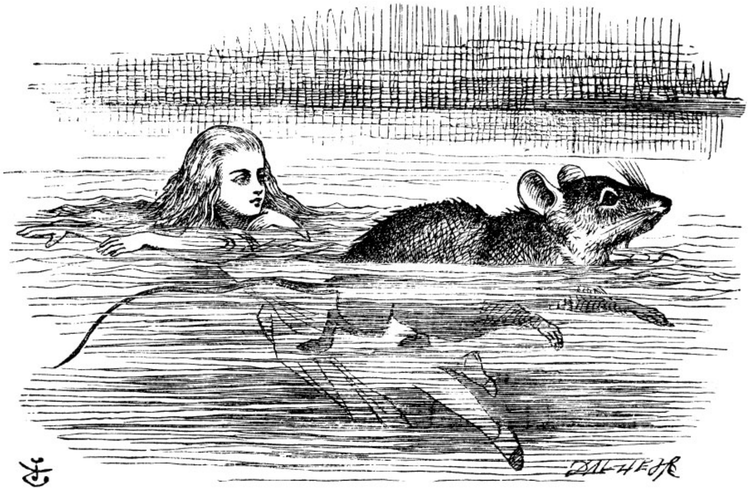
# Economics PhD Math Camp

University of Missouri

First edition

Saku Aura
David M. Kaplan





*1st edition, Summer 2024*
*Updated July 22, 2024*

*However, she soon made out that she was in the pool of tears which she had wept when she was nine feet high. 'I wish I hadn't cried so much!' said Alice, as she swam about, trying to find her way out. . . . So she began: 'O Mouse, do you know the way out of this pool? I am very tired of swimming about here, O Mouse!'*

Lewis Carroll, *Alice's Adventures in Wonderland*

# Contents

# Preface

This text was prepared for the intensive two-week (30-hour) "math camp" for incoming Economics PhD students at the University of Missouri. The material was refined over the first author's many years of teaching, with the second author revising, editing, and making everything look pretty. In math camp, students review (or learn) the mathematical concepts and tools to help them succeed in the first-year core Economics PhD courses in microeconomics, macroeconomics, and econometrics.

This text's source files are freely available. Instructors may modify them as desired, or copy and paste LaTeX code into their own lecture notes, subject to the Creative Commons license linked on the copyright page. The text was written in Overleaf, an online (free) LaTeX environment. You may see, copy, and download the entire project from Overleaf[1] or from Kaplan's website.[2]

Other free resources are linked at the end of each chapter. Wikipedia is also a good reference for most topics, so we do not explicitly link to every single page. Besides asking your instructor, you can also ask question on one of the Stack Exchange websites, or often you will find somebody already asked (and answered) your question when you search the site. There are sites for Economics (https://economics.stackexchange.com/), Mathematics (https://math.stackexchange.com/), and Statistics (https://stats.stackexchange.com/), among others.

We provide learning objectives for the overall book and for each chapter. This follows current best practices for course design.

Thanks to our colleagues for feedback, and special thanks to Wei Zhao (Mizzou PhD alumna) for typing up the majority of the initial hand-written notes.

Saku Aura and David M. Kaplan
Summer 2024
Columbia, Missouri, USA

[1] https://www.overleaf.com/read/vtphdxkssxpg#7f8b91
[2] https://kaplandm.github.io/teach.html

# Textbook Learning Objectives

For good reason, it has become standard practice to list learning objectives for a course as well as each unit within the course. Below are the overall learning objectives for math camp. Each chapter lists more specific learning objectives that map to one or more of these overall objectives.

The textbook learning objectives (TLOs) are the following.

1. Define terms from calculus, linear algebra, and other mathematical fields used in economics, both mathematically and intuitively.

2. Solve certain types of (systems of) equations common in economics.

3. Characterize the properties of mathematical (systems of) equations common in economics.

# Notation

## Variables

Unless otherwise specified, a plain lowercase letter like $x$ is a scalar, while bold lowercase $\boldsymbol{x}$ is a vector (specifically a column vector unless otherwise specified), and bold uppercase $\boldsymbol{A}$ is a matrix. The individual elements of a vector or matrix are scalars; for example, $\boldsymbol{x}$ has $j$th element $x_j$, and $\boldsymbol{x} = (x_1, \ldots, x_J)'$; and the row $i$, column $j$ element of matrix $\boldsymbol{A}$ is $A_{ij}$.

The following symbols are used.

*Functions*

| | |
|---|---|
| $f\colon X \to Y$ | function $f$ assigns a value in $Y$ to each value in $X$ |
| $f'(x)$ or $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$ | first derivative of $f$ evaluated at $x$ |
| $\frac{\partial f(x,y)}{\partial x}$ or $f_x$ | first partial derivative of $f$ with respect to $x$ |
| $\nabla f$ | gradient (Jacobian) of function $f$ |
| $f \in C^k$ | function $f$ is $k$ times continuously differentiable |
| $\circ$ | function composition |

*Correspondences*

| | |
|---|---|
| $f\colon X \rightrightarrows Y$ | correspondence $f$ assigns a subset of $Y$ to each value in $X$ |
| $f\colon X \twoheadrightarrow Y$ | correspondence $f$ assigns a subset of $Y$ to each value in $X$ |

*Logic*

| | |
|---|---|
| $\Longrightarrow$ | implies |
| $\Longleftarrow$ | is implied by |
| $\Longleftrightarrow$ | if and only if |
| iff | if and only if |

*Convergence*

| | |
|---|---|
| $\lim\limits_{n\to\infty}$ | limit (like in pre-calculus) |
| $\to$ | converges to (like in pre-calculus) |

*Other binary relations*

| | |
|---|---|
| $\equiv$ | is defined as |
| $\approx$ | approximately equals (informal) |
| $\sim$ | is distributed as (or "follows") |

*Probability*

| | |
|---|---|
| $\mathrm{N}(\mu, \sigma^2)$ | normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $\mathrm{N}(0,1)$ | standard normal distribution |
| $F_Y(\cdot)$ | cumulative distribution function (CDF) of $Y$ |
| $f_Y(\cdot)$ | PMF of $Y$ if discrete; PDF if continuous |
| $\mathbb{1}\{\cdot\}$ | indicator function |
| $\mathrm{P}(A)$ | probability of event $A$ |
| $\mathrm{P}(A \mid B)$ | conditional probability of $A$ given $B$ |
| $\mathrm{E}(Y)$ | expectation (mean) of $Y$ |
| $\mathrm{Var}(Y)$ | variance of $Y$ |
| $\mathrm{Cov}(Y, X)$ | covariance |
| $\mathrm{Corr}(Y, X)$ | correlation |

*Sets*

| | |
|---|---|
| $\{a, b, \ldots\}$ | a set (containing elemnts $a$, $b$, etc.) |
| $i = 1, \ldots, n$ | same as $i \in \{1, \ldots, n\}$ (integers from 1 to $n$) |
| $j = 1, \ldots, J$ | same as $j \in \{1, \ldots, J\}$ (integers from 1 to $J$) |
| $x \in A$ | element $x$ is in set $A$ |
| $x \notin A$ | element $x$ is not in set $A$ |
| $\emptyset$ | the empty set |
| $\mathbb{R}$ | the set of real numbers |
| $\mathbb{N}$ | the set of natural numbers |
| $A \cup B$ | union of sets $A$ and $B$ |
| $A \cap B$ | intersection of sets $A$ and $B$ |
| $A^{\complement}$ | complement of set $A$ |
| $A \setminus B$ | set difference of sets $A$ and $B$ |
| $A \times B$ | Cartesian product of sets $A$ and $B$ |
| $\overline{A}$ | closure of set $A$ |
| $\partial A$ | boundary of set $A$ |
| $\mathrm{int}\, A$ | interior of set $A$ |
| $\subseteq$ | subset |
| $\supseteq$ | superset |
| $\subsetneq$ | proper subset |
| $\supsetneq$ | proper superset |

*Optimization*

| | |
|---|---|
| $\min_{x \in S} f(x)$ | the lowest possible value of $f(x)$, i.e., $\min\{f(x) : x \in S\}$ |
| $\max_{x \in S} f(x)$ | the highest possible value of $f(x)$, i.e., $\max\{f(x) : x \in S\}$ |
| $\arg\min_{x \in S} f(x)$ | the value of $x$ that minimizes $f(x)$ |
| $\arg\max_{x \in S} f(x)$ | the value of $x$ that maximizes $f(x)$ |
| s.t. | subject to |

*Matrices and vectors*

| | |
|---|---|
| $A'$ or $A^\top$ or $A^T$ | transpose of $A$ |
| $x \cdot y$ | dot product of vectors $x$ and $y$ |
| $A^{-1}$ | inverse of matrix $A$ |
| $\mathrm{adj}(A)$ | adjoint of matrix $A$ |
| $\det(A)$ or $\lvert A \rvert$ | determinant of matrix $A$ |
| $I_n$ | the $n \times n$ identity matrix |
| $\mathrm{tr}(A)$ | trace of matrix $A$ |

# Chapter 1
# Sets, Functions, and Sequences

This chapter introduces notation and concepts for working with sets, functions, and sequences, all important mathematical objects in economics.

*Unit learning objectives for this chapter*

1.1. Define vocabulary words (in **bold**) related to sets, functions, and sequences, both mathematically and intuitively [TLO 1]

1.2. Perform various operations on sets [TLOs 2 and 3]

1.3. Assess whether or not specific (pairs of) sets, functions, or sequences possess certain properties [TLO 3]

## 1.1 Sets

In this section, we introduce set notation and concepts, operations, properties, and relations. Sometimes sets are written in script like $\mathcal{A}$, but we just use uppercase like $A$.

### 1.1.1 Describing sets

**Definition 1.1.** A **set** is a collection of objects, called **elements** or **members**. A set does not count how many times an element appears (i.e., a set is not a multiset), nor does a set keep track of order (i.e., a set is not a tuple). The notation $x \in A$ means that $x$ "is an element of" (or "is in" or "belongs to") set $A$. Similarly, $x \notin A$ means that $x$ is not an element of $A$.

Notationally, there are a few ways to show a set's elements. This is called **set-builder notation**. Most simply, the elements can be written in list inside curly braces $\{\}$ and separated by commas, like in Example 1.1. If the elements follow an obvious pattern, then you can replace some with an ellipsis $\ldots$ like in Example 1.2. In both cases, this is defining sets by **enumeration**. Sets can also be defined by a **predicate**, meaning a condition that is true if and only if an object is an element of the set; usually the domain is explicitly specified, like $x \in \mathbb{R}$ or $n \in \mathbb{N}$. This is indicated using a colon : or vertical bar | as in Example 1.3.

**Example 1.1.** The following are sets of numbers, animals, and functions, respectively: $A = \{1, 2, 3\}$, $B = \{\text{ant}, \text{bat}, \text{cat}\}$, $C = \{\sin(\cdot), \cos(\cdot), \tan(\cdot)\}$. ■

**Example 1.2.** The set $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ can be written as $A = \{1, \dots, 9\}$. The set of natural numbers can be written as $\mathbb{N} = \{1, 2, 3, \dots\}$. Verbal descriptions help when the pattern is less obvious; for example, "Define the set of square numbers as $S = \{1, 4, 9, \dots\}$" is more clear than "Let $S = \{1, 4, 9, \dots\}$." ■

**Example 1.3.** The following sets are defined by a predicate: $A = \{\, x \in \mathbb{N} : x \leq 3 \,\}$, $B = \{\, x \in \mathbb{R} \mid x^2 \in \mathbb{N} \,\}$. ■

The two special sets in Definitions 1.2 and 1.3 are often useful.

**Definition 1.2.** The **empty set** is denoted $\emptyset$ (or $\varnothing$ or other variations) and contains no elements: $\emptyset \equiv \{\}$. (Note: this differs from a null set.)

**Definition 1.3.** The **universal set** is commonly denoted $\Omega$ (but sometimes $U$ or $S$ or others); it depends on the context and contains all possible elements (the universe) relevant for that context.

**Example 1.4.** If we are considering the outcome of a six-sided die being rolled, then we may want to define the universal set as $\Omega = \{1, 2, 3, 4, 5, 6\}$, or equivalently $\Omega = \{1, \dots, 6\}$, and the set of possible outcomes that are both odd and even is $\emptyset$. If instead we are looking at a country's annual economic growth, then we might define $\Omega = \{\text{expansion}, \text{contraction}, \text{zero growth}\}$. ■

**Definition 1.4.** The **cardinality** of set $A$ is commonly denoted $|A|$ and is a measure of the "size" of $A$ (the number of elements). A **finite set** has a finite number of elements. A **countably infinite set** has an infinite number of elements, but they have a one-to-one mapping to the natural numbers. An **uncountable set** (or uncountably infinite set) has cardinality greater than the set of natural numbers.

**Example 1.5.** The set $A = \{-2, 0, 2\}$ has $|A| = 3$. The set of all square numbers $B = \{1, 4, 9, \dots\}$ is countably infinite because we can uniquely map each element to one of the natural numbers; for example, we can label the elements of $B$ as $b_n$ for $n = 1, 2, \dots$, with $b_n = n^2$. (This should be at least somewhat surprising or confusing because clearly $B \subsetneq \mathbb{N}$, yet $|B| = |\mathbb{N}|$!) The set of real numbers $\mathbb{R}$ is uncountable, as is the continuum of real numbers $[0, 1] = \{\, x \in \mathbb{R} : 0 \leq x \leq 1 \,\}$, as shown by Cantor's diagonal argument: essentially, if we think we succeeded in writing out every real number with one in each row (so we can label the rows $1, 2, \dots$), then we can always create a new real number that differs from all of these by making the first decimal digit differ from the first number's first digit, making the second decimal digit differ from the second number's, making the third digit differ from the third number's, etc.; for example, if we started with 0.935, 0.748, and 0.123, then 0.876 changes the first number's 9 to an 8, changes the second's 4 to a 7, and changes the third's 3 to a 6. ■

### 1.1.2 Operations

The union, intersection, complement, and set difference are the most basic set operations.

**Definition 1.5.** The **union** of sets $A$ and $B$ is $A \cup B \equiv \{ x : x \in A \text{ or } x \in B \}$.

**Definition 1.6.** The **intersection** of sets $A$ and $B$ is $A \cap B \equiv \{ x : x \in A \text{ and } x \in B \}$.

**Definition 1.7.** The **complement** of set $A$ is $A^{\mathsf{C}} \equiv \{ x \in \Omega : x \notin A \}$, where $\Omega$ is the universal set. That is, $A^{\mathsf{C}}$ is all elements in the universal set that are not in $A$. Notational variations include $A^{\mathsf{c}}$ and $\overline{A}$.

**Definition 1.8.** The **set difference** of sets $A$ and $B$ is $A \setminus B \equiv \{ x \in A : x \notin B \}$.

The Cartesian product is often useful for defining domains.

**Definition 1.9.** The **Cartesian product** of sets $A$ and $B$ is $A \times B \equiv \{ (a, b) : a \in A \text{ and } b \in B \}$. The Cartesian product of a set with itself is often written with an exponent, like $A^2 \equiv A \times A$, or $A^4 \equiv A \times A \times A \times A$.

**Example 1.6.** If $A = \{1, 2\}$ and $B = \{3, 4\}$, then $A \times B = \{(1, 3), (1, 4), (2, 3), (2, 4)\}$. ■

**Example 1.7.** The set of all real two-dimensional vectors can be written as the Cartesian product $\mathbb{R} \times \mathbb{R}$ or $\mathbb{R}^2$. That is, $(x, y) \in \mathbb{R}^2$ means $x \in \mathbb{R}$ and $y \in \mathbb{R}$. ■

### 1.1.3 Set relationships

The following definitions concern basic relationships between two sets.

**Definition 1.10.** Sets $A$ and $B$ are equal when $x \in A \iff x \in B$; that is, each element of $A$ is an element of $B$, and each element of $B$ is an element of $A$.

**Definition 1.11.** Two sets are **disjoint** when they do not have any element in common: $A \cap B = \emptyset$.

**Definition 1.12.** The notation $A \subseteq B$ indicates that $A$ is a **subset** of $B$, meaning $x \in A \implies x \in B$. This is equivalent to saying that $B$ is a **superset** of $A$, written $B \supseteq A$.

**Definition 1.13.** The notation $A \subsetneq B$ indicates that $A$ is a **proper subset** of $B$, meaning that $A \subseteq B$ and $A \neq B$ ($A$ is a subset of $B$, but they are not equal). This is equivalent to saying that $B$ is a **proper superset** of $A$, written $B \supsetneq A$.

The symbols $\subset$ and $\supset$ can be confusing because they are used with different meaning by different authors. Sometimes they mean subset and superset; sometimes they mean proper subset and proper superset. Be careful when you see them, and try to avoid them in your own work.

**Example 1.8.** Set equality $A = B$ can equivalently be defined as: $A \subseteq B$ and $B \subseteq A$. ■

**Example 1.9.** $\{1, 2, 3\} = \{3, 2, 1\}$ (recall elements in a set do not have an order). ■

**Example 1.10.** $\{\, x : x \text{ is prime factor of } 8 \,\} = \{\, x : x \text{ is prime factor of } 1024 \,\}$ (recall elements do not appear multiple times). ■

**Example 1.11.** If $A = \{\, x \in \mathbb{R} : x < 0 \,\}$ and $B = \{\, x \in \mathbb{R} : x \leq 0 \,\}$, then it is true that $A \subsetneq B$, $A \subseteq B$, $B \supsetneq A$, and $B \supseteq A$. ■

### 1.1.4 Properties

The following properties are helpful for understanding more complex sets.

**Proposition 1.1.** *The* ***distributive law*** *says that for sets $A, B, C$,*

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

**Proposition 1.2.** ***De Morgan's laws*** *say that*

$$(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement},$$
$$(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}.$$

*This generalizes to*

$$A_1^{\complement} \cup A_2^{\complement} \cup \cdots \cup A_N^{\complement} = \bigcup_{i=1}^{N} A_i^{\complement} = \left( \bigcap_{i=1}^{N} A_i \right)^{\complement} = (A_1 \cap A_2 \cap \cdots \cap A_N)^{\complement},$$
$$A_1^{\complement} \cap A_2^{\complement} \cap \cdots \cap A_N^{\complement} = \bigcap_{i=1}^{N} A_i^{\complement} = \left( \bigcup_{i=1}^{N} A_i \right)^{\complement} = (A_1 \cup A_2 \cup \cdots \cup A_N)^{\complement}.$$

### 1.1.5 Relations

Although the formal definition of a relation may seem confusing, the examples after should make it clear.

**Definition 1.14.** A **relation** $R$ describes whether or not values $a \in A$ and $b \in B$ satisfy a certain condition; specifically, this is a **binary relation** because it is between elements from two sets ($A$ and $B$). Formally, this can be written as $aRb$ if the relation holds for $a$ and $b$, sometimes read as "$a$ is $R$-related to $b$." Alternatively, $R$ may be defined as the subset of $A \times B$ for which the relation holds; that is, $R \subseteq A \times B$ and the relation holds if and only if $(a, b) \in R$.

**Example 1.12.** Consider the "less than" relationship between two natural numbers: it holds when $a < b$. Alternatively, define $R = \{ (a, b) \in \mathbb{N}^2 : a < b \}$; then $a < b$ is equivalent to $(a, b) \in R$. ■

**Example 1.13.** Consider the relation "is the same species as" for the universal set $\Omega = \{\text{poodle}, \text{husky}, \text{frog}\}$. Then $R = \{ (a, b) \in \Omega^2 : a \text{ same species as } b \}$, or more specifically $R = \{(\text{poodle}, \text{pug}), (\text{pug}, \text{poodle})\}$. (Those are dog breeds, in the same species *Canis familiaris*.) ■

**Definition 1.15.** Let $R \subseteq A \times B$ be a relation. The **domain** of $R$ is $A$, and the **active domain** (or "domain of definition") is $\{ a \in A : (a, b) \in R \text{ for some } b \in B \}$, although sometimes the latter is referred to as the domain (so be careful). The **codomain** of $R$ is $B$, and the **range** (or **image**, or less commonly "active codomain") is $\{ b \in B : (a, b) \in R \text{ for some } a \in A \}$.

The following are properties that a relation may (or may not) have. Let $R$ be a relation defined on $X \times X$.

1. **Reflexive** means $xRx$ for all $x \in X$.
2. **Transitive** means that for all $x, y, z \in X$, if $xRy$ and $yRz$, then $xRz$.
3. **Symmetric** means that for all $x, y \in X$, if $xRy$, then $yRx$.
4. **Antisymmetric** means that for all $x, y \in X$, if $xRy$ and $yRx$, then $x = y$.
5. **Complete** means that for all $x, y \in X$, $xRy$ or $yRx$ (or both).
6. **Partial order** means that $R$ is reflexive, transitive, and antisymmetric. (Note: this differs from a strict partial order!)
7. **Total order** (or linear order) means that $R$ is a complete partial order. (Note: this differs from a strict total order!)

**Example 1.14.** On the set of real numbers $\mathbb{R}$, the relation $\leq$ is a total order. It is complete because for any $x, y \in \mathbb{R}$, $x \leq y$ or $y \leq x$ (or both). It is a partial order because it is reflexive, transitive, and antisymmetric. It is reflexive because for any $x \in \mathbb{R}$, $x \leq x$. It is transitive because $x \leq y$ and $y \leq z$ implies $x \leq z$. It is antisymmetric because if $x \leq y$ and $y \leq x$, then $x = y$. ■

**Example 1.15.** Consider the relation in two-dimensional real Euclidean space that holds iff both coordinates of the first vector are less than the respective coordinates of the second vector. That is, this is the elementwise inequality relation that holds for vectors $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$ and $\boldsymbol{y} = (y_1, y_2) \in \mathbb{R}^2$ iff $x_1 \leq y_1$ and $x_2 \leq y_2$. Equivalently, $R = \{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^2 \times \mathbb{R}^2 : x_1 \leq y_1 \text{ and } x_2 \leq y_2 \}$. This is not complete because it does not hold in either direction for elements like $\boldsymbol{x} = (1, 9)$ and $\boldsymbol{y} = (8, 2)$, for which $x_1 \leq y_1$ but $x_2 > y_2$, and similarly $y_2 \leq x_2$ but $y_1 > x_1$. However, it is a partial order because it is reflexive, transitive, and antisymmetric. It is reflexive because for any $\boldsymbol{x} \in \mathbb{R}^2$, $x_1 \leq x_1$ and $x_2 \leq x_2$. It is transitive because $x_1 \leq y_1$ and $x_2 \leq y_2$ combined with $y_1 \leq z_1$ and $y_2 \leq z_2$ implies $x_1 \leq z_1$ and $x_2 \leq z_2$. It is antisymmetric because if $x_1 \leq y_1$ and $x_2 \leq y_2$, and similarly $y_1 \leq x_1$ and $y_2 \leq x_2$, then $x_1 = y_1$ and $x_2 = y_2$, meaning $\boldsymbol{x} = \boldsymbol{y}$. ■

## 1.2 Functions

First we formally define what a function is, before discussing continuity and limits.

### 1.2.1 Definitions

**Definition 1.16.** A **function** $f: X \to Y$ is a rule that assigns one element of $Y$ to each element of $X$: for each $x \in X$, there is a single value $f(x) = y \in Y$. Alternatively, from the set theory perspective, a function $F$ is a relation (Definition 1.14) whose domain is $X$, and for each $x \in X$, there is exactly one $y \in Y$ such that $xFy$, or equivalently such that $(x, y) \in F$.

**Example 1.16.** Let $f: \{0,1\} \to \{0,1\}$ be $f(x) = 1 - x$. Note that each $x \in \{0,1\}$ generates a single value of $f(x)$ because $f(0) = 1$ and $f(1) = 0$. Using the set theoretic definition, $f = \{(0,1),(1,0)\}$. ■

**Example 1.17.** Let $f: \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$. Set theoretically, this is equivalent to $f = \{\, (x, y) \in \mathbb{R}^2 : y = x^2 \,\}$. ■

**Definition 1.17.** A **correspondence** $f: X \rightrightarrows Y$ maps each value $x \in X$ to a subset of $Y$. Other terms for a correspondence include **set-valued function** and **multifunction**.

**Example 1.18.** Consider standard function $f: \{0,1\} \to \{0,1\}$ compared with correspondence $g: \{0,1\} \rightrightarrows \{0,1\}$. For $x = 0$, the standard function's value must be either $f(0) = 0$ or $f(0) = 1$. Note these are both elements of $\{0,1\}$, not subsets. Instead of $f(0) = 0$ or $f(0) = 1$, a correspondence would have $g(0) = \{0\}$ or $g(0) = \{1\}$, where $\{0\}$ and $\{1\}$ are subsets of $\{0,1\}$ rather than elements. The correpondence could also have $g(0) = \{0,1\}$. (Also possible is $g(0) = \emptyset$, although such $x \in X$ with $g(x) = \emptyset$ are not part of the "domain" of $g$.) ■

Here we focus on functions; correspondences are described further in Section 9.1.

### 1.2.2 Continuity and limits

We first formally define limits, for the purpose of defining continuity. We start with the univariate case, followed by the multivariate generalization.

### Limits

**Definition 1.18.** Given univariate function $f\colon S \to \mathbb{R}$ (for some $S \subseteq \mathbb{R}$), the **limit** $\lim_{x\to x_0} f(x) = y_0$ means that for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in S$, $0 < |x - x_0| < \delta$ implies $|f(x) - y_0| < \epsilon$.

Figure 1.1 illustrates this epsilon–delta definition of a limit.

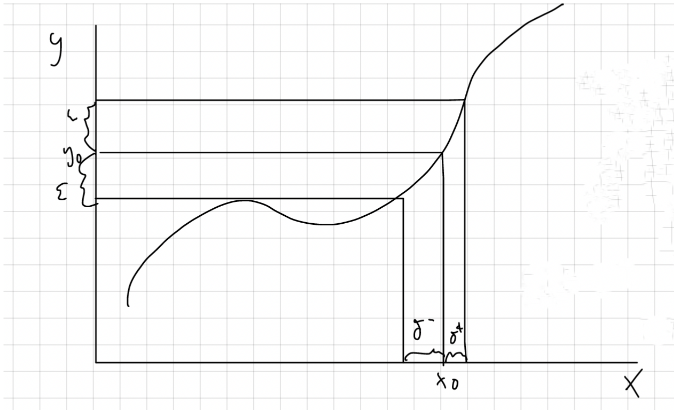



Figure 1.1: Definition of limit, trying to find $\delta$ for each arbitrarily small $\epsilon > 0$.

**Example 1.19.** For $f(x) = x^2$, we can formally show that $\lim_{x\to 0} f(x) = 0$. Given any $\epsilon > 0$, the following are all equivalent:

$$|f(x) - 0| < \epsilon \iff |x^2| < \epsilon \iff x^2 < \epsilon \iff -\sqrt{\epsilon} < x < \sqrt{\epsilon} \iff |x| < \sqrt{\epsilon}$$
$$\iff |x - 0| < \sqrt{\epsilon}.$$

Thus, writing $\delta$ as a function of $\epsilon$, $\delta(\epsilon) = \sqrt{\epsilon}$ satisfies the condition. That is, using Definition 1.18 with $x_0 = y_0 = 0$ and $\delta = \sqrt{\epsilon}$, we have proved that $0 < |x - x_0| < \delta = \sqrt{\epsilon}$ implies $|f(x) - y_0| < \epsilon$. ■

### Continuity

**Definition 1.19.** A function $f$ is **continuous** at point $x_0$ when $\lim_{x\to x_0} f(x) = f(x_0)$ (which requires that the limit exists).

Continuity over a region essentially means continuous at every point in the region (with one-sided limits for boundary points). (Note: there is also a stronger notion of uniform continuity.)

**Example 1.20.** Consider $f(x) = x^2$ for $x > 0$. We show that $\lim_{x\to x_0} = x_0^2$. First, note the equivalence $|f(x) - x_0^2| < \epsilon \iff |x^2 - x_0^2| < \epsilon$.

For the case $x > x_0$, with $0 < x - x_0 < \delta_+$:

$$\begin{aligned} x^2 - x_0^2 < \epsilon &\iff x^2 < x_0^2 + \epsilon \iff x < \sqrt{x_0^2 + \epsilon} \iff x_0 + \delta_+ < \sqrt{x_0^2 + \epsilon} \\ &\iff \delta_+ < \sqrt{x_0^2 + \epsilon} - x_0. \end{aligned}$$

For the case $x < x_0$, with $0 < x_0 - x < \delta_-$:

$$\begin{aligned} x^2 - x_0^2 < \epsilon &\iff x^2 > x_0^2 - \epsilon \iff x > \sqrt{x_0^2 - \epsilon} \iff x_0 - \delta_- > \sqrt{x_0^2 - \epsilon} \\ &\iff \delta_- < x_0 - \sqrt{x_0^2 - \epsilon}. \end{aligned}$$

Combining these,

$$\delta(\epsilon, x_0) = \min\left\{\sqrt{x_0^2 + \epsilon} - x_0,\, x_0 - \sqrt{x_0^2 + \epsilon}\right\} = \sqrt{x_0^2 + \epsilon} - x_0$$ ■

Continuity is generalized to multivariate functions in Section 8.3.1.

## 1.3 Sequences

In economics, we are often interested in whether or not a particular sequence converges, and if so then what value it converges to.

