = ||f(\mathbf{x}_0) - f(\mathbf{x}_1)|| \leq \beta ||\mathbf{x}_0 - \mathbf{x}_1||$$

Induction step: We want to prove

$$||\mathbf{x}_n - \mathbf{x}_{n+1}|| \leq \beta^n ||\mathbf{x}_0 - \mathbf{x}_1|| \implies ||\mathbf{x}_{n+1} - \mathbf{x}_{n+2}|| \leq \beta^{n+1} ||\mathbf{x}_0 - \mathbf{x}_1||.$$

We have

$$\begin{aligned} ||\mathbf{x}_{n+1} - \mathbf{x}_{n+2}|| &= ||f(\mathbf{x}_n) - f(\mathbf{x}_{n+1})|| && \text{(by construction)} \\ &\leq \beta ||\mathbf{x}_n - \mathbf{x}_{n+1}|| && (f \text{ is a contraction}) \\ &\leq \beta^{n+1} ||\mathbf{x}_0 - \mathbf{x}_1|| && \text{(by induction hypothesis)} \end{aligned}$$

□

Now we want to prove $\{\mathbf{x}_n\}$ satisfies the Cauchy criterion. For any $k \geq n$,

$$\begin{aligned} ||\mathbf{x}_k - \mathbf{x}_n|| &\leq ||\mathbf{x}_k - \mathbf{x}_{k-1}|| + ||\mathbf{x}_{k-1} - \mathbf{x}_{k-2}|| + \cdots + ||\mathbf{x}_{n+1} - \mathbf{x}_n|| && \text{(triangle inequality)} \\ &\leq (\beta^k + \beta^{k-1} + \cdots + \beta^n) ||\mathbf{x}_0 - \mathbf{x}_1|| && \text{(from lemma)} \end{aligned}$$

Now we can write the sum on the right-hand side as

$$\begin{aligned}\beta^n(\beta^{k-n}+\beta^{k-n-1}+\cdots+1) &= \frac{\beta^n}{1-\beta}(1-\beta)(\beta^{k-n}+\beta^{k-n-1}+\cdots+1)\\ &= \beta^n\frac{(\beta^{k-n}+\beta^{k-n-1}+\cdots+1)-(\beta^{k-n+1}+\beta^{k-n}+\beta^{k-n-1}+\cdots+\beta)}{1-\beta}\\ &= \beta^n\frac{1-\beta^{k-n+1}}{1-\beta}\\ &\leq \beta^n\frac{1}{1-\beta}\end{aligned}$$

So for all n, k with $k \geq n$ we have $||\mathbf{x}_k - \mathbf{x}_n|| \leq \beta^n \frac{1}{1-\beta}||\mathbf{x}_0 - \mathbf{x}_1||$.

Now for any $\epsilon > 0$, choose N such that $\beta^N < \frac{1-\beta}{||\mathbf{x}_0-\mathbf{x}_1||}\epsilon$; then for all n, k with $k \geq n \geq N$ we have

$$||\mathbf{x}_k - \mathbf{x}_n|| \leq \beta^n\frac{1}{1-\beta}||\mathbf{x}_0 - \mathbf{x}_1|| \leq \beta^N\frac{1}{1-\beta}||\mathbf{x}_0 - \mathbf{x}_1|| < \epsilon,$$

and $\{\mathbf{x}_n\}$ satisfies the Cauchy criterion, and so must converge to some point $\mathbf{x}^*$.

Since f is continuous, $\mathbf{x}_{n+1} = f(\mathbf{x}_n) \to f(\mathbf{x}^*)$. Since $\{\mathbf{x}_{n+1}\}$ is a subsequence of $\{\mathbf{x}_n\}$, $\mathbf{x}_n \to \mathbf{x}^*$. Then $f(\mathbf{x}^*) = \mathbf{x}^*$.

To see that $\mathbf{x}^*$ has to be unique, note that if there were some other fixed point y^*, we would have

$$||\mathbf{x}^* - \mathbf{y}^*|| = ||f(\mathbf{x}^*) - f(\mathbf{y}^*)|| \leq \beta||\mathbf{x}^* - \mathbf{y}^*||$$

which is only possible if $\mathbf{x}^* = \mathbf{y}^*$.

□

Chapter 2

Matrix Algebra

2.1 Matrices

A **matrix** is a rectangular array of numbers. When this array has k rows and n columns, we call it a k by n matrix. We denote the entry in the ith row and jth column of a matrix A by A_{ij}.

An $n \times 1$ matrix is called a (column) **vector**. A $1 \times n$ matrix is similarly referred to as a **row vector**.

2.1.1 Matrix Operations

- Addition: To add two matrices A and B, we add them entrywise: $(A + B)_{ij} = A_{ij} + B_{ij}$.

- Scalar multiplication: We can multiply a matrix by a scalar $c \in \mathbb{R}$ by simply multiplying each of its entries by that scalar: $(cA)_{ij} = cA_{ij}$.

- Matrix multiplication: For an $m \times k$ matrix A and a $k \times n$ matrix B, we can form their product AB. This product will be a $m \times n$ matrix. For each entry $(AB)_ij$ we take the ith row of A and the jth column of B — each of which have k entries — and we multiply them entrywise, and sum the result. I.e.,

$$(AB)_{ij} = \begin{bmatrix} A_{i1} & A_{i2} & \cdots & A_{ik} \end{bmatrix} \begin{bmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{kj} \end{bmatrix} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{ik}B_{kj}$$

When matrices are the correct size to be added or multiplied together, we call them **conformable.**

2.1.2 Laws of Matrix Algebra

Identity: Let I be the $n \times n$ matrix with ones on the diagonal and zeros everywhere else, and 0 be the $n \times n$ matrix of zeros. Then for every $n \times k$ matrix A and $m \times n$ matrix B,

- $IA = A$, $0A = 0$;
- $BI = B$, $B0 = 0$.

Associativity of multiplication and addition: For conformable A, B, and C,

- $(A + B) + C = A + (B + C)$
- $(AB)C = A(BC)$

Distributivity of multiplication and addition: For conformable A, B, and C,

- $A(B + C) = AB + AC$
- $(A + B)C = AC + BC$

Commutativity of addition: For conformable A and B,

- $A + B = B + A$
- In general, $AB \neq BA$. (In fact, if A and B are not square, AB and BA may not even have the same dimensions.)

2.1.3 Transpose

Definition. The transpose of a $k \times n$ matrix A is the $n \times k$ matrix obtained by flipping A about the diagonal, i.e., switching A's rows with its columns. This matrix is written A^T or A'.

Rules of Transposition

- $(A+B)^T = A^T + B^T$
- $(rA)^T = rA^T$, where r is a scalar
- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
  - See proof on p. 158 of the book (Theorem 8.1)

2.1.4 Inverse

Definition. Suppose A is a $n \times n$ matrix. Then the $n \times n$ matrix B is an inverse of A if $AB = BA = I$, where I is the $n \times n$ identity matrix.

If such a matrix B exists, then we say that A is **invertible** or **nonsingular**. Multiplying by A^{-1} is the analog of dividing by the matrix A. We will see how to calculate the inverse of a matrix shortly.

Useful Results

Suppose A and B are invertible matrices.

- A has a unique inverse. (See proof on p.165 of the book, Theorem 8.5.)
- $(A^{-1})^{-1} = A$
- $(A^T)^{-1} = (A^{-1})^T$
- $(AB)^{-1} = B^{-1}A^{-1}$

Rules of transposition and inversion will be used extensively in Econometrics, especially in the beginning when you study Least Squares. Several proofs will require knowledge of these properties, especially:

$$(AB)^T = B^T A^T, \quad (A^T)^{-1} = (A^{-1})^T, \quad (AB)^{-1} = B^{-1}A^{-1}$$

2.1.5 Special Kinds of Matrices

For a (usually square) matrix A:

- A is **upper triangular** if each of its elements below the main diagonal (i.e., A_{ij} for $i > j$) are 0.
- A is **lower triangular** if each of its elements above the main diagonal (i.e., A_{ij} for $i < j$) are 0.
- A is **diagonal** if each of its elements off the main diagonal (i.e., A_{ij} for $i \neq j$) are 0. A matrix is diagonal if and only if it is upper- and lower-triangular.
- A is **symmetric** if $A' = A$.
- A is **idempotent** if $AA = A$. Can you think of any examples? Least Squares analysis in Econometrics will make use of idempotent matrices — especially the projection matrix.

2.1.6 Partitioned Matrices

Frequently, in order to make keeping track of the elements of a matrix easier, we may wish to partition matrices into blocks. Below, we partition a matrix S into four blocks:

$$S = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{bmatrix}$$

can be written as

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}, B = \begin{bmatrix} a_{13} & a_{14} & a_{15} \end{bmatrix}, C = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} D = \begin{bmatrix} a_{23} & a_{24} & a_{25} \\ a_{33} & a_{34} & a_{35} \end{bmatrix}$$

If we have two conformable matrices which have been partitioned in the same way, we can add them block-wise the same way we would add unpartitioned matrices element-wise. Moreover, we can do the same thing (treat blocks like they are matrix entries) for conformable matrices that are partitioned into conformable blocks.

2.1.7 Kronecker Product

Definition. If A is a $m \times n$ matrix and B is a $p \times q$ matrix, then the Kronecker product $A \otimes B$ is the $mp \times nq$ partitioned matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

More explicitly, we have

$$A \otimes B = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\ a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\ \vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\ a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\ a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq} \end{bmatrix}.$$

When you study Seemingly Unrelated Regression (SUR) in Econometrics, you will need to recall what the Kronecker operator does and some properties, including:

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}, \quad (A \otimes B)^T = A^T \otimes B^T$$

Example 12. Find

$$\begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 3 \\ 2 & 1 \end{bmatrix}$$

Useful Rules

- $A \otimes (B + C) = A \otimes B + A \otimes C$
- $(A + B) \otimes C = A \otimes C + B \otimes C$
- For any scalar r, $(rA) \otimes B = A \otimes (rB) = r(A \otimes B)$
- $(A \otimes B) \otimes C = A \otimes (B \otimes C)$

- In general, $A \otimes B \neq B \otimes A$
- $(A \otimes B)(C \otimes D) = AC \otimes BD$
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
- $(A \otimes B)^T = A^T \otimes B^T$

2.1.8 Matrices Homework

1. Simon and Blume Exercise 8.8.

2.2 Determinants

The **determinant** of a square matrix is useful for several things, the main one being that it is a sufficient statistic for **determining** whether the matrix is invertible.

The determinant is defined recursively as follows:

- The determinant of a 1×1 matrix is given by its only entry.
- The determinant of an $n \times n$ matrix A, denoted $\det(A)$, is defined from the determinants of its $n-1 \times n-1$ submatrices as follows:
  1. Choose $i \in \{1, 2, \ldots, n\}$.
  2. For each $j \in \{1, 2, \ldots, n\}$, take the determinant of the submatrix of A formed by deleting row i and column j. (This is referred to as the (i, j)**th minor** of A.) Multiply the result by $A_{ij}(-1)^{i+j}$. ($(-1)^{i+j}$ times the (i, j)th minor of A is referred to as the (i, j)**th cofactor** of A.)
  3. Add all of the results together.

Example 13. Find an expression for the determinant of an arbitrary 2×2 matrix.

Example 14. Let

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ 3 & 13 & 2 \end{bmatrix}$$

Find the determinant of A.