First, we define some properties of sets that translate to properties of sequences.

**Definition 1.20.** Let $S \subsetneq \mathbb{R}$ be a bounded set. If value $b \in \mathbb{R}$ satisfies $b \geq x$ for all $x \in S$, then $b$ is an upper bound of $S$. The **supremum** of $S$ is the least upper bound, denoted $\sup S$ or sometimes $\sup(S)$; that is, $\sup S$ is an upper bound of $S$, and $\sup S \leq b$ for any other upper bound $b$. If $\sup S \in S$, then the **maximum** of $S$ is well-defined, with $\max S = \sup S$. Similarly, the **infimum** of $S$ is the greatest lower bound, and if $\inf S \in S$, then the **minimum** of $S$ is well-defined, with $\min S = \inf S$.

**Example 1.21.** Let $S = \{1, 2, \ldots, 10\}$. The values 99, 34, and 15 are all upper bounds. The least upper bound is $\sup S = 10$ because $x \leq 10$ for all $x \in S$ and there is no value $b < 10$ that satisfies $x \leq b$ for all $x \in S$. Further, $\sup S \in S$, so $\max S = 10$. The greatest lower bound is $\inf S = 1$, which is in $S$, so $\min S = 1$. ■

**Example 1.22.** Let $S = \{1/z : 0 < z < \infty, z \in \mathbb{R}\}$. The values $-99$, $-5$, and 0 are all lower bounds. The greatest lower bound is $\inf S = 0$ because $0 \leq x$ for all $x \in S$. Because $0 \notin S$, the minimum $\min S$ does not exist. Further, there is no value $b \in \mathbb{R}$ that is an upper bound of $S$, because $S$ also includes the value $1/[(1/b)/2] = 2b > b$. ■

We now define sequences and their properties, focusing on convergence.

**Definition 1.21.** A **sequence** of real numbers is a function $x\colon \mathbb{N} = \{1, 2, 3, \ldots\} \to \mathbb{R}$, meaning $k \mapsto x(k)$ with all $x(k) \in \mathbb{R}$, often denoted by $x_k$ for $k = 1, 2, \ldots$. More generally, we can also have sequences of vectors, functions, sets, and other mathematical objects.

**Definition 1.22.** A **monotone sequence** satisfies either $x_{k+1} \geq x_k\ \forall k$ or $x_{k+1} \leq x_k\ \forall k$.

**Definition 1.23.** A **convergent sequence** $\{x_k\}_{k=1}^{\infty}$ converges to value $x \in \mathbb{R}$, denoted $\lim_{k\to\infty} x_k = x$, when $\forall \epsilon > 0$, there exists $N_\epsilon$ such that $k > N_\epsilon \implies |x_k - x| < \epsilon$.

**Theorem 1.3.** *Every bounded monotonic sequence converges.*

**Definition 1.24.** A sequence $\{x_k\}$ is a **Cauchy sequence** iff for every $\epsilon > 0$ there exists $N_\epsilon$ such that $|x_n - x_m| < \epsilon$ for all $n > N_\epsilon$ and $m > N_\epsilon$.

**Theorem 1.4.** *A sequence convergences if and only if it is a Cauchy sequence.*

**Example 1.23.** Let $x_k = 1 + \frac{1}{k}$. We can show that this is a convergent sequence by three different approaches.

1. We can see immediately that $x_k \geq 1$ for all $k$, and that $x_k \leq 2$ for all $k$; thus, the sequence is bounded. Also, $x_{k+1} = 1 + \frac{1}{k+1} < 1 + \frac{1}{k} = x_k$, so it is monotonically decreasing. Thus, by Theorem 1.3, the sequence converges.

2. We can verify that Definition 1.23 applies to our guess that the limit is $x = 1$. Given that the sequence is decreasing and bounded below by 1: given any $N$, for all $k > N$, $1 \leq x_k < x_N$, so $|x_k - 1| < |x_N - 1|$. Thus, if for any $\epsilon > 0$ we can find $N_\epsilon$ such that $|x_{N_\epsilon} - 1| < \epsilon$, it follows that $|x_k - 1| < \epsilon$ for all $k > N_\epsilon$, which would satisfy Definition 1.23. We can solve for $N_\epsilon$ by plugging in $x_{N_\epsilon} = 1 + 1/N_{epsilon}$ to get $|x_{N_\epsilon} - 1| = 1/N_\epsilon$. Thus, the condition $|x_{N_\epsilon} - 1| < \epsilon$ is equivalent to $1/N_\epsilon < \epsilon$, or equivalently $N_\epsilon > 1/\epsilon$, which is well-defined for any $\epsilon > 0$, and passes the sanity check that we should need larger $N_\epsilon$ to get the sequence within a smaller $\epsilon$ of the limit.

3. We can use the Cauchy criterion in Theorem 1.4. Without loss of generality, let $m > n$ and calculate

$$|x_n - x_m| = |1 + (1/n) - [1 + (1/m)]| = |1/n - 1/m| = 1/n - 1/m < 1/n$$

because $0 < 1/m < 1/n$. Thus, if we let $N_\epsilon > 1/\epsilon$, then given $n, m > N_\epsilon$, $|x_n - x_m| < 1/n < 1/N_\epsilon < \epsilon$, satisfying the condition in Theorem 1.4. ■

**Example 1.24.** This simple example shows why it is important to verify convergence instead of only calculating a limit. Consider the geometric series $x_k = 1 + q + q^2 + \cdots + q^{k-1}$. If we try to compute the limit $x = \lim_{k\to\infty} x_k$,

$$\begin{aligned} x &= 1 + q + q^2 + \cdots \\ &= 1 + q(1 + q + q^2 + \cdots) \\ &= 1 + qx, \end{aligned}$$

implying $x(1 - q) = 1$, or $x = 1/(1 - q)$. This is true for certain values of $q$; for example, if $q = 1/2$, then $\lim_{k\to\infty} x_k = 2$. However, for other values of $q$, our formula for $x$ is well-defined yet clearly wrong. For example, if $q = 2$, then our formula says the limit is $1/(1-2) = -1$, but clearly this is not the limit of $1+2+4+8+\cdots$! This nonsense is because the sequence only converges for $-1 < q < 1$, so we can only apply our formula for such $q$. ■

## Optional resources

Optional resources for this chapter

- Wikipedia: sets, set-builder notation, limits (of a function, including $\epsilon$–$\delta$), inf/sup
- KC Border's Intro to Correspondences: `https://web.archive.org/web/20240415203621/https://healy.econ.ohio-state.edu/kcb/Notes/Correspondences.pdf`
- Dartmouth sequences notes and problems: `https://web.archive.org/web/20240308013726/https://math.dartmouth.edu/opencalc2/dcsbook/c1pdf/sec12.pdf`
- `https://web.archive.org/web/20240713110834/https://hummedia.manchester.ac.uk/schools/soss/economics/pg/psmaths/pre-sessiomathsbook.pdf` pages 2–23

## Exercises

**Exercise E1.1.** In econometrics, you want to learn about a parameter $\theta \in \Theta$, where the parameter space $\Theta \subseteq \mathbb{R}$. Different economic theories imply $\theta$ should be in different subsets of the parameter space: $\Theta_1 \subset \Theta$, $\Theta_2 \subset \Theta$, or $\Theta_3 \subset \Theta$, which may (or may not) overlap.

a. You decide to test the joint hypothesis that $\theta \in \Theta_1$ and $\theta \in \Theta_2$, i.e., $H_0\colon \theta \in (\Theta_1 \cap \Theta_2)$. Describe the subset of $\Theta$ where $\theta$ can be if you reject this hypothesis. That is, if $H_0$ is false, then $\theta \in R$; what is $R$? Describe $R$ in two different ways, using Proposition 1.2.

b. Now you test $H_0\colon \theta \in (\Theta_1 \cap \Theta_2 \cap \Theta_3)$. Again, describe set $R$ in two different ways, where $\theta \in R$ is implied by $H_0$ being false.

**Exercise E1.2.** Let $S = \{1,2,3\}$ and consider the relation $R = \{(1,1),(1,3),(2,2),(3,1),(3,3)\}$ on $S \times S$. Say whether or not this relation is reflexive, transitive, symmetric, antisymmetric, and complete, and for each briefly explain why.

**Exercise E1.3.** Let $x_i \in \mathbb{R}$ for $i = 1,\ldots,n$. In econometrics, the "empirical distribution function" based on such observations is $\hat{F}(r) = (1/n)\sum_{i=1}^{n} \mathbb{1}\{x_i \le r\}$, where $\mathbb{1}\{\cdot\}$ is the indicator function, meaning $\mathbb{1}\{x_i \le r\} = 1$ when $x_i \le r$ and otherwise $\mathbb{1}\{x_i \le r\} = 0$. Is $\hat{F}(\cdot)$ a continuous function? Explain either analytically ($\epsilon$–$\delta$) or visually (draw a graph).

**Exercise E1.4.** You may know the famous result that $\sum_{n=1}^{\infty}(1/n)$ is not finite, even though $1/n$ is monotonic and bounded inside the interval $[0, 1]$ (over $n = 1, 2, \ldots$). But then, does this violate Theorem 1.3? Explain.

# Chapter 2

# Linear Algebra: Matrix Math

Matrices are often use in economics, both to store values in an organized way and to define operations on other values.

*Unit learning objectives for this chapter*

2.1. Define vocabulary words (in **bold**) related to matrices, both mathematically and intuitively [TLO 1]

2.2. Perform various operations on a matrix [TLOs 2 and 3]

2.3. Assess whether or not a matrix possesses certain properties [TLO 3]

Note: most of this chapter's material is explained better and with plentiful examples in this excellent online textbook: https://web.archive.org/web/20240516182445/https://textbooks.math.gatech.edu/ila/overview.html

## 2.1 Matrix notation

For this chapter, we use the following notation. A matrix is denoted by a capital letter like $A$ or $B$. Its elements are referred to using the corresponding lowercase letter with subscripts for the row and column, like $a_{ij}$ for the row $i$, column $j$ entry in matrix $A$. The matrix is a rectangular array; if there are $m$ rows and $n$ columns in $A$, then we say $A$ has dimensions $m \times n$, or $A$ is an $m$-by-$n$ matrix, or we write $A_{m\times n}$. Alternatively, we write $(a_{ij})_{m\times n}$, where $a_{ij}$ is the representative matrix entry, for generic row $i$ and column $j$. In some cases, a comma is inserted between the two subscripts for clarity; for example, to denote the entry in row 1 and column 11, $a_{1,11}$ is much more clear than $a_{111}$. To denote a particular row or column of $A$, sometimes a dot is used, like $a_{i\cdot} = (a_{i1}, a_{i2}, \ldots, a_{in})$ for the $i$th row of the matrix, or $a_{\cdot j} = (a_{1j}, \ldots, a_{mj})'$ for the $j$th column. Combining some of these notations,

$$A = (a_{ij})_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [a_{\cdot 1}, \ldots, a_{\cdot n}] = \begin{bmatrix} a_{1\cdot} \\ \vdots \\ a_{m\cdot} \end{bmatrix}. \tag{2.1}$$

**Example 2.1.** In the 2-by-4 matrix $B = \begin{bmatrix} 0 & 9 & 4.1 & 2 \\ 7 & 7 & 4 & 4 \end{bmatrix}$, the entry in row 1 and column 3 has value $b_{1,3} = 4.1$. ■

**Example 2.2.** To indicate a $2 \times 2$ matrix whose entries sum their row and column index, $(i+j)_{2\times 2} = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}$. However, it may be more clear to write $A_{2\times 2}$ with $a_{ij} = i + j$. ■

The **identity matrix** is a special matrix with the following properties and notation. It has the same number of rows and columns. Each entry is either 0 or 1; specifically, the row-$i$ column-$j$ entry is $\mathbb{1}\{i=j\}$, which equals 1 if $i = j$ and zero otherwise. The $n \times n$ identity matrix is

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = (\mathbb{1}\{i=j\})_{n\times n}, \quad \text{where } \mathbb{1}\{i=j\} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.2}$$

## 2.2 Matrix properties

The following are basic properties of a matrix.

**Definition 2.1.** A **square** matrix has the same number of rows as columns. For a square $n \times n$ matrix $A$, the **main diagonal** consists of the elements $a_{ii}$ from $i = 1, \ldots, n$.

**Example 2.3.** The identity matrix from (2.2) is a square matrix. ■

**Definition 2.2.** A **diagonal matrix** is a square matrix with all entries not on the main diagonal equal to zero: $a_{ij} = 0$ for all $i \neq j$. Sometimes a diagonal matrix is specified by giving only its main diagonal elements: $A = \text{diag}(d_1, \ldots, d_n)$ is a diagonal matrix with $a_{ii} = d_i$ and $a_{ij} = 0$ for $i \neq j$.

**Example 2.4.** The matrix $A = \text{diag}(1, 2, 3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ is diagonal. ■

**Definition 2.3.** A square $n \times n$ matrix $A$ is **symmetric** iff $a_{ij} = a_{ji}$ for all $(i, j) \in \{1, \ldots, n\}^2$.

**Example 2.5.** The matrix $A = \begin{bmatrix} 2 & 7 \\ 7 & 2 \end{bmatrix}$ is symmetric, but the matrix $B = \begin{bmatrix} 2 & 1 \\ 7 & 2 \end{bmatrix}$ is not because $b_{21} = 7 \neq 1 = b_{12}$. ■

**Definition 2.4.** For a square matrix $A$, the **trace** $\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$ is the sum of the entries on the main diagonal.

**Example 2.6.** Given $A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$, the main diagonal consists of the elements $a_{11} = 2$ and $a_{22} = 5$, and the trace is $\operatorname{tr}(A) = a_{11} + a_{22} = 2 + 5 = 7$. ■

## 2.3 Matrix operations

This section contains operations involving either one matrix or two matrices.

### 2.3.1 Basics: transpose and arithmetic

**Definition 2.5.** The **transpose** of matrix $A = (a_{ij})_{m\times n}$ switches the rows and columns: $A' = (a_{ji})_{n\times m}$. Alternative notation to $A'$ includes $A^\top$ and $A^T$.

**Example 2.7.** Matrix $A$ being symmetric (Definition 2.3) is equivalent to $A' = A$. ■

**Example 2.8.** The transpose of $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ is $A' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$; the first row of $A$ becomes the first column of $A'$, the second row of $A$ becomes the second column of $A'$, and the dimensions change from $2 \times 3$ (two rows by three columns) to $3 \times 2$ (three rows by two columns). ■

Matrix addition and subtraction operate entry-wise. To add or subtract two matrices requires their dimensions to be the same. If $C = A + B$, then $c_{ij} = a_{ij} + b_{ij}$; similarly, if $D = A - B$, then $d_{ij} = a_{ij} - b_{ij}$. In different notation,

$$A + B = (a_{ij} + b_{ij})_{m\times n}, \quad A - B = (a_{ij} - b_{ij})_{m\times n}. \tag{2.3}$$

Multiplying a matrix by a scalar also operates entry-wise: simply multiply each entry by the scalar. For example, given scalar $c \in \mathbb{R}$ and matrix $A$ with entries $a_{ij}$, then $cA$ has entries $ca_{ij}$.

Matrix multiplication is less simple, but equally important. To compute the product $AB$ requires that the number of columns in $A$ equals the number of rows in $B$. Thus, in the following, consider the product of generic $m \times n$ matrix $A$ and $n \times k$ matrix $B$. Note that unlike scalar multiplication, matrix multiplication is not commutative: if $m \neq k$, then the product $BA$ is not even defined. The dimensions of $AB$ take the number of rows from $A$ and the number of columns from $B$. Here, multiplying $A_{m\times n}B_{n\times k}$ produces an

$m \times k$ matrix. The resulting matrix's row-$i$ column-$j$ entry equals the dot product of row $i$ from $A$ with column $j$ from $B$, denoted $a_{i\cdot} \cdot b_{\cdot j}$ below. This can be written as

$$A_{m\times n} B_{n\times k} = \left( \sum_{h=1}^{n} a_{ih} b_{hj} \right)_{m\times k} = (a_{i\cdot} \cdot b_{\cdot j})_{m\times k}. \tag{2.4}$$

To clarify, if $C = AB$, then $C$ is matrix with $m$ rows and $k$ columns whose row-$i$ column-$j$ entry is

$$c_{ij} = \sum_{h=1}^{n} a_{ih} b_{hj}. \tag{2.5}$$

**Example 2.9.** $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} (1,2)\cdot(1,2) \\ (3,4)\cdot(1,2) \end{bmatrix} = \begin{bmatrix} (1)(1)+(2)(2) \\ (3)(1)+(4)(2) \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$. ■

The name of the identity matrix from (2.2) is due to its multiplicative properties. For any $m \times n$ matrix $A$,

$$AI_n = A, \quad I_m A = A. \tag{2.6}$$

Multiplying by the identity matrix is analogous to multiplying a scalar by 1.

Besides the lack of commutative property, it can also be confusing that $AB$ being a matrix of zeros does not imply that either $A = 0$ or $B = 0$. As a simple example: if $A = \begin{bmatrix} -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, then $AB = (-1)(3) + (1)(3) = 0$.

A useful property of matrix multiplication is

$$(AB)' = B'A'. \tag{2.7}$$

If you ever forget whether it's $B'A'$ or $A'B'$, think about the dimensions: $B'A'$ requires that the number of columns in $B'$ equals the number of rows in $A'$, which is equivalent to the number of columns in $A$ equalling the number of rows in $B$, which must be true if $AB$ exists. In contrast, the existence of $AB$ does not generally imply the existence of $A'B'$. More generally, for longer products, you reverse the order when "distributing" the transpose, like

$$(ABC\cdots YZ)' = Z'Y'\cdots C'B'A'. \tag{2.8}$$

### 2.3.2 Determinant

The **determinant** is a particular scalar summary of a square matrix that relates to other matrix properties.

#### Geometric interpretation

A geometric definition that is not particularly useful for economics is: a matrix's determinant equals the (signed) volume of the hyper-parallelepiped defined by its row or column vectors. Notationally, the following are all commonly used:

$$\det(A) = \det A = |A|. \tag{2.9}$$

**Example 2.10.** Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Section 2.3.2 shows geometrically that the determinant is 1; the "volume" is the area of the square (a special case of parallelogram) that has sides from the origin to $(1, 0)$ and to $(0, 1)$. ■

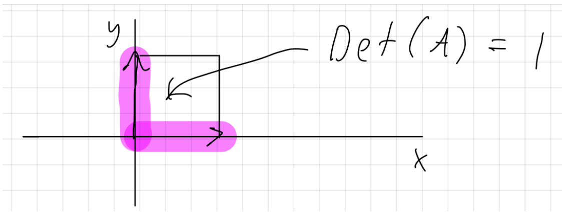



Figure 2.1: Geometric interpretation of determinant in Example 2.10.

### Small matrices

Determinants are not usually calculated by hand, but the formulas for $1 \times 1$ and $2 \times 2$ matrices are simple. For a $1 \times 1$ matrix,

$$\det(A_{1\times 1}) = a_{11}. \tag{2.10}$$

For a $2 \times 2$ matrix,

$$\det(A_{2\times 2}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{2.11}$$

The $2 \times 2$ case is often written with elements $a, b, c, d$ as $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.

### Cofactor expansion

For larger matrices, there is a recursive way to compute the determinant. It uses the following two definitions.

**Definition 2.6.** Given $n \times n$ matrix $A$, its $\boldsymbol{(i, j)}$ **minor** is denoted $A_{ij}$ and is formed by deleting row $i$ and column $j$ of $A$, leaving an $(n-1) \times (n-1)$ matrix.

**Definition 2.7.** Given $n \times n$ matrix $A$, its $\boldsymbol{(i, j)}$ **cofactor** is the scalar value $c_{ij} = (-1)^{i+j} \det(A_{ij})$, where $A_{ij}$ is from Definition 2.6. The $n \times n$ **cofactor matrix** $C$ has $c_{ij}$ as its row-$i$, column-$j$ entries: $C = (c_{ij})_{n\times n}$.

The following cofactor expansion can be done for any row or column in $n \times n$ matrix $A$; they all give the same result. All of the formulas express the determinant of the $n \times n$

matrix in terms of determinants of $(n-1)\times(n-1)$ matrices. Applying it recursively, these can then be expressed in terms of determinants of $(n-2)\times(n-2)$ matrices, etc., until either (2.11) or (2.10) can be used. Recall $a_{ij}$ is the row-$i$, column-$j$ element of $A$, and $c_{ij}$ is the cofactor from Definition 2.7. For any row $i \in \{1,\dots,n\}$,

$$\det(A) = \sum_{j=1}^{n} a_{ij}c_{ij} = a_{i1}c_{i1} + a_{i2}c_{i2} + \cdots + a_{in}c_{in}. \tag{2.12}$$

Similarly, for any column $j \in \{1,\dots,n\}$,

$$\det(A) = \sum_{i=1}^{n} a_{ij}c_{ij} = a_{1j}c_{1j} + a_{2j}c_{2j} + \cdots + a_{nj}c_{nj}. \tag{2.13}$$

**Example 2.11.** Applying (2.13) with $j=1$ (first column),

$$\begin{aligned}
\begin{vmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{vmatrix} &= \overbrace{1}^{a_{11}}\overbrace{(-1)^{1+1}\underbrace{\begin{vmatrix} 5 & 8 \\ 6 & 9 \end{vmatrix}}_{\det(A_{11})}}^{c_{11}} + \overbrace{2}^{a_{21}}\overbrace{(-1)^{2+1}\underbrace{\begin{vmatrix} 4 & 7 \\ 6 & 9 \end{vmatrix}}_{\det(A_{21})}}^{c_{21}} + \overbrace{3}^{a_{31}}\overbrace{(-1)^{3+1}\underbrace{\begin{vmatrix} 4 & 7 \\ 5 & 8 \end{vmatrix}}_{\det(A_{31})}}^{c_{31}} \\
&= \overbrace{\begin{vmatrix} 5 & 8 \\ 6 & 9 \end{vmatrix}}^{\text{use (2.11)}} -2 \overbrace{\begin{vmatrix} 4 & 7 \\ 6 & 9 \end{vmatrix}}^{\text{use (2.11)}} +3 \overbrace{\begin{vmatrix} 4 & 7 \\ 5 & 8 \end{vmatrix}}^{\text{use (2.11)}} \\
&= (45-48) - 2(36-42) + 3(32-35) \\
&= -3 - (-12) + (-9) = 0.
\end{aligned}$$

■

### Some determinant properties

The following properties can be helpful.

**Proposition 2.1.** *Let $A$ be a square matrix.*

1. *Equations (2.12) and (2.13) hold for any row $i$ or column $j$.*
2. *Taking the transpose does not change the determinant: $|A| = |A'|$.*
3. *Interchanging (switching) two rows changes the sign of the determinant, though not its absolute value.*
4. *Multiplying every entry in one row by a real number $c$ multiplies the determinant by $c$.*
5. *Adding $c$ times a row to another row does not change the value of the determinant; nor does adding $c$ times one column to another column.*
6. *The determinant of any identity matrix is one: for any $n$, $\det(I_n) = 1$.*

7. *For $n \times n$ square matrices $A$ and $B$,* $\det(AB) = \det(A)\det(B)$.

**Example 2.12.** To illustrate Proposition 2.1(3),

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = -\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \quad \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = -\begin{vmatrix} a & b & c \\ g & h & i \\ d & e & f \end{vmatrix}. \qquad \blacksquare$$

**Example 2.13.** To illustrate Proposition 2.1(4),

$$\begin{vmatrix} 1 & 2 \\ 3c & 4c \end{vmatrix} = c\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}, \quad \begin{vmatrix} a & b \\ 5c & 5d \end{vmatrix} = 5\begin{vmatrix} a & b \\ c & d \end{vmatrix}. \qquad \blacksquare$$

**Example 2.14.** To illustrate Proposition 2.1(5), adding twice the second row to the first row:

$$1 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = \begin{vmatrix} 1+0 & 0+2 \\ 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} = (1)(1) - (2)(0) = 1. \qquad \blacksquare$$

**Example 2.15.** To illustrate Proposition 2.1(7), if $A = I_n$, then this says $\det(I_nB) = \det(I_n)\det(B) = \det(B)$ because $\det(I_n) = 1$; we can verify this result by $\det(AB) = \det(I_nB) = \det(B)$. ■

**Example 2.16.** Proposition 2.1 can be applied to derive the determinant of a seemingly complicated matrix without calculating the determinant directly:

$$\begin{vmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 7 \\ 2 & 3 & 8 \\ 3 & 3 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 6 \\ 2 & 3 & 6 \\ 3 & 3 & 6 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 0 \\ 2 & 3 & 0 \\ 3 & 3 & 0 \end{vmatrix} = 0,$$

where the first equality is by property Item 5 when adding $-1$ times the first column to the second column, and similarly for the second equality adding $-1$ times the first column to the third column, and similarly for the third equality adding $-2$ times the second column to the third column. The final equality uses the cofactor expansion for column $j = 3$: in (2.13), each $a_{i3} = 0$, so every term in the sum is zero, and thus the sum itself is zero. ■

### 2.3.3 Matrix inverse

Here we define the matrix inverse, describe how to compute it, and give some useful properties.

**Definition 2.8.** Given $n \times n$ square matrix $A$, its **inverse** $A^{-1}$ is the $n \times n$ matrix satisfying $A^{-1}A = AA^{-1} = I_n$, the identity matrix. If such a matrix does not exist, then $A^{-1}$ is undefined. If such a matrix does exist, then $A$ is called **invertible** or **nonsingular**, and $A^{-1}$ is unique.

The steps to calculate the inverse use the adjoint, related to the cofactor matrix.

**Definition 2.9.** Given $n \times n$ matrix $A$, its adjoint is $\text{adj}(A) = C'$, the transpose of $A$'s cofactor matrix (Definition 2.7).

Given Definition 2.9, we can compute

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A). \tag{2.14}$$

The adjoint is always defined, so we can see that $A$ is invertible if and only if $\det(A) \neq 0$.

**Example 2.17.** We apply (2.14) to invert $A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$. First, by (2.11), $|A| = (1)(3) - (2)(0) = 3$. Second, using Definition 2.6, the minors are $A_{11} = [1]$, $A_{12} = [0]$, $A_{21} = [2]$, and $A_{22} = [1]$; these are $1 \times 1$ matrices. Third, using Definition 2.7, the corresponding cofactors are $c_{11} = (-1)^{1+1} \det([3]) = 3$, $c_{12} = (-1)^{1+2} \det([0]) = 0$, $c_{21} = (-1)^{2+1} \det([2]) = -2$, and $c_{22} = (-1)^{2+2} \det([1]) = 1$. Fourth, putting these into a matrix and taking the transpose, Definition 2.9 gives $\text{adj}(A) = \begin{bmatrix} 3 & -2 \\ 0 & 1 \end{bmatrix}$. Fifth, using (2.14),

$$A^{-1} = \frac{1}{3} \begin{bmatrix} 3 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -2/3 \\ 0 & 1/3 \end{bmatrix}. \tag{2.15}$$

To conclude, we can directly verify that this satisfies $AA^{-1} = I_2$:

$$\overbrace{\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} 1 & -2/3 \\ 0 & 1/3 \end{bmatrix}}^{A^{-1}} = \begin{bmatrix} (1)(1) + (2)(0) & (1)(-2/3) + (2)(1/3) \\ (0)(1) + (3)(0) & (0)(-2/3) + (3)(1/3) \end{bmatrix} = \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}^{I_2}. \qquad \blacksquare$$

The following matrix inverse properties can be helpful.

**Proposition 2.2.** *Assume $A$ and $B$ are invertible matrices with the same dimensions.*

1. $(A^{-1})^{-1} = A$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(A')^{-1} = (A^{-1})'$ *(the inverse of the transpose equals the transpose of the inverse)*
4. $\det(A^{-1}) = 1/\det(A)$

## 2.4 Matrices, linear equation systems, and Cramer's rule

In economics, we model relationships between difference variables. Such relationships are complex and interrelated, so often we study a system of multiple equations. Often certain parts are known or observed, and we wish to solve for values of the remaining variables (or "parameters") that satisfy the relationships. Matrices help us do this.

Consider the following **system of equations** and its solution. Let $b = (b_1, \ldots, b_n)'$ be a known $n \times 1$ vector of values. Let $A = (a_{ij})_{n \times n}$ be a known matrix. Let $x = (x_1, \ldots, x_n)'$ be an unknown $n \times 1$ vector whose values satisfy the relationship $Ax = b$, or equivalently

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n. \end{aligned} \tag{2.16}$$

If $A$ is invertible, then $Ax = b$ implies $A^{-1}Ax = A^{-1}b$. Recalling $A^{-1}A = I_n$ and $I_n x = x$, this gives the solution $x = A^{-1}b$.

Further, **Cramer's rule** provides a formula for the individual $x_j$. Let $A_j$ denote matrix $A$ with column $j$ replaced by vector $b$. Then, for each $j \in \{1, \ldots, n\}$,

$$x_j = \frac{\det(A_j)}{\det(A)}. \tag{2.17}$$

**Example 2.18.** We apply (2.17) with $j = 1$ to the following system of linear equations:

$$\left.\begin{aligned} x_1 + 2x_2 &= 1 \\ 2x_1 + \phantom{2}x_2 &= 3 \end{aligned}\right\} \implies A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},\ b = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Solving for $x_1$ with (2.17),

$$|A| = (1)(1) - (2)(2) = -3, \tag{2.18}$$

$$A_1 = \begin{bmatrix} \overbrace{\begin{matrix} 1 \\ 3 \end{matrix}}^{b} & \begin{matrix} 2 \\ 1 \end{matrix} \end{bmatrix} \implies |A_1| = (1)(1) - (3)(2) = -5, \tag{2.19}$$

so $x_1 = (-5)/(-3) = 5/3$. ■

## 2.5 Linear dependence and rank

The rank of a matrix is an important property, defined in terms of linear independence.

**Definition 2.10.** Consider the set of $n$ vectors, $a_i$ for $i = 1, \ldots, n$, each $a_i$ having dimension $k \times 1$. Let $0_k$ be the $k \times 1$ vector of zeros. For $i = 1, \ldots, n$, let $c_i \in \mathbb{R}$ be scalar values. The set of vectors $\{a_i\}$ is **linearly independent** iff $\sum_{i=1}^n c_i a_i = 0_k$ implies all the $c_i = 0$. Conversely, the set of vectors is **linearly dependent** iff there is a solution with at least one $c_i \neq 0$, which implies that one of the vectors can be expressed as a linear combination of the others (or that one of the $a_i$ contains all zeros).

**Definition 2.11.** The **rank** of a matrix is the maximum number of its columns that are linearly independent. Equivalently, the rank is the dimension of the vector space spanned by the columns of the matrix, also called the dimension of the column space of the matrix.

The rank can also be defined in terms of minors, which must be defined first.

**Definition 2.12.** Given $m \times n$ matrix $A$, a **minor of order $k$** is a $k \times k$ matrix formed by deleting $m-k$ rows and deleting $n-k$ columns of $A$. "Minor of order $k$" or "minor determinant of order $k$" may also refer to the determinant of such a matrix.

**Proposition 2.3.** *Let $A$ be an $m \times n$ matrix.*

1. $\text{rank}(A) = \text{rank}(A')$
2. $\text{rank}(A) \leq \min\{m, n\}$
3. $\text{rank}(A)$ *is the largest $k$ such that $A$ has a non-zero minor determinant of order $k$*
4. *Adding a multiple of one column of $A$ to another of its columns does not change the rank, nor does adding a multiple of one row to another row*

Although it can be helpful to think specifically about "column rank" (number of linearly independent columns) or "row rank" (number of linearly independent rows), $\text{rank}(A) = \text{rank}(A')$ means they are equal.

**Example 2.19.** Consider the rank of the matrix from Example 2.16,

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}.$$

From Example 2.16, we know $|A| = 0$. Given this and Proposition 2.3(3), the rank cannot be 3 because the (only) minor determinant of order 3 is zero. However, there is a non-zero $2 \times 2$ minor determinant:

$$\underbrace{\begin{vmatrix} 1 & 4 \\ 2 & 5 \end{vmatrix}}_{\text{2-minor}} = (1)(5) - (4)(2) = -3 \neq 0, \tag{2.20}$$

so $\text{rank}(A) = 2$. Alternatively, we can use Proposition 2.3(4), applying the same column operations from Example 2.16:

$$\text{rank}\left(\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}\right) = \text{rank}\left(\begin{bmatrix} 1 & 3 & 6 \\ 2 & 3 & 6 \\ 3 & 3 & 6 \end{bmatrix}\right) = \text{rank}\left(\begin{bmatrix} 1 & 3 & 0 \\ 2 & 3 & 0 \\ 3 & 3 & 0 \end{bmatrix}\right) = 2. \tag{2.21}$$

By Definition 2.11, the rank of the final matrix is two because the two non-zero columns are linearly independent (because there is no $c \in \mathbb{R}$ such that $(c, 2c, 3c) = (3, 3, 3)$). ■

## 2.6 Eigenvalues and eigenvetors

In first-year economics, eigenvalues and eigenvectors are primarily useful in macroeconomics, for rewriting a matrix such that it can be easily multiplied by itself repeatedly. They also appear in more advanced econometrics, in methods like principal components analysis.

### 2.6.1 An example

The general idea is that multiplying a matrix by a certain type of vector yields a scalar multiple of that vector, as seen in the following example. Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix}, \quad x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; x_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \; x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \; z = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}. \tag{2.22}$$

For each $x_j$, $Ax_j$ is proportional to $x_j$ (can be written as a scalar multiple of $x_j$), but this is not true for $Az$:

$$Ax_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = x_1, \quad Ax_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = 2x_2, \quad Ax_3 = \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix} = -3x_3, \quad Az = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \tag{2.23}$$

the last of which cannot be written as $cz$ for any $c \in \mathbb{R}$.

The above also holds for any vector proportional to $x_1$, $x_2$, or $x_3$. That is, for vector $v_1 = \rho x_1$, $Av_1 = A\rho x_1 = \rho(Ax_1) = \rho x_1 = v_1$, and similarly $Av_2 = 2v_2$ for any $v_2 = \rho x_2$ and $Av_3 = -3v_3$ for any $v_3 = \rho x_3$. Although the vectors are not unique, they are unique up to scale, and the multipliers are unique.

It can be useful to think of $A$ as generating a basis of a space generated by vectors $x_1$, $x_2$, and $x_3$. More complicated matrices can be understood to generate a basis for:

1. A proper subspace of $\mathbb{R}^n$ (above, the space generated by $A$ is the whole $\mathbb{R}^3$)
2. The basis does not have to be orthogonal (above, $x_1'x_2 = 0$, $x_2'x_3 = 0$, and $x_1'x_3 = 0$)
3. Unfortunately, even starting with real-valued matrices,
   (a) We might require a complex-valued basis
   (b) For some pathological matrices, this intrepretation is not valid

### 2.6.2 Finding eigenvalues and eigenvectors

Here, we consider the goal of finding eigenvalues and the corresponding eigenectors. The set of eigenvalues is also called the **spectrum**. Relatedly, analysis of eigenvalues and eigenvectors is also called **spectral analysis**, and the corresponding matrix decomposition can be called either eigendecomposition or spectral decomposition (see Section 2.7).

Given $n \times n$ matrix $A$, its eigenvalues and eigenvectors are defined as follows. Let $x$ be an $n \times 1$ vector, and let $\lambda \in \mathbb{R}$. If

$$Ax = \lambda x, \tag{2.24}$$

then $\lambda$ is an **eigenvalue** of $A$, and $x$ is a corresponding **eigenvector**. Because there are an infinite number of eigenvectors corresponding to a single $\lambda$, often they are normalized to have unit length, $x'x = 1$. Generally, there are $n$ pairs of eigenvalue and (normalized) eigenvector, but some eigenvalues may be repeated. For example, if $A = I_2$, then $x_1 = (0, 1)'$ and $x_2 = (1, 0)'$ are the normalized eigenvectors, but both eigenvalues are one: $Ax_1 = x_1$ and $Ax_2 = x_2$

These eigenvalue and eigenvector pairs can be characterized as follows. We want a solution to $Ax = \lambda x$, which is equivalent to $Ax = \lambda I_n x$, which can be rewritten as $Ax - \lambda I_n x = 0$ or $(A - \lambda I_n)x = 0$. This equation has a non-trivial solution ($x \neq 0$) only if the matrix $A - \lambda I_n$ is not full rank, which is equivalent to having a determinant of zero. Thus, a value $\lambda$ is an eigenvalue of $A$ if and only if it is a solution to the **characteristic equation**,

$$\det(A - \lambda I) = 0, \tag{2.25}$$

where the left-hand side is the **characteristic polynomial**, which is an $n$th-degree polynomial in $\lambda$. Given an eigenvalue, we can then solve for the corresponding eigenvector.

**Example 2.20.** Consider $A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$. The characteristic equation

$$\begin{vmatrix} 1-\lambda & 2 \\ 0 & 2-\lambda \end{vmatrix} = 0 \implies (1-\lambda)(2-\lambda) = 0 \tag{2.26}$$

has two solutions: the eigenvalues $\lambda = 1$ and $\lambda = 2$. The eigenvector corresponding to $\lambda = 1$ must satisfy $Ax = x$:

$$\overbrace{\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}^{x} = \overbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}^{x} \implies \begin{aligned} x_1 + 2x_2 &= x_1 \\ 2x_2 &= x_2 \end{aligned} \implies x_2 = 0 \text{ and any } x_1 \in \mathbb{R}. \tag{2.27}$$

To normalize the eigenvector to have unit length, $x_1 = 1$ (or $x_1 = -1$), so the first eigenvector is $x = (1, 0)$. For the second eigenvalue $\lambda = 2$,

$$\begin{aligned} x_1 + 2x_2 &= 2x_1 \\ 2x_2 &= 2x_2 \end{aligned} \implies x_1 = 2x_2$$

from the first equation, while the second equation does not add any information (it is a tautology). Again normalizing to unit length, the eigenvector is $x = (2/\sqrt{5}, 1/\sqrt{5})'$. ■

If a particular eigenvalue appears twice, then there are more degrees of freedom in solving for the eigenvectors, and there are multiple (linearly independent, unit-length) eigenvectors corresponding to that eigenvalue.

Interestingly, the matrix determinant (Section 2.3.2) and trace (Definition 2.4) can be expressed in terms of eigenvalues.

**Proposition 2.4.** *For any $n \times n$ matrix $A$ with eigenvalues $\lambda_i$ (for $i = 1, \ldots, n$; not necessarily unique), the determinant equals the product of the $n$ eigenvalues, and the trace equals the sum of the $n$ eigenvalues:*

$$\det(A) = \prod_{i=1}^{n} \lambda_i = \lambda_1 \lambda_2 \cdots \lambda_n,$$
$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

## 2.7 Diagonalization and decomposition

Eigenvalues and eigenvectors are especially useful for rewriting a matrix in a way that makes it easier to manipulate. Specifically, a matrix is "decomposed" into a product of multiple matrices (with certain properties), defined below. This decomposition is closely related to the concept of diagonalization.

**Definition 2.13.** A square matrix $A$ is **diagonalizable** iff there exists invertible matrix $P$ such that $P^{-1}AP = D$ is a diagonal matrix (Definition 2.2). The process of finding such $P$ and $D$ is called **diagonalization**.

**Definition 2.14.** When it exists, the **eigendecomposition** or **spectral decomposition** of square matrix $A$ is $A = PDP^{-1}$, where each column of $P$ is an eigenvector of $A$ (conventionally normalized to have unit length), and $D$ is a diagonal matrix whose entries are the corresponding eigenvalues; that is, $d_{jj}$ is the eigenvalue corresponding to the eigenvector in column $j$ of $P$.

In Definition 2.13, the entries of diagonal matrix $D$ being eigenvalues of $A$ follows from two facts. First, a diagonal matrix's eigenvalues are simply its main diagonal entries. Second, for any invertible matrix $P$, the eigenvalues of $P^{-1}AP$ equal those of $A$:

$$\begin{aligned} P^{-1}APx = \lambda x \iff P^{-1}APx = \lambda I_n x \iff (P^{-1}AP - \lambda I_n)x = 0 \\ \iff (P^{-1}AP - \lambda P^{-1}P)x = 0 \iff P^{-1}(A - \lambda I_n)Px = 0. \end{aligned}$$

Given that $P$ and $P^{-1}$ are invertible (full rank), this can only have a solution with $x \neq 0$ if $(A - \lambda I_n)$ is not full rank, which is equivalent to (2.25), the characteristic equation for $A$ itself.

**Proposition 2.5.** *If matrix $A$ is diagonalizable as $P^{-1}AP = D$, then it has the eigendecomposition $A = PDP^{-1}$. If matrix $A$ has eigendecomposition $A = PDP^{-1}$, then it is diagonalizable as $P^{-1}AP = D$.*

*Proof.* Starting from $P^{-1}AP = D$: left-multiply each side by $P$, and right-multiply by $P^{-1}$: $PP^{-1}APP^{-1} = PDP^{-1}$, and the $PP^{-1} = I_n$ disappear. Starting from $A = PDP^{-1}$: left-multiply each side by $P^{-1}$ and right-multiply by $P$, yielding $P^{-1}AP = P^{-1}PDP^{-1}P$. □

For $n \times n$ matrix $A$, the eigendecomposition and diagonalization exist if and only if $A$ has $n$ linearly independent eigenvectors. Although this is not a full proof, consider the matrix $P$ whose columns are the eigenvectors of $A$. If these are all linearly independent, then $P$ is full rank (Definition 2.11), so it has non-zero determinant (by Proposition 2.3(3) with $k = n$), implying it is invertible by (2.14). Thus, $P^{-1}$ exists, and we can write the decomposition $A = PDP^{-1}$.

**Example 2.21.** Not all matrices are diagonalizable. For example, the matrix $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ is not diagonalizable. ■

The formula can look different if $A$ is symmetric (Definition 2.3). In that case, $P^{-1} = P'$, so the decomposition is $A = PDP'$.

**Example 2.22.** To compute the $k$th power of decomposed matrix $A = PDP^{-1}$,

$$\begin{aligned} A^k &= \underbrace{A \times \cdots \times A}_{k \text{ times}} = (PD\overbrace{P^{-1}) \times (P}^{=I_n}DP^{-1}) \times \cdots \times (PDP^{-1}) \\ &= P \times \underbrace{D \times \cdots \times D}_{k \text{ times}} \times P^{-1} = PD^kP^{-1} = P\,\mathrm{diag}(\lambda_1^k, \ldots, \lambda_n^k)P^{-1}, \end{aligned}$$

where the $\lambda_j$ are eigenvalues of $A$ and diag$(\cdot)$ creates a diagonal matrix with those entries $(\lambda_j^k)$ on the main diagonal. ■

**Example 2.23.** Similar to Example 2.22, the decompositions helps us find the $k$th "root" of matrix $A$: the matrix $B$ such that $B^k = A$. Write $A = PDP^{-1}$, assume $k$ is odd and everything in $\mathbb{R}$. Then, $B = PD^{1/k}P^{-1}$, where $D^{1/k} = \mathrm{diag}(\lambda_1^{1/k}, \ldots, \lambda_n^{1/k}$; by Example 2.22, $B^k = A$. ■

## 2.8 Quadratic forms and definiteness

A **quadratic form** of symmetric matrix $A$ is

$$Q = x'Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} a_{ij}x_ix_j. \tag{2.28}$$

Given $A$, this is a scalar-valued function of vector $x \in \mathbb{R}^n$. That is, we can think of it as a function $q_A\colon \mathbb{R}^n \to \mathbb{R}$.

### 2.8.1 Definitions

**Definition 2.15.** The following terms apply to matrix $A$ or the corresponding quadratic form $Q$ in (2.28). As usual, $x \neq 0$ means $x \neq (0, 0, \ldots, 0)$, and it is assumed $x \in \mathbb{R}^n$.

1. **Positive definite** (sometimes written $A > 0$): for all $x \neq 0$, $x'Ax > 0$.
2. **Positive semidefinite** (sometimes written $A \geq 0$): for all $x$, $x'Ax \geq 0$.
3. **Negative definite** (sometimes written $A < 0$): for all $x \neq 0$, $x'Ax < 0$.
4. **Negative semidefinite** (sometimes written $A \leq 0$): for all $x$, $x'Ax \leq 0$.
5. If none of the above apply, then it is called **indefinite**.

From the definition, positive definite implies positive semidefinite, like how for scalars $z > 0$ implies $z \geq 0$. Similarly, negative definite implies negative semidefinite, like how scalar $z < 0$ implies $z \leq 0$.

For a given matrix, Definition 2.15 can be checked as in the following two subsections.

### 2.8.2 Minor determinant test of definiteness

First, we need to define certain types of minors, extending Definition 2.6.

**Definition 2.16.** A **principal minor** of $n \times n$ matrix $A$ is a submatrix derived by removing both row $i$ and column $i$ from $A$ for $i \in S \subseteq \{1, \ldots, n\}$. Alternatively, the principal minor can refer to the determinant of this matrix. The **order** of the principal minor is how many columns and rows are left: if $d$ are removed, then the order $k = n - d$. The **leading principal minor** of order $k$ refers to the $k \times k$ submatrix (or its determinant) in the top-left of $A$ (after removing rows $k+1, k+2, \ldots, n$ and removing columns $k+1, \ldots, n$).

**Proposition 2.6.** *Let $D_k$ be the leading principal minor determinant of order $k$. Let $\Delta_k$ be a principal minor determinant of order $k$. Consider $n \times n$ matrix $A$, and let $k \in \{1, \ldots, n\}$.*

1. *Matrix $A$ is positive definite if and only if $D_k > 0$ for all $k$.*
2. *Matrix $A$ is positive semidefinite if and only if $\Delta_k \geq 0$ for all $k$ and all $\Delta_k$.*
3. *Matrix $A$ is negative definite if and only if $(-1)^k D_k > 0$ for all $k$.*
4. *Matrix $A$ is negative semidefinite if and only if $(-1)^k \Delta_k \geq 0$ for all $k$ and all $\Delta_k$.*

**Example 2.24.** Applying Proposition 2.6 shows that $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is positive semidefinite. For $k = 1$, the principal minors are all the determinant of $1 \times 1$ matrix $[1]$, which is $\det([1]) \geq 0$. For $k = 2$, the (only) principal minor is the determinant of $A$ itself, which by (2.11) is $(1)(1) - (1)(1) = 0 \geq 0$. Although $D_1 = 1 > 0$, $D_2 = 0$, so the matrix is not positive definite. ■

**Example 2.25.** Applying Proposition 2.6 shows that $B = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{bmatrix}$ is negative definite. For $k = 1$, $D_k = \det([-1]) = -1$, so $(-1)^k D_k = (-1)^1(-1) > 0$. For $k = 2$,

$$D_k = \det\left(\begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}\right) = (-1)(-2) - (0)(0) = 2, \tag{2.29}$$

so $(-1)^k D_k = (-1)^2(2) = 2 > 0$. For $k = 3$, $D_k = |B| = (-1)(-2)(-3) = -6$, so $(-1)^k D_k = (-1)^3(-6) = 6 > 0$. Equivalently, we can notice the alternating sign pattern of the $D_k$ over $k = 1, 2, 3$: $-1 < 0$, $2 > 0$, $-6 < 0$. ■

### 2.8.3 Eigenvalue characterization of definiteness

**Proposition 2.7.** *Let $A$ be a symmetric $n \times n$ matrix with eigenvalues $\lambda_i$ for $i = 1, \ldots, n$.*

1. *Matrix $A$ is positive definite if $\lambda_i > 0$ for all $i = 1, \ldots, n$.*
2. *Matrix $A$ is positive semidefinite if $\lambda_i \geq 0$ for all $i = 1, \ldots, n$.*
3. *Matrix $A$ is negative definite if $\lambda_i < 0$ for all $i = 1, \ldots, n$.*
4. *Matrix $A$ is negative semidefinite if $\lambda_i \leq 0$ for all $i = 1, \ldots, n$.*

*Proof.* Given $A$ is symmetric, it is diagonalizable with $PAP' = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = D$, where $PP' = I_n$. Consider general $x \in \mathbb{R}^n$, and let $y \equiv P'x$, implying $x = Py$. Then,

$$x'Ax = y'P'APy = y'P'\overbrace{PDP'}^{=A}Py = y'Dy = \sum_{i=1}^{n} \lambda_i y_i^2. \tag{2.30}$$

Because $y_i^2 \geq 0$ for any $y_i \in \mathbb{R}$, the signs of the $\lambda_i$ determine the signs of $x'Ax$ related to definiteness in Definition 2.15 as stated in Proposition 2.7. □

## Optional resources

Optional resources for this chapter

- Very (very!) nice online "Interactive Linear Algebra" textbook: `https://web.archive.org/web/20240516182445/https://textbooks.math.gatech.edu/ila/overview.html`
- Wikipedia `https://en.wikipedia.org/wiki/Matrix_(mathematics)` has matrix notation, properties, operations, etc.; there are also Wikipedia pages for related topics like matrix rank, Cramer's rule, etc.
- `https://web.archive.org/web/20240713110834/https://hummedia.manchester.ac.uk/schools/soss/economics/pg/psmaths/pre-sessiomathsbook.pdf` pages 46–56

## Exercises

**Exercise E2.1.** Prove: if $A$ is a diagonal matrix (Definition 2.2), then it is symmetric (Definition 2.3).

**Exercise E2.2.** Use (2.4) to prove (2.6).

**Exercise E2.3.** Use (2.7) to prove that $(ABC)' = C'B'A'$. (Hint: it may help to define $D = BC$.)

**Exercise E2.4.** Let $\boldsymbol{x}$ be a column vector with $k$ entries. Let $\boldsymbol{C}$ be an $n \times k$ matrix. For each of the following: give the dimensions of the resulting matrix or else say that it is not well defined.

a. $\boldsymbol{x}\boldsymbol{x}'$

b. $\boldsymbol{x}'\boldsymbol{x}$

c. $\boldsymbol{C}\boldsymbol{C}'$

d. $\boldsymbol{C}'\boldsymbol{C}$

e. $\boldsymbol{C}\boldsymbol{x}$

f. $\boldsymbol{x}'\boldsymbol{C}$

g. $\boldsymbol{w}\boldsymbol{C}'$ with $\boldsymbol{w} = [\boldsymbol{x}\ \boldsymbol{x}]'$

**Exercise E2.5.** Let $\boldsymbol{x} = (x_1, \ldots, x_n)'$, $\boldsymbol{w} = (w_1, \ldots, w_n)'$, and $\boldsymbol{e} = (1, 1, \ldots, 1)'$ an $n \times 1$ column vector with every entry equal to one. Given scalar $c \in \mathbb{R}$, consider the matrix $\boldsymbol{A} = [\boldsymbol{e}, \boldsymbol{x}, c\boldsymbol{x}, \boldsymbol{w}]$. Using Proposition 2.3, explain why $\text{rank}(\boldsymbol{A}) < 4$.

## Appendix 2.A Quadratics with linear constraints

Warning: the following has not been revised or edited.

**Quadratics with Linear Constraints:**

Consider a quadratic form

$\underbrace{\boldsymbol{x}^T}_{1\times n}\underbrace{A}_{n\times n}\underbrace{\boldsymbol{x}}_{n\times 1}$ subject to $\underbrace{B}_{m\times n}\underbrace{\boldsymbol{x}}_{n\times 1} = \underbrace{\boldsymbol{0}}_{m\times 1}$

Define a "Bordered" Matrix:

$$\tilde{B} = \begin{bmatrix} 0_{m\times m} & B_{m\times n} \\ B^T_{n\times m} & A_{n\times n} \end{bmatrix}$$

**Minor determinant tests:**

$$\text{Define as } \tilde{B}_r = \begin{vmatrix} 0 & \ldots & 0 & b_{11} & \ldots & b_{1r} \\ \vdots & & & \vdots & & \\ 0 & \cdots & 0 & b_{m_1} & \cdots & b_{mr} \\ b_{11} & \cdots & b_{m_1} & a_{11} & \ldots & a_{1r} \\ \vdots & & & \vdots & & \\ b_{1r} & \ldots & b_{m_r} & a_{r1} & \cdots & a_{rr} \end{vmatrix}$$

Minor determinant with $r$ first rows of constraints & rows of $A$ (also colomns)
Positive Definite $\Leftrightarrow (-1)^m B_r > 0$ for $r = m+1, \ldots, n$
Negative Definite $\Leftrightarrow (-1)^r B_r > 0$ for $r = m+1, \ldots, n$

**Example 2.26.** $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{x}^T A \boldsymbol{x} = x_1^2 + x_2^2$
subject to $x_1 = x_2 \implies x_1 - x_2 = 0$
$\Leftrightarrow \underbrace{\begin{bmatrix} 1 & -1 \end{bmatrix}}_{B} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Leftrightarrow B\boldsymbol{x} = 0$

$$\tilde{B} = \left[\begin{array}{c|cc} 0 & 1 & -1 \\ \hline 1 & 1 & 0 \\ -1 & 0 & 1 \end{array}\right]$$

$$|\tilde{B}| = -1 \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & -1 \\ 1 & 0 \end{vmatrix} = -1 - 1 = -2 < 0$$

Because $\underbrace{(-1)^1}_{-} \underbrace{B_2}_{-} > 0 \Leftrightarrow A$ is positive definite with respect to the constraint

$n =$ number of variables
$m =$ number of constraints
$r =$ running counter ■

# Chapter 3

# Calculus

Calculus is used heavily in microeconomics, macroeconomics, and econometrics. This chapter covers topics that will be useful for first-year PhD economics courses.

*Unit learning objectives for this chapter*

3.1. Define vocabulary words (in **bold**) related to multivariate calculus, both mathematically and intuitively [TLO 1]

3.2. Derive a local approximation of a function [TLO 2]

3.3. Apply the Implicit Function Theorem to compute derivatives of one variable with respect to another, when the former is not written as an explicit function of the latter [TLO 2]

3.4. Assess whether or not a function possesses certain properties [TLO 3]

## 3.1 Derivatives

To start, here are some of the most commonly used derivative rules in economics. Here $a, b \in \mathbb{R}$ are constants, $f, g, h$ are functions (from $\mathbb{R} \to \mathbb{R}$), and notationally $f'(x) = \frac{\mathrm{d}}{\mathrm{d}x} f(x)$ is the first derivative (evaluated at $x$).

- Linearity: $\frac{\mathrm{d}}{\mathrm{d}x}[af(x) + bg(x)] = af'(x) + bg'(x)$.
- Power rule: $\frac{\mathrm{d}}{\mathrm{d}x} ax^b = bax^{b-1}$, for any $a, b \in \mathbb{R}$ (and recall $1/x = x^{-1}$, etc.).
- Product rule: $\frac{\mathrm{d}}{\mathrm{d}x}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$.
- Chain rule: $\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$.
- Quotient rule: $\frac{\mathrm{d}}{\mathrm{d}x}[f(x)/g(x)] = [f'(x)g(x) - f(x)g'(x)]/[g(x)]^2$, or just apply the product rule to $f(x)h(x)$ with $h(x) = [g(x)]^{-1}$, using the chain rule to get $h'(x) = -[g(x)]^{-2}g'(x)$.
- Log: $\frac{\mathrm{d}}{\mathrm{d}x} \log(x) = 1/x$. (Recall log always means natural log.)
- Exponential: $\frac{\mathrm{d}}{\mathrm{d}x} e^x = e^x$.

- Second derivative: just the derivative of the derivative, $\frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x) = \frac{\mathrm{d}}{\mathrm{d}x} f'(x)$.

**Example 3.1.** Here are some examples of the above rules:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}x}[3x^2 + 2x - 1] = 6x + 2, && \frac{\mathrm{d}}{\mathrm{d}x}[x \log(x)] = \log(x) + (x/x) = \log(x) + 1, \\ &\frac{\mathrm{d}}{\mathrm{d}x} e^{-2x} = -2e^{-2x}, && \frac{\mathrm{d}}{\mathrm{d}x} \log(1 + x^2) = \frac{1}{1 + x^2}(2x). \end{aligned}$$ ■

The following definitions are for a function $f\colon \mathbb{R}^n \to \mathbb{R}$. For simplicity, here we assume all relevant derivatives exist. To put this more formally, we first define the corresponding sets of functions.

**Definition 3.1.** The set $C^1$ contains all functions that have a continuous first derivative. More generally, the set $C^n$ contains all functions that have a continuous derivative of order $n$ (which implies existence and continuity of derivatives of order $n - 1$ and lower, too). Such functions are also called $n$ times continuously differentiable.

**Example 3.2.** Consider $f(x) = x^2$. The first derivative is $f'(x) = 2x$, which is a continuous function, so $f \in C^1$. The second derivative is $f''(x) = 2$, also continuous, so $f \in C^2$, too. Any higher derivative is just zero, so $f \in C^n$ for any $n > 0$. ■

**Example 3.3.** Consider $g(x) = \mathrm{sgn}(x)x^2$, where the sign function is defined as $\mathrm{sgn}(x) = -1$ if $x < 0$, $\mathrm{sgn}(x) = 1$ if $x > 0$, and $\mathrm{sgn}(x) = 0$ if $x = 0$. The first derivative is $g'(x) = |2x|$, which is a continuous function, so $g \in C^1$. For the second derivative,

$$g''(x) = \begin{cases} -2 & \text{if } x < 0 \\ \ \ 2 & \text{if } x > 0 \\ \text{undefined} & \text{if } x = 0, \end{cases}$$

which is not a continuous function (nor even defined at zero), so $g$ is not in $C^2$ (it is not twice continuously differentiable). However, if we restrict the domain to only strictly positive (or only strictly negative) values of $x$, then $g$ would be twice (and more) continuously differentiable. ■

**Definition 3.2.** Given $f\colon \mathbb{R}^n \to \mathbb{R}$, its **gradient** (or **Jacobian**) $\nabla f\colon \mathbb{R}^n \to \mathbb{R}^n$ is the $1 \times n$ vector of first-order partial derivative functions, assuming they exist. Define $f_1$ be the partial derivative of $f$ with respect to its first argument, and generally $f_j$ is the partial derivative of $f$ with respect to its $j$th argument. That is, evaluated at point $x \in \mathbb{R}^n$,

$$f_j(x) \equiv \lim_{h \to 0} \frac{f(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f(x)}{h}.$$

Notationally, sometimes instead of (or in addition to) $x$ being an evaluation point, $x$ represents the vector of "dummy" arguments as in $f(x) = f(x_1, \ldots, x_n)$, and the partial derivative functions are written as $\frac{\partial f}{\partial x_j}(\cdot)$, but this can be confusing when $x$ is also the point of evaluation. Writing the $(\cdot)$ to make clear these are functions, the gradient of $f(\cdot)$ is

$$\nabla f(\cdot) \equiv (f_1(\cdot), \dots, f_n(\cdot)),$$

and the gradient evaluated at point $p \in \mathbb{R}^n$ is $\nabla f(p) = (f_1(p), \dots, f_n(p))$.