2.2.1 Properties of Determinants

- A is nonsingular $\Leftrightarrow \det(A) \neq 0$.
- The determinant of a lower-triangular, upper-triangular, or diagonal matrix is the product of its diagonal entries.
- $\det(A) = \det(A^T)$
- $\det(AB) = \det(A)\det(B)$
- In general, $\det(A + B) \neq \det(A) + \det(B)$.
- $\det(rA) = r^n \det(A)$, where A is a $n \times n$ matrix and r is a scalar.
- $\det(A^{-1}) = (\det(A))^{-1}$
- If one forms the matrix B by interchanging two rows (or columns) of a $n \times n$ matrix A, then $\det(B) = -\det(A)$.
- If two rows (or columns) of A are equal, $\det(A) = 0$.
- If a row (column) of a $n \times n$ matrix A is multplied by a scalar r to form a matrix B, then $\det(B) = r \cdot \det(A)$.

Example 15. Find the determinant of the matrix V,

$$V = \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}$$

2.2.2 The Trace of a Matrix

Definition. The trace of a square matrix A, denoted $\mathrm{tr}(A)$, is the sum of its diagonal elements: $\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}$.

Properties of the Trace

- $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$.
- For all $c \in \mathbb{R}$, $\mathrm{tr}(cA) = c\,\mathrm{tr}(A)$.
- $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

- $\text{tr}(A') = \text{tr}(A)$.

Some proofs in Econometrics will require you to recall the properties of the trace of a matrix.

2.3 Systems of Linear Equations

2.3.1 Preliminaries

Consider a system of n linear equations:

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &=& b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &=& b_2 \\ \vdots & \vdots & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &=& b_n \end{array}$$

This system can be represented in matrix form by the equation $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

2.3.2 Gauss-Jordan Elimination

There are three **elementary operations** that we can perform on both A and the vector $\mathbf{b}$ that leave the equation $A\mathbf{x} = \mathbf{b}$ intact:

1. Switching two rows.
2. Multiplying a row by a scalar.
3. Add a scalar multiple of one row to another row.

The reason why we can do this to both sides becomes clear if we write

$$A\mathbf{x} = \begin{bmatrix} A_{1\cdot}\mathbf{x} \\ A_{2\cdot}\mathbf{x} \\ \vdots \\ A_{n\cdot}\mathbf{x} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

where $A_{i\cdot}$ is the ith row of n. This reduces the system to n scalar equations, and we know we can use the elementary operations on those.

Note that we can also think of the row operations as being performed on A and the identity matrix, rather than A and $\mathbf{b}$, since we can write

$$A\mathbf{x} = \begin{bmatrix} A_{1\cdot}\mathbf{x} \\ A_{2\cdot}\mathbf{x} \\ \vdots \\ A_{n\cdot}\mathbf{x} \end{bmatrix} = \begin{bmatrix} I_{1\cdot}\mathbf{b} \\ I_{2\cdot}\mathbf{b} \\ \vdots \\ I_{n\cdot}\mathbf{b} \end{bmatrix} = \mathbf{b}.$$

Gauss-Jordan elimination is the process of solving a linear system using these elementary row operations. To start, form the partitioned matrix $[A|I]$. Then, apply elementary row operations to the partitioned matrix until the matrix is in **reduced row echelon form**.

Definition. A matrix B is in reduced row echelon form if whenever B_{ij} is the first nonzero entry of row i, then (1) $B_{ij} = 1$ and (2) $B_{\ell j} = 0$ for all $\ell \neq i$.

That B is in reduced row echelon form does not mean that it is equal to the identity matrix I, or even that its upper-left block is equal to I. For instance, the matrix

$$B = \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 57 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is in reduced row echelon form.

One can show that every matrix has a unique reduced row echelon form, so this process is always feasible. Moreover, the right-hand block can be multiplied by $\mathbf{b}$ to give a solution to $A\mathbf{x} = \mathbf{b}$.

Example 16. Solve the system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & -1 \\ 1 & 3 & 4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 5 \\ 1 \\ 6 \end{bmatrix}$$

If the reduced row echelon form of A is equal to I, then the resulting right-hand block is A^{-1}. To see why, note that we can represent elementary row operations as matrix multiplication:

- Let E_{ij} be formed by switching the ith and jth rows of I. Left-multiplying a matrix A by E_{ij} is equivalent to switching its ith and jth rows.

- Let $E_{ij}(r)$ be formed by replacing I_{ji}, the entry in the ith column and jth row of I, with r.
- Let $E_i(r)$ be formed by replacing I_{ii}, the ith diagonal entry of I, with r.

Each of these matrices are formed by conducting an elementary row operation on I. When we left-multiply A by one of them, we do the corresponding operation to A. (Check this for yourself.)

So really, what we are doing is taking our original equation $A\mathbf{x} = \mathbf{b}$ and multiplying each side by a sequence of elementary matrices $E_1, E_2, \ldots, E_m$ such that $E_1 E_2 \ldots E_m A = I$. Thus, $E_1 E_2 \ldots E_m = A^{-1}$.

This means that we can find A^{-1} using Gauss-Jordan elimination. (In fact, this is probably the easiest way to find A^{-1} by hand.)

Example 17. Solve the system above for arbitrary $\mathbf{b}$.

2.3.3 Rank and Linear Independence

But we can only do this if the reduced row echelon form of A is I. One other way of putting this is that A is of **full rank**.

Definition. The **rank** of a matrix A is the number of nonzero rows in the reduced row echelon form of A.

Example 18. Find the rank of the matrix

$$A = \begin{bmatrix} 0 & 2 & 2 \\ 1 & -3 & -1 \\ -2 & 0 & -4 \\ 4 & 6 & 14 \end{bmatrix}$$

Hint: Use matrix operations to get matrix A in row echelon form.

To understand why full rank is so important for a matrix to be nonsingular, consider that when a square matrix A is not full rank — and hence some rows become zeroes in the reduced row echelon form — there are some equations in the system $A\mathbf{x} = \mathbf{b}$ that are either redundant (if Gaussian elimination reduces the corresponding entry of $\mathbf{b}$ to zero) or impossible for $\mathbf{x}$ to satisfy (otherwise). This redundancy is called **linear dependence.**

A vector $\mathbf{x}$ is a **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n$ if for some scalars $a_1, a_2, \ldots, a_n$,

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n.$$

A set of vectors $V = \{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n\}$ is **linearly independent** if none are a linear combination of the others, and **linearly dependent** otherwise.

Proposition 2.1. *The rank of A is equal to the number of linearly independent rows of A.*

Proof. (Sketch) The only elementary matrix operation that can eliminate a row of A is addition of a scalar multiple of another row. Thus, if a row of A is zero in its reduced row echelon form, that row must have been eliminated through a succession of operations adding scalar multiples of other rows to it. In other words, it must be equal to a linear combination of the remaining rows. So the rank of A is at least the number of linearly independent rows of A.

Moreover, if a row of A is linearly dependent on the other rows, i.e., if

$$A_{i\cdot} = a_1 A_{1\cdot} + a_2 A_{2\cdot} + \cdots + a_{i-1} A_{i-1\cdot} + a_{i+1} A_{i+1\cdot} + \cdots + a_n A_{n\cdot},$$

then we can conduct a sequence of elementary row operations on $A_{i\cdot}$, adding $-a_1 A_{1\cdot}$, $-a_2 A_{2\cdot}$, et cetera, in order to get a row of zeroes. Since the reduced row echelon form of A is unique, putting A in reduced row echelon form *requires* us to continue zeroing out rows in this way until the remaining vectors are not linearly dependent.

□

The notion of rank will be important in least squares analysis. In least squares analysis, we assume the data matrix X is an $n \times k$ matrix with rank k, the number of columns in the matrix. This means there are no linear dependencies among the columns. In the context of econometrics, we call the existence of a linear dependency (i.e., the absence of full rank) perfect collinearity.

2.3.4 Classification of Linear Systems

See Fact 7.11 in the book (p.147).

2.3.5 Inverting Matrices

Definition. For any $n \times n$ matrix A, let B be the matrix whose (i, j)th entry is the (i, j)th cofactor of A. We call the B' the **adjoint** matrix of A, denoted $\text{adj}(A)$.

Theorem 2.1. *If A is nonsingular, then $A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$.*

If we have a linear system $A\mathbf{x} = \mathbf{b}$, and if A is a nonsingular matrix, then we can solve the system by determining the inverse of A, since $A\mathbf{x} = \mathbf{b}$ and nonsingular $A \implies \mathbf{x} = A^{-1}\mathbf{b}$.

Theorem 2.2. *The inverse of the 2×2 matrix*

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

is given by

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

2.3.6 Cramer's Rule

Theorem 2.3. *For a nonsingular matrix A and a conformable vector $\boldsymbol{b}$, let B_i be the matrix formed by replacing the ith column of A with $\boldsymbol{b}$. Then $A\boldsymbol{x} = \boldsymbol{b} \Leftrightarrow x_i = \frac{\det(B_i)}{\det(A)}$ for each $i \in \{1, 2, \ldots, n\}$.*

Example 19. Simon and Blume, Example 9.4.

Example 20. Find all solutions of the system of linear equations:

$$\begin{cases} 2x + 3y - z &= 3 \\ 3x + 3y + 2z &= 7 \\ 7x + 5y &= 13 \end{cases}$$

Note there are several ways to do this (Cramer's Rule, inverting the coefficient matrix, direct substitution, etc).

2.3.7 Systems of Linear Equations Homework

1. Simon and Blume Exercise 9.7.
2. Simon and Blume Exercise 9.11 (a) and (b).
3. Simon and Blume Exercise 7.10.
4. Simon and Blume Exercise 7.12.

Chapter 3

Multivariable Calculus and Implicit Functions

Before we get started, I want to remind you of an important conceptual point. Up until now, the main way we've thought about n-dimensional vectors $\mathbf{x}$ is as points in $\mathbb{R}^n$. But we can also think about them geometrically as being *arrows* between the origin and the point we've identified them with. These arrows have a **length** given by the distance between the point $\mathbf{x}$ and the point $\mathbf{0}$, which is just $||\mathbf{x}||$. This is referred to as the **norm** of $\mathbf{x}$. They also have a **direction** which is given by the vector that points in the same direction as $\mathbf{x}$, but has length one: $\frac{\mathbf{x}}{||\mathbf{x}||}$. (We sometimes call vectors of length one **unit vectors**).

Another important property is that two vectors are perpendicular if and only if their dot product is zero. In fact, their dot product can be written in terms of the angle between the vectors and their length:

$$\mathbf{x} \cdot \mathbf{y} = ||\mathbf{x}||||\mathbf{y}|| \cos\theta,$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. Note that, holding $\mathbf{y}$ and the length of $||\mathbf{x}||$ constant, $\mathbf{x} \cdot \mathbf{y}$ is largest when $\mathbf{x}$ points in the same direction as $\mathbf{y}$.

This is important so that when I talk about the direction of this or that vector, you have some intuitive understanding of what I am talking about.

3.1 Partial Derivatives and Gradients

In single-variable calculus, the derivative of a function (when it exists) tells us its slope (or rather that of its tangent line) at a point. In other words, if we increase x from x_0 to $x_0 + \Delta$ for small Δ, we increase $f(x)$ from $f(x_0)$ to approximately $f(x_0) + f'(x_0)\Delta$.

With functions of more than one variable, there is no longer a single number that can

represent slope, because the argument $\mathbf{x}_0 \in \mathbb{R}^n$ can change in more than two directions. Instead, the slope of a function $f : \mathbb{R}^n \to \mathbb{R}$ is represented by n different **partial derivatives**, written $\frac{\partial f}{\partial \mathbf{x}_i}$, which tell us how $f(\mathbf{x})$ will change in response to a small perturbation in one of the entries of $\mathbf{x}$: if Δ is a vector in $\mathbb{R}^n$ whose ith entry is Δ_i and whose other entries are zero, we can approximate

$$f(\mathbf{x}_0 + \Delta) \approx f(\mathbf{x}_0) + \Delta_i \frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{x}).$$

If we want to consider a perturbation along more than one axis at once — i.e., if the other entries of Δ are allowed to be nonzero, then we can make the approximation

$$f(\mathbf{x} + \Delta) \approx f(\mathbf{x}) + \sum_{i=1}^{n} \Delta_i \frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{x}).$$

We can write this more cleanly by representing the partial derivatives of f as a vector. Define the **gradient** of f, written ∇f, as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}_1}(\mathbf{x}) \\ \frac{\partial f}{\partial \mathbf{x}_2}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial \mathbf{x}_n}(\mathbf{x}) \end{bmatrix}$$

Then we can write our approximation as

$$f(\mathbf{x} + \Delta) \approx f(\mathbf{x}) + \Delta \cdot \nabla f(\mathbf{x}).$$

Much as the equation $y = f(x_0) + f'(x_0)(x - x_0)$ gives the tangent line to f at $\mathbf{x}_0$, the equation $y = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$ describes the **tangent plane** to f.

Just as we say that a function $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable on an open set $A \subseteq \mathbb{R}$ when its derivative exists and is continuous for all points on A, we say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is **continuously differentiable** on an open $A \subseteq \mathbb{R}^n$ when its *partial* derivatives each exist and are continuous on A.

Aside from its role in defining the tangent plane to f, the gradient has some useful geometric properties. For any vector $\mathbf{v}$ of unit length (i.e., with $||\mathbf{v}|| = 1$) the **directional derivative** at $\mathbf{x}$ in direction $\mathbf{v}$ is given by $\nabla f(\mathbf{x}) \cdot \mathbf{v}$. The directional derivative gives the rate of change of f along the parameterized curve $a(t) = \mathbf{x} + t\mathbf{v}$. (You can verify this with the Chain Rule when we discuss it later on.)

This tells us that the direction in which f is increasing fastest is the one in which ∇f is pointing: recall that $\mathbf{x} \cdot \mathbf{y} = ||x||||y|| \cos\theta$ (where θ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$) and that the cosine function is maximized at zero. Hence, the directional derivative $\nabla f(\mathbf{x}) \cdot \mathbf{v} = ||\nabla f(\mathbf{x})||||\mathbf{v}|| \cos\theta$ is largest when ∇f and $\mathbf{v}$ point in the same direction.

3.2 Jacobian and Hessian Matrices

What about vector-valued functions $f : \mathbb{R}^n \to \mathbb{R}^k$? Essentially, we can do the same thing for each of the k components of f as we did for single-valued functions of more than one variable. That is, we can write

$$\begin{aligned} f_1(\mathbf{x}+\Delta) &\approx f_1(\mathbf{x}) + \Delta \cdot \nabla f_1(\mathbf{x}) \\ f_2(\mathbf{x}+\Delta) &\approx f_2(\mathbf{x}) + \Delta \cdot \nabla f_2(\mathbf{x}) \\ &\vdots \\ f_k(\mathbf{x}+\Delta) &\approx f_k(\mathbf{x}) + \Delta \cdot \nabla f_k(\mathbf{x}) \end{aligned}$$

Wouldn't it be nice if we could write this system of equations in matrix form? We can do just that with the **Jacobian** derivative matrix $Df(\mathbf{x})$, written as follows:

$$Df(\mathbf{x}) = \begin{bmatrix} \nabla f_1(\mathbf{x})' \\ \nabla f_2(\mathbf{x})' \\ \vdots \\ \nabla f_k(\mathbf{x})' \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{x}_1}(\mathbf{x}) & \frac{\partial f_1}{\partial \mathbf{x}_2}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial \mathbf{x}_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial \mathbf{x}_1}(\mathbf{x}) & \frac{\partial f_2}{\partial \mathbf{x}_2}(\mathbf{x}) & \cdots & \frac{\partial f_2}{\partial \mathbf{x}_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial \mathbf{x}_1}(\mathbf{x}) & \frac{\partial f_k}{\partial \mathbf{x}_2}(\mathbf{x}) & \cdots & \frac{\partial f_k}{\partial \mathbf{x}_n}(\mathbf{x}) \end{bmatrix}$$

The rows of the $k \times n$ Jacobian matrix are the (transposes of) the gradient vectors of the component functions. This allows us to write our approximation as $f(\mathbf{x}+\Delta) \approx f(\mathbf{x}) + Df(\mathbf{x})\Delta$.

Once again, we say that a function $f : \mathbb{R}^n \to \mathbb{R}^k$ is continuously differentiable on an open $A \subseteq \mathbb{R}^n$ when the partial derivatives of each of its components each exist and are continuous on A.

If the function we are taking a derivative of is a gradient ∇f of a function $f : \mathbb{R}^n \to \mathbb{R}$, then the derivative matrix is instead called a **Hessian** matrix, denoted $D^2 f(\mathbf{x})$. Its i,jth

entries are the **second-order partial derivatives** $\frac{\partial}{\partial \mathbf{x}_j}\frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{x})$, denoted $\frac{\partial^2 f}{\partial \mathbf{x}_j \partial \mathbf{x}_i}(\mathbf{x})$:

$$Df(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial \mathbf{x}_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial \mathbf{x}_2 \partial \mathbf{x}_1}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial \mathbf{x}_n \partial \mathbf{x}_1}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial \mathbf{x}_1 \mathbf{x}_2}(\mathbf{x}) & \frac{\partial^2 f}{\partial \mathbf{x}_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial \mathbf{x}_n \partial \mathbf{x}_2}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \mathbf{x}_1 \partial \mathbf{x}_n}(\mathbf{x}) & \frac{\partial^2 f}{\partial \mathbf{x}_2 \partial \mathbf{x}_n}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial \mathbf{x}_n^2}(\mathbf{x}) \end{bmatrix}$$

If f is twice continuously differentiable (i.e., if its second-order partial derivatives exist and are continuous), **Young's Theorem** tells us that the order of differentiation is irrelevant: $\frac{\partial^2 f}{\partial \mathbf{x}_i \partial \mathbf{x}_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial \mathbf{x}_j \partial \mathbf{x}_i}(\mathbf{x})$. Thus, the Hessian matrix of such a function is always symmetric.

Example 21. Compute the Hessian matrix of the Cobb-Douglas utility function $U(x_1, x_2, x_3) = a x_1^{\alpha} x_2^{\beta} x_3^{1-\alpha-\beta}$.

3.3 The Chain Rule

Suppose that we have two functions: $f : \mathbb{R}^n \to \mathbb{R}^k$ and $g : \mathbb{R}^m \to \mathbb{R}^n$. We want to calculate the $k \times m$ Jacobian matrix of the composition of f and g.

If $n, m, k = 1$, we know how to do this: use the **chain rule**, $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$. For higher dimensions, let's think for a moment to get some intuition. We said that for a small change in $\mathbf{x}$ given by Δ, the change in $g(\mathbf{x})$ would be approximated by $Dg(\mathbf{x})\Delta$. Accordingly, for a small change in $g(\mathbf{x})$ given by $Dg(\mathbf{x})\Delta$, the change in $f(g(\mathbf{x}))$ is approximated by $Df(g(\mathbf{x}))Dg(\mathbf{x})\Delta$. This suggests that $Df \circ g(\mathbf{x}) = Df(g(\mathbf{x}))Dg(\mathbf{x})$ — which turns out to be correct. (This is shown formally in Theorem 14.4 in the book.)

Example 22. Simon and Blume Example 14.13.

Example 23. Simon and Blume Example 14.14.

3.4 Implicit Functions

Suppose that we have some implicit relationship $G(\mathbf{x}, \mathbf{y}) = \mathbf{c}$ for continuously differentiable $G : \mathbb{R}^{k+n} \to \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^k$ and $\mathbf{y}, \mathbf{c} \in \mathbb{R}^n$. For instance, perhaps we have a bunch of first-order conditions (which we'll discuss later) given by

$$\nabla_{\mathbf{q}} u(\mathbf{q}^*, \theta) = \mathbf{0}$$

for some vector of parameters θ. (When we are taking a gradient/Jacobian/Hessian with respect to a subvector of the arguments taken by the function, we put a subscript on it.)

We want to find a function $f : A \to \mathbb{R}^n$ for some $A \subseteq \mathbb{R}^k$ such that for all $\mathbf{x} \in A$, $G(\mathbf{x}, f(\mathbf{x})) = \mathbf{c}$, if such a function exists. Moreover, we want to know its derivative $Df(\mathbf{x})$. In the context of our previous example, we perhaps want to know how the solution $\mathbf{q}^*(\theta)$ changes in response to a change in our parameters, i.e., $Dq^*(\theta)$.

Suppose for a moment that such a function exists, and moreover, that it's continuously differentiable. Then let $\hat{f} : A \to \mathbb{R}^{k+n}$ be defined by $\hat{f}(\mathbf{x}) = \begin{bmatrix} \mathbf{x} \\ f(\mathbf{x}) \end{bmatrix}$. Then from the chain rule,

$$\begin{aligned} DG \circ \hat{f} &= DG(\mathbf{x}, f(\mathbf{x}))D\hat{f}(\mathbf{x}) \\ &= [D_{\mathbf{x}}G(\mathbf{x}, f(\mathbf{x}))|D_{\mathbf{y}}G(\mathbf{x}, f(\mathbf{x}))] \begin{bmatrix} I \\ Df(\mathbf{x}) \end{bmatrix} \\ &= D_{\mathbf{x}}G(\mathbf{x}, f(\mathbf{x})) + D_{\mathbf{y}}G(\mathbf{x}, f(\mathbf{x}))Df(\mathbf{x}) \end{aligned}$$

where

$$D_{\mathbf{x}}G(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \frac{\partial G_1}{\partial \mathbf{x}_1}(\mathbf{x}, \mathbf{y}) & \frac{\partial G_1}{\partial \mathbf{x}_2}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial G_1}{\partial \mathbf{x}_k}(\mathbf{x}, \mathbf{y}) \\ \frac{\partial G_2}{\partial \mathbf{x}_1}(\mathbf{x}, \mathbf{y}) & \frac{\partial G_2}{\partial \mathbf{x}_2}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial G_2}{\partial \mathbf{x}_k}(\mathbf{x}, \mathbf{y}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_n}{\partial \mathbf{x}_1}(\mathbf{x}, \mathbf{y}) & \frac{\partial G_n}{\partial \mathbf{x}_2}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial G_n}{\partial \mathbf{x}_k}(\mathbf{x}, \mathbf{y}) \end{bmatrix}$$

and

$$D_{\mathbf{y}}G(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \frac{\partial G_1}{\partial \mathbf{y}_1}(\mathbf{x}, \mathbf{y}) & \frac{\partial G_1}{\partial \mathbf{y}_2}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial G_1}{\partial \mathbf{y}_n}(\mathbf{x}, \mathbf{y}) \\ \frac{\partial G_2}{\partial \mathbf{y}_1}(\mathbf{x}, \mathbf{y}) & \frac{\partial G_2}{\partial \mathbf{y}_2}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial G_2}{\partial \mathbf{y}_n}(\mathbf{x}, \mathbf{y}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_n}{\partial \mathbf{y}_1}(\mathbf{x}, \mathbf{y}) & \frac{\partial G_n}{\partial \mathbf{y}_2}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial G_n}{\partial \mathbf{y}_n}(\mathbf{x}, \mathbf{y}) \end{bmatrix}$$

Now, if $G(\mathbf{x}, f(\mathbf{x})) = G \circ \hat{f}(\mathbf{x}) = \mathbf{c}$ for all $\mathbf{x} \in A$, then its derivative matrix has to be zero on A: otherwise we could move $\mathbf{x}$ in some direction and G would no longer equal $\mathbf{c}$. So we must have

$$D_{\mathbf{y}}G(\mathbf{x}, f(\mathbf{x}))Df(\mathbf{x}) + D_{\mathbf{x}}G(\mathbf{x}, f(\mathbf{x})) = 0.$$

If $D_{\mathbf{y}}G(\mathbf{x}, f(\mathbf{x}))$ is nonsingular on A, we have

$$\begin{aligned} D_{\mathbf{y}}G(\mathbf{x}, f(\mathbf{x}))Df(\mathbf{x}) &= -D_{\mathbf{x}}G(\mathbf{x}, f(\mathbf{x})) \\ Df(\mathbf{x}) &= -D_{\mathbf{y}}G(\mathbf{x}, f(\mathbf{x}))^{-1}D_{\mathbf{x}}G(\mathbf{x}, f(\mathbf{x})) \end{aligned}$$

So now the question is: when are we guaranteed that such a function f exists and is continuously differentiable on some A? The answer is provided by the **implicit function theorem**.