**Example 3.4.** Let $f(x) = 5x_1 + x_2 x_3^2$; note $f \colon \mathbb{R}^3 \to \mathbb{R}$. The partial derivatives are $\frac{\partial f}{\partial x_1}(x) = 5$, $\frac{\partial f}{\partial x_2}(x) = x_3^2$, and $\frac{\partial f}{\partial x_3}(x) = 2x_2 x_3$. Thus, $\nabla f(x) = (5, x_3^2, 2x_2 x_3)$. ■

**Definition 3.3.** Given $f \colon \mathbb{R}^n \to \mathbb{R}$, its **Hessian** $H \colon \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ is the $n \times n$ matrix of second-order partial derivative functions, assuming they exist. Let $f_{ij}$ be the second-order partial derivative with respect to the $i$th and $j$th arguments; this is often called the "cross partial" when $i \neq j$, and if $i = j$, then $f_{jj}$ is the second partial derivative with respect to the $j$th argument. Notationally, this is often written in terms of dummy argument vector $x$ as $f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ and $f_{ii} = \frac{\partial^2 f}{\partial x_i^2}$. Again writing $(\cdot)$ to make explicit that these are functions,

$$H(\cdot) \equiv \begin{bmatrix} f_{11}(\cdot) & f_{12}(\cdot) & \cdots & f_{1n}(\cdot) \\ f_{21}(\cdot) & f_{22}(\cdot) & \cdots & f_{2n}(\cdot) \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1}(\cdot) & f_{n2}(\cdot) & \cdots & f_{nn}(\cdot) \end{bmatrix}.$$

If $f \in C^2$, then $f_{ij} = f_{ji}$ and the Hessian is symmetric.

**Example 3.5.** Continuing Example 3.4 with $f(x) = 5x_1 + x_2 x_3^2$, we can compute the Hessian by taking the partial derivatives of the gradient $\nabla f(x) = (5, x_3^2, 2x_2 x_3)$:

$$H(x) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 2x_3 \\ 0 & 2x_3 & 2x_2 \end{bmatrix}.$$ ■

## 3.2 Local approximation: Taylor expansion

Often the truth is too complex for us to analyze, so we use an approximation. Here, we consider approximations of a function near a particular point. Such a point might represent a market equilibrium (macro) or a parameter's true value (econometrics).

**Theorem 3.1** (mean value theorem). *Given open set $G \subseteq \mathbb{R}^n$, let $f \colon G \to \mathbb{R}$ with $f \in C^1$. Consider points $x, y \in \mathbb{R}^n$ such that the (closed) line segment from $x$ to $y$ is contained in $G$. Then, there exists point $\omega \in \mathbb{R}^n$ on that line segment (excluding the endpoints) such that $f(y) - f(x) = \nabla f(\omega) \cdot (y - x)$, where $\cdot$ is the dot product. (See Figure 3.1.)*

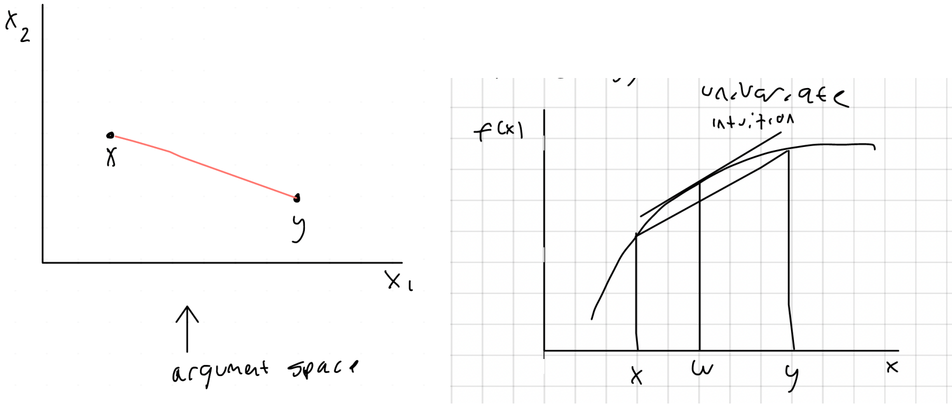



Figure 3.1: Illustrations for mean value theorem, Theorem 3.1.

**Example 3.6.** In the univariate case, given continuously differentiable $f$ on an open interval containing $[x, y]$, there exists $\omega \in (x, y)$ such that $f(y) - f(x) = f'(\omega)(y - x)$. Let $f(t) = t^2$, so $f : \mathbb{R} \to \mathbb{R}$ and $f \in C^1$ with $f'(t) = 2t$. Let $x = 1$ and $y = 3$. Then, $f(y) = 9$, $f(x) = 1$, $f(y) - f(x) = 8$, and $y - x = 2$. Thus, with $\omega = 2$ (so $x < \omega < y$), $f'(\omega) = 2\omega = 4$, and $f(y) - f(x) = f'(\omega)(y - x)$. ■

Taylor's theorem implicitly uses the mean value theorem to quantify the error of a polynomial approximation of a function around a point. This type of approximation is also called a **Taylor expansion**. There are various ways to write the theorem; the following version is often sufficient for economics.

**Theorem 3.2** (Taylor's theorem, univariate)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ with $f \in C^n$. Then,*

$$f(x) = f(a) + \sum_{i=1}^{k} \frac{f^{(i)}(a)(x-a)^i}{i!} + \frac{f^{(k+1)}(\varepsilon)(x-a)^n}{n!}$$

*for some $\varepsilon$ between $a$ and $x$, where $f^{(i)}(\cdot)$ denotes the $i$th derivative of $f(\cdot)$. This is called a $k$th-order Taylor expansion.*

**Theorem 3.3** (Taylor's theorem, multivariate, first-order)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ with $f \in C^2$. Then, for column vectors $a, x \in \mathbb{R}^n$,*

$$f(x) = f(a) + \nabla f(a) \cdot (x - a) + \frac{1}{2}(x - a)' H(\varepsilon)(x - a)$$

*for some $\varepsilon$ between $a$ and $x$, where $\nabla f(\cdot)$ is the gradient of $f(\cdot)$, $\nabla f(a) \cdot (x - a)$ is the dot product of $x - a$ and the gradient evaluated at $a$, and $H(\cdot)$ is the Hessian of $f(\cdot)$. (Dimensions note: $x - a$ is an $n \times 1$ column vector, so $(x - a)'$ is $1 \times n$, so $(x - a)' H(\varepsilon)(x - a)$ is $(1 \times n)(n \times n)(n \times 1) = 1 \times 1$.)*

The $k$th-order multivariate result is analogous to Theorems 3.2 and 3.3 but includes complicated notation, so is omitted. Hopefully you won't need it this year!

## 3.3 Convexity and concavity

This section concerns convexity (and concavity) of both sets and functions.

### 3.3.1 Convex sets

Besides being useful itself, the idea of a convex set is also used to define quasiconcavity and quasiconvexity of a function in Section 3.4.

**Definition 3.4.** Set $C \subseteq \mathbb{R}^n$ is **convex** iff for any two points $x, y \in C$, the point $\lambda x + (1-\lambda)y$ is also in $C$ for all $\lambda \in [0,1]$. That is, a set is convex if and only if the line segment joining any two points in the set is fully contained in the set.

**Example 3.7.** In Figure 3.2, sets $A$, $B$, and $C$ are all convex, but $D$ is not. ■

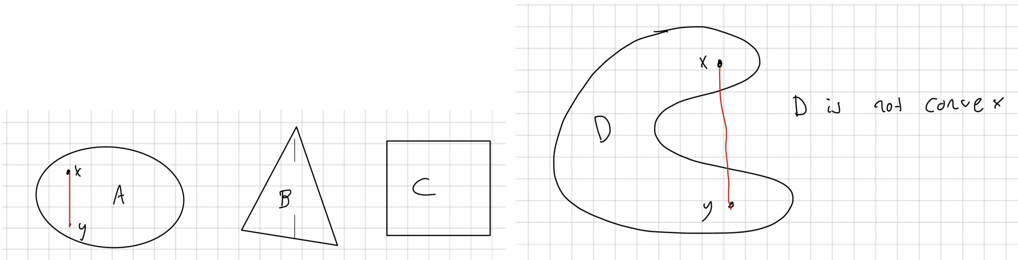



Figure 3.2: Sets for Example 3.7.

**Example 3.8.** Hyperplanes are convex. Given fixed vector $p \in \mathbb{R}^n$ and fixed $m \in \mathbb{R}$, a hyperplane can be written as $H = \{x \in \mathbb{R}^n : p \cdot x = m\}$. Given $\lambda \in [0,1]$ and any $x, y \in H$,

$$p \cdot [\lambda x + (1-\lambda)y] = \lambda \overbrace{p \cdot x}^{=m} + (1-\lambda) \overbrace{p \cdot y}^{=m} = \lambda m + (1-\lambda)m = m,$$

so the point $\lambda x + (1-\lambda)y$ is in the hyperplane $H$, too. ■

**Example 3.9.** Half-spaces are convex. Again given vector $p \in \mathbb{R}^n$ and scalar $m \in \mathbb{R}$, an open half-space can be written as $\tilde{H} = \{x \in \mathbb{R}^n : p \cdot x < m\}$. Similar to Example 3.8, given $\lambda \in [0,1]$ and any $x, y \in \tilde{H}$,

$$p \cdot [\lambda x + (1-\lambda)y] = \lambda \overbrace{p \cdot x}^{<m} + (1-\lambda) \overbrace{p \cdot y}^{<m} < \lambda m + (1-\lambda)m = m,$$

so the point $\lambda x + (1-\lambda)y$ is in the half-space $\tilde{H}$, too. The same results hold for a closed half-space, replacing all $<$ with $\leq$. ■

**Proposition 3.4.** *The intersection of convex sets is convex. That is, if sets $A$ and $B$ are convex, then $A \cap B$ is convex.*

*Proof.* Let $A$ and $B$ be convex. Consider points $x, y \in A \cap B$. Let $\tilde{x} = \lambda x + (1 - \lambda)y$ for $\lambda \in [0, 1]$. Because $x, y \in A$ and $A$ is convex, then $\tilde{x} \in A$. Because $x, y \in B$ and $B$ is convex, then $\tilde{x} \in B$. Thus, $\tilde{x} \in (A \cap B)$. □

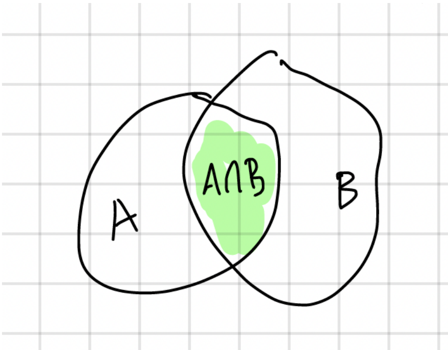



Figure 3.3: Illustration of Proposition 3.4.

### 3.3.2 Convex and concave functions

Recall from the univariate case that a $C^2$ function is concave if $f''(x) \leq 0$ for all $x$, or convex if $f''(x) \geq 0$ for all $x$. The following definitions are more general because they do not require $f \in C^2$.

**Definition 3.5.** Let function $f: S \to \mathbb{R}$ for convex set $S$. Then, $f$ is **concave** iff $f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$ for all $x, y \in S$ and all $\lambda \in (0, 1)$, and **strictly concave** if $\geq$ is replaced by $>$. Similarly, $f$ is **convex** iff $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ for all $x, y \in S$ and all $\lambda \in (0, 1)$, and **strictly convex** if $\leq$ is replaced by $<$.

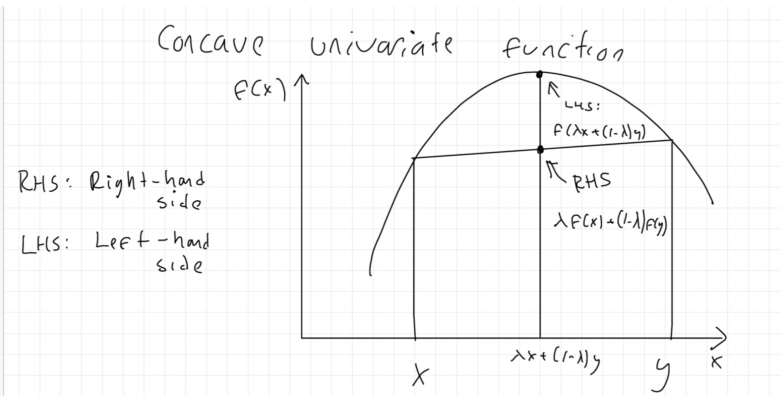



Figure 3.4: Illustration of Definition 3.5 for a concave univariate function.

Figure 3.4 illustrates Definition 3.5 for a concave univariate function. We can see that $f(\cdot)$ evaluated at a convex combination of $x$ and $y$ is always greater than the corresponding convex combination of $f(x)$ and $f(y)$, i.e., $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$.

**Example 3.10.** Given $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, let $f\colon \mathbb{R}^n \to \mathbb{R}$ with $f(x) = \max_i x_i$ (the single greatest component of the vector). Given any $x, y \in \mathbb{R}^n$ and any $\lambda \in (0,1)$,

$$\begin{aligned} f\big(\lambda x + (1-\lambda)y\big) &= \max_i \{\lambda x_i + (1-\lambda)y_i\} \\ &\leq \max_i \{\lambda x_i\} + \max_i \{(1-\lambda)y_i\} \\ &= \lambda \max_i \{x_i\} + (1-\lambda)\max_i \{y_i\} \\ &= \lambda f(x) + (1-\lambda)f(y), \end{aligned}$$

satisfying the definition of convexity in Definition 3.5. ■

The following properties help assess the convexity or concavity of more complex functions.

1. If $f_1, \ldots, f_n$ are all concave functions, then $a_1 f_1 + \cdots + a_n f_n$ is concave given $a_i \geq 0$.
2. If $f$ is concave and $F\colon \mathbb{R} \to \mathbb{R}$ is concave and increasing, then $U(\cdot) = F(f(\cdot))$ is a concave function. (Equivalently: $U = F \circ f$.)
3. If $f$ is concave and $F\colon \mathbb{R} \to \mathbb{R}$ is convex and decreasing, then $U(\cdot) = F(f(\cdot))$ is a convex function.
4. If $f$ is concave, then $-f(\cdot)$ is convex.

### 3.3.3 Derivative-based definitions

Previous sections defined "global" convexity and concavity; here, we focus on a function's shape at a particular point. The global versions simply require the local definition to hold at every point in the domain (or in some subset of the domain).

The following are *sufficient* conditions for strict convexity and concavity at a point, but they are not *necessary*. This can be seen even in the familiar univariate case, where $f''(x) > 0$ implies strict convexity at point $x$, but strict convexity does not require $f''(x) > 0$. Although this may seem surprising, consider $f(x) = x^4$, so $f'(x) = 4x^3$ and $f''(x) = 12x^2$. At $x = 0$, $f''(0) = 0$, which fails the condition $f''(x) > 0$, even though the function is strictly convex (as you can see from a graph).

Consider $f \in C^2$ and $H(x)$, the Hessian matrix evaluated at $x$, and its leading principal minor determinants $D_r(x)$. Then, the following hold.

1. If $D_r(x) > 0$ for $r = 1, \ldots, n$, then $f$ is strictly convex at $x$. (By Proposition 2.6, this condition is equivalent to $H(x)$ being positive definite.)

2. If $(-1)^r D_r(x) > 0$ for $r = 1, \ldots, n$, then $f$ is strictly concave at $x$. (By Proposition 2.6, this condition is equivalent to $H(x)$ being negative definite.)

**Example 3.11.** Consider $f(x, y) = \log x + \log y$ for $x > 0$ and $y > 0$. The gradient, Hessian, and principal minor deterimants are

$$\begin{aligned} \nabla f(x, y) &= (1/x, 1/y), \\ H &= \begin{bmatrix} -1/x^2 & 0 \\ 0 & -1/y^2 \end{bmatrix}, \\ (-1)^1 D_1(x, y) &= 1/x^2 > 0 \text{ (given } x > 0), \\ (-1)^2 D_2(x, y) &= \frac{1}{x^2 y^2} > 0 \text{ (given } x, y > 0). \end{aligned}$$

Thus, for any $x, y > 0$, $f$ is strictly concave. ■

For (non-strict) convexity and concavity, the following Hessian-based conditions are both sufficient and necessary. Let $f \colon S \to \mathbb{R}$ for convex $S \subseteq \mathbb{R}^n$ and $f \in C^2$. Let $H(x)$ be its Hessian evaluated at point $x$. Let $\Delta_r(x)$ be a principal minor determinant of $H(x)$ of order $r$.

1. Function $f$ is convex at point $x$ if and only if $\Delta_r(x) \geq 0$ for all $r = 1, \ldots, n$ and all $\Delta_r(x)$ of order $r$. (By Proposition 2.6, this condition is equivalent to $H(x)$ being positive semidefinite.)

2. Function $f$ is concave at point $x$ if and only if $(-1)^r \Delta_r(x) \geq 0$ for all $r = 1, \ldots, n$ and all $\Delta_r(x)$ of order $r$. (By Proposition 2.6, this condition is equivalent to $H(x)$ being negative semidefinite.)

## 3.4 Quasiconcavity and quasiconvexity

Like concavity and convexity, quasiconcavity and quasiconvexity describe the shape of a function.

**Definition 3.6.** Let $f \colon S \to \mathbb{R}$ for convex set $S \subseteq \mathbb{R}^n$. Function $f$ is **quasiconcave** iff the upper level sets $P_a^{\geq} = \{x \in S : f(x) \geq a\}$ are convex for all $a \in \mathbb{R}$. Function $f$ is is **quasiconvex** iff the lower level sets $P_a^{\leq} = \{x \in S : f(x) \leq a\}$ are convex for all $a \in \mathbb{R}$.

**Example 3.12.** Figure 3.5 shows a picture of a quasiconcave function. ■

**Example 3.13.** The classic "bell curve" function $f(x) = e^{-x^2}$ is neither concave nor convex, but it is quasiconcave. This is apparent visually from Figure 3.6 but can also be shown analytically. Consider the upper level sets of $f$. For $a \leq 0$, $P_a^{\geq} = \mathbb{R}$ (which is convex), and for $a > 1$, $P_a^{\geq}$ is the empty set (which is convex). For $0 < a < 1$ (so $-\infty < \log(a) < 0$), the condition $f(x) \geq a$ has the following equivalent representations:

$$e^{-x^2} \geq a \iff -x^2 \geq \log(a) \iff x^2 \leq -\log(a) \iff -\sqrt{-\log(a)} \leq x \leq \sqrt{-\log(a)},$$

which is a closed interval and thus convex. ■

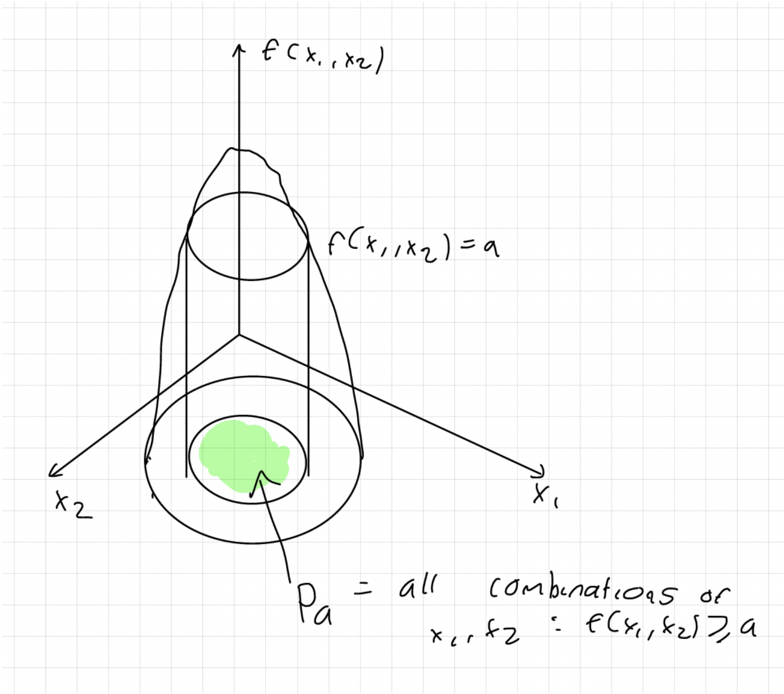



Figure 3.5: Illustration of a quasiconcave function for Example 3.12.

**Example 3.14.** Any monotonic univariate function is both quasiconcave and quasiconvex. ■

**Example 3.15.** Any linear multivariate function is both quasiconcave and quasiconvex. ■

**Example 3.16.** The sum of two quasiconcave functions is not necessarily quasiconcave. For example, both $f(x) = e^{-x^2}$ and $g(x) = e^{-(x-9)^2}$ are quasiconcave, but $h(x) = f(x) + g(x)$ is not quasiconcave, as can be seen clearly from a graph. For example, $h(\cdot)$ has values near zero for $2 \leq x \leq 7$, so the upper level set where $h(x) \geq 0.5$ includes values around $x = 0$ and around $x = 9$ but not all the values in between (so is not convex). ■

Note increasing transformations preserve quasiconcavity and quasiconvexity, whereas decreasing transformations reverse the property. For example, the exponential function $\exp(\cdot)$ is an increasing transformation; $f(x) = -x^2$ is quasiconcave, and $\exp(f(x)) = \exp(-x^2)$ is also quasiconcave, as in Example 3.13. (In contrast: $f(x) = -x^2$ is concave, but $\exp(f(x))$ is not concave.) As a simple example of a decreasing transformation, if $f(x)$ is quasiconcave, then $g(x) = -f(x)$ is quasiconvex.

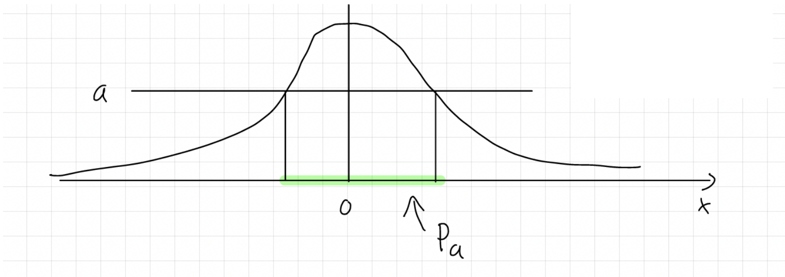



Figure 3.6: Illustration of the quasiconvex bell curve $f(x) = e^{-x^2}$ for Example 3.13.

## 3.5 Implicit function theorem

Our goal is to derive the relationship between variables $y$ and $x$ given an equation that only "implicitly" states their relationship. That is, we do not already know $y = f(x)$, but rather something like $f(x, y) = 0$. Assume $x, y \in \mathbb{R}$. Intuitively, it seems we might be able to differentiate to get $f_x \mathrm{d}x + f_y \mathrm{d}y = 0$, where $f_x = \frac{\partial f}{\partial x}$ is the partial derivative of $f$ with respect to its first argument, and $f_y = \frac{\partial f}{\partial y}$ is the partial derivative of $f$ with respect to its second argument. This would imply

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{f_x}{f_y}. \tag{3.1}$$

The implicit function theorem makes this precise.

### 3.5.1 Two-variable case

The following are the formal conditions and results like (3.1). Given open set $A \subseteq \mathbb{R}^2$, let $f\colon A \to \mathbb{R}$ with $f \in C^1$. Additionally, $(x_0, y_0) \in A$ with $f(x_0, y_0) = 0$ and $f_y(x_0, y_0) \neq 0$, i.e., the partial derivative with respect to the second argument evaluated at $(x_0, y_0)$ is non-zero. Then, given $y$ and $x$ satisfy $f(x, y) = 0$, there exists a neighborhood of $(x_0, y_0)$ in which we can write $y$ as an explicit function of $x$. Specifically, the "neighborhood" is the rectangle $I_0 \times I_1$ for intervals $I_0 = (x_0 - \delta, x_0 + \delta)$ and $I_1 = (y_0 - \epsilon, y_0 + \epsilon)$ and small enough $\delta, \epsilon > 0$, with $I_0 \times I_1 \subseteq A$. Further,

1. for every $x \in I_0$, the equation $f(x, y) = 0$ defines a unique solution in $I_1$, $y = \phi(x)$; and

2. $\phi \in C^1$ with

$$\phi'(x) = -\frac{f_x(x, \phi(x))}{f_y(x, \phi(x))}. \tag{3.2}$$

**Example 3.17.** Let $f(x,y) = y\log(2-xy) = 0$ and $(x_0, y_0) = (1,1)$. Note $f \in C^1$ in a neighborhood of $(1,1)$, with partial derivatives

$$
\begin{aligned}
f_x(x,y) &= y\frac{1}{2-xy}(-y) = -\frac{y^2}{2-xy}, \\
f_y(x,y) &= \log(2-xy) + y\frac{1}{2-xy}(-x) = \log(2-xy) - \frac{xy}{2-xy}, \\
f_y(1,1) &= \log(2-(1)(1)) - (1)(1)/(2-(1)(1)) = 0 - (1/1) = -1 \neq 0.
\end{aligned}
$$

The assumptions are satisfied, so $\phi(\cdot)$ exists. Further, its derivative at $x = 1$ is

$$\phi'(1) = -\frac{f_x(1,1)}{f_y(1,1)} = -\frac{-1^1/(2-1)}{-1} = -\frac{-1}{-1} = -1, \tag{3.3}$$

using $f_y(1,1) = -1$ from above and the fact that $1 = y_0 = \phi(x_0) = \phi(1)$. ■

With further assumptions about differentiability of $f$, we can similarly find higher-order derivatives.

### 3.5.2 General case

Now consider vector-valued $y \in \mathbb{R}^n$ with $n$ equations: $f\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$. Previously we considered the special case $n = 1$ (scalar $y$, scalar-valued $f$). Assume each component of $f$ is $C^1$; that is, $f(x,y) = (f_1(x,y), \ldots, f_n(x,y))$ and each $f_j \in C^1$ for $j = 1, \ldots, n$.

Intuitively, things are similar to before, but now we have vectors and matrices. Previously, the intuition was that we could differentiate to get $f_x \mathrm{d}x + f_y \mathrm{d}y = 0$, then rearrange to $f_y \frac{\mathrm{d}y}{\mathrm{d}x} = -f_x$ and finally $\frac{\mathrm{d}y}{\mathrm{d}x} = -f_x/f_y$, our object of interest. Now, in the intermediate form $f_y \frac{\mathrm{d}y}{\mathrm{d}x} = -f_x$, we have a matrix times a vector equal to a vector, so instead of "dividing" we pre-multiply by the matrix inverse to get $\frac{\mathrm{d}y}{\mathrm{d}x} = -[f_y]^{-1} f_x$.

We use the following vectors and matrix. The object of interest is the $n \times 1$ column vector

$$\frac{\mathrm{d}y}{\mathrm{d}x} \equiv \left(\frac{\partial y_1}{\partial x}, \ldots, \frac{\partial y_n}{\partial x}\right)'. \tag{3.4}$$

The $f_x$ is also an $n \times 1$ column vector, whose $i$th element is the partial derivative of the $i$th element of $f$ with respect to $x$,

$$f_x \equiv \left(\frac{\partial f_1}{\partial x}, \ldots, \frac{\partial f_n}{\partial x}\right)'. \tag{3.5}$$

Because $y$ also has $n$ elements, the object $f_y$ is an $n \times n$ matrix, whose row $i$, column $j$ entry is the partial derivative of the $i$th element of $f$ with respect to the $j$th element of $y$:

$$f_y \equiv \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_1} & \cdots & \frac{\partial f_n}{\partial y_n} \end{bmatrix}. \tag{3.6}$$

The implicit function theorem tells us about the relationship between $x$ and $y$ that satisfy $f(x,y) = 0$ in a neighborhood of point $(x_0, y_0)$ that also satisfies $f(x_0, y_0) = 0$. Besides assuming differentiability of $f$, the matrix $f_y$ evaluated at $(x_0, y_0)$ must be invertible, which is usually written in terms of a non-zero determinant, $\det(f_y) \neq 0$. Then,

$$\frac{\mathrm{d}y}{\mathrm{d}x}(x_0, y_0) = -[f_y(x_0, y_0)]^{-1} f_x(x_0, y_0). \tag{3.7}$$

**Example 3.18.** Let $n = 2$ with $f(x,y) = (y_1 - xy_2, y_1^2 + y_2^2 - 2)$ and $(x_0, y_0) = (1,1,1)$. Then,

$$f_x = (-y_2, 0)', \quad f_x(1,1,1) = (-1,0)', \tag{3.8}$$

$$f_y = \begin{bmatrix} 1 & -x \\ 2y_1 & 2y_2 \end{bmatrix}, \quad f_y(1,1,1) = \begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}, \tag{3.9}$$

$$\det(f_y(1,1,1)) = \begin{vmatrix} 1 & -1 \\ 2 & 2 \end{vmatrix} = (2)(1) - (-1)(2) = 4. \tag{3.10}$$

Let the implicitly defined function be $y = \phi(x) = (\phi_1(x), \ldots, \phi_n(x))$. We can use Cramer's rule as in (2.17) to compute

$$\phi_1'(1) = \frac{\begin{vmatrix} 1 & -1 \\ 0 & 2 \end{vmatrix}}{4} = \frac{(2)(1) - (-1)(0)}{4} = \frac{1}{2}, \tag{3.11}$$

where the numerator has replaced column $i = 1$ of matrix $f_y$ with the vector $-f_x(1,1,1) = (1,0)'$. ■

## Optional resources

Optional resources for this chapter

- Khan Academy calculus: https://www.khanacademy.org/math/old-ap-calculus-ab
- Dartmouth "Open Calculus": https://web.archive.org/web/20220702043656/https://math.dartmouth.edu/opencalc2/
- Dartmouth sequences notes and problems: https://web.archive.org/web/20240308013726/https://math.dartmouth.edu/opencalc2/dcsbook/c1pdf/sec12.pdf
- https://web.archive.org/web/20240713110834/https://hummedia.manchester.ac.uk/schools/soss/economics/pg/psmaths/pre-sessiomathsbook.pdf pages 24–32 and 57–62 and 77–85
- Wikipedia https://en.wikipedia.org/wiki/Taylor%27s_theorem Taylor's theorem

## Exercises

**Exercise E3.1.** Let $f(x_1, x_2) = (x_1 + 2x_2)^2$.

a. Compute the gradient $\nabla f(x_1, x_1)$.

b. Compute the Hessian $H(x_1, x_2)$.

c. Is the function strictly convex, strictly concave, convex, concave, or none of these?