Theorem 3.1 (Implicit Function Theorem)**.** *Suppose $G : \mathbb{R}^{k+n} \to \mathbb{R}^n$ is continuously differentiable. For all $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $G(\boldsymbol{x}^*, \boldsymbol{y}^*) = \boldsymbol{c}$ and the Jacobian derivative $D_{\boldsymbol{y}}G(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is nonsingular, there exists an open set $A \subseteq \mathbb{R}^k$ containing $\boldsymbol{x}^*$ and a continuously differentiable function $f : A \to \mathbb{R}^n$ such that $G(\boldsymbol{x}, f(\boldsymbol{x})) = \boldsymbol{c}$ for all $\boldsymbol{x} \in A$. Its Jacobian matrix is given by*

$$Df(\boldsymbol{x}) = -D_{\boldsymbol{y}}G(\boldsymbol{x}, f(\boldsymbol{x}))^{-1}D_{\boldsymbol{x}}G(\boldsymbol{x}, f(\boldsymbol{x})).$$

Important things to note about the theorem:

- k is your number of exogenous variables ($\mathbf{x}$).
- n is your number of endogenous variables ($\mathbf{y}$).
- When $n = 1$, y is a scalar, and the derivative expression collapses to

$$Df(\mathbf{x}) = -\frac{1}{\frac{\partial G}{\partial y}(\mathbf{x}, f(\mathbf{x}))}D_{\mathbf{x}}G(\mathbf{x}, f(\mathbf{x})).$$

  Note that since G and f are single-valued, their Jacobians are just the transpose of their gradients (in the case of G, with respect to $\mathbf{x}$). Thus, we can write

$$\nabla f(\mathbf{x}) = -\frac{1}{\frac{\partial G}{\partial y}(\mathbf{x}, f(\mathbf{x}))}\nabla_{\mathbf{x}}G(\mathbf{x}, f(\mathbf{x})).$$

- When $n = 1$, x is a scalar, and so $D_xG(x, f(x))$ is a k-entry column vector whose ith entry is the derivative of the ith component of G with respect to x.
- When $k = n = 1$, both x and y are scalars, and the derivative expression collapses to

$$f'(\mathbf{x}) = -\frac{\frac{\partial G}{\partial x}(x, f(x))}{\frac{\partial G}{\partial y}(x, f(x))}.$$

- The technique we used to apply the chain rule when G depends on $\mathbf{x}$ in multiple ways — here, both directly and via f — can be applied more generally. Suppose that $G : \mathbb{R}^{k+m} \to \mathbb{R}$. We can decompose a vector $\mathbf{y} \in \mathbb{R}^{k+m}$ into two subvectors, i.e., partition the vector as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

where $\mathbf{y}_1 \in \mathbb{R}^k$ and $\mathbf{y}_2 \in \mathbb{R}^m$. Accordingly, we can write $G(\mathbf{y})$ as $G(\mathbf{y}_1, \mathbf{y}_2)$. Now suppose we have two functions $f : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^n \to \mathbb{R}^m$, and we want to know $D_{\mathbf{x}} G(f(\mathbf{x}), h(\mathbf{x}))$. Using the same trick as we did in the above derivation, we can define

$$g(\mathbf{x}) = \begin{bmatrix} f(\mathbf{x}) \\ h(\mathbf{x}) \end{bmatrix}$$

Then $D_{\mathbf{x}} G(f(\mathbf{x}), h(\mathbf{x})) = DG \circ g(\mathbf{x})$, which by the chain rule is given by

$$\begin{aligned} DG(g(\mathbf{x}))Dg(\mathbf{x}) &= \begin{bmatrix} D_{\mathbf{y}_1} G(f(\mathbf{x}), h(\mathbf{x})) & D_{\mathbf{y}_2} G(f(\mathbf{x}), h(\mathbf{x})) \end{bmatrix} \begin{bmatrix} Df(\mathbf{x}) \\ Dh(\mathbf{x}) \end{bmatrix} \\ &= D_{\mathbf{y}_1} G(f(\mathbf{x}), h(\mathbf{x}))Df(\mathbf{x}) + D_{\mathbf{y}_2} G(f(\mathbf{x}), h(\mathbf{x}))Dh(\mathbf{x}) \end{aligned}$$

Example 24. Suppose we have $f(x, y) = x^2 + y^2 = c$. We want to find $y(x)$ such that $x^2 + [y(x)]^2 = c$.

Example 25. Simon and Blume Exercise 15.8.

Example 26.

- Find the slope of the indifference curve $u(x, y) = ax^\alpha y^{1-\alpha} = c$.
- Find the gradient of the utility function.
- Show that the tangent line to the indifference curve is perpendicular to the utility function's gradient.

3.4.1 Tangents to Level Curves

Suppose that we consider the level curve $G(\mathbf{x}, y) = c$ at a point $(\mathbf{x}_0, y_0)$ where $\frac{\partial G}{\partial y} \neq 0$. Then the implicit function theorem says we can write this curve as $f(\mathbf{x}) = y$ in some open set containing $\mathbf{x}_0$, and that

$$\nabla f(\mathbf{x}) = \frac{-1}{\frac{\partial G}{\partial y}(\mathbf{x}, f(\mathbf{x}))} \nabla_{\mathbf{x}} G(\mathbf{x}, f(\mathbf{x})).$$

So we can write the tangent plane to f at $(\mathbf{x}_0, y_0)$ as

$$\hat{f}(\mathbf{x}) = y_0 + \frac{-1}{\frac{\partial G}{\partial y}(\mathbf{x}_0, y_0)} \nabla_{\mathbf{x}} G(\mathbf{x}_0, y_0) \cdot (\mathbf{x} - \mathbf{x}_0)$$

This tangent plane is perpendicular to the gradient of G: Consider a typical vector in the tangent plane given by

$$\begin{bmatrix} \mathbf{x} \\ \hat{f}(\mathbf{x}) \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \mathbf{x} - \mathbf{x}_0 \\ \frac{-1}{\frac{\partial G}{\partial y}(y_0, \mathbf{x}_0)} \nabla_{\mathbf{x}} G(\mathbf{x}_0, y_0) \cdot (\mathbf{x} - \mathbf{x}_0) \end{bmatrix}$$

If we take the dot product of this vector with $\nabla G(\mathbf{x}_0, y_0) = \begin{bmatrix} \nabla_{\mathbf{x}} G(\mathbf{x}_0, y_0) \\ \frac{\partial G}{\partial y}(\mathbf{x}_0, y_0) \end{bmatrix}$, we get

$$\nabla_{\mathbf{x}} G(\mathbf{x}_0, y_0)(\mathbf{x} - \mathbf{x}_0) + \frac{-\frac{\partial G}{\partial y}(\mathbf{x}_0, y_0)}{\frac{\partial G}{\partial y}(\mathbf{x}_0, y_0)} \nabla_{\mathbf{x}} G(\mathbf{x}_0, y_0) \cdot (\mathbf{x} - \mathbf{x}_0) = 0$$

This makes sense: vectors in the tangent plane to a level curve are those directions in which G is staying (to a first-order approximation) constant, whereas the gradient is the direction of greatest increase for G.

3.4.2 Inverse Function Theorem

One important application of the Implicit Function Theorem is the **Inverse Function Theorem**.

Consider a continuously differentiable function $g : \mathbb{R}^n \to \mathbb{R}^n$. We want to use this function's Jacobian derivative to find out whether it is one-to-one, and if so, what the Jacobian derivative of its inverse g^{-1} is.

In other words, we want to find g^{-1} such that $g(g^{-1}(\mathbf{x})) - \mathbf{x} = 0$, or equivalently, such that $G(\mathbf{x}, g^{-1}(\mathbf{x})) = 0$, where $G(\mathbf{x}, \mathbf{y}) \equiv g(\mathbf{y}) - \mathbf{x}$.

Observe that $DG_{\mathbf{y}}(\mathbf{x}, \mathbf{y}) = Dg(\mathbf{y})$, and $DG_{\mathbf{x}}(\mathbf{x}, \mathbf{y}) = I$. Plugging this into the implicit function theorem yields the following corollary:

Corollary 3.1 (Inverse Function Theorem)**.** *Suppose $g : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. For all $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $g(\boldsymbol{y}^*) = \boldsymbol{x}^*$ and the Jacobian derivative $Dg(\boldsymbol{y}^*)$ is nonsingular, there exists an open set $A \subseteq \mathbb{R}^n$ containing $\boldsymbol{x}^*$ on which g is one-to-one. Its inverse on A, $g^{-1} : A \to \mathbb{R}^n$, is continuously differentiable, and its Jacobian matrix is given by*

$$Dg^{-1}(\boldsymbol{x}) = (Dg(g^{-1}(\boldsymbol{x})))^{-1}.$$

The easy way to remember this last part is *the derivative of the inverse is the inverse of the derivative.*

Chapter 4

Unconstrained Optimization

4.1 Quadratic Forms and Definite Matrices

The two most basic classes of functions from $\mathbb{R}$ into itself are probably the linear functions (those whose graph is a line), given by $f(x) = mx + b$, and the quadratic functions, given by $f(x) = ax^2$. Each of these has a generalization to functions on $\mathbb{R}^n$.

- An **affine function** $f : \mathbb{R}^n \to \mathbb{R}^m$ is any function such that for each $1 \leq i \leq m$, $f_i(\mathbf{x}) = \mathbf{b}_i + \sum_{j=1}^n a_{ij}\mathbf{x}_j$. These can be represented in matrix form: $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, where A is the matrix whose i, jth element is a_{ij}.
- A **quadratic form** $Q : \mathbb{R}^n \to \mathbb{R}$ is any real-valued function such that $Q(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^i a_{ij}\mathbf{x}_i\mathbf{x}_j$.

This section considers quadratic forms.

It is straightforward to see that any quadratic form can be represented as $Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$, where A is a symmetric matrix: Let A's ith diagonal element be given by the coefficient a_{ii}, and let its off-diagonal elements A_{ij} are given by $a_{ij}/2$ if $i \leq j$ and $a_{ji}/2$ if $j \leq i$.

A quadratic form Q is characterized by the values it takes away from zero. (Observe that $Q(\mathbf{0}) = 0$ for any quadratic form Q.) If $Q(\mathbf{x}) > 0$ for $\mathbf{x} \neq \mathbf{0}$, and so Q is minimized at $\mathbf{0}$, we say that Q is **positive definite**. If $Q(\mathbf{x}) < 0$ for $\mathbf{x} \neq \mathbf{0}$, we say it is **negative definite**. If the inequalities are weak, because $Q(\mathbf{x}) = 0$ for some $\mathbf{x} \neq \mathbf{0}$, we instead say that Q is positive or negative **semidefinite**, respectively. And if Q takes on both positive and negative values, we say that it is **indefinite**.

These descriptors can also be applied to symmetric matrices, indicating the definiteness of the quadratic form they are associated with. We say that a symmetric matrix A is

- Positive definite if $\mathbf{x}'A\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$;
- Positive semidefinite if $\mathbf{x}'A\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$;
- Negative definite if $\mathbf{x}'A\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$; and
- Negative semidefinite if $\mathbf{x}'A\mathbf{x} \leq 0$ for all $\mathbf{x} \neq \mathbf{0}$.

4.1.1 Test for Definiteness

Recall that the (i, j)th minor of a matrix A is the determinant of the submatrix formed by deleting the ith row and jth column of A.

The k**th order leading principal minor** of A is the determinant of the submatrix formed by deleting the last $n - k$ rows and $n - k$ columns.

Example 27. Let A be a 3×3 matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

The leading principal minors are:

$$|a_{11}|, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad |A|.$$

It turns out that the signs of a matrix's leading principal minors determine its definiteness.

Theorem 4.1. *An $n \times n$ symmetric matrix is positive definite if and only if all of its leading principal minors are positive. It is negative definite if and only if its first order leading principal minor is negative and its leading principal minors alternate in sign (i.e., the second order leading principal minor is positive, the third is negative, etc.)*

We can turn this into a test for semidefiniteness by inserting the word **weakly** at key points throughout the statement.

Theorem 4.2. *An $n \times n$ symmetric matrix is positive semidefinite if and only if all of its leading principal minors are **weakly** positive. It is negative definite if and only if its first order leading principal minor is **weakly** negative and its leading principal minors **weakly** alternate in sign (i.e., the second order leading principal minor is **weakly** positive, the third is **weakly** negative, etc.)*

Example 28. Simon and Blume Exercise 16.1.

4.2 Necessary and Sufficient Conditions

Let's return to our optimization problem,

$$\max_{x \in T} f(x).$$

When $T \subseteq \mathbb{R}^n$, and f is continuously differentiable on T, we can use calculus to find a solution $\mathbf{x}^*$. The way we do so depends on whether T is open (in which case we call our problem **unconstrained**) or closed (in which case we call our problem **constrained**).

This chapter considers the former case.

4.2.1 First Order Conditions

First-order conditions — conditions on the first-order derivatives of a function — narrow down our search for a maximum to several points that *might be* maxima or minima. These conditions essentially state that the tangent plane to f is flat at $\mathbf{x}^*$. Intuitively, if we are at a maximum of f on an open set T, we must be at the top of a "hill" in the graph of T; and at the top of such a hill, the graph must be flat.

Formally, we say that $\mathbf{x}^*$ is a **critical point** of f if $\nabla f(\mathbf{x}^*) = 0$. If $\mathbf{x}^*$ maximizes f on T, then it must be a critical point:

Theorem 4.3 (from Simon & Blume, Theorem 17.1)**.** *Suppose that T is an open subset of $\mathbb{R}^n$, and $f : T \to \mathbb{R}$ is C^1. If $\boldsymbol{x}^* \in \arg\max_{x \in T} f(\boldsymbol{x})$, then $\nabla f(\boldsymbol{x}^*) = 0$.*

Example 29. Find the critical points of $F(x, y) = x^3 - y^3 + 9xy$.

4.2.2 Second Order Conditions

But just because $\mathbf{x}^*$ is a critical point of f doesn't mean that it solves our optimization problem. All it means is that the graph of f is flat there — and while that's true of points that are on top of a hill in the graph of f, it's also true of points that are at the bottom of a valley, and of what are called "saddle points": points that are, say, at the bottom of a valley in one direction and at the top of a hill in another.

Fortunately, when our function is C^2, we have a way to differentiate between these: a **second-order condition** which ensures that a critical point $\mathbf{x}^*$ is a **unique local maximum**

of f: i.e., there is some open set $U \ni \mathbf{x}^*$ such that $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \in U \setminus \mathbf{x}^*$. This condition is simply that f is locally strictly concave.

Theorem 4.4 (from Simon & Blume, Theorem 17.2)**.** *Suppose that $f : T \to \mathbb{R}$ is C^2. If $\boldsymbol{x}^*$ is a critical point of f, and $D^2 f(\boldsymbol{x}^*)$ is negative definite, then $\boldsymbol{x}^*$ is a unique local maximum of f.*

If f is concave on all of T, and T is a convex set, then this statement about *local* maxima extends to a statement about *global* maxima:

Theorem 4.5 (from Simon & Blume, Theorem 17.8)**.** *Suppose that T is a convex open subset of $\mathbb{R}^n$, and $f : T \to \mathbb{R}$ is C^2 and concave. If $\boldsymbol{x}^* \in T$ is a critical point of f, then $\boldsymbol{x}^* \in \arg\max_{x \in T} f(\boldsymbol{x})$.*

Example 30. Take the critical values from Example 29 and determine whether they are max, min or saddle points.

Example 31. Simon and Blume Exercise 17.1 (a).

Example 32. Simon and Blume Exercise 17.2 (a).

Example 33. Simon and Blume Example 17.3.

4.2.3 Unconstrained Optimization Homework

Simon and Blume, Exercises 17.1 and 17.2.

4.3 The Envelope Theorem

Envelope theorems (of which there are many) tell us about how the maximized value of a function changes as a function of a parameter; i.e., they tell us about $\nabla f^*(\theta)$, where (recall) $f^*(\theta) \equiv \max_{x \in T} f(x, \theta)$. In particular, they tell us that under certain conditions,

$$\nabla f^*(\theta) = \nabla_\theta [f(x, \theta)]_{x = x^*(\theta)}.$$

That is, the derivative of the maximized objective function f^* with respect to a parameter θ is just the derivative of the objective f with respect to the parameter, evaluated at the maximizer $x^*(\theta)$ for that parameter.

The easiest version of this theorem is when $x^*(\theta)$ is C^1 and T is open.

Theorem 4.6 (Envelope Theorem, easy version)**.** *Suppose that $f : \mathbb{R}^{n+m} \to \mathbb{R}$, $T \subseteq \mathbb{R}^n$ is open, and $\Theta \subseteq \mathbb{R}^m$. If $\boldsymbol{x}^* : \Theta \to T$ is C^1 and for each $\theta \in \Theta$,*

$$f^*(\theta) \equiv \max_{\boldsymbol{x} \in T} f(\boldsymbol{x}, \theta) = f(\boldsymbol{x}^*(\theta), \theta),$$

then

$$\nabla f^*(\theta) = \nabla_\theta [f(x, \theta)]_{\boldsymbol{x} = \boldsymbol{x}^*(\theta)}.$$

Proof. By the chain rule,

$$\begin{aligned}[\nabla f^*(\theta)]' = Df^*(\theta) &= D_{\mathbf{x}} f(\mathbf{x}^*(\theta), \theta) Dx^*(\theta) + D_\theta \left[f(\mathbf{x}, \theta)\right]_{\mathbf{x} = \mathbf{x}^*(\theta)} \\ &= [\nabla_{\mathbf{x}} f(\mathbf{x}^*(\theta), \theta)]' Dx^*(\theta) + \left[\nabla_\theta \left[f(\mathbf{x}, \theta)\right]_{\mathbf{x} = \mathbf{x}^*(\theta)}\right]'\end{aligned}$$

By Theorem 4.3, $\nabla_{\mathbf{x}} f(\mathbf{x}^*(\theta), \theta) = 0$ for each θ. The proposition follows. □

4.4 Taylor Series and Necessary and Sufficient Conditions

In the last chapter, we discussed linear approximations of the form

$$f(\mathbf{x} + \Delta) \approx f(\mathbf{x}) + \Delta \cdot \nabla f(\mathbf{x}),$$

without actually discussing (other than intuitively) why this approximation works, or how we might improve it.

It turns out that the linear approximation is the most basic version of what's called a **Taylor series approximation**, and that for a C^k function, we can give a better approximation with k terms corresponding to the first k derivatives.

Let's do this with functions of a single variable first, because we haven't talked about derivatives of multivariable functions of order greater than two. (Nor do we want to.)

Definition. Suppose $f : \mathbb{R} \to \mathbb{R}$ is C^k. The kth order Taylor polynomial (or Taylor approximation) for f at a point $a \in \mathbb{R}$ is given by

$$P_k(x) = f(a) + f'(a)(x - a) + \frac{1}{2!} f''(a)(x - a)^2 + \frac{1}{3!} f'''(a)(x - a)^3 + \cdots + \frac{1}{k!} f^{(k)}(a).$$

As is probably unsurprising, higher order Taylor polynomials give better approximations to f. **Taylor's theorem** tells us exactly how good these approximations are.

Theorem 4.7 (Taylor's Theorem)**.** *Suppose $f : \mathbb{R} \to \mathbb{R}$ is C^k. Consider $P_k(x)$, the kth order Taylor polynomial for f at $a \in \mathbb{R}$. Then*

$$\frac{P_k(a+h_n) - f(a+h_n)}{(h_n)^k} \to 0 \text{ for all } \{h_n\} \subset \mathbb{R} \text{ with } h_n \to 0.$$

Taylor's theorem says that the approximation error is of **order** h^k: that is, the remainder vanishes as h becomes small, and does so *faster* than h^k does.

Example 34. Compute the first and second order Taylor polynomial of the exponential function $f(x) = e^x$ at $x = 0$.

Taylor's theorem extends to the $\mathbb{R}^n$ case. The first-order Taylor polynomial at $\mathbf{a}$ is just the tangent plane:

$$P_1(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}).$$

Theorem 4.8 (Taylor's Theorem for 1st-order Approximations in $\mathbb{R}^n$)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is C^k. Consider $P_1(\boldsymbol{x})$, the 1st order Taylor polynomial for f at $\boldsymbol{a} \in \mathbb{R}^n$. Then*

$$\frac{P_1(\boldsymbol{a}+\boldsymbol{h}_n) - f(\boldsymbol{a}+\boldsymbol{h}_n)}{||\boldsymbol{h}_n||} \to 0 \text{ for all } \{\boldsymbol{h}_n\} \subset \mathbb{R}^n \text{ with } \boldsymbol{h}_n \to 0.$$

Note that now, the order of the approximation error is the *norm* of $\mathbf{h}$: we can't divide by a vector, we can only divide by its length.

The second-order Taylor polynomial includes a term for the Hessian:

$$P_2(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})' D^2 f(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

Just as the error of the second-order approximation in $\mathbb{R}$ was of order h^2, the error of the second-order approximation in $\mathbb{R}^n$ is of order $||\mathbf{h}||^2$.

Theorem 4.9 (Taylor's Theorem for 2nd-order Approximations in $\mathbb{R}^n$)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is C^k. Consider $P_2(\boldsymbol{x})$, the 2nd order Taylor polynomial for f at $\boldsymbol{a} \in \mathbb{R}^n$. Then*

$$\frac{P_2(\boldsymbol{a}+\boldsymbol{h}_n) - f(\boldsymbol{a}+\boldsymbol{h}_n)}{||\boldsymbol{h}_n||^2} \to 0 \text{ for all } \{\boldsymbol{h}_n\} \subset \mathbb{R}^n \text{ with } \boldsymbol{h}_n \to 0.$$

Example 35. Compute the Taylor polynomial of order 2 of $F(x, y) = x^{1/4} y^{3/4}$ at $(1, 1)$.

These results tell us that it's reasonable to use a linear specification when you run a regression: though the true relationship between the dependent and independent variables

may be nonlinear, one can approximate that relationship linearly. Moreover, it's why that approximation improves when you add a second-order polynomial in the independent variables to your specification.

But they are also useful for understanding **why** our first and second order conditions are necessary and sufficient for a (local) maximum. Let's use it to prove Theorem 4.3.

Proof. Suppose that $\mathbf{x}^* \in \arg\max_{x \in T} f(\mathbf{x})$. Since T is open, there is some $\epsilon > 0$ such that $B_\epsilon(\mathbf{x}^*) \subseteq T$. Our first-order Taylor approximation at $\mathbf{x}^*$ is given by

$$P_1(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*).$$

Suppose that $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$. We want to define a sequence of vectors $\mathbf{s}_n \to \mathbf{0}$ such that $\mathbf{x}^* + \mathbf{s}_n \in T$, and then show that the sequence $\frac{f(\mathbf{x}^* + \mathbf{s}_n) - f(\mathbf{x}^*)}{||\mathbf{s}_n||}$ converges to something positive; hence, $f(\mathbf{x}^* + \mathbf{s}_n) > f(\mathbf{x}^*)$ for some n.