**Exercise E3.2.** Let $f(x) = x^2$ and $x_0 = 1$.

a. Write the first-order Taylor approximation of $f(x)$ around $x_0$, including the remainder.

b. Write the second-order Taylor approximation of $f(x)$ around $x_0$, including the remainder.

**Exercise E3.3.** Let $X = [0, 1] \subset \mathbb{R}$, the closed unit interval.

a. Prove that $X$ is a convex set.

b. Prove that $X \times X = [0, 1]^2$ is a convex set.

c. Prove that $X \times X \times X = [0, 1]^3$ is a convex set.

**Exercise E3.4.** Let $g(x) = 4 - 4x$, $h(x) = 2 - x$, and $f\colon [0, 2] \to [0, 4]$ with $f(x) = \max\{g(x), h(x)\}$.

a. Use Definition 3.5 to prove that $f$ is convex.

b. Use Definition 3.6 (not Example 3.14) to prove that $f$ is quasiconcave.

**Exercise E3.5.** Let $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with $f(x, y) = y - x^2 = 0$ and $(x_0, y_0) = (2, 4)$. We can directly see $y = x^2$ and $\frac{\mathrm{d}y}{\mathrm{d}x} = 2x$, so $\frac{\mathrm{d}y}{\mathrm{d}x}(x_0, y_0) = 4$. Instead, use the implicit function theorem, specifically (3.2), to show that $\frac{\mathrm{d}y}{\mathrm{d}x}(x_0, y_0) = 4$.

**Exercise E3.6.** This is a higher-dimensional version of E3.5. Let $x \in \mathbb{R}$ and $y = (y_1, y_2) \in \mathbb{R}^2$, with $f(x, y) = (y_1 - x^2, y_2 - x^2)$, and $x_0 = 2$ and $y_0 = (4, 4)'$. Use the implicit function theorem, specifically (3.7), to compute the $2\times 1$ column vector $\frac{\mathrm{d}y}{\mathrm{d}x}(x_0, y_0)$.

## Appendix 3.A A second-derivative test for strict quasiconcavity and convexity

Warning: the following has not been revised or edited.

### 3.A.1 Definition

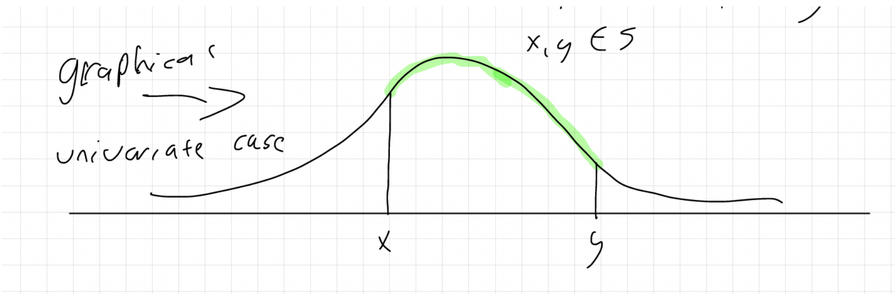
Strict version $\iff$ level sets strictly convex

Strictly convex set: $x, y \in S \implies \lambda x + (1-\lambda)y \in \text{interior}(S)$ for all $\lambda \in (0,1)$.

Strict quasiconcavity: $f\colon s \to \mathbb{R}$ is **strictly quasiconcave** if

$$f(\lambda x + (1-\lambda)y) > \min\{f(x), f(y))$$

for $\lambda \in (0,1), x \neq y, x, y \in S$




### 3.A.2 Test

Let's form a Bordered Hessian $\underbrace{B}_{(n+1)\times(n+1)}$.

$$B = \begin{bmatrix} 0 & \nabla f(x) \\ (\nabla f(x))^T & H(x) \end{bmatrix}$$

$$B_r = \begin{vmatrix} 0 & f_1 & \dots & f_r \\ f_1 & f_{11} & \dots & f_{1r} \\ \vdots & \vdots & \dots & \vdots \\ f_r & f_{r1} & \dots & f_{rr} \end{vmatrix}$$

gradient intuitively comes from linearizing a constraint

$$f(x) = a$$

1. Function $F \neq$ is strictly quasiconcave if $(-1)^r B_r > 0 \quad \forall x \in S, x = 2, \dots, n$
2. Function $F$ is strectly quasiconcave if $B_r < 0 \quad \forall x \in S, r = 2 \dots, n$

**Example 3.19.** Consider $f(x,y) = -x^2 - y^2$ for $x > 0, y > 0$, $B = \begin{bmatrix} 0 & -2x & -2y \\ -2x & -2 & 0 \\ -2y & 0 & -2 \end{bmatrix}$,

$n = 2 \to$ only one determinant:

$$B_2 = \begin{vmatrix} 0 & -2x & -2y \\ -2x & -2 & 0 \\ -2y & 0 & -2 \end{vmatrix} = 2x\begin{vmatrix} -2x & -2y \\ 0 & -2 \end{vmatrix} - 2y\begin{vmatrix} -2x & -2y \\ -2 & 0 \end{vmatrix} = 8x^2 + 8y^2 > 0$$

for all $x, y > 0$. This is what is needed for quasiconcavity because $(-1)^2 B_2 > 0$. ■

# Chapter 4

# Static Optimization

Optimization appears in microeconomics, macroeconomic, and econometrics. In this chapter, we consider the simplest setting, without constraints or anything changing over time. Chapter 5 adds constraints.

*Unit learning objectives for this chapter*

4.1. Define vocabulary words (in **bold**) related to static optimization, both mathematically and intuitively [TLO 1]

4.2. Solve for maximum or minimum points of a function [TLO 2]

4.3. Apply the envelope thoerem [TLO 2]

4.4. Characterize when a function has an interior maximum or minimum [TLO 3]

4.5. Characterize points that are maxima or minima [TLO 3]

## 4.1 Conditions for interior maximum or minimum

We first consider conditions on partial derivatives related to the point where a function takes its maximum or minimum value. We use the setting of A4.1.

**Assumption A4.1.** Let $f\colon S \to \mathbb{R}$ for set $S \subseteq \mathbb{R}^n$, and assume $f \in C^1$. Let $f_i$ be the partial derivative of $f$ with respect to the $i$th component, for $i \in \{1, \dots, n\}$.

The following definitions are used throughout the chapter.

**Definition 4.1.** Point $x_0$ is a **global maximum point** of $f$ iff $f(x_0) \geq f(x)$ for all $x \in S$, and $f(x_0)$ is then the **maximum value** of $f$.

**Definition 4.2.** Point $x_0$ is a **global minimum point** of $f$ iff $f(x_0) \leq f(x)$ for all $x \in S$, and $f(x_0)$ is then the **minimum value** of $f$.

**Definition 4.3.** Point $x_0$ is a **local maximum point** of $f$ iff $f(x_0) \geq f(x)$ for all $x$ in a neighborhood of $x_0$; that is, there exists $\epsilon > 0$ such that $f(x_0) \geq f(x)$ for all $\|x - x_0\| < \epsilon$.

**Definition 4.4.** Point $x_0$ is a **local minimum point** of $f$ iff $f(x_0) \leq f(x)$ for all $x$ in a neighborhood of $x_0$; that is, there exists $\epsilon > 0$ such that $f(x_0) \leq f(x)$ for all $\|x - x_0\| < \epsilon$.

**Definition 4.5.** Point $x_0$ is a (global or local) **extreme point** iff it is either a (global or local) maximum point or minimum point.

### 4.1.1 Necessary condition

The condition that all first derivatives equal zero at an extreme point is often called the first-order condition (Definition 4.6). However, note that this assumes that such derivatives exist, and Proposition 4.1 also assumes the point is in the interior of the domain $S$.

**Definition 4.6.** Given A4.1, the **first-order condition** (**FOC**) at a point $x_0$ is that all first-order partial derivatives equal zero: $f_i(x_0) = 0$ for all $i \in \{1, \ldots, n\}$.

**Proposition 4.1.** *Given A4.1, the first-order condition (Definition 4.6) is a necessary but not sufficient condition for an interior global extreme point at $x_0$.*

*Proof.* Saying the zero derivative is a "necessary" condition is equivalent to saying, "If $x_0$ is an interior extreme point, then the partial derivatives are all zero." Using proof by contradiction, assume there is some non-zero derivative at $x_0$; we will show this implies $x_0$ is not an extreme point. By the mean value theorem (Theorem 3.1),

$$f(x) = f(x_0) + \nabla f(\tilde{x})(x - x_0) \tag{4.1}$$

where $\tilde{x}$ is a point between $x$ and $x_0$. The gradient function $\nabla f(\cdot)$ is continuous because $f \in C^1$. Thus, if $\nabla f(x_0) \neq 0$, there exists $x$ sufficiently close to $x_0$ such that $\nabla f(x) \neq 0$ and for any $\tilde{x}$ between $x_0$ and $x$, $\nabla f(\tilde{x}) \neq 0$. Without loss of generality, assume the first element of the partial derivative vector is strictly positive, $f_1(x_0) > 0$. (You can always re-order the elements of $x$ to make the first partial non-zero; the extension to strictly negative is straightforward.) Consider $x = x_0 + (\delta, 0, \ldots, 0)$, adding $\delta > 0$ to only the first component. Let $\delta$ be small enough that $f_1(\tilde{x}) > 0$ for all $x_0 \leq \tilde{x} \leq x$. Then,

$$f(x) = f(x_0) + \nabla f(\tilde{x})(x - x_0) = f(x_0) + \overbrace{f_1(\tilde{x})}^{>0}\overbrace{\delta}^{>0} > f(x_0). \tag{4.2}$$

With $\delta < 0$, the analogous argument yields $f(x) < f(x_0)$. Thus, $x_0$ is neither a maximum point nor minimum point. This proves the contrapositive of Proposition 4.1, which (as always) is equivalent to Proposition 4.1 itself. $\square$

A point satisfying the first-order condition is called a **stationary point** (or sometimes **critical point**). Proposition 4.1 says that all (interior) extreme points are stationary points, but not all stationary points are extreme points.

### 4.1.2 Sufficient conditions

Practically, in general, we solve the first-order condition to get the set of stationary points, and then we check the sufficient conditions for each stationary point. The sufficient conditions tell us if we know the point is a maximum, we know it is a minimum, or we do not know. Logically, the sufficient condition for a maximum point implies the point is a maximum, but the logical inverse is not true (the "fallacy of the inverse" or "denying the antecedent"). That is, if the sufficient condition does not hold, then we simply remain uncertain.

Further assuming $S$ is a convex set,

- if $f$ is concave on $S$, then $x_0$ is an interior global maximum point if and only if $f_i(x_0) = 0$ for all $i \in \{1, \ldots, n\}$; and
- if $f$ is convex on $S$, then $x_0$ is an interior global minimum point if and only if $f_i(x_0) = 0$ for all $i \in \{1, \ldots, n\}$.

## 4.2 General conditions for existence

Perhaps surprisingly, not all functions $f : S \to \mathbb{R}$ have global maximum and minimum points. Theorem 4.2 provides conditions for the existence of such points. Formal definitions of "bounded" and "closed" are given later; for now, just know that economics often involves closed, bounded sets.

**Theorem 4.2.** *Let $f : S \to \mathbb{R}$ be a continuous function ($f \in C^0$) on a closed and bounded set $S$. Then, $f$ attains a maximum and minimum value in the set $S$. That is, there exist points $\bar{x}_0 \in S$ and $\underline{x}_0 \in S$ such that $\bar{x}_0$ is a global maximum point (Definition 4.1) and $\underline{x}_0$ is a global minimum point (Definition 4.2).*

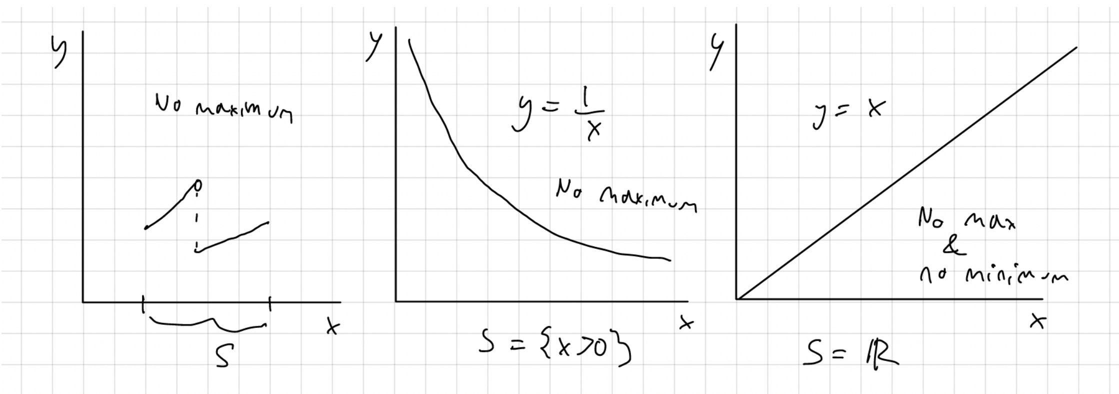



Figure 4.1: Illustration of the need for the conditions in Theorem 4.2.

Figure 4.1 illustrates why all three assumptions are needed for Theorem 4.2 (continuity of $f$, closed $S$, bounded $S$).

## 4.3 Envelope theorem

The envelope theorem provides a shortcut when considering the value of a function at its maximum, when varying other parameters. For example, this is useful for analyzing the profit of a firm when the firm makes profit-maximizing decisions.

Consider the following maximization problem. Let $x$ be the choice vector, and let $r$ be a vector of parameters. The maximizing choice of $x$ is a function of $r$,

$$x^*(r) \equiv \arg\max_{x \in S} f(x, r), \tag{4.3}$$

assumed to be in the interior of $S$. We are interested in the relationship between $r$ and the corresponding maximum value

$$f^*(r) \equiv \max_{r \in S} f(x, r) = f(x^*(r), r). \tag{4.4}$$

This is an example of a **value function**.

Generally, the parameter vector $r$ can influence $f^*(r) = f(x^*(r), r)$ both directly and indirectly. The direct effect is through the second $(r)$ argument to $f(x^*(r), r)$. The indirect effect is through the first argument of $f(x^*(r), r)$, where the value of $r$ can affect the maximizing point $x^*(r)$. However, under some regularity conditions (like for every $r$, $f(\cdot, r)$ is strictly concave and twice differentiable), the indirect effects are all zero due to certain partial derivatives being zero at the maximizing $x^*(r)$, as in Proposition 4.1. Specifically,

$$\frac{\partial f^*(r)}{\partial r_j} = \left.\frac{\partial f(x, r)}{\partial r_j}\right|_{x=x^*(r)}. \tag{4.5}$$

That is, we can simply plug $x = x^*(r)$ into $f(x, r)$ when taking the partial derivative with respect to $r_j$, without worrying about the derivative of $x^*(r)$ with respect to $r$.

The result of (4.5) can be proved with an argument like the following. First, by the multivariate chain rule,

$$\frac{\partial f^*(r)}{\partial r_j} = \frac{\partial}{\partial r_j} f(x^*(r), r) = \left.\frac{\partial f}{\partial r_j}\right|_{x=x^*(r)} + \sum_{i=1}^{n} \overbrace{\left.\frac{\partial f}{\partial x_i}\right|_{x=x^*(r)}}^{=0 \text{ by Proposition 4.1}} \frac{\partial x_i^*(r)}{\partial r_j} = \left.\frac{\partial f}{\partial r_j}\right|_{x=x^*(r)}. \tag{4.6}$$

That is, the key is that the first-order conditions (Proposition 4.1) set all the "indirect effect" terms to zero.

**Example 4.1.** Consider a profit-maximizing price-taking firm who chooses output $x$ given price $p$ and cost function $c(x)$. We are interested in how the maximized profit changes with the market price $p$. (Here $c(\cdot)$ is fixed and cannot change; you can imagine just replacing $c(x) = x^2$ or something, so there is no more $c(\cdot)$.) The profit function is $\pi(x, p) = px - c(x)$. The firm chooses $x^*(p) = \arg\max_x \pi(x, p)$ and earns profit $\pi^*(p) = \max_x \pi(x, p) = \pi(x^*(p), p)$. The envelope theorem (4.5) says

$$\frac{\mathrm{d}\pi^*(p)}{\mathrm{d}p} = \left.\frac{\partial \pi(x, p)}{\partial p}\right|_{x=x^*(p)} = x^*(p) \tag{4.7}$$

because $\frac{\partial \pi(x,p)}{\partial p} = \frac{\partial}{\partial p}[px - c(x)] = x$. ■

## 4.4 Local and global second-order conditions

Recall from Proposition 4.1 that the first-order condition (all first-order partial derivatives equal to zero) is only a necessary condition for a global maximum or minimum, meaning that a variety of other types of points also satisfy that condition. Specifically, for an interior point with partial derivatives $f_i(x_0) = 0$ for all $i$, we can have a local minimum (Definition 4.4), local maximum (Definition 4.3), or saddle point. Figure 4.2 illustrates all three types. In the first plot, the "global max" is also a local maximum, but the "local max" is below the "global max" and hence not a global maximum. The second plot shows a univariate saddle point for a function like $y = 6 + (x - x_0)^3$: the first derivative is $\frac{\mathrm{d}y}{\mathrm{d}x} = 3(x - x_0)^2$, which equals zero at $x_0$, but $x_0$ is not even a local minimum or maximum because $y$ is strictly increasing in $x$. The third plot shows a higher-dimensional saddle point (which actually looks a little like a saddle for a horse): deviating from $x_0$ along the $x_1$ direction increases $y$, making it seem like a local minimum, but deviating from $x_0$ along the $x_2$ direction decreases $y$, seeming like a local maximum.

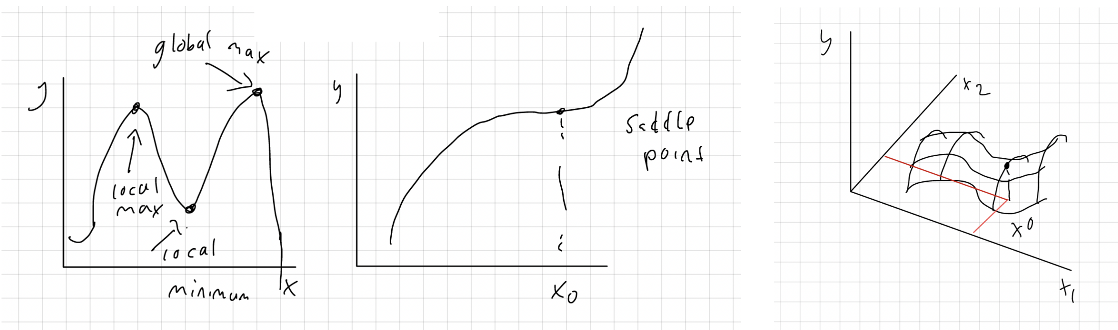



Figure 4.2: Illustration of different types of points satisfying the first-order condition.

The rest of this section characterizes additional conditions to help distinguish among these different types of points. These "second-order conditions" relate to the second-order partial derivatives, just as the first-order condition used the first-order partial derivatives.

### 4.4.1 Sufficient second-order conditions

A function's second derivatives can help distinguish among types of points, through a **second-order condition** (**SOC**).

**Theorem 4.3.** *Assume A4.1. Assume that the first-order condition (Definition 4.6) holds at point $x^*$. Assume $f \in C^2$ in a neighborhood of $x^*$. Let $H(x^*)$ denote the Hessian matrix of second derivatives (Definition 3.3) evaluated at $x^*$. Then, the following hold.*

1. *If $H(x^*)$ is positive definite, then $x^*$ is a local minimum point.*

2. *If $H(x^*)$ is negative definite, then $x^*$ is a local maximum point.*

3. *If one of the above conditions holds for all $x \in S$ (not only $x^*$), then $x^*$ is a* global *minimum or maximum point.*

*Proof.* Recall the definitions of positive definite and negative definite from Definition 2.15 (and related tests in Proposition 2.6).

1. Consider a second-order Taylor expansion around $x^*$: treating all the vectors as $n \times 1$ column vectors, and for some $\tilde{x}$ between $x^*$ and $x$,

$$f(x) = f(x^*) + \overbrace{\nabla f(x^*)'}^{=0 \text{ by FOC}}(x - x^*) + (1/2)(x - x^*)'H(\tilde{x})(x - x^*) \tag{4.8}$$

$$= f(x^*) + (1/2) \overbrace{(x - x^*)'H(\tilde{x})(x - x^*)}^{>0 \text{ if } H \text{ positive definite and } x \neq x^*} . \tag{4.9}$$

   In a small enough neighborhood of $x^*$, it holds by continuity of second derivatives that if $H(x^*)$ is positive definite, then so is $H(\tilde{x})$. Thus, $f(x) > f(x^*)$ for $x \neq x^*$ in that neighborhood.

2. The Taylor expansion is identical, but now $(x - x^*)'H(\tilde{x})(x - x^*) < 0$ if $H(\tilde{x})$ is negative definite and $x \neq x^*$.

3. If the positive or negative definiteness of $H(x)$ holds for all $x \in S$, then the local maximum or minimum argument extends globally. For example, for the minimum, the expansion is the same as before, but now $H(\tilde{x})$ is positive definite for any $\tilde{x} \in S$, so $f(x) > f(x^*)$ for any $x \in S$. □

### 4.4.2 Necessary second-order conditions

The following result parallels Theorem 4.4.

**Theorem 4.4.** *Given the assumptions and notation of Theorem 4.3, the following hold.*

1. *If $x^*$ is a local minimum, then $H(x^*)$ is positive semidefinite.*

2. *If $x^*$ is a local maximum, then $H(x^*)$ is negative semidefinite.*

**Example 4.2.** Consider some univariate examples first. If $f(x) = x^2$, then the FOC is satisfied at $x^* = 0$, and the second derivative is $f''(x) = 2 > 0$ everywhere, which corresponds to a positive definite Hessian $H(0)$. By Theorem 4.3, we conclude that

$x^* = 0$ is a local minimum. The necessary condition $H(0) \geq 0$ (positive semidefinite) is also satisfied. With $g(x) = x^4$, like in Section 3.3.3, $g''(x) = 12x^2$ so $g''(0) = 0$, so Theorem 4.3 does not apply. However, $x^* = 0$ is still a local (and global) minimum point, so by Theorem 4.4 we know $g''(0) \geq 0$, which we already verified. ■

**Example 4.3.** Here we try to find the maximum point of $f(x_1, x_2) = \log(x_1) + \log(x_2) + \log(3 - x_1 - x_2)$ over $A = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 > 0, x_2 > 0, 3 - x_1 - x_2 > 0\}$. That is, we consider the maximization problem

$$\max_{x_1, x_2} \log(x_1) + \log(x_2) + \log(3 - x_1 - x_2). \tag{4.10}$$

First, we solve the first-order conditions, setting each partial derivative equal to zero when evaluated at $x^*$. From the FOC for $x_1$,

$$0 = \left.\frac{\partial f}{\partial x_1}\right|_{x=x^*} = (1/x_1^*) - 1/(3 - x_1^* - x_2^*) \implies 3 - x_1^* - x_2^* = x_1^* \implies x_2^* = 3 - 2x_1^*. \tag{4.11}$$

From the FOC for $x_2$,

$$\begin{aligned} 0 = \left.\frac{\partial f}{\partial x_2}\right|_{x=x^*} &= 0 + (1/x_2^*) - 1/(3 - x_1^* - x_2^*) \\ \implies 3 - x_1^* - x_2^* = x_2^* &\implies x_1^* = 3 - 2x_2^* = 3 - 2(3 - 2x_1^*) = 3 - 6 + 4x_1^* \\ \implies 3x_1^* = 3 &\implies x_1^* = 1 \implies x_2^* = 3 - 2x_1^* = 1. \end{aligned} \tag{4.12}$$

Thus, our candidate point for a local maximum is $x^* = (1, 1)$.

We now show that the SOC in Theorem 4.3 holds, so $x^*$ is a local maximum point. The gradient and Hessian are

$$\nabla f(x_1, x_2) = \left[\frac{1}{x_1} - \frac{1}{3 - x_1 - x_2}, \frac{1}{x_2} - \frac{1}{3 - x_1 - x_2}\right], \tag{4.13}$$

$$H(x_1, x_2) = \begin{bmatrix} -\frac{1}{x_1^2} - \frac{1}{(3-x_1-x_2)^2} & -\frac{1}{(3-x_1-x_2)^2} \\ -\frac{1}{(3-x_1-x_2)^2} & \frac{-1}{x_2^2} - \frac{1}{(3-x_1-x_2)^2} \end{bmatrix}. \tag{4.14}$$

For the local sufficient condition, we use Proposition 2.6 to show that $H(1, 1)$ is negative definite:

$$H(1, 1) = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}, \tag{4.15}$$

$$(-1)^1 D_1 = (-1)(-2) = 2 > 0, \tag{4.16}$$

$$(-1)^2 D_2 = \begin{vmatrix} -2 & -1 \\ -1 & -2 \end{vmatrix} = 4 - 1 = 3 > 0, \tag{4.17}$$

so $H(1, 1)$ is negative definite and $x^* = (1, 1)$ is a local maximum.

Globally, too, $H(x_1, x_2)$ is negative definite. For any $(x_1, x_2) \in A$,

$$D_1(x) = -\frac{1}{x_1^2} - \frac{1}{(3 - x_1 - x_2)^2} < 0 \implies (-1)D_1(x) > 0. \tag{4.18}$$

Letting $a = -1/x_1^2$, $b = -1/(3 - x_1 - x_2)^2$, and $c = -1/x_2^2$, and noting $a < 0$, $b < 0$, and $c < 0$ for all $(x_1, x_2) \in A$,

$$D_2(x) = \begin{vmatrix} a+b & b \\ b & c+b \end{vmatrix} = (a+b)(c+b) - b^2 = ac + ab + bc + b^2 - b^2 = ac + ab + bc > 0, \tag{4.19}$$

implying $(-1)^2 D_2(x) > 0$ for all $x \in A$. Thus, $H(x) < 0$ for any $x \in A$, and $x^* = (1, 1)$ is a global maximum by Theorem 4.3. ■

## Optional resources

Optional resources for this chapter

- `https://web.archive.org/web/20240713110834/https://hummedia.manchester.ac.uk/schools/soss/economics/pg/psmaths/pre-sessiomathsbook.pdf` pages 33–36 and 66–68

## Exercises

**Exercise E4.1.** Let $f(x) = x^3 - 3x$, and consider $\max_{x \in [-3,3]} f(x)$.

a. Compute $f'(x)$ and solve the first-order condition for the two critical points $x_1^*$ and $x_2^*$.

b. For both points: use Theorem 4.3 to judge whether each point is a local maximum or local minimum or neither.

c. What is the global maximum, $\max_{x \in [-3,3]} f(x)$? (Hint: it must be either a local max or one of the boundary points, $x = -3$ or $x = 3$.)

**Exercise E4.2.** Let $f(x) = 2\log(x) + [(x/2) - 2]^2$, and consider $\max_{x \in [0.4,4]} f(x)$.

a. Compute $f'(x)$ and solve the first-order condition for the critical point $x^*$.

b. Compute the second derivative at the critical point, $f''(x^*)$; can you tell if it's a local max or min?

c. Show that over $x \in [0.4, 4]$, $f'(x) \geq 0$. Given that, what's the global maximizer $x^* = \arg\max_{x \in [0.4,4]} f(x)$?

## Appendix 4.A Homogeneity of a function

Warning: the following has not been revised or edited.

See also pages 63–65 of https://web.archive.org/web/20240713110834/https://hummedia.manchester.ac.uk/schools/soss/economics/pg/psmaths/pre-sessiomathsbook.pdf.