Define the sequence $\{\mathbf{s}_n\} \subset \mathbb{R}^n$ by the rule

$$\mathbf{s}_n = \frac{\epsilon}{2} \frac{1}{n} \frac{\nabla f(\mathbf{x}^*)}{||\nabla f(\mathbf{x}^*)||}.$$

In other words, we are choosing a vector of length $\frac{\epsilon}{2n}$ in the direction of the gradient. Since $||\mathbf{s}_n|| = \frac{\epsilon}{2n} < \epsilon$, $\mathbf{x}^* + \mathbf{s}_n \in T$.

Now we have

$$P_1(\mathbf{x}^* + \mathbf{s}_n) = f(\mathbf{x}^*) + ||\nabla f(\mathbf{x}^*)|| \frac{\epsilon}{2n}.$$

Then we have

$$\begin{aligned}
\frac{f(\mathbf{x}^* + \mathbf{s}_n) - f(\mathbf{x}^*)}{||s_n||} &= \frac{f(\mathbf{x}^* + \mathbf{s}_n) - P(\mathbf{x}^* + \mathbf{s}_n) + P(\mathbf{x}^* + \mathbf{s}_n) - f(\mathbf{x}^*)}{||\mathbf{s}_n||} \\
&\frac{f(\mathbf{x}^* + \mathbf{s}_n) - P(\mathbf{x}^* + \mathbf{s}_n)}{||s_n||} \to 0 && \text{(by Taylor's Theorem)} \\
&\frac{P(\mathbf{x}^* + \mathbf{s}_n) - f(\mathbf{x}^*)}{||\mathbf{s}_n||} = ||\nabla f(\mathbf{x}^*)|| && \forall n \\
&\Rightarrow \frac{f(\mathbf{x}^* + \mathbf{s}_n) - f(\mathbf{x}^*)}{||\mathbf{s}_n||} \to ||\nabla f(\mathbf{x}^*)|| > 0
\end{aligned}$$

so there is some N such that for all $n \geq N$,

$$\begin{aligned}
\left| \frac{f(\mathbf{x}^* + \mathbf{s}_n) - f(\mathbf{x}^*)}{||\mathbf{s}_n||} - ||\nabla f(\mathbf{x}^*)|| \right| &< ||\nabla f(\mathbf{x}^*)|| \\
\Rightarrow \frac{f(\mathbf{x}^* + \mathbf{s}_n) - f(\mathbf{x}^*)}{||\mathbf{s}_n||} &> 0
\end{aligned}$$

and so $f(\mathbf{x}^* + \mathbf{s}_n) > f(\mathbf{x}^*)$ — a contradiction, since $\mathbf{x}^* \in \arg\max_{x \in T} f(\mathbf{x})$. □

We can do something similar for Theorem 4.4; the proof is in the book if you are curious. The key is that when $\nabla f(\mathbf{x}^*) = 0$ and $D^2 f(\mathbf{x}^*)$ is negative definite, then the second-order Taylor approximation to f at $\mathbf{x}^*$ is just

$$P_2(\mathbf{x}) = f(\mathbf{x}^*) + (\mathbf{x} - \mathbf{x}^*)' D^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*)$$

and so $P_2(\mathbf{x}) < f(\mathbf{x}^*)$ for all $\mathbf{x} \neq \mathbf{x}^*$.

Chapter 5

Constrained Optimization

Let's return once again to our optimization problem,

$$\max_{x \in T} f(x),$$

with $T \subseteq \mathbb{R}^n$, and f continuously differentiable on $\mathbb{R}^n$. Unlike the previous chapter, when we assumed that T was open, we'll now assume that T is closed. And more specifically, we'll assume that T is defined by constraint functions; i.e.,

$$T = \{\mathbf{x} \in \mathbb{R}^n | h_1(\mathbf{x}) = c_1, \cdots, h_m(\mathbf{x}) = c_m, g_1(\mathbf{x}) \leq b_1, \cdots, g_k(\mathbf{x}) \leq b_k\}$$

The constraints $h_1(\mathbf{x}) = c_1, \cdots, h_m(\mathbf{x}) = c_m$ are called **equality constraints**, whereas $g_1(\mathbf{x}) \leq b_1, \cdots, g_k(\mathbf{x}) \leq b_k$ are called **inequality constraints**.

5.1 Equality Constraints

For illustrative purposes, consider the problem with one equality constraint:

$$\max f(x_1, x_2) \text{ such that } h_1(x_1, x_2) = c_1$$

At the maximizer point $\mathbf{x}^* = (x_1^*, x_2^*)$ the level curve of f and the constraint are tangent to each other, i.e. both have a common slope. We will explore this observation in order to find a characterization of the solution for this class of problems and its generalization to m restrictions.

In order to find the derivative of the level curve at the optimum $\mathbf{x}^*$, recall from the implicit function theorem that for a function $G(y_1, y_2) = c$, $\frac{dy_2}{dy_1} = -\frac{\partial G}{\partial y_1}(y_1, y_2) / \frac{\partial G}{\partial y_2}(y_1, y_2)$. In particular, the slope $\frac{dx_2}{dx_1}$ for f and h must be the same at $\mathbf{x}^*$:

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*)/\frac{\partial f}{\partial x_2}(\mathbf{x}^*) = \frac{\partial h_1}{\partial x_1}(\mathbf{x}^*)/\frac{\partial h_1}{\partial x_2}(\mathbf{x}^*)$$

which can be written as

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*)/\frac{\partial h_1}{\partial x_1}(\mathbf{x}^*) = \frac{\partial f}{\partial x_2}(\mathbf{x}^*)/\frac{\partial h_1}{\partial x_2}(\mathbf{x}^*) = \lambda \tag{5.1}$$

where λ is the common value of the slope at $\mathbf{x}^*$. We are assuming that the ratios above do not have zero denominator.

Now rewrite (5.1) as two equations

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) - \lambda\frac{\partial h_1}{\partial x_1}(\mathbf{x}^*) = 0 \tag{5.2}$$

$$\frac{\partial f}{\partial x_2}(\mathbf{x}^*) - \lambda\frac{\partial h_1}{\partial x_2}(\mathbf{x}^*) = 0 \tag{5.3}$$

and note that the system (5.2),(5.3) has three unknowns: (x_1^*, x_2^*, λ). Thus, we need to take into account a third equation in order to solve the system. This equation is the constraint itself:

$$h_1(\mathbf{x}^*) - c_1 = 0 \tag{5.4}$$

Hence, the pair $(\mathbf{x}^*, \lambda)$ that maximizes $f(x_1, x_2)$ subject to $h_1(x_1, x_2) = c_1$ solves the system (5.2-5.4).

This system can be written in the form of a function L with three arguments: x_1, x_2, and λ. This function is called the **Lagrangian** of the problem:

$$L(x_1, x_2, \lambda) \equiv f(x_1, x_2) - \lambda[h_1(x_1, x_2) - c_1]$$

The variable λ is called the **Lagrange multiplier**. The system of equations (5.2-5.4) can be recovered from the first-order conditions for *unconstrained* maximization of L: $\frac{\partial L(\cdot)}{\partial x_1} = 0$, $\frac{\partial L(\cdot)}{\partial x_2} = 0$, $\frac{\partial L(\cdot)}{\partial \lambda} = 0$.

This method — the **method of Lagrange** — transforms a constrained maximization

problem into an unconstrained maximization problem (which we already know how to solve). The transformation requires the introduction of one more variable into the problem: the multiplier. It is important to note that the transformation requires $\frac{\partial h_1(\mathbf{x}^*)}{\partial x_1} \neq 0$ or $\frac{\partial h_1(\mathbf{x}^*)}{\partial x_2} \neq 0$ or both, otherwise the ratios (4) are not well defined. This is called **constraint qualification.** There are several criteria by which a constraint can be deemed qualified. The most basic is the one we use most often in economics: the constraint functions are affine.

A more general constraint qualification relies upon the rank of the Jacobian matrix for the (vector-valued) constraint function $h(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_m(\mathbf{x})]'$. We say that $h_1, \ldots h_m$ satisfy the **nondegenerate constraint qualification (NDCQ)** at $\mathbf{x}^*$ if the rank of the Jacobian matrix $Dh(\mathbf{x}^*)$is m. Thinking back to the implicit function theorem, this means that if $h(\mathbf{x}^*) = \mathbf{c}$, then the surface described by the constraint functions can be written as an explicit function from $\mathbb{R}^{n-m}$ to $\mathbb{R}^m$ in the neighborhood of $\mathbf{x}^*$.

(To see why: Decompose $\mathbf{x}$ into a subvector $\mathbf{x}_1$ corresponding to the nonzero columns of the reduced row echelon form of $Dh(\mathbf{x}^*)$, and $\mathbf{x}_2$ corresponding to the other $n - m$ entries. It follows that $D_{\mathbf{x}_1}(\mathbf{x}_1, \mathbf{x}_2)$ is nonsingular. Then by the implicit function theorem, there is an open set $A \subseteq \mathbb{R}^{n-m}$ containing $\mathbf{x}_2^*$ and a function $\hat{h} : A \to \mathbb{R}^m$ such that $h(\hat{h}(\mathbf{x}_2), \mathbf{x}_2) = \mathbf{c}$ for all $\mathbf{x}_2 \in A$.)

Theorem 5.1 (Simon and Blume Theorem 18.2 (Constrained Maximization with Equality Constraints: First-Order Conditions))**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable, and that $\boldsymbol{c} \in \mathbb{R}^m$. Then if*

$$\boldsymbol{x}^* \in \arg\max_{h(\boldsymbol{x})=\boldsymbol{c}} f(\boldsymbol{x})$$

and $\text{rank}(Dh(\boldsymbol{x}^*)) = m$, there exists $\lambda^* \in \mathbb{R}^m$ such that $(\boldsymbol{x}^*, \lambda^*)$ is a critical point of

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda \cdot (h(\boldsymbol{x}) - \boldsymbol{c}).$$

An aside: Sometimes we write the problem $\max_{h(\mathbf{x})=\mathbf{c}} f(\mathbf{x})$ as

$$\max f(\mathbf{x}) \text{ s.t. } h(\mathbf{x}) = \mathbf{c}$$

Example 36. Simon and Blume Example 18.4.

Example 37. Simon and Blume Example 18.5.

Example 38. Simon and Blume Exercise 18.7.

Example 39. Simon and Blume Exercise 18.4.

5.2 Second Order Sufficient Conditions

For unconstrained optimization, we used a second-order condition to ensure that at a given critical point, the objective function is locally concave. For constrained optimization, the second-order condition serves the same function, but only needs to ensure that the function is hill-shaped **on the constraint set**. More specifically, we no longer must require that the Hessian matrix $D^2 f$ is negative definite at our critical point (for a unique local maximum) or negative semidefinite everywhere (for a global maximum) — i.e., that $\mathbf{v}' D^2 f(\mathbf{x}^*)\mathbf{v} < 0 \forall \mathbf{v}$ (in the former case), or $\mathbf{v}' D^2 f(\mathbf{x})\mathbf{v} < 0 \forall \mathbf{v}, \mathbf{x}$ (in the latter case). Instead, we need to ensure that such inequalities hold for the Hessian **of the Lagrangian with respect to x**, $D^2_{\mathbf{x}} L$, and for vectors $\mathbf{v}$ that are in the constraint set (when h is linear) or tangent to it (more generally).