**Definition 4.7.** Function $f\colon S \subseteq \mathbb{R}^n \to \mathbb{R}$ is **homogeneous of degree $\boldsymbol{k}$** iff

$$f(\lambda x) = \lambda^k f(x) \text{ for all } x \in S, \lambda > 0.$$

**Example 4.4.**

- Firm cost function is homogeneous of degree 1 in input prices:

$$c(w; q) = \min_{\boldsymbol{x}} w\boldsymbol{x}$$
$$\text{subject to } f(\boldsymbol{x}) \geq q$$

  where $w$ denotes inout, $q$ output level prices, $\boldsymbol{x}$ = input vector, $f(\boldsymbol{x})$ = production function

- Consumer demands are homogeneous of degree zero in prices and income:

$$x(p, m) = \arg\max_{x} v(\boldsymbol{x})$$
$$\text{subject to } p\boldsymbol{x} \leq m$$

  $\boldsymbol{x}$ = demand vector, $p$ = prices, $m$ = income.

- We often assume that production functions are homogeneous of degree 1. ∎

Euler's Theorem states that for a homogeneous function of degree $k$,

$$\sum_{i=1}^{n} x_i \overbrace{f_i(x)}^{\text{partial derivatives}} = kf(x).$$

**Example 4.5.** Cost function: $w = (\text{wage}, r)$. Euler's Theorem:

$$\text{wage} \cdot \frac{\partial c}{\partial \text{ wage}} + r \cdot \frac{\partial c}{\partial r} = 1 \cdot \text{cost}.$$ ∎

### 4.A.1 Homogeneity and partial derivatives

If $f(x)$ is homogeneous of degree $k$ then $f_i(x)$ (partial derivative) is homogeneous of degree $k-1$.

*Proof.* Let $y = \lambda x \iff x = \frac{1}{\lambda} y \Leftrightarrow \frac{\partial f(x)}{\partial x_i} = \frac{\partial f(y)}{\partial y_i} \cdot \frac{\partial y_i}{\partial x_i}$. Consider equation $f(\lambda x) = \lambda^k f(x)$.

- Take derivative with respect to $\lambda$: $\sum_{i=1}^{n} x_i f_i(\lambda x) = k \cdot \lambda^{k-1} \cdot f(x)$
- Evaluate at $\lambda = 1$: $\sum_{i=1}^{n} X_i f_i(x) = kf(x)$ □

*Proof.* Consider the derivative result. Let $y = \lambda x$; more importantly, $y_i = \lambda x_i \implies x_i = \frac{1}{\lambda}$.

$$\begin{aligned}\frac{\partial f(y)}{\partial y_i} &= \frac{\partial f(y)}{\partial x_i} \cdot \frac{\partial x_i}{\partial y_i} \\ &= \frac{\partial \lambda^k \cdot f(x)}{\partial x_i} \cdot \frac{1}{\lambda} = \lambda^{k-1} \cdot \frac{\partial f(x)}{\partial x_i}\end{aligned}$$ □

### 4.A.2 Implication of homogeneity for level curves

An isocurve or level curve plots $f(x) = \alpha$ for a given constant $\alpha$. For example, if $x = (x_1, x_2)$ and $f(x) = x_1 x_2$, then $f(x) = \alpha$ implies $x_2 = \alpha / x_1$, which can be plotted on the two-dimensional $(x_1, x_2)$ plane.

Homogeneity implies that isocurves satisfy a scaling property along rays from the origin, as illustrated in Figure 4.3. Thus, for a homogeneous function, all isocurves can be generated from any single isocurve.

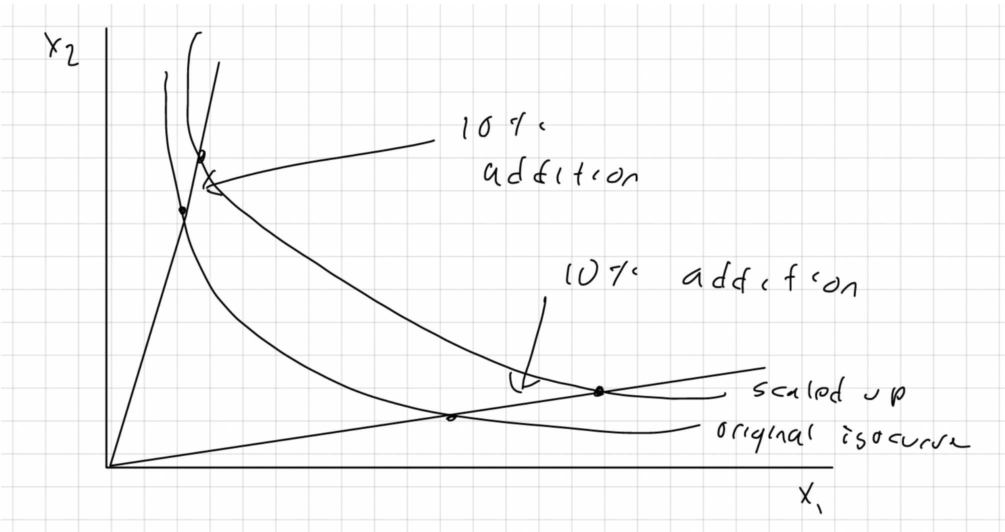



Figure 4.3: Example isocurves for a homogeneous function.

# Chapter 5

# Constrained Optimization

This chapter extends Chapter 4 to the case where there are constraints. This is common in economics. For example, a consumer wants to maximize utility subject to a budget constraint, or a firm wants to maximize profit subject to their production function.

*Unit learning objectives for this chapter*

5.1. Define vocabulary words (in **bold**) related to constrained optimization, both mathematically and intuitively [TLO 1]

5.2. Solve a constrained maximization problem [TLO 2]

5.3. Characterize a point that solves a constrained maximization problem [TLO 3]

## 5.1 Equality constraints

Consider the maximization problem

$$\begin{aligned} &\max_x f(x) \\ &\text{s.t. } g_j(x) = b_j, \ j = 1, \ldots, m, \end{aligned} \tag{5.1}$$

where $x \in \mathbb{R}^n$ with $n > m$, and "s.t." stands for "subject to." Note we could always define $\tilde{g}_j(x) = g_j(x) - b_j$ to make the constraints $\tilde{g}_j(x) = 0$, which is simpler and commonly seen. This (5.1) can be solved by the substitution method (solving for some of the $x_j$ using the constraints and substituting those $x_j$ into an unconstrained maximization of $f(x)$), the implicit substitution method, or the method we describe below: the Lagrange method.

### 5.1.1 Lagrange method

The Lagrange method "adds" the constraints to the objective function before solving the first-order condition and checking second-order condtiions. The **Lagrangian** takes the objective function from (5.1) and subtracts a **Lagrange multiplier** $\lambda_j$ times each term $g_j(x) - b_j$ that equals zero when the constraint holds:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{j=1}^{m} \lambda_j (g_j(x) - b_j). \tag{5.2}$$

Sometimes the $\lambda_j$ terms are added instead of subtracted; mathematically, it is equivalent.

### First-order condition

Similar to the case without constraints, we solve the first-order condition (FOC) to compute a candidate point $x^*$ that may be a maximum point (if a second-order condition holds). The FOC is not actually a necessary condition for a maximum point, unless a "constraint qualification condition" holds; we assume one does.

As before (Definition 4.6), the FOC sets all first-order partial derivatives equal to zero, but now we consider the partials of $\mathcal{L}(x, \lambda)$ with respect to each element of $x$ or $\lambda$. For the $\lambda$ components, for each $j \in \{1, \dots, m\}$,

$$0 = \left.\frac{\partial \mathcal{L}(x, \lambda)}{\partial \lambda_j}\right|_{\substack{x=x^* \\ \lambda=\lambda^*}} = -(g_j(x^*) - b_j) \implies g_j(x^*) = b_j, \tag{5.3}$$

so each constraint is satisfied at the solution $x^*$. For the $x$ components, for each $i \in \{1, \dots, n\}$,

$$0 = \left.\frac{\partial \mathcal{L}(x, \lambda)}{\partial x_i}\right|_{\substack{x=x^* \\ \lambda=\lambda^*}} = \frac{\partial f(x^*)}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(x^*)}{\partial x_i}, \tag{5.4}$$

Sometimes this is written more compactly as $\nabla f(x^*) = \lambda' \nabla g(x^*)$, where $\lambda' = (\lambda_1, \dots, \lambda_m)$ and the matrix $\nabla g(x^*)$ has $\frac{\partial g_j(x^*)}{\partial x_i}$ as the entry in row $j$ and column $i$.

### Necessary and sufficient conditions

Warning: there are many related but different results for necessary and sufficient conditions for equality-constrained maximization. However, generally the necessary condition focuses on the first-order condition being satisfied, while the sufficient conditions add a negative definite Hessian (local maximum) or concavity (global maximum).

- (*Necessary*) Suppose $f$ and $g$ are $C^1$ functions on a convex set $S \subseteq \mathbb{R}^n$, $x^*$ solves the constrained maximization problem in (5.1), and constraint qualification holds at $x^*$, meaning the vectors $\nabla g_j$ over $j = 1, \dots, m$ are all non-zero and mutually linearly independent (or equivalently, the matrix $\nabla g(x^*)$ has rank $m$, with $m \leq n$). Then, there exists a unique vector $\lambda^*$ such that $x^*$ solves the Lagrange first-order conditions.

- (*Sufficient*) If there exist $\lambda^*$ and $x^*$ satisfying the Lagrangian first-order condition, and if $\mathcal{L}(\cdot, \lambda^*)$ is concave, then $x^*$ solves the maximization problem. (Or for a local maximum: the Hessian of $\mathcal{L}(\cdot, \lambda^*)$ evaluated at $x^*$ is negative definite.)

**Example 5.1.** We solve the first-order condition and then check the Hessian for the maximization problem

$$\begin{aligned} &\max_{x,y} \ \log x + \log y \\ &\text{s.t. } x + y = M, \end{aligned} \tag{5.5}$$

where $M$ is a constant and $x, y > 0$. (You can imagine this as a simplistic consumer's problem where $M$ is the amount of money the consumer spends on two goods that yield utility $\log x + \log y$.) The Lagrangian is

$$\mathcal{L}(x, y, \lambda) = \log x + \log y - \lambda(x + y - M). \tag{5.6}$$

The first-order conditions are

$$\left.\begin{aligned} 0 &= \left.\frac{\partial \mathcal{L}(x,y,\lambda)}{\partial x}\right|_{\substack{x=x^*\\ y=y^*\\ \lambda=\lambda^*}} = (1/x^*) - \lambda^* \\ 0 &= \left.\frac{\partial \mathcal{L}(x,y,\lambda)}{\partial y}\right|_{\substack{x=x^*\\ y=y^*\\ \lambda=\lambda^*}} = (1/y^*) - \lambda^* \\ 0 &= \left.\frac{\partial \mathcal{L}(x,y,\lambda)}{\partial \lambda}\right|_{\substack{x=x^*\\ y=y^*\\ \lambda=\lambda^*}} = -(x^* + y^* - M) \end{aligned}\right\} \implies \begin{cases} \lambda^* = 1/x^* \\ \lambda^* = 1/y^* \implies x^* = y^* \\ M = x^* + y^* = 2x^* \\ \qquad \implies x^* = y^* = M/2 \\ \qquad \implies \lambda^* = 2/M. \end{cases} \tag{5.7}$$

For the second-order condition, plugging $\lambda^* = 2/M$ into the Lagrangian, $\mathcal{L}(x, y, \lambda^*) = \log x + \log y - (2/M)(x + y - M)$. The Hessian is

$$H = \begin{bmatrix} -1/x^2 & 0 \\ 0 & -1/y^2 \end{bmatrix}, \tag{5.8}$$

which is negative definite for all $x, y > 0$ (because it is a diagonal matrix with all entries strictly negative), so $\mathcal{L}(\cdot, \cdot, \lambda^*)$ is strictly concave and $x^* = y^* = M/2$ is the global maximum. ■

### 5.1.2 Envelope theorem and shadow values

The Lagrange multiplier $\lambda_j$ can be interpreted as a **shadow value** (or in some contexts **shadow price**): the marginal value of "relaxing" the corresponding constraint. "Relaxing" is more appropriate with an inequality constraint, but consider the following qualitative idea. A consumer is trying to maximize utility subject to a budget constraint. For now, we assume the consumer spends the whole budget and write the budget constraint as an equality constraint, although more realistically it's an inequality. For example, if they spend $x$ and $y$ on two goods (respectively) and spend $M$ money total, then we write $x + y = M$ like in Example 5.1, although really their constraint is $x + y \leq M$. The Lagrange multiplier $\lambda$ tells us the marginal utility of the consumer when we increase their budget $M$.

This interpretation can be derived using an extension of the Envelope Theorem. Generally, let $\theta$ be a parameter vector that can enter both the objective function and constraints, like $f(x, \theta)$ and $g(x, \theta)$. For example, in (5.1), the $b_j$ values in the constraints are the elements of vector $\theta$; in Example 5.1, $M$ is the only element in $\theta$. Let $x^*$ be the solution to the maximization problem; because it can depend on $\theta$, we write $x^*(\theta)$. Similarly, we write $\lambda^*(\theta)$ for the Lagrange multiplier vector at the solution. Consider the maximized value of the objective function as a function of $\theta$ (the **value function**, hence $V$):

$$\begin{aligned} V(\theta) &= \max_x f(x, \theta) \text{ s.t. } g_j(x, \theta) = 0 \ (j = 1, \ldots, m) \\ &= f(x^*(\theta), \theta) = f(x^*(\theta), \theta) - \sum_{j=1}^{m} \lambda_j^*(\theta) g_j(x^*, \theta) \end{aligned} \tag{5.9}$$

because $g_j(x^*, \theta) - b_j = 0$ Similar to the unconstrained Envelope Theorem in (4.5), only the direct effect of $\theta$ on $V(\theta)$ enters the derivative, and the indirect effects (through $x^*$ and $\lambda^*$) are zero due to the first-order condition. In this case, skipping the proof,

$$\frac{\partial V(\theta)}{\partial \theta_j} = \left.\frac{\partial f(x, \theta)}{\partial \theta_j}\right|_{x=x^*(\theta)} - \sum_{j=1}^{m} \lambda_j^*(\theta) \left.\frac{\partial g(x, \theta)}{\partial \theta_j}\right|_{x=x^*(\theta)}. \tag{5.10}$$

**Example 5.2.** In Example 5.1, the only parameter is $M$, which does not enter $f$. Replacing $M$ with $\theta$, the constraint $x + y = M$ can be rewritten as $g(x, y, \theta) = 0$ for $g(a, b, c) = a + b - c$, so $\frac{\partial g(x,y,\theta)}{\partial \theta} = -1$. Using (5.10), the marginal value of increasing the "budget" $\theta$ is $V'(\theta) = 0 - \lambda^*(\theta)(-1) = \lambda^*(\theta)$, which is the Lagrange multiplier that was shown to have value $\lambda^*(\theta) = 2/\theta$. Note: you may worry that writing the Lagrangian with $+\lambda(x + y - \theta)$ instead of $-\lambda(x + y - \theta)$ would cause the marginal value of money to be negative, but the changed sign in canceled out by the flipped sign of $\lambda^*(\theta) = -2/M$ in that case; so whether we write $-\lambda$ or $+\lambda$ in the Lagrangian, or $x + y - \theta$ or $\theta - x - y$, the marginal value of money will be the same. ■

### 5.1.3 Local second-order conditions

To see if $x^*$ corresponds to a local minimum or local maximum (or neither), the **bordered Hessian** matrix can be constructed and analyzed. "Bordered" only refers to having a particular structure; it's really just a normal Hessian matrix. Specifically, here considering the problem $\max_x f(x)$ s.t. $g_j(x) = 0$ $(j = 1, \ldots, m)$, it's the Hessian matrix of the

Lagrangian $\mathcal{L}(\lambda, x) = f(x) - \sum_{j=1}^{m} \lambda_j g_j(x)$, which has the structure

$$B = \begin{bmatrix} 0_{m\times m} & \nabla g_{m\times n} \\ (\nabla g_{m\times n})' & \mathcal{L}'' \end{bmatrix}. \tag{5.11}$$

The top left of $B$ is an $m \times m$ matrix with every entry equal to zero because the second derivative with respect to $\lambda$ is zero: the first derivative wrt any $\lambda_j$ is the corresponding constraint like $g_j(x)$, which does not contain any $\lambda$, hence taking another derivative wrt $\lambda$ yields zero. The bottom right of $B$ is an $n \times n$ matrix that is the second derivative matrix of $\mathcal{L}(\lambda, x)$ with respect to vector $x$. That is, the row $i$, column $j$ entry of $\mathcal{L}''$ is $\frac{\partial^2 \mathcal{L}(\lambda,x)}{\partial x_i \partial x_j}$. The top right of $B$ is an $m \times n$ matrix with the cross-partials (second derivative wrt one element of $\lambda$ and one element of $x$). It is easier to think about taking the partial wrt $\lambda_j$ first, which yields the constraint, $g_j(x)$. Thus, the row $j$, column $i$ element in matrix $\nabla g$ is $\frac{\partial g_j(x)}{\partial x_i}$; that is, $\nabla g$ is the gradient (first derivative) matrix of $g$. Finally, the bottom left of $B$ is simply the transpose of the top right.

In the special case where the constraint functions are linear in $x$, then the constraints disappear from $\mathcal{L}''$, leaving $\mathcal{L}''$ equal to the Hessian of the objective function $f$. Then, (5.11) is the Hessian of $f$ "bordered" by the gradient of the constraints, hence the term "bordered Hessian." However, with nonlinear $g_j$, the constraints may also appear in $\mathcal{L}''$.

Recall the second-order conditions for unconstrained optimization from Theorem 4.3 that depend on the positive or negative definiteness of the Hessian, along with the relationship between definiteness and leading principal minors from Proposition 2.6. Here, with $m$ equality constraints, intuitively we only effectively have $n - m$ choice variables. That is, if we fix the values of $(x_1, \ldots, x_{n-m})$, then the other $(x_{n-m+1}, \ldots, x_n)$ are fully determined by the constraints. Correspondingly, the second-order conditions depend on only the last $n - m$ leading principal minors of the $(m + n) \times (m + n)$ Hessian of the Lagrangian. That is, the conditions only use the leading principal minors of order $2m + 1, \ldots, m + n$. Let $B_r(x^*)$ denote the determinant of the leading principal minor of order $r$ of the bordered Hessian in (5.11) evaluated at $x^*$. The first-order condition solution $x^*$ is a:

- local minimum if $(-1)^m B_r(x^*) > 0$ for $r = 2m + 1, \ldots, m + n$;
- local maximum if $(-1)^{r-m} B_r(x^*) > 0$ for $r = 2m + 1, \ldots, m + n$;
- possible local minimum, local maximum, or saddlepoint (neither min nor max), if the above conditions do not hold.

**Example 5.3.** Check the bordered Hessian condition for Example 5.1, maximizing $\log x + \log y$ given $x + y = M$. Note $n = 2$ and $m = 1$. In (5.11), the top right is $\nabla g = (1, 1)$. The bottom right is a $2 \times 2$ matrix with the following entries. Given

$$\frac{\partial \mathcal{L}}{\partial x} = (1/x) - \lambda, \quad \frac{\partial \mathcal{L}}{\partial y} = (1/y) - \lambda, \tag{5.12}$$

the second derivatives are

$$\frac{\partial^2 \mathcal{L}}{\partial x^2} = -x^{-2}, \quad \frac{\partial^2 \mathcal{L}}{\partial y^2} = -y^{-2}, \quad \frac{\partial^2 \mathcal{L}}{\partial x \partial y} = 0. \tag{5.13}$$

Putting everything together, the bordered Hessian is

$$B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -x^{-2} & 0 \\ 1 & 0 & -y^{-2} \end{bmatrix}. \tag{5.14}$$

Evaluating this at the first-order condition solution $(x^*, y^*) = (M/2, M/2)$,

$$B(x^*, y^*) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -4/M^2 & 0 \\ 1 & 0 & -4/M^2 \end{bmatrix}. \tag{5.15}$$

We need to check the last $n - m = 2 - 1 = 1$ leading principal minor determinant, which is just the determinant of the full matrix:

$$\det(B(x^*, y^*)) = \begin{vmatrix} 0 & 1 & 1 \\ 1 & -4/M^2 & 0 \\ 1 & 0 & -4/M^2 \end{vmatrix} \tag{5.16}$$

$$= 0 \times \begin{vmatrix} -4/M^2 & 0 \\ 0 & -4/M^2 \end{vmatrix} - 1 \times \begin{vmatrix} 1 & 0 \\ 1 & -4/M^2 \end{vmatrix} + 1 \times \begin{vmatrix} 1 & -4/M^2 \\ 1 & 0 \end{vmatrix} \tag{5.17}$$

$$= 0 - (1)(-4/M^2 - 0) + (1)(0 - (-4/M^2)) = (4/M^2) + (4/M^2), \tag{5.18}$$

so for $r = 2m + 1 = 3$, $(-1)^{r-m} B_r(x^*, y^*) = (-1)^2(8/M^2) > 0$ (if $M \neq 0$). Thus, $(x^*, y^*) = (M/2, M/2)$ is a local maximum. ∎

## 5.2 Inequality constraints

In economics, constraints are often inequalities, like a budget constraint. Consider the maximization problem of (5.1) but with inequality constraints:

$$\begin{aligned} &\max_x f(x) \\ &\text{s.t. } g_j(x) \leq b_j, \; j = 1, \ldots, m. \end{aligned} \tag{5.19}$$

Sometimes we can replace the $\leq$ with $=$ if $f$ and $g$ are increasing in $x$. For example, if we have a budget constraint $x + y \leq M$ and utility $f$ is increasing in $x$ and $y$, then we know the maximum must have $x^* + y^* = M$. However, there are more general solutions, discussed below.

It helps to distinguish constraints that are binding (equalities) at the optimum from constraints that are not.

**Definition 5.1.** An **active constraint** is binding at the optimum; that is, $g_j(x) \leq b_j$ is an active constraint iff $g_j(x^*) = b_j$.

**Definition 5.2.** A **passive constraint** is not binding (or **slack**) at the optimum; that is, $g_j(x) \leq b_j$ is a passive constraint iff $g_j(x^*) < b_j$.

**Example 5.4.** As alluded to, an example of (5.19) is utility maximization subject to a budget constraint. Given utility function $f(\cdot)$, budget $M$, goods vector $x$, and corresponding price vector $p$, the consumer's problem is $\max_x f(x)$ subject to $p \cdot x \leq M$. (Also as noted, assuming $f$ is increasing in all its arguments, we know this will have an active constraint, so we could just use $p \cdot x = M$ for the constraint.) ■

### 5.2.1 First-order conditions (Karush–Kuhn–Tucker)

Similar to before, solving the first-order conditions yields a set of points that contains all maxima, but may also include minima or saddle points. Here, these necessary conditions are called the **Karush–Kuhn–Tucker (KKT) conditions**. (They are also called Kuhn–Tucker (KT), even though it was discovered that Karush's work was done a decade earlier.)

The Lagrangian for (5.19) is the same as in the equality-constrained case:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{j=1}^{m} \lambda_j [g_j(x) - b_j]. \tag{5.20}$$

The KKT conditions are

- For all $i = 1, \ldots, n$: $\frac{\partial \mathcal{L}}{\partial x_i}(x^*, \lambda^*) = \frac{\partial f}{\partial x_i}(x^*) - \sum_{j=1}^{m} \lambda_j^* \frac{\partial g_j}{\partial x_i}(x^*) = 0$.
- For all $j = 1, \ldots, m$: $\lambda_j^* \geq 0$ with $\lambda_j^* = 0$ if $g_j(x^*) < b_j$ (passive constraint).
- For all $j = 1, \ldots, m$: $g_j(x^*) \leq b_j$ (the constraints are satisfied).

Sometimes the part about $\lambda_j^* = 0$ is omitted from the second condition, and instead (equivalently) a fourth condition is added that $\lambda_j^*[g_j(x^*) - b_j] = 0$ for all $j = 1, \ldots, m$, which is called **complementary slackness**.

**Theorem 5.1** (Karush–Kuhn–Tucker Theorem)**.** *If constraint qualification holds for all active constraints, then the Karush–Kuhn–Tucker conditions are necessary conditions. That is, given a maximum point $x^*$, there exists $\lambda^*$ such that the KKT conditions hold. (One version of constraint qualification:*[1] *labeling the active constraints as $j = 1, \ldots, \bar{m}$, the $\bar{m} \times n$ matrix with $\nabla g_j$ as its $j$th row has rank $\bar{m}$.)*

[1] Others: https://en.wikipedia.org/wiki/Karush%E2%80%93Kuhn%E2%80%93Tucker_conditions#Regularity_conditions_(or_constraint_qualifications)

### 5.2.2 Sufficient conditions

The following are sufficient conditions for a point satisfying the first-order conditions to be a global maximum point. Remember that even if no sufficient condition holds, the point could still be a maximum, but we would need another way to prove it.

- Simplest but strongest (most restrictive): $\mathcal{L}$ is concave in $x$ for the $\lambda^*$ that satisfies the KKT conditions; that is, $\mathcal{L}(\cdot, \lambda^*)$ is a concave function.

- Or: $f$ is concave, and all $\lambda_i^* g_i$ are quasiconvex (which implies the constraint set is convex).

- Or: $f$ is quasiconcave, all $\lambda_i g_i$ are quasiconvex, and $\nabla f(x^*) \neq 0$.

A typical procedure is to check the second or third variation, and convert to equality constraints if possible.

Sufficient conditions for a local maximum point are the same as for the equality constrained case when we consider only the active constraints.

## 5.3 Mixed constraint problems

Although beyond our scope, Karush–Kuhn–Tucker can extend to maximization problems with a mix of both equality and inequality constraints, like

$$\begin{aligned} \max_x \; & f(x) \\ \text{s.t. } & g(x) = a \\ & h(x) \leq b. \end{aligned}$$

## Optional resources

Optional resources for this chapter

- Notes from Peter Ireland (BC): https://web.archive.org/web/20240709190024/http://irelandp.com/econ7720/notes/notes1.pdf
- https://web.archive.org/web/20240713110834/https://hummedia.manchester.ac.uk/schools/soss/economics/pg/psmaths/pre-sessiomathsbook.pdf pages 37–43 and 69–76
- https://en.wikipedia.org/wiki/Hessian_matrix#Bordered_Hessian (bordered Hessian sufficient conditions)
- https://en.wikipedia.org/wiki/Nonlinear_programming

## Exercises

**Exercise E5.1.** Consider $\max_{x,y} xy$ s.t. $px + qy = M$, assuming $x, y \geq 0$ (which you don't need to add to the Lagrangian, just double-check your solution). Here $p$ and $q$ are the respective prices of goods $x$ and $y$, and $M$ is the total budget.

a. Write out the Lagrangian.

b. Solve the first-order conditions for $x^*$, $y^*$, and $\lambda^*$.

c. Compute the bordered Hessian matrix.

d. Compute the determinant of the bordered Hessian; what does this imply for the bordered Hessian criteria?

e. In the $x$–$y$ plane, draw a graph with some utility level curves ($xy = c$ for different values of $c$) and the budget constraint ($px + qy = M$) and explain why you think the FOC solution is a global maximum, and how this is logically possible even though the bordered Hessian condition was not satisfied.

# Chapter 6

# Difference Equations

**Difference equations** describe the next value in a sequence in terms of the current and previous values in the sequence. In economics, our sequences are usually over time: $x_t$ is the value in time period $t$.

Given a difference equation, we usually want to solve for $x_t$ in terms of parameters and initial value $x_0$, find equilibrium values, and/or assess stability.

There are many types of difference equations, categorized by various properties. For example, the following is a non-stochastic, time-homogeneous, first-order, nonlinear difference equation:

$$x_{t+1} = f(x_t). \tag{6.1}$$

- **Non-stochastic**: there are no random variables (no noise, no error term); $x_t$ fully determines $x_{t+1}$
- **Time-homogeneous**: the function $f(\cdot)$ does not change over time (no $t$ subscript)
- **First-order**: $x_{t+1}$ only depends on the immediately previous value $x_t$; it does not further depend on $x_{t-1}$, $x_{t-2}$, etc.
- **Nonlinear**: the function $f(\cdot)$ may be nonlinear

Later, we will also study linear and/or higher-order difference equations, as well as a time-varying example.

*Unit learning objectives for this chapter*

6.1. Define vocabulary words (in **bold**) related to difference equations, both mathematically and intuitively [TLO 1]

6.2. Solve difference equations by writing an element in the sequence in terms of (only) the initial value of the sequence (and other parameters) [TLO 2]

6.3. Solve for equilibrium values of a difference equation [TLO 2]

6.4. Characterize the stability (or instability) of a difference equation [TLO 3]

## 6.1 Solving a difference equation

One of our main goals is to **solve** a difference equation: instead of writing the relationship between $x_{t+1}$ and previous values, write it in terms of the initial value $x_0$.

### 6.1.1 Nonlinear, first-order

The difference equation in (6.1) can be solved numerically with the **insertion method** (or **forward solution**) if we know a starting value $x_0$:

$$\begin{aligned}
x_1 &= f(x_0), \\
x_2 &= f(x_1) = f\big(f(x_0)\big), \\
&\vdots \\
x_{t+1} &= f(x_t) = \overbrace{(f \circ f \circ \cdots \circ f)}^{t+1 \text{ times}}(x_0).
\end{aligned}$$

However, in practice this approach is often computationally inefficient and numerically inaccurate.

### 6.1.2 Linear, first-order

Consider the special case of (6.1) with linear $f(x) = ax + b$:

$$x_{t+1} = ax_t + b. \tag{6.2}$$

For a given $x_0$, and assuming $a \neq 1$ (to avoid zero denominator), we can solve this as

$$\begin{aligned}
x_t &= \overbrace{a^t x_0 + (a^{t-1} + a^{t-2} + \cdots + a + 1)b}^{\text{via insertion method}} \\
&= a^t x_0 + \frac{1-a^t}{1-a} b \\
&= a^t \left( x_0 - \frac{b}{1-a} \right) + \frac{b}{1-a}.
\end{aligned} \tag{6.3}$$

## 6.2 Equilibrium and stability

An **equilibrium point** (or **steady state** or **fixed point**) $x^*$ satisfies the following: if $x_t = x^*$, then $x_{t+s} = x^*$ for all $s > 0$. That is, if a sequence reaches an equilibrium point, then it stays at that point forever after. However, some equilibrium points are "unstable": even if $x_t = x^* + \epsilon$ for very small $\epsilon > 0$, the sequence may diverge away from $x^*$. Thus, we want both to derive equilibrium points and to characterize their stability.

### 6.2.1 Linear, first-order

Below, we consider equilibrium and stability in the linear, first-order difference equation in (6.2).

#### Deriving the equilibrium

To find an equilibrium, replace the $x_{t+1}$ and $x_t$ with $x^*$, and then solve for $x^*$. Applying this procedure to (6.2):

$$x^* = ax^* + b \implies x^*(1-a) = b \implies x^* = b/(1-a) \text{ given } a \neq 1. \tag{6.4}$$

To verify that this is an equilibrium point: if $x_t = x^* = b/(1-a)$, then

$$x_{t+1} = ab/(1-a) + b = \frac{ab + b(1-a)}{1-a} = \frac{b}{1-a} = x^*.$$

#### Stability

A difference equation has **asymptotic stability** when

$$\lim_{t\to\infty} x_t = x^*. \tag{6.5}$$

When is the linear first-order difference equation in (6.2) asymptotically stable? Assuming $a \neq 1$ and using $x^* = b/(1-a)$ from (6.4), (6.3) becomes

$$x_t = a^t(x_0 - x^*) + x^*.$$

Thus, to get $x_t \to x^*$, we need either $x_0 = x^*$ (start at the equilibrium) or $a^t \to 0$. The latter condition is equivalent to $|a| < 1$. So,

$$|a| < 1 \implies \lim_{t\to\infty} x_t = x^*.$$

That is, if $|a| < 1$, then the sequence $x_t$ converges to the equilibrium point $x^*$, regardless of the initial value $x_0$.

Instead, the sequence diverges if $|a| \geq 1$ (unless $x_0 = x^*$). For example, if $a = b = 1$, then $x_{t+1} = x_t + 1$, so $x_t = x_0 + t$. Or, if $a = 2$, $b = -1$, and $x_0 = 0$, then $x_1 = -1$, $x_2 = -3$, $x_3 = -7$, etc., diverging away from the equilibrium point $x^* = b/(1-a) = 1$. (If we start exactly at $x_0 = x^* = 1$, then all $x_t = x^* = 1$, but otherwise the sequence diverges.)

Figure 6.1 shows some examples of different types of behavior.

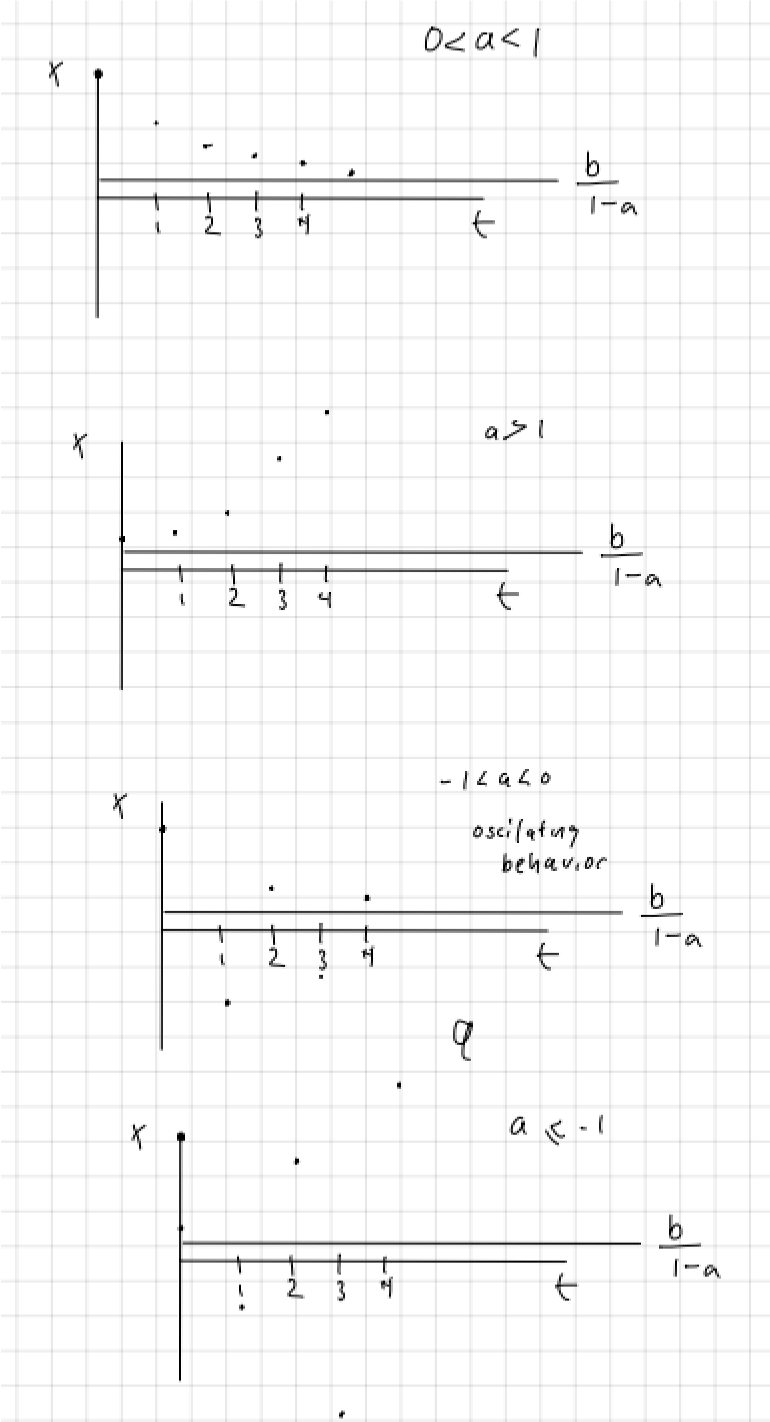



Figure 6.1: Behavior of different linear first-order difference equations.

**Extension to time-varying parameter**

The following difference equation is not time-homogeneous because the time-constant parameter $b$ is replaced by time-varying $b_t$:

$$x_{t+1} = ax_t + b_{t+1}.$$

This is a special case of the more general linear difference equation $x_{t+1} = a_{t+1}x_t + b_{t+1}$, which is a special case of the general nonlinear $x_{t+1} = f_{t+1}(x_t)$.

Despite time-varying $b_t$, this can be solved by the insertion method like in (6.3):

$$x_t = ax_{t-1} + b_t = a\overbrace{(ax_{t-2} + b_{t-1})}^{x_{t-1}} + b_t = \cdots = a^t x_0 + \sum_{k=1}^{t} a^{t-k} b_k.$$

### 6.2.2 Systems of difference equations

Instead of a scalar $x_t$, consider a vector $\boldsymbol{x}_t$ with $k$ elements. Thus, $\boldsymbol{b}$ must also be a vector of $k$ elements, and $\boldsymbol{A}$ is a $k$-by-$k$ coefficient matrix:

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{b}. \tag{6.6}$$

An equilibrium of (6.6) can be derived using the strategy of (6.4) but with matrix math. Substituting $\boldsymbol{x}^*$ for $\boldsymbol{x}_{t+1}$ and $\boldsymbol{x}_t$ in (6.6), with $\boldsymbol{I}_k$ the $k$-by-$k$ identity matrix,

$$\boldsymbol{x}^* = \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b} \implies (\boldsymbol{I}_k - \boldsymbol{A})\boldsymbol{x}^* = \boldsymbol{b} \implies \boldsymbol{x}^* = (\boldsymbol{I}_k - \boldsymbol{A})^{-1}\boldsymbol{b} \tag{6.7}$$

if the matrix inverse exists.

The invertibility of $\boldsymbol{I}_k - \boldsymbol{A}$ depends on the eigenvalues of $\boldsymbol{A}$. Generally, a matrix is invertible if and only if all of its eigenvalues are non-zero. Let $(\lambda_1, \ldots, \lambda_k)$ denote the eigenvalues of $\boldsymbol{A}$. The eigenvalues of the identity matrix $\boldsymbol{I}_k$ all equal one. Thus, the eigenvalues of $\boldsymbol{I}_k - \boldsymbol{A}$ are $1 - \lambda_i$ for $i \in \{1, \ldots, k\}$, so $\boldsymbol{I}_k - \boldsymbol{A}$ is invertible (and an equilibrium exists) if and only if $\lambda_i \neq 1$ for all $i \in \{1, \ldots, k\}$. Put differently, an eigenvalue of $\lambda_i = 1$ in the coefficient matrix $\boldsymbol{A}$ in (6.6) indicates non-stationary behavior, lacking any convergence over time. However, recall from Section 6.2.1 that a difference equation can be unstable and diverge even if it has an equilibrium.

Consider the system results above in the special case of a scalar from Section 6.2.1. The condition that matrix $\boldsymbol{A}$ does not have any eigenvalue $\lambda_i = 1$ simplifies to the condition that $a \neq 1$; that is, interpreting $a$ as a 1-by-1 matrix, its only eigenvalue is $\lambda_1 = a$. This matches the condition from (6.4) for an equilibrium to exist. However, we also saw that $a > 1$ generates an unstable difference equation that diverges unless $x_0 = x^* = b/(1 - a)$ exactly. Thus, additional conditions are required for stability of the system difference equation, too.

### 6.2.3 Higher-order difference equations

In contrast to the first-order difference equation in which $x_{t+1}$ depends only on $x_t$, a **higher-order** difference equation also depends on $x_{t-1}$ and/or other previous values in the sequence. For example, in a second-order difference equation, $x_{t+1}$ depends on $x_t$ and $x_{t-1}$; third-order means $x_{t+1}$ depends on $x_t$, $x_{t-1}$, and $x_{t-2}$; etc.

It is convenient to model a higher-order difference equation as a first-order system of difference equations. For example, consider the second-order linear difference equation

$$x_{t+1} = a_1 x_t + a_2 x_{t-1} + b, \tag{6.8}$$

where all variables are scalars. Define vector $\boldsymbol{y}_t \equiv (y_{1,t}, y_{2,t})' \equiv (x_t, x_{t-1})'$. Then,

$$\boldsymbol{y}_{t+1} = \begin{bmatrix} x_{t+1} \\ x_t \end{bmatrix} = \begin{bmatrix} a_1 x_t + a_2 x_{t-1} + b \\ x_t \end{bmatrix} = \begin{bmatrix} a_1 y_{1,t} + a_2 y_{2,t} + b \\ y_{1,t} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} + \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

Thus, the second-order difference equation in (6.8) is equivalent to the first-order difference equation system

$$\boldsymbol{y}_{t+1} = \boldsymbol{A}\boldsymbol{y}_t + \boldsymbol{b}, \quad \boldsymbol{A} = \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix}, \ \boldsymbol{b} = \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

### 6.2.4 Stability of nonlinear difference equations

Given interval $\mathcal{I} \subset \mathbb{R}$, consider the first-order nonlinear difference equation

$$x_{t+1} = f(x_t), \quad f\colon \mathcal{I} \mapsto \mathcal{I}.$$

Equilibrium $x^*$ solves $x^* = f(x^*)$.

Concepts of stability are defined as follows. First, $x^*$ is **locally asymptotically stable** if (and only if) there exists $\epsilon > 0$ such that $|x_0 - x^*| < \epsilon$ implies $\lim_{t\to\infty} x_t = x^*$. That is, if the initial value $x_0$ is close enough to the equilibrium point $x^*$ ("locally"), then as $t \to \infty$ ("asymptotically") $x_t$ converges to the equilibrium ("stable"). Second, $x^*$ is unstable if there exists $\epsilon > 0$ such that for every $x$ with $0 < |x - x^*| < \epsilon$, $|f(x) - x^*| > |x - x^*|$. That is, if $x_t$ gets close to the equilibrium $x^*$, then $x_{t+1} = f(x)$ actually gets farther from $x^*$.

Local stability is related to the derivative of $f(\cdot)$ as follows. Consider the equilbrium point $x^*$ such that $f(x^*) = x^*$. (There may be multiple; just consider one at a time.) Assume $f(\cdot)$ is once differentiable in a neighborhood of $x^*$ (a small open interval around $x^*$). Then,

$$|f'(x^*)| < 1 \implies x^* \text{ is locally asymptotically stable,} \tag{6.9}$$

$$|f'(x^*)| > 1 \implies x^* \text{ is unstable.} \tag{6.10}$$

Note that $x_{t+1} = f(x_t) = x_t$ has $f'(x) = 1$ for all $x$, so $f'(x^*) = 1$, which does not satisfy either derivative condition. It also does not satisfy either the stable or unstable

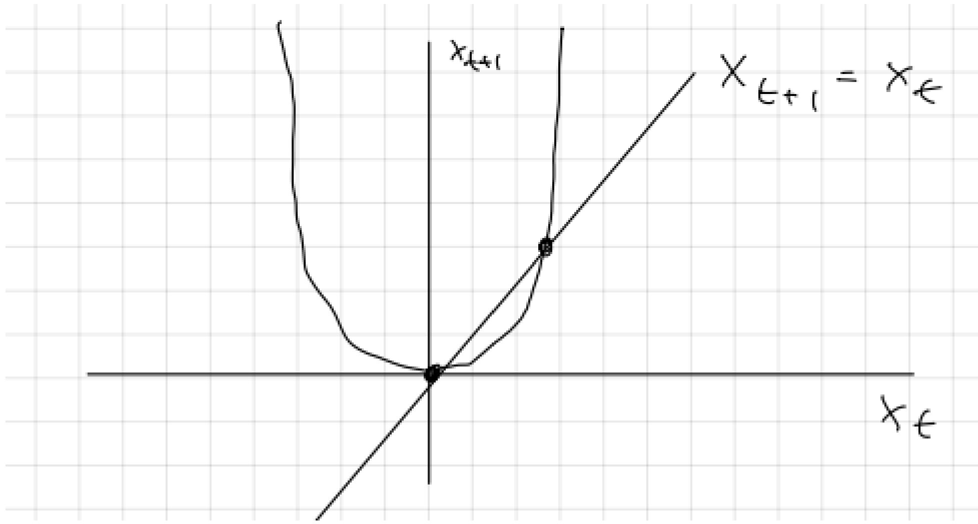



Figure 6.2: Nonlinear difference equation function in Example 6.1.

definitions. For example, if we consider equilibrium point $x^* = 5$ and $x_0 = 5 + \epsilon$, we get $x_t \to 5 + \epsilon \neq 5$ for any $\epsilon > 0$, violating local asymptotic stability. Further, because $f(x) = x$, then $|f(x) - x^*| = |x - x^*|$, violating the definition of unstable, too.

**Example 6.1.** Let $x_{t+1} = f(x_t) = x_t^2$. First, solve for the potential steady states: $x^* = (x^*)^2 \implies x^*(x^* - 1) = 0$, so $x^* \in \{0, 1\}$. Second, compute the derivative: $f(x) = x^2$, so $f'(x) = 2x$. Third, evaluate the conditions in (6.9) and (6.10). At $x^* = 0$, $|f'(0)| = 0$, satisfying (6.9) and thus local asymptotic stability. At $x^* = 1$, $|f'(1)| = 2$, satisfying (6.10), so $x^* = 1$ is not stable. ■

One particular type of instability is **cyclical behavior**, where the sequence comes back to the same value repeatedly, only to leave for different values again. That is, $x_{t+1} \neq x_t$, but given **period** $p$, $x_{t+p} = x_t$, and further $x_{t+cp} = x_t$ for any positive integer (natural number) $c \in \mathbb{N}$.

**Example 6.2.** Let $f(x) = 1 - x$ and $x_{t+1} = f(x_t) = 1 - x_t$. If $x_t = 1$, then $x_{t+1} = 1 - 1 = 0$, and $x_{t+2} = 1 - x_{t+1} = 1 - 0 = 1 = x_t$. Further, for any $c \in \mathbb{N}$, $x_{t+2c} = x_t$. ■

## 6.3 Implicit difference equations

Consider the **implicit difference equation**

$$g(x_t, x_{t+1}) = 0. \tag{6.11}$$

It is a "difference equation" because it characterizes the relationship between $x_{t+1}$ and $x_t$; it is "implicit" because it does not explicitly write $x_{t+1} = \dots$ as a function of $x_t$.

The previous results about asymptotic stability and instability apply directly if we consider (6.11) to define an implicit function $x_{t+1} = f_(x_t)$ in a neighborhood (small open interval) of the fixed point $x^*$ that solves $g(x^*, x^*) = 0$. We can use the implicit function theorem to calculate the derivatives needed to evaluate the conditions in (6.9) and (6.10):

$$f'(x^*) = \frac{-g_1(x^*, x^*)}{g_2(x^*, x^*)}, \text{ where } g_j(x^*, x^*) \equiv \left. \frac{\partial g(v_1, v_2)}{\partial v_j} \right|_{(v_1,v_2)=(x^*,x^*)}. \tag{6.12}$$

## Optional resources

Optional resources for this chapter

- Dartmouth linear difference equation notes and problems: https://web.archive.org/web/20240307160154/https://math.dartmouth.edu/opencalc2/dcsbook/c1pdf/sec14.pdf
- Dartmouth nonlinear difference equation notes and problems: https://web.archive.org/web/20240308013346/https://math.dartmouth.edu/opencalc2/dcsbook/c1pdf/sec15.pdf

## Exercises

**Exercise E6.1.** As in (6.6), let $\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{b}$, with $\boldsymbol{b} = (2, 3)'$ and $\boldsymbol{A} = \text{diag}(0.2, 0.8) = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.8 \end{bmatrix}$. Plugging in $\boldsymbol{x}_{t+1} = \boldsymbol{x}_t = \boldsymbol{x}^*$ like in (6.7), solve for a possible solution $\boldsymbol{x}^*$.

**Exercise E6.2.** Let $x_{t+1} = f(x_t) = x_t^3 - 3x_t^2 + 3x_t$.

a. Solve for all the potential steady-state $x^*$ values. (Hint: they are all integers; if you have trouble factoring, then just graph it on your computer or use Wolfram Alpha or something.)
b. For each of the three $x^*$ values, compute the first derivative $f'(x^*)$, and say whether or not each is stable.

**Exercise E6.3.** Redo Example 6.1 after framing the difference equation implicitly like in Section 6.3. That is, instead of $x_{t+1} = x_t^2$, let $0 = g(x_t, x_{t+1}) = x_{t+1} - x_t^2$.

a. Solve for the two $x^*$ points from $g(x^*, x^*) = 0$.
b. What is the general partial derivative wrt the first argument, $g_1(x_t, x_{t+1})$?
c. What is the general partial derivative wrt the second argument, $g_2(x_t, x_{t+1})$?
d. For each $x^*$: use (6.12) to compute $f'(x^*)$, and assess stability using (6.9) and (6.10)

# Chapter 7

# Discrete Time Dynamic Optimization

In economics, often we model agents (individuals, firms, governments) making decisions that affect outcomes in the current time period as well as choices available in future periods. For example, if I spend lots of money this year on consumption, it may increase my utility this year, but reduce the amount of money I can spend next year and in future years beyond that. A **discrete time** model has distinct time periods $t = 1, 2, 3, \ldots$ (like years, for example), rather than a continuous time model (where time is a continuum). As usual, **optimization** refers to maximization (of utility or profit) or minimization (of cost), by choosing the value of some variable, possibly subject to some constraints (like consumption being non-negative). **Dynamic** refers to the fact that the period $t$ optimization decision affects future periods (in contrast to a static optimization problem). Combining these elements gives us **discrete time dynamic optimization**.

For simplicity, we consider **non-stochastic** problems, where nothing is random (everything is deterministic).

*Unit learning objectives for this chapter*

7.1. Define vocabulary words (in **bold**) related to dynamic optimization, both mathematically and intuitively [TLO 1]

7.2. Solve dynamic optimization problems using multiple different strategies [TLO 2]

## 7.1 Basics

We use the following notation and definitions throughout.

The **state variable** $x_t$ contains information relevant for our decision at time period $t$, and our choice at time $t$ cannot change it (although our choice may affect future $x_{t+1}$. For example, my current wealth at the beginning of this year (period $t$) is important for my choice of how much to consume this year, but at this point I cannot go back in time to change how much wealth I started the year with. Scalar notation $x_t$ is used for simplicity, but more generally the state variable can be a vector.

The **control variable** $u_t$ is what we get to choose (control). There may be constraints on the available choices; the constraints may depend on the state variable $x_t$. For example, if $u_t$ is my consumption spending in year $t$, then I am constrained to $u_t \geq 0$. If $x_t$ is my starting wealth level and debt spending is not allowed, then also $u_t \leq x_t$. Generally, let $\mathcal{U}_t$ denote the set of possible choices at time $t$: $u_t \in \mathcal{U}_t$. As with $x_t$, scalar $u_t$ is used for simplicity.

The **law of motion** describes how the state variable evolves over time. Specifically,

$$x_{t+1} = g_t(x_t, u_t), \tag{7.1}$$

where in general the function $g_t$ may change over time and may depend on the current state ($x_t$) and choice ($u_t$) nonlinearly. (Alternative notation: $g(t, x_t, u_t)$.)

Similar to Chapter 5, the goal is to make choices to maximize an objective function, subject to constraints. Specifically, the goal is to

$$\begin{aligned} &\max_{\{u_t\}_{t=0}^{T}} \sum_{t=0}^{T} f_t(x_t, u_t) \\ &\text{s.t. } \underbrace{x_{t+1} = g_t(x_t, u_t)}_{\text{law of motion}}, \quad \underbrace{x_0 = \bar{x}_0}_{\text{initial condition}}, \quad \underbrace{x_{T+1} \geq \bar{x}_{T+1}}_{\text{terminal condition}}. \end{aligned} \tag{7.2}$$

The constraint $x_0 = \bar{x}_0$ is the **initial condition**, and $x_{T+1} \geq \bar{x}_{T+1}$ is the **terminal condition**, like ending with a non-negative amount of money. Sometimes the terminal condition is an equality, or there is no terminal condition. Note the solution is the full sequence of choices $(u_0, u_1, \ldots, u_T)$. Usually the choice $u_t$ depends on state $x_t$ and may be called a **policy**, **policy rule**, or **policy function** (the policy is a function of the state).

Like in (4.4) and (5.9), the **value function** is the maximum possible objective given the optimal choice. Here it is more complex because the value function can change over time, so we include subscript $s$ for the time period, and maximization is wrt $u_s$ and all future choices:

$$J_s(x_s) = \max_{u_s,\ldots,u_T} \sum_{t=s}^{T} f_t(x_t, u_t) \text{ s.t. law of motion and initial/terminal conditions.} \tag{7.3}$$

The dependence on $x_s$ is separate from the time period dependence of the function $J_s(\cdot)$ itself. For example, if $x$ is good (like capital), then $J_s(x_s + 1) > J_s(x_s)$: if we have more capital in time period $s$, then we can get more profit. This is different than say $J_1(5) > J_T(5)$: given five units of capital at time $t = 1$, we can get more profit through period $T$ than if we're given five units of capital at time $T$ and only have one period to earn profit.

## 7.2 Finite-horizon problems: backward induction

With a **finite-horizon** problem (finite $T$), we can start from the optimization problem in period $T$ and then work through $t = T-1, T-2, \ldots, 0$, a process called **backward induction**. To set this up, we write the value function in terms of the current period's objective plus next period's value function, and note that in period $T$ the future value is zero (because there is no future):

$$J_s(x_s) = \max_u f_s(x_s, u) + J_{s+1}\big(g_s(x_s, u)\big) \text{ for } s = 0, 1, \ldots, T-1, \tag{7.4}$$

$$J_T(x_T) = \max_u f_T(x_T, u) \text{ s.t. terminal condition on } x_{T+1} = g_T(x_T, u). \tag{7.5}$$

**Example 7.1.** Consider (7.2) with $T = 2$ time periods, objective $f_t(x, u) = f(x, u) = \log(u)$, law of motion $g_t(x, u) = g(x, u) = x - u$, initial condition $x_0 = \bar{x}$, and terminal condition $x_{T+1} = x_3 = 0$. To use backward induction, we first solve the final period problem, $\max_u f_T(x_T, u)$ s.t. $x_{T+1} = 0$, or more specifically $\max_u \log(u)$ s.t. $0 = x_3 = g(x_2, u) = x_2 - u$. The constraint alone determines $u_2^* = x_2$. The period-$T$ value function is thus $J_2(x_2) = \log(x_2)$. Now, plug $J_2$ into the $t = 1$ problem:

$$\max_u \ f_1(x_1, u) + J_2(g_1(x_1, u)) = \max_u \ \log(u) + \log(x_1 - u). \tag{7.6}$$

Solving the FOC, set zero equal to the derivative wrt $u$ (treating $x_1$ as given) evaluated at the optimum $u_1^*$:

$$0 = \frac{1}{u_1^*} - \frac{1}{x_1 - u_1^*} \implies u_1^* = x_1/2 \implies J_1(x_1) = \log(x_1/2) + \log(x_1/2). \tag{7.7}$$

Plugging $J_1$ into the $t = 0$ problem and taking the FOC, using $2\log((x_0 - u)/2) = 2\log(x_0 - u) - 2\log(2)$,

$$0 = \frac{1}{u_0^*} - \frac{2}{x_0 - u_0^*} \implies u_0^* = x_0/3 = \bar{x}_0/3, \tag{7.8}$$

and $J_0(x_0) = \log(x_0/3) + 2\log((x_0 - x_0/3)/2) = 3\log(x_0/3)$, with optimal policy $u_0^* = u_1^* = u_2^* = \bar{x}_0/3$. ■

## 7.3 Euler equations

An **Euler equation** is a first-order condition (necessary but not sufficient) in dynamic optimization that describes intertemporal tradeoffs (between periods $t$ and $t+1$) given optimal choices. Deriving Euler equations under uncertainty can help connect theory with data, either to test the theory or to estimate economic parameters, but we do not consider uncertainty here.

Assuming the law of motion allows us to solve for $u_t$ as a function of $(t, x_t, x_{t+1})$, we can rewrite (7.2) to derive the Euler equation. Specifically, if we can solve $g_t(x_t, u_t) = x_{t+1}$

for $u_t$, then we can rewrite the objective function in terms of $x_t$ and $x_{t+1}$: $f_t(x_t, u_t) = F_t(x_t, x_{t+1})$. The value function becomes

$$J_t(x_t) = \max_{x_{t+1}} F_t(x_t, x_{t+1}) + J_{t+1}(x_{t+1}). \tag{7.9}$$

That is, instead of choosing $u_t$ (which determines $x_{t+1}$), we can think of choosing $x_{t+1}$ (which determines $u_t$). The FOC of (7.9) wrt $x_{t+1}$ is

$$0 = \frac{\partial F_t}{\partial x_{t+1}} + \frac{\partial J_{t+1}}{\partial x_{t+1}}. \tag{7.10}$$

The Envelope Theorem wrt $x_t$ gives

$$\frac{\partial J_t}{\partial x_t} = \frac{\partial F_t}{\partial x_t} \implies \frac{\partial J_{t+1}}{\partial x_{t+1}} = \frac{\partial F_{t+1}}{\partial x_{t+1}}. \tag{7.11}$$

Combining these,

$$0 = \frac{\partial F_t}{\partial x_{t+1}} + \frac{\partial J_{t+1}}{\partial x_{t+1}} = \frac{\partial F_t}{\partial x_{t+1}} + \frac{\partial F_{t+1}}{\partial x_{t+1}}. \tag{7.12}$$

**Example 7.2.** Here is another example of an Euler equation. Let $a_{t+1} = g(a_t, u_t)$ be the law of motion, but now we cannot solve for $u_t$. For simplicity, let $f_t = f$ for all $t$. The value function is

$$J_t(a_t) = \max_{u_t, a_{t+1}} f(a_t, u_t) + J_{t+1}(a_{t+1}) \text{ s.t. } a_{t+1} = g(a_t, u_t) \tag{7.13}$$

The Lagrangian is

$$\mathcal{L}_t(a_t) = f(a_t, u_t) + J_{t+1}(a_{t+1}) + \lambda_t[a_{t+1} - g(a_t, u_t)]. \tag{7.14}$$

By the Envelope Theorem,

$$\frac{\partial J_t}{\partial a_t}(a_t^*) = \frac{\partial f}{\partial a_t}(a_t^*) - \lambda_t \frac{\partial g}{\partial a_t}(a_t^*) \implies \frac{\partial J_{t+1}}{\partial a_{t+1}}(a_{t+1}^*) = \frac{\partial f}{\partial a_{t+1}}(a_{t+1}^*) - \lambda_{t+1} \frac{\partial g}{\partial a_{t+1}}(a_{t+1}^*). \tag{7.15}$$

The first-order conditions (wrt $u_t$ and $a_{t+1}$) are

$$0 = \frac{\partial f}{\partial u_t} - \lambda_t \frac{\partial g}{\partial u_t}, \quad 0 = \frac{\partial J_{t+1}}{\partial a_{t+1}} + \lambda_t = 0. \tag{7.16}$$

These can be solved for one Euler equation in terms of the given functions. ■

## 7.4 Infinite-horizon problems

Now consider a variation of (7.2) with $T = \infty$:

$$\max_{\{u_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t f(x_t, u_t) \tag{7.17}$$
$$\text{s.t. } \underbrace{x_{t+1} = g_t(x_t, u_t)}_{\text{law of motion}}, \quad \underbrace{x_0 = \bar{x}_0}_{\text{initial condition}},$$

and possibly a condition on the limiting behavior of $x_t$ as $t \to \infty$. Here $\beta$ is a **discount factor** that says we value our utility next period less than this period (and so on), with $0 < \beta < 1$.