The test for negative definiteness of an $n \times n$ matrix A on the **subspace** of $\mathbb{R}^n$ associated with an $k \times n$ matrix B (the collection of vectors $\mathbf{x}$ for which $B\mathbf{x} = \mathbf{0}$) involves the principal minors of a **bordered matrix**.

$$\begin{bmatrix} 0 & B \\ B' & A \end{bmatrix}.$$

Theorem 5.2 (Simon and Blume Theorem 16.4)**.** *For an $n \times n$ matrix A and a $k \times n$ matrix B:*

- *A is positive definite on $\{\boldsymbol{x} | B\boldsymbol{x} = \boldsymbol{0}\}$ — i.e., $\boldsymbol{v}' A \boldsymbol{v} > 0$ for all $\boldsymbol{v} \in \{\boldsymbol{x} | B\boldsymbol{x} = \boldsymbol{0}\}$ if and only if the last $n - k$ principal minors of the bordered matrix $\begin{bmatrix} 0 & B \\ B' & A \end{bmatrix}$ are negative if m is odd and positive if m is even.*

- *A is negative definite on $\{\boldsymbol{x} | B\boldsymbol{x} = \boldsymbol{0}\}$ — i.e., $\boldsymbol{v}' A \boldsymbol{v} < 0$ for all $\boldsymbol{v} \in \{\boldsymbol{x} | B\boldsymbol{x} = \boldsymbol{0}\}$ if and only if the last $n - k$ principal minors of the bordered matrix $\begin{bmatrix} 0 & B \\ B' & A \end{bmatrix}$ alternate in sign, and the $n + k$th principal minor is negative if n is odd and positive if n is even.*

This leads us to our second-order condition for a local maximum.

Theorem 5.3. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable, and that $\boldsymbol{c} \in \mathbb{R}^m$. If $(\boldsymbol{x}^*, \lambda^*)$ is a critical point of*

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda \cdot (h(\boldsymbol{x}) - \boldsymbol{c}),$$

and $D_{\boldsymbol{x}}^2 L(\boldsymbol{x}^*, \lambda^*)$ is negative definite on $\{\boldsymbol{x} | D^2 h(\boldsymbol{x}^*)\boldsymbol{x} = \boldsymbol{0}\}$, then $(\boldsymbol{x}^*, \lambda^*)$ is a unique local maximum of f on $\{\boldsymbol{x} | h(\boldsymbol{x}) = \boldsymbol{c}\}$.

Example 40. Simon and Blume Example 19.7.

Example 41. Simon and Blume Example 19.8.

5.3 Meaning of the Multiplier

Theorem 5.4 (Meaning of the Multiplier)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. If there exist continuously diferentiable functions $\boldsymbol{x}^* : \mathbb{R}^m \to \mathbb{R}^n$ and $\lambda^* : \mathbb{R}^m \to \mathbb{R}^m$ such that for each $\boldsymbol{c} \in \mathbb{R}^m$, $\boldsymbol{x}^*(\boldsymbol{c}) \in \arg\max_{h(\boldsymbol{x})=\boldsymbol{c}} f(\boldsymbol{x})$ and $(\boldsymbol{x}^*(\boldsymbol{c}), \lambda^*(\boldsymbol{c}))$ is a critical point of $L(\boldsymbol{x}, \lambda, \boldsymbol{c})$, then*

$$\nabla_{\mathbf{c}} f(\boldsymbol{x}^*(\boldsymbol{c})) = \lambda^*(\boldsymbol{c}).$$

Proof. Since $h(\mathbf{x}^*(\mathbf{c})) = \mathbf{c}$ for all $\mathbf{c}$,

$$f(\mathbf{x}^*(\mathbf{c})) = L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c}) \text{ for all } \mathbf{c}.$$

It follows that $\nabla_{\mathbf{c}} f(\mathbf{x}^*(\mathbf{c})) = \nabla_{\mathbf{c}} [L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c})]$. By the chain rule,

$$\begin{aligned} D_{\mathbf{c}} [L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c})] =& D_{\mathbf{x}} L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c}) D\mathbf{x}^*(\mathbf{c}) + D_{\lambda} L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c}) D\lambda^*(\mathbf{c}) \\ &+ D_{\mathbf{c}} [L(\mathbf{x}, \lambda, \mathbf{c})]_{\mathbf{x}=\mathbf{x}^*(\mathbf{c}), \lambda=\lambda^*(\mathbf{c})} \end{aligned}$$

Since $(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}))$ is a critical point of $L(\mathbf{x}, \lambda, \mathbf{c})$, we have

$$D_{\mathbf{x}} L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c}) = [\nabla_{\mathbf{x}} L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c})]' = \mathbf{0}'.$$

Since it also satisfies the constraint $h(\mathbf{x}^*(\mathbf{c})) = \mathbf{c}$, we can write

$$D_{\lambda} L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c}) = [\nabla_{\lambda} L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c})]' = -(h(\mathbf{x}^*(\mathbf{c})) - \mathbf{c})' = \mathbf{0}'.$$

Since $D_{\mathbf{c}} L(\mathbf{x}, \lambda, \mathbf{c}) = \lambda$, we thus have

$$\begin{aligned} D_{\mathbf{c}} [L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c})] =& \mathbf{0}' D\mathbf{x}^*(\mathbf{c}) + \mathbf{0}' D\lambda^*(\mathbf{c}) + \lambda^*(\mathbf{c})' \\ \Rightarrow \nabla_{\mathbf{c}} f(\mathbf{x}^*(\mathbf{c})) = \nabla_{\mathbf{c}} [L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c})] =& D_{\mathbf{c}} [L(\mathbf{x}^*(\mathbf{c}), \lambda^*(\mathbf{c}), \mathbf{c})]' = \lambda^*(\mathbf{c}) \end{aligned}$$

as desired. □

Thus, we can interpret the Lagrange multiplier as the marginal benefit of relaxing the constraint. In consumer theory, when our consumer faces a budget constraint taking the form of $\mathbf{p} \cdot \mathbf{x} \leq w$, this means that the Lagrange multiplier from a utility maximization problem represents the consumer's **marginal utility of wealth**.

5.4 Inequality Constraints

While equality constraints are easier to handle, optimization in economics frequently takes place on sets defined by **inequality constraints**: i.e., on

$$T = \{\mathbf{x} \in \mathbb{R}^n | g_1(\mathbf{x}) \leq b_1, \cdots, g_k(\mathbf{x}) \leq b_k\}.$$

To develop some intuition for how we can proceed with solving this problem, consider a problem with one inequality constraint:

$$\max f(\mathbf{x}) \text{ such that } g_1(\mathbf{x}) \leq b_1.$$

Now, if the constraint **binds** at our solution $\mathbf{x}^*$ — that is, if we are on the boundary of the constraint region, and $g_1(\mathbf{x}^*) = b_1$ — we once again need the level set of f at $\mathbf{x}^*$ to be tangent to the level set of g_1 at $\mathbf{x}^*$; otherwise they cross, and we can achieve a higher value of f along the curve $g_1(\mathbf{x}) = b_1$.

Since we said the tangent plane of a level set of f is perpendicular to the gradient ∇f, this means the gradients of f and g_1 must either point in the same direction, or in opposite directions. So far, this has the same implications as an equality constraint: we need

$$\nabla f(\mathbf{x}^*) = \lambda \nabla g_1(\mathbf{x}^*) \text{ for some } \lambda.$$

But now we can rule out the case where they point in opposite directions. Recall that the gradient is a function's direction of greatest increase. Then if $\nabla f(\mathbf{x}^*)$ points in the opposite direction from $\nabla g(\mathbf{x}^*)$, the direction of f's greatest *increase* is the same as the direction of g's greatest *decrease*; thus, $\nabla f(\mathbf{x}^*)$ points **into** the set $\{\mathbf{x} | g_1(\mathbf{x}) \leq 0\}$, and f takes on a greater value for some $\mathbf{x}$ with $g_1(\mathbf{x}) < 0$. In other words, when the constraint g_1 binds, we need

$$\nabla f(\mathbf{x}^*) = \lambda^* \nabla g_1(\mathbf{x}^*) \text{ for some } \lambda^* \geq 0.$$

However, this condition is only necessary for points on the boundary of the constraint set. For points $\mathbf{x}^*$ on the interior of the set (i.e., those with $g_1(\mathbf{x}^*) = b_1$) to be maxima, our necessary condition must come from *unconstrained* optimization. That is, we must have

$$\nabla f(\mathbf{x}^*) = \mathbf{0}.$$

We can reconcile these two conditions into one convenient specification by writing

$$\begin{aligned} \nabla f(\mathbf{x}^*) &= \lambda^* \nabla g_1(\mathbf{x}^*) \text{ for some } \lambda^* \geq 0 \\ 0 &= \lambda^*(g_1(\mathbf{x}^*) - b_1) \end{aligned}$$

Notice that the second condition says that either $g_1(\mathbf{x}^*) = b_1$ (so the constraint binds), or $\lambda^* = 0$ (so the first-order condition becomes $\nabla f(\mathbf{x}^*) = \mathbf{0}$). This is called a **complementary slackness** condition. If we add one of these for each inequality constraint, along with our restriction on the sign of the multiplier, to the first-order conditions for the Lagrangean $L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda \cdot (g(\mathbf{x}) - \mathbf{b})$, we can give the appropriate necessary conditions for a maximum on a region defined by inequalities. Note that we only need the constraint qualification to hold for the **active** constraints (those that bind); we denote the set of active constraints at $\mathbf{x}$ by $L(\mathbf{x})$ and its number of elements by $\ell(\mathbf{x})$.

Theorem 5.5. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^k$ are continuously differentiable, and that $\boldsymbol{b} \in \mathbb{R}^k$. Then if*

$$\boldsymbol{x}^* \in \arg \max_{g(\boldsymbol{x}) \leq \boldsymbol{b}} f(\boldsymbol{x})$$

and $\text{rank}(Dg_{L(\boldsymbol{x}^*)}(\boldsymbol{x}^*)) = \ell(\boldsymbol{x}^*)$, there exists $\lambda^* \in \mathbb{R}^k$ such that

1. *$\boldsymbol{x}^*$ is a critical point of*
$$L(\boldsymbol{x}, \lambda^*) = f(\boldsymbol{x}) - \lambda^* \cdot (g(\boldsymbol{x}) - \boldsymbol{b});$$
2. *$\lambda^* \geq \boldsymbol{0}$;*
3. *$g(\boldsymbol{x}) \leq \boldsymbol{b}$: and*
4. *$\lambda^* \cdot (g(\boldsymbol{x}) - \boldsymbol{b}) = 0$.*

Note that we need to explicitly specify $g(\mathbf{x}) \leq \mathbf{b}$ now rather than getting it as a consequence of the FOC of L with respect to λ (which we no longer take).

Example 42. Simon and Blume Example 18.9.

Example 43. Simon and Blume Exercise 18.10.

5.4.1 Inequality Constraints Homework

Simon and Blume Exercises 18.12 and 18.14.

Chapter 6

Probability: Basics

A **probability space** is a triple $(\Omega, \mathcal{F}, P)$.

- Ω is called a **sample space** or **state space**. The sample space is the space of possible **outcomes** or **states of the world** that we are interested in. For instance, if we flip a coin twice, Ω is given by $\{HH, HT, TH, TT\}$.

  For now, we'll assume that Ω has finitely many elements.

- $\mathcal{F}$ is a collection of subsets of Ω that we can assign a probability to. I.e., for each $E \in \mathcal{F}$, we can assign a number between 0 and 1 to E representing how likely we think it is that the outcome is in E. (Formally, the term for $\mathcal{F}$ is a σ**-algebra**.) When Ω is finite, we'll generally let $\mathcal{F}$ contain every subset of Ω.