The following subsections describe different approaches to solving this infinite-horizon dynamic optimization problem.

### 7.4.1 Bellman equation

The **Bellman equation** is essentially the value function in (7.4) applied to our special case of (7.17):

$$J(x) = \max_u f(x, u) + \beta J(g(x, u)). \tag{7.18}$$

Similar to Section 7.3, we can try to solve this by combining the first-order condition and the Envelope Theorem, as in Example 7.3.

**Example 7.3.** We will solve

$$\max \sum_{t=0}^{\infty} \beta^t \log c_t \text{ s.t. } x_{t+1} = x_t - c_t, \tag{7.19}$$

given initial value $x_0$. You can imagine start with $x_0$ amount of some consumption good, and deciding how much to consume in each period, and how much to leave for next period ($x_{t+1} = x_t - c_t$: the amount available next period is how much you start with minus how much you consume).

We can solve this using the Bellman equation,

$$J(x) = \max_c \log(c) + \beta J(x - c). \tag{7.20}$$

The FOC is

$$0 = \frac{\partial J}{\partial c}(c^*) = (1/c^*) - \beta J'(x - c^*) \implies \beta c_t^* J'(x_{t+1}^*) = 1. \tag{7.21}$$

Similar to Section 7.3, we can rewrite the Bellman equation in terms of choosing next period's $x$ instead of this period's $c$:

$$J(x_t) = \max_{x_{t+1}} \log(x_t - x_{t+1}) + \beta J(x_{t+1}). \tag{7.22}$$

Applying the Envelope Theorem to this formulation,

$$J'(x_t) = \frac{1}{x_t - x_{t+1}^*} = \frac{1}{c_t^*} \implies J'(x_{t+1}) = \frac{1}{c_{t+1}^*}. \tag{7.23}$$

Plugging this into (7.21),

$$\beta c_t^*(1/c_{t+1}^*) = 1 \implies c_{t+1}^* = \beta c_t^*, \tag{7.24}$$

our consumption Euler equation. (Note this is specific to our choice of a log utility function.) Applying the Euler equation starting at $c_0^*$,

$$c_1^* = \beta c_0^*, \quad c_2^* = \beta c_1^* = \beta^2 c_0^*, \quad \ldots, \quad c_t^* = \beta^t c_0^*. \tag{7.25}$$

This determines the full path of $\{c_t^*\}_{t=1}^{\infty}$ as a function of $c_0^*$, but it does not determine $c_0^*$. Here, given that the objective function $\log(\cdot)$ is strictly increasing, we know we should eventually consume all of the good: $\lim_{t\to\infty} x_t = 0$. Combining this with the Euler equation, we can solve for $c_0^*$:

$$x_t^* = x_0 - \sum_{k=0}^{t-1} c_k^* = x_0 - \sum_{k=0}^{t-1} \beta^k c_0^*,$$
$$0 = \lim_{t\to\infty} x_t = x_0 - c_0^* \sum_{k=0}^{\infty} \beta^k = x_0 - c_0^* \frac{1}{1-\beta} \implies c_0^* = (1-\beta)x_0,$$

which combined with the Euler equation implies $c_t^* = \beta^t(1-\beta)x_0$.

We can also solve for $x_t^*$ as a function of $x_0$, and for $c_t^*$ as a function of $x_t^*$. First, plugging in for $c_0^*$ in the expression for $x_t^*$, and dividing both sides by $x_0$ (to simplify the RHS),

$$\begin{aligned} x_t^*/x_0 &= 1 - \sum_{k=0}^{t-1} \beta^k (1-\beta) \\ &= 1 - \sum_{k=0}^{t-1} \beta^k + \beta \sum_{k=0}^{t-1} \beta^k \\ &= 1 - 1 - \sum_{k=1}^{t-1} \beta^k + \sum_{k=1}^{t-1} \beta^k + \beta^t = \beta^t, \end{aligned}$$

so $x_t^* = \beta^t x_0$. Given $c_t^* = \beta^t(1-\beta)x_0$ from before, this implies $c_t^* = (1-\beta)\beta^t x_0 = (1-\beta)x_t^*$: the optimal policy is to consume fraction $1-\beta$ of the available good. ∎

### 7.4.2 Guess and verify

In the alternative approach of **guess and verify**, you

1. guess the functional form of the value function,
2. plug the guess into the Bellman equation,
3. solve the maximization problem, and
4. verify that the (solved) value function has the form that we guessed.

We use guess and verify to solve Example 7.3. Recall the Bellman equation in (7.20) was $J(x) = \max_c \log(c) + \beta J(x-c)$. This time, let $c = \theta x$: we choose $\theta$, the proportion of the available good $x$ that we consume, which leaves $(1-\theta)x$ for the next period. The corresponding Bellman equation is

$$J(x) = \max_\theta \log(\theta x) + \beta J((1-\theta)x). \tag{7.26}$$

Our guess is that the value function has the functional form

$$J(x) = \psi + \gamma \log(x) \tag{7.27}$$

for some (currently unknown) constants $\psi$ and $\gamma$ that do not include $x$. We then plug this into the right-hand side of (7.26) and solve the FOC wrt $\theta$:

$$\begin{aligned}
J(x) &= \max_\theta \ \log(\theta x) + \beta \overbrace{[\psi + \gamma\log((1-\theta)x)]}^{J((1-\theta)x)} \\
&= \max_\theta \ \log(\theta) + \beta\gamma\log(1-\theta) + \overbrace{\log(x) + \beta\psi + \beta\gamma\log(x)}^{\text{no } \theta}, && (7.28)\\
0 &= \frac{\partial}{\partial\theta}[\log(\theta) + \beta\gamma\log(1-\theta) + \log(x) + \beta\psi + \beta\gamma\log(x)]\big|_{\theta=\theta^*} = \frac{1}{\theta^*} - \frac{\beta\gamma}{1-\theta^*}, && (7.29)\\
\theta^* &= \frac{1}{1+\beta\gamma}.
\end{aligned}$$

Plugging this into the value function,

$$\begin{aligned}
J(x) &= \log(\theta^*) + \beta\gamma\log(1-\theta^*) + \log(x) + \beta\psi + \beta\gamma\log(x) \\
&= -\log(1+\beta\gamma) + \beta\gamma\log\left(\frac{\beta\gamma}{1+\beta\gamma}\right) + \beta\psi + \log(x)(1+\beta\gamma) \\
&= \overbrace{\beta\psi - \log(1+\beta\gamma) + \beta\gamma\log\left(\frac{\beta\gamma}{1+\beta\gamma}\right)}^{\psi} + \overbrace{(1+\beta\gamma)}^{\gamma}\log(x).
\end{aligned}$$

First, note this matches the structure that we initially guessed. (As will be show, neither the $\gamma$ nor $\psi$ have $x$ inside.) Second, this lets us solve for $\psi$ and $\gamma$:

$$\begin{aligned}
&\gamma = 1 + \beta\gamma \implies \gamma = 1/(1-\beta), \\
&\psi = \beta\psi - \log(1+\beta\gamma) + \beta\gamma\log\left(\frac{\beta\gamma}{1+\beta\gamma}\right) \implies \psi = \frac{\beta\gamma\log(\beta\gamma) - (1+\beta\gamma)\log(1+\beta\gamma)}{1-\beta},
\end{aligned}$$

and we could further plug in for $\gamma$ to solve for $\psi$ in terms of only $\beta$.

### 7.4.3 Value function iteration

Yet another approach is to solve the Bellman equation with iterative consecutive "guesses," a procedure called **value function iteration**. With more iterations, assuming the problem is well-defined, the value function guesses should converge to the true value function. As you'll see below, doing this by hand is as terribly tedious as it sounds. However, computers can be used (after discretizing the state space), although writing code is beyond our scope.

As in Example 7.3, consider our simple $J(x) = \max_{c\in[0,x)} \log c + \beta J(x-c)$, but let $J^i$ denote the $i$th iteration guess of the value function. Convergence here means $J^i(x) = J^{i+1}(x)$ for all $x$. (In practice, it may suffice that $|J^i(x) - J^{i+1}(x)| < \delta$ for some small tolerance $\delta > 0$.)

The remainder of this subsection shows a few iterations and then the convergence (as $i \to \infty$) to the same solution from Example 7.3.

Our initial guess is simple and clearly wrong: $J^0(x) = 0$ for all $x$.

In the first iteration, we solve for $J^1$ using $J^0$:

$$J^1(x) = \max_{c\in[0,x)} \log c + \beta \overbrace{J^0(x-c)}^{=0} \implies c^* = x \implies J^1(x) = \log(x). \tag{7.30}$$

The solution $c^* = x$ is not from an FOC, but just taking the maximum possible value to maximize $\log c$ over $c \in [0, x)$. (Yes, technically $x$ is excluded from $[0, x)$, but it'll be ok.)

In the second iteration, we solve for $J^2$ using $J^1$:

$$J^2(x) = \max_{c\in[0,x)} \log c + \beta \overbrace{J^1(x-c)}^{\log(x-c)}, \tag{7.31}$$

$$\text{FOC:} \quad 0 = \frac{1}{c^*} - \frac{\beta}{x-c^*} \implies c^* = x/(1+\beta), \tag{7.32}$$

$$\begin{aligned}\implies J^2(x) &= \log(x/(1+\beta)) + \beta\log(\beta x/(1+\beta)) \\ &= (1+\beta)\log(x) + \beta\log(\beta) - (1+\beta)\log(1+\beta). \end{aligned}\tag{7.33}$$

In the third iteration,

$$J^3(x) = \max_{c\in[0,x)} \log c + \beta \overbrace{[(1+\beta)\log(x-c) + \beta\log(\beta) - (1+\beta)\log(1+\beta)]}^{J^2(x-c)}, \tag{7.34}$$

$$0 = \frac{1}{c^*} - \frac{\beta(1+\beta)}{x-c^*} \implies c^* = x/(1+\beta+\beta^2), \tag{7.35}$$

$$\begin{aligned}
J^3(x) &= \log\left(\frac{x}{1+\beta+\beta^2}\right) \\
&\quad + \beta\left[(1+\beta)\log\left(\frac{x(\beta+\beta^2)}{1+\beta+\beta^2}\right) + \beta\log(\beta) - (1+\beta)\log(1+\beta)\right] \\
&= (1+\beta+\beta^2)\log(x) + \beta^2\log(\beta) - \beta(1+\beta)\log(1+\beta) \\
&\quad + \beta(1+\beta)\log(\underbrace{\beta+\beta^2}_{\beta(1+\beta)}) - (1+\beta+\beta^2)\log(1+\beta+\beta^2) \\
&= (1+\beta+\beta^2)\log(x) + (\beta+2\beta^2)\log(\beta) - (1+\beta+\beta^2)\log(1+\beta+\beta^2).
\end{aligned} \tag{7.36}$$

For general iteration $i+1$, by induction it can be shown that

$$J^{i+1}(x) = \sum_{h=0}^{i}\beta^h\log(x) + \sum_{h=1}^{i}h\beta^h\log(\beta) + \left(\sum_{h=0}^{i}\beta^h\right)\log\left(\sum_{h=0}^{i}\beta^h\right), \tag{7.37}$$

$$c^{*i} = \frac{x}{\sum_{h=0}^{i}\beta^h}. \tag{7.38}$$

Given the general iteration's value function, we can find the limit to which it converges. Recall that for any $0<\beta<1$, $\sum_{h=0}^{\infty}\beta^h = 1/(1-\beta)$; also, $\sum_{h=1}^{\infty}h\beta^h = \beta/(1-\beta)^2$.[1] Thus,

$$\begin{aligned}
\lim_{i\to\infty} J^{i+1}(x) &= \sum_{h=0}^{\infty}\beta^h\log(x) + \sum_{h=1}^{\infty}h\beta^h\log(\beta) + \left(\sum_{h=0}^{\infty}\beta^h\right)\log\left(\sum_{h=0}^{\infty}\beta^h\right) \\
&= \frac{1}{1-\beta}\log(x) + \frac{\beta}{(1-\beta)^2}\log(\beta) + \frac{1}{1-\beta}\log\left(\frac{1}{1-\beta}\right).
\end{aligned}$$

That is, the value function converges. Additionally, $c^* = \lim_{i\to\infty} c^{*i} = x/[1/(1-\beta)] = (1-\beta)x$, just as in Example 7.3.

The remainder of this subsection shows the induction used to derive $J^{i+1}(x)$ above. From the initial $i=0,1,2,3$ iterations, our starting point is an iteration $i$ with

$$c^{*i} = \frac{x}{\sum_{h=0}^{i-1}\beta^h}, \qquad J^i(x) = \log(x)\sum_{h=0}^{i-1}\beta^h + w_i, \tag{7.39}$$

[1] https://www.wolframalpha.com/input?i=sum+of+h*b%5Eh+from+h%3D1+to+infinity

where $w_i$ depends on $\beta$ but not $c$ (or $x$), so $\frac{\partial w_i}{\partial c} = 0$. For iteration $i+1$,

$$J^{i+1}(x) = \max_c \log(c) + \beta J^i(x) = \max_c \log(c) + \beta \log(x-c) \sum_{h=0}^{i-1} \beta^h + \beta w_i. \tag{7.40}$$

The first-order condition is

$$0 = (1/c^*) - \frac{\beta}{x - c^*} \sum_{h=0}^{i-1} \beta^h \implies c^* + c^* \beta \sum_{h=0}^{i-1} \beta^h = x \implies c^* = \frac{x}{\sum_{h=0}^{i} \beta^h}.$$

Plugging this in, the value function is

$$\begin{aligned}
J^{i+1}(x) &= \log(c^*) + \beta \log(x - c^*) \sum_{h=0}^{i-1} \beta^h + \beta w_i \\
&= \log\left(\frac{x}{\sum_{h=0}^{i} \beta^h}\right) + \log\left(x - \frac{x}{\sum_{h=0}^{i} \beta^h}\right) \sum_{h=1}^{i} \beta^h + \beta w_i \\
&= \log(x) - \log\left(\sum_{h=0}^{i} \beta^h\right) + \left[\log(x) + \log\left(\sum_{h=1}^{i} \beta^h\right) - \log\left(\sum_{h=0}^{i} \beta^h\right)\right] \sum_{h=1}^{i} \beta^h + \beta w_i \\
&= \log(x) \sum_{h=0}^{i} \beta^h + w_{i+1},
\end{aligned}$$

where $w_{i+1}$ is only a function of $\beta$ (and includes $w_i$). In principle, we could solve explicitly for the $w_i$ in terms of $\beta$, but it would be of limited value.

### 7.4.4 Policy function iteration

Instead of generating iteratively better guesses of the value function, **policy function iteration** generates iteratively better guesses of the **policy function** that describes the choice variable(s) as a function of the state variable(s). In our running example, the state variable is $x$, the choice variable is $c$, and the policy function is $c(x)$. In iteration $i$, the guessed policy function is $c^i$, and there is convergence if $c^i(x) = c^{i+1}(x)$ for all $x$.

Before making our first guess, we derive the value function corresponding to the policy $c(x) = \gamma x$. First, we derive the sequence $x_t$. Given $x_0$, $c_0 = c(x_0) = \gamma x_0$, so $x_1 = x_0 - c_0 = x_0(1-\gamma)$. Similarly, $c_1 = \gamma x_1 = \gamma(1-\gamma)x_0$, leaving $x_2 = x_1 - c_1 = x_1(1-\gamma) = x_0(1-\gamma)^2$. Inductively, given any $x_t = x_0(1-\gamma)^t$, the policy function gives $c_t = c(x_t) = \gamma x_t$, so $x_{t+1} = x_t - c_t = x_t(1-\gamma) = x_0(1-\gamma)^{t+1}$. Also, $c_t = \gamma x_t = \gamma(1-\gamma)^t x_0$. Plugging this into the maximization problem (from the perspective of $t = 0$), and using $\sum_{t=0}^{\infty} \beta^t = 1/(1-\beta)$

and $\sum_{t=0}^{\infty} t\beta^t = \beta/(1-\beta)^2$,

$$\begin{aligned}
J(x_0) &= \sum_{t=0}^{\infty} \beta^t \log(c_t) \\
&= \sum_{t=0}^{\infty} \beta^t \log(\gamma(1-\gamma)^t x_0) \\
&= \sum_{t=0}^{\infty} \beta^t \log(\gamma) + \sum_{t=0}^{\infty} \beta^t \overbrace{\log((1-\gamma)^t)}^{=t\log(1-\gamma)} + \sum_{t=0}^{\infty} \beta^t \log(x_0) \\
&= \frac{\log(x_0)}{1-\beta} + \frac{\log(\gamma)}{1-\beta} + \log(1-\gamma) \overbrace{\sum_{t=0}^{\infty} \beta^t t}^{=\beta/(1-\beta)^2} .
\end{aligned}$$

Our initial guess is simple but wrong: $c^0(x) = x/2$. Given this policy function, we can solve for the corresponding value function $J^0(x)$. Using the above with $\gamma = 0.5$,

$$J^0(x) = \frac{\log(x)}{1-\beta} + \overbrace{\frac{\log(0.5)}{1-\beta} + \frac{\log(1-0.5)\beta}{(1-\beta)^2}}^{\equiv d^0 \text{ (constant)}} . \tag{7.41}$$

The first iteration derives $c^1$ using $J^0$ and the FOC:

$$\begin{aligned}
&\max_c \log c + \beta J^0(x) = \max_c \log c + \frac{\beta}{1-\beta} \log(x-c) + d^0 \\
&\implies 0 = \frac{\partial}{\partial c} = \frac{1}{c^1(x)} - \frac{\beta}{(1-\beta)(x-c^1(x))} \implies c^1(x) = (1-\beta)x
\end{aligned}$$

is the policy function in iteration $i = 1$. Note this is of the same general form $\gamma x$, but now $\gamma = 1-\beta$ instead of $\gamma = 0.5$. But, this means we can again use our general value function for policy $\gamma x$:

$$J^1(x) = \frac{\log(x)}{1-\beta} + \overbrace{\frac{\log(1-\beta)}{1-\beta} + \frac{\log(1-(1-\beta))\beta}{(1-\beta)^2}}^{\equiv d^1 \text{ (constant)}} .$$

The second iteration derives $c^2$ using $J^1$ and the FOC:

$$\begin{aligned}
&\max_c \log c + \beta J^1(x) = \max_c \log c + \frac{\beta}{1-\beta} \log(x-c) + d^1 \\
&\implies 0 = \frac{\partial}{\partial c} = \frac{1}{c^2(x)} - \frac{\beta}{(1-\beta)(x-c^2(x))} \implies c^2(x) = (1-\beta)x
\end{aligned}$$

is the policy function in iteration $i = 2$. We have converged: $c^2 = c^1$. And, this should seem familiar as the optimal policy from solving this problem using the other approaches.

## Optional resources

Optional resources for this chapter

- Paul Bergin's lecture notes on dynamic programming: https://web.archive.org/web/20240313023135/https://faculty.econ.ucdavis.edu/faculty/bergin/ECON260D/dynprog.pdf
- Euler equations: https://web.archive.org/web/20240426015627/https://mitsloan.mit.edu/shared/ods/documents?DocumentID=4171
- Dynamic programming on Wikipedia
- Bellman equation on Wikipedia

## Exercises

**Exercise E7.1.** Solve the following by backward induction in terms of initial value $x_0$: $\max_{c_0,c_1,c_2} \sum_{t=0}^{2} \beta^t c_t$ such that $x_{t+1} = 1 + x_t - c_t$ and $x_3 \geq 0$, where $0 < \beta < 1$ is a discount factor. (Hint: the FOC won't help, but it should be simple enough to figure out directly.)

**Exercise E7.2.** Redo Example 7.3 with a different utility function, so the maximization problem is $\max \sum_{t=0}^{\infty} \beta^t \sqrt{c_t}$ subject to $x_{t+1} = x_t - c_t$ (and $c_t \geq 0$), given initial value $x_0 > 0$.

# Chapter 8

# Set Properties: Metrics and Topology

In this chapter, we will generalize and formalize our Euclidean intuition about properties of sets. Despite the title "topology," we focus on definitions based on a notion of distance, so these apply more to "metric spaces" than "topological spaces." However, we can still discuss the topology of a metric space.[1]

*Unit learning objectives for this chapter*

8.1. Define vocabulary words (in **bold**) related to topology of a vector space, both mathematically and intuitively [TLO 1]

8.2. Characterize the properties of particular points, sets, and functions [TLO 3]

## 8.1 Distance

Studying sets often requires a notion of distance to quantify how far apart two points are. That is, "distance" is a function that maps every pair of points to a non-negative real number. Such a function has many possible names, like **metric**, distance function, distance metric, or distance measure. We say "distance metric" as a reminder of what it measures ("metric" more generally just means "measurement") while helping you get used to the word "metric."

For intuition, think about different possible distance metrics in $\mathbb{R}^n$. We are familiar with Euclidean distance: for $x, y \in \mathbb{R}^n$,

$$d_2(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{8.1}$$

[1]Like https://en.wikipedia.org/wiki/Metric_space#The_topology_of_a_metric_space

But there are other possibilities (which are equivalent in terms of defining other properties of sets/shapes in $\mathbb{R}^n$). For example,

$$d_1(x, y) = \sum_{i=1}^{n} |x_i - y_i|, \quad d_\infty(x, y) = \max_i |x_i - y_i|. \tag{8.2}$$

These three distance metrics are also known respectively as the $L^2$ (or Euclidean), $L^1$ (or taxicab), and $L^\infty$ (or Chebyshev) distance. In terms of "topology," these three are actually equivalent, even though they clearly differ quantitatively.

There are some requirements for a function $d$ to be a valid distance metric. (You can think of Euclidean space for intuition, but distance can also be defined on other spaces, like spaces of functions; the combination of a set and a metric is called a metric space, but beyond our scope.)

**Definition 8.1.** Given set $S$, a valid distance metric $d \colon S \times S \to \mathbb{R}_{\geq 0}$ satisfies the following properties for any points $x, y, z \in S$.

- **Symmetry**: $d(x, y) = d(y, x)$
- **Non-negativity**: $d(x, y) \geq 0$
- **Identity of indiscernibles**: $d(x, y) = 0$ if and only if $x = y$
- **Triangle inequality** (or "subaddivitiy"): $d(x, y) \leq d(x, z) + d(z, y)$

(Beyond our scope, there are other almost-metrics: a semimetric does not require the triangle inequality, a pseudometric does not require the identity of indiscernibles, a divergence does not require symmetry or triangle inequality, etc.)

## 8.2 Set properties

Given a distance metric, various properties of sets and points can be defined.

### 8.2.1 Preliminaries

Before describing open, closed, and compact sets, we define the following useful terms, which use a distance metric.

**Definition 8.2.** Given set $\Omega$ (like $\mathbb{R}^n$), an **open ball** of radius $r \in \mathbb{R}$ around point $p \in \Omega$ is the set of points whose distance from $p$ is strictly less than $r$: $B_r(p) = \{x \in \Omega \mid d(p, x) < r\}$.

**Definition 8.3.** With respect to set $S$ (like $S \subseteq \mathbb{R}^n$), point $p$ is an **interior point** iff there exists $r > 0$ such that $B_r(p) \subseteq S$.

**Definition 8.4.** Denoted $\operatorname{int} S$ (or $S^{\circ}$), the **interior** of set $S$ is the set of all interior points of $S$.

**Definition 8.5.** With respect to set $S$ (like $S \subseteq \mathbb{R}^n$), point $p$ is an **exterior point** iff there exists $r > 0$ such that $B_r(p) \cap S = \emptyset$.

**Definition 8.6.** With respect to set $S \subseteq \Omega$ (like $S \subseteq \mathbb{R}^n$), point $p$ is a **boundary point** iff for all $r > 0$, $B_r(p)$ includes points from both $S$ and its complement $S^\complement$. (Note this does not require $p \in S$.)

**Definition 8.7.** Denoted $\partial S$, the **boundary** of set $S$ is the set of all boundary points of $S$.

**Definition 8.8.** Denote $\bar{S}$, the **closure** of set $S$ is the union of its interior and boundary: $\bar{S} = \text{int}(S) \cup \partial S$.

### 8.2.2 Open and closed sets

Now we can define open sets and closed sets, and their properties.

**Definition 8.9.** An **open set** $S$ contains only interior points ($S = \text{int}\, S$).

**Proposition 8.1** (open set properties)**.** *The sets $\emptyset$ and $\mathbb{R}^n$ are open. The union of any number of (including infinite) open sets is open. The intersection of finitely many open sets is open.*

*Proof.* Special cases: $\text{int}\,\emptyset = \emptyset$ (it has no points, so it has no interior points), so it is open. For any $x \in \mathbb{R}^n$ and any $r > 0$, $B_r(x) \subset \mathbb{R}^n$, so it contains only interior points and is an open set.

Union: let $A = \bigcup A_i$ with each $A_i$ open. Then for any $x \in A$, there exists $i$ such that $x \in A_i$, and because $A_i$ is open, then there exists ball $B_r(x) \subset A_i$, which implies $B_r(x) \subset A$ because $A_i \subseteq A$.

Intersection: let $A = \bigcap_{i=1}^n A_i$ with each $A_i$ open. For any $x \in A$, $x \in A_i$ for all $i$, which implies that for all $i = 1, \ldots, n$, there exists $r_i > 0$ such that $d_{r_i}(x) \subset A_i$ because $A_i$ is open. Because there are finitely many $r_i$, the minimum $\min_i r_i$ is well defined (which is not true for an infinite number), and the ball with such radius is thus contained in all $A_i$: $d_{\min_i r_i}(x) \subset A_i$ for all $i = 1, \ldots, n$. Consequently, the ball is also contained in $A$, satisfying the definition. $\square$

The "finitely many" is binding, as seen in the following example. In $\mathbb{R}$, define the sequence of open sets (open intervals) $A_i = (-1/i, 1 + 1/i)$ for $i = 1, 2, \ldots$. Then, $\bigcap_{i=1}^\infty A_i = [0, 1]$, the closed unit interval. (But for any finite $n < \infty$, $\bigcap_{i=1}^n A_i$ is open.)

**Definition 8.10.** Set $S \subseteq \Omega$ (like $S \subseteq \mathbb{R}^n$) is **closed** iff its complement $S^\complement$ is open. (Equivalently: $S$ contains all of its boundary points.) (Also equivalent: $S$ equals its closure, $S = \bar{S}$.)

**Proposition 8.2** (closed set properties)**.** *The sets $\emptyset$ and $\mathbb{R}^n$ are closed. The union of finitely many closed sets is closed. The intersection of any number of (including infinite) closed sets is closed.*

*Proof.* Special cases: these often seem surprising given that they are both open sets, and in English "open" and "closed" are mutually exclusive opposites. However, they can be understood when looking at the definition of a closed set. In $\mathbb{R}^n$, $\emptyset^{\complement} = \mathbb{R}^n$ and $(\mathbb{R}^n)^{\complement} = \emptyset$, which are both open by Proposition 8.1, so both are closed. (For the second equivalent definition: note that neither $\mathbb{R}^n$ nor $\emptyset$ has any boundary point, so they trivially contain all of their boundary points.)

Union/intersection: the complement of a union is the intersection of the complements, and the complement of an intersection is the union of the complements, so the properties derive from the open set properties in Proposition 8.1. That is, if $A = \bigcup A_i$, then $A^{\complement} = \bigcap A_i^{\complement}$; and if $A = \bigcap A_i$, then $A^{\complement} = \bigcup A_i^{\complement}$; and $A$ is closed iff $A^{\complement}$ is open. $\square$

### 8.2.3 Compact sets

The fully general definition of compactness is less intuitive, so we present the definition specific to Euclidean space $\mathbb{R}^n$. To do so requires one more definition first.

**Definition 8.11.** A set $S \subseteq \mathbb{R}^n$ is **bounded** iff there exists $M \in \mathbb{R}$ such that $d(x, y) < M