- The **probability measure** $P : \mathcal{F} \to [0, 1]$ tells us how likely each event is. We place two requirements on P:

  - For all $E \in \mathcal{F}$, $P(E^C) = 1 - P(E)$.
  - If $E_1, E_2, \ldots, E_n \in \mathcal{F}$ are *disjoint* sets (i.e., they have no elements in common), then $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$.

Example 44. Show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. We can decompose A as $(A \setminus B) \cup (A \cap B)$: the first of these terms contains all the elements of A that **are not** in B, the other contains all the elements of A that **are** in B. These sets are disjoint, so $P(A) = P(A \setminus B) + P(A \cap B)$. Likewise, $P(B) = P(A \setminus B) + P(A \cap B)$. Further, we have

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A).$$

The three sets on the right-hand side are disjoint; thus

$$\begin{aligned} P(A \cup B) &= P(A \setminus B) + P(A \cap B) + P(B \setminus A) \\ &= P(A) + P(B \setminus A) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

□

6.1 Conditional Probability

Given an event $A \in \mathcal{F}$, we can define a **conditional probability measure** $P(\cdot|A)$ that tells us what the probabilities of other events in $\mathcal{F}$ are, given that the outcome is in E. The conditional probability measure has to be consistent with P: for all $B \in \mathcal{F}$, we have

$$P(A \cap B) = P(B|A)P(A).$$

This is called the **law of total probability**. Thus, for an event A for which $P(A) \neq 0$, the conditional probability measure on A is uniquely defined: For any $B \in \mathcal{F}$,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Note that, unless $P(A) = P(B)$, $P(A|B) \neq P(B|A)$. So how do we find $P(B|A)$ (for instance, the probability that one has a disease given that one tests positive for it) from $P(A|B)$ (for instance, the probability that the test is positive given that a person has the disease)? We use what's called **Bayes' rule**. Since $P(A \cap B) = P(A|B)P(B)$, we have

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)},$$

and since $P(A) = P(A \cap B) + P(A \cap B^C)$,

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^C)P(B^C)}.$$

Example 45. Suppose you are a doctor. Your patients routinely undergo screening for a certain disease. This disease is relatively common: of those recommended to undergo screening, about 1 in 100 have the disease. The test has false positive and false negative rates of 1%. That is, of every 100 people who have the disease, about 99 will test positive,

and of every 100 people who do not have the disease, about 1 will test positive. Of every 100 people who test positive, how many have the disease?

For each individual patient, we can write the relevant state space as

$$\Omega = \{(+, \text{has disease}), (+, \text{no disease}), (-, \text{has disease}), (-, \text{no disease})\}.$$

We can write some of the relevant events more succinctly as follows:

$$\begin{aligned} + &\equiv \{(+, \text{has disease}), (+, \text{no disease})\} \\ - &\equiv \{(-, \text{has disease}), (-, \text{no disease})\} \\ D &\equiv \{(+, \text{has disease}), (-, \text{has disease})\} \\ N &\equiv \{(+, \text{no disease}), (-, \text{no disease})\} \end{aligned}$$

In essence, the question asks: if we assign probabilities to these events according to their frequency — i.e., let $P(D) = .01$, $P(+|D) = .99$, and $P(+|N) = .01$ — what is $P(D|+)$?

According to Bayes' rule,

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|N)P(N)} = \frac{.99 \times .01}{.99 \times .01 + .01 \times (1 - .01)} = \frac{1}{2}$$

So if we assign probabilities to these events according to their frequency, then the probability that a patient who tests positive has the disease is .5. In other words, of every 100 people that test positive, only 50 have the disease — despite false positive and false negative rates of only 1%. This is because **prior probabilities** (here, the 1 in 100 who actually have the disease) continue to matter even after receiving more information: a positive test should only cause us to **update** our beliefs about a patient's disease status, not overturn our priors entirely.

Example 46 (Monty Hall Problem)**.** You are a contestant on a game show. There are three doors in front of you. Behind one of them — randomly chosen with equal probability — is a prize. Behind the other two are goats.

You get to pick one of them — let's call your choice door 1. Then, the host picks one of the doors you **did not** pick — either door 2 or door 3 — which does not have a prize behind it, and opens it to reveal a goat.

Then, you get to choose between the remaining doors, and receive whatever is behind it. Should you switch doors?

Let's figure out the probability that the prize is behind door 1 versus behind whichever door the host did not open. Let's let an element of the state space be (x, y) where x is the

door that the prize is behind and y is the door that the host opens. We want to find $P(x = 2|y = 3)$ and $P(x = 3|y = 2)$.

From Bayes' rule, $P(x = 2|y = 3) = \frac{P(2,3)}{P(y=3)}$ and $P(x = 3|y = 2) = \frac{P(3,2)}{P(y=2)}$.

We have $P(\{(1,2),(1,3)\}) = P(\{(2,2),(2,3)\}) = P(\{(3,2),(3,3)\}) = \frac{1}{3}$: the x values are equally likely. But the host is never going to reveal a prize: $P(3,3) = P(2,2) = 0$. Since $\frac{1}{3} = P(x = 2) = P(2,2) + P(2,3)$, we have $P(2,3) = \frac{1}{3}$; similarly, $P(3,2) = \frac{1}{3}$.

Now we need to find $P(y = 3)$. This depends on how the host chooses between doors when the prize is behind door 1. Let's let $P(1,3) = z$ and $P(1,2) = \frac{1}{3} - z$, for $z \in [0, \frac{1}{3}]$. (They have to sum to $\frac{1}{3}$, since $P(x = 1) = \frac{1}{3}$.) Then

$$\begin{aligned} P(y = 3) &= P(1,3) + P(2,3) + P(3,3) = z + \frac{1}{3} + 0 \\ P(y = 2) &= P(1,2) + P(2,2) + P(3,2) = \frac{1}{3} - z + 0 + \frac{1}{3} \\ \Rightarrow P(x = 2|y = 3) &= \frac{1/3}{z + 1/3} \in [\frac{1}{2}, 1] \\ P(x = 3|y = 2) &= \frac{1/3}{2/3 - z} \in [\frac{1}{2}, 1] \end{aligned}$$

So the probability of winning after switching is always at least as high as after staying on door 1, and sometimes higher.

6.1.1 Independence

We say that two events A and B are **independent** if $P(A \cap B) = P(A)P(B)$. Thus, conditioning on one does not affect the probability of the other: if $P(A)$ and $P(B)$ are nonzero, we have $P(A) = P(A|B)$ and $P(B|A) = P(B)$.

We say a collection of events $A_1, A_2, \ldots, A_n$ are **pairwise independent** if **all** pairs of those events are independent. (Independence isn't transitive: Just because A and B are independent and B and C are indepenedent, doesn't mean A and C are.)

We say a collection of events $A_1, A_2, \ldots, A_n$ are **mutually independent** if each event is independent of **any intersection of two or more of the other events.** (This is stronger than pairwise independence, which I unfortunately misremembered in class.)

Example 47. Show that if A and B are independent, then so are A^C and B^C.

Proof. We have $P(A^C \cap B^C) = P((A \cup B)^C) = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B)) = 1 - P(A) - P(B) + P(A)P(B) = (1 - P(A))(1 - P(B))$. □

6.2 Random Variables and Expected Values

Suppose we have a function $X : \Omega \to \mathbb{R}$. When Ω is finite, we call such a function a **discrete random variable**: a function that associates a real number with each of finitely many outcomes.

There are lots of things that we can do with discrete random variables. One of the most useful is to take their **expected value**: the average value taken on by $X(\omega)$. We denote this as $E_P[X(\omega)]$, or if it is not ambiguous, just $E[X(\omega)]$ or $E[X]$, and define it as

$$E_P[X(\omega)] \equiv \sum_{\omega \in \Omega} P(\omega)X(\omega).$$

Example 48 (Expected Value of a Fair Dice Roll)**.** Suppose I want to know the expected roll of a (fair) 6-sided die. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$. Since the die is fair, $P(\omega) = \frac{1}{6}$ for each $\omega \in \Omega$. Then I would take the expected value of $X(\omega) = \omega$:

$$E_P[X] = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = 3.5.$$

Likewise, I could find the expected roll of a loaded die: consider the probability measure $\hat{P}$ with $\hat{P}(6) = \frac{1}{2}$ and $\hat{P}(\omega) = \frac{1}{10}$ for $\omega \neq 6$: in this case,

$$E_{\hat{P}}[X] = \frac{1}{10} + \frac{2}{10} + \frac{3}{10} + \frac{4}{10} + \frac{5}{10} + \frac{6}{2} = 4.5.$$

This is the basic way that we think about the way agents evaluate uncertain situations: they have some utility function $u : \Omega \to \mathbb{R}$ that takes on different values in different states of the world, and they prefer one probability measure over outcomes P to another $\hat{P}$ if and only if $E_P[u(\omega)] \geq E_{\hat{P}}[u(\omega)]$.

Example 49 (Expected Utility from a Fair Dice Roll)**.** Now suppose, for the sake of example, that my utility from a dice roll is given by

$$u(x) = 12x - x^2$$

So that means that

$$\begin{aligned} u(1) &= 11 & \qquad u(4) &= 32 \\ u(2) &= 20 & \qquad u(5) &= 35 \\ u(3) &= 27 & \qquad u(6) &= 36 \end{aligned}$$

Let P represent the distribution over outcomes generated by a fair die and $\hat{P}$ represent the distribution generated by a loaded die. We have

$$\begin{aligned} E_P[u] &= \frac{11}{6} + \frac{20}{6} + \frac{27}{6} + \frac{32}{6} + \frac{35}{6} + \frac{36}{6} \\ &= \frac{161}{6} \approx 29.8 \\ E_{\hat{p}}[u] &= \frac{11}{10} + \frac{20}{10} + \frac{27}{10} + \frac{32}{10} + \frac{35}{10} + \frac{36}{2} \\ &= \frac{125}{10} + 18 = 30.5 \end{aligned}$$

We may also want to measure how dispersed the values are that f takes on. We do this by taking the **variance** of x, defined as

$$\text{Var}(X) \equiv E[(X(\omega) - E[X])^2]$$

We sometimes denote $\text{Var}(X)$ as σ_X^2.

Likewise, if we have another random variable, $Y : \Omega \to \mathbb{R}$, we may want to measure how much the values taken by these variables move together. We do this with the **covariance** of X and Y:

$$\text{Cov}(X, Y) \equiv E[(X(\omega) - E[X])(Y(\omega) - E[Y])]$$

We sometimes denote this as σ_{XY}.

Sometimes, we want to think about discrete random variables without reference to the underlying sample space. The **probability mass function** lets us do this: define $p_x : \mathbb{R} \to [0, 1]$ as

$$p_X(x) \equiv P(X^{-1}(x))$$

That is, $p_X(x)$ gives the probability of the preimage of the number z under x. Then we can write

$$E[X] = \sum_{x \in X(\Omega)} p_X(x)x.$$

What about continuous random variables? That is, what if the underlying sample space is no longer finite, and X takes on a continuum of values? In this case, X may no longer have a probability mass function: it may be (in fact, it is likely to be) the case that the probability that X is in some interval (a, b) is positive while the probability that X is *any individual number* in that interval is zero. Instead, when we want to avoid working with the sample space, we use the **cumulative distribution function** of X, $F_X : \mathbb{R} \to [0, 1]$.

The cumulative distribution function tells us the probability that X takes on values less than or equal to a number:

$$F_X(z) \equiv P(X^{-1}((-\infty, z]))$$

If F_x is continuously differentiable, then there is a continuous analogue to the probability mass function, called the **probability density function** $f_X : \mathbb{R} \to \mathbb{R}$, defined as

$$f_X(x) \equiv \frac{d}{dx} F_X(x).$$

This allows us to write an expected value as

$$E[x] = \int_{-\infty}^{\infty} x f(x) dx.$$

6.2.1 Probability Homework

- Simon and Blume Exercise A5.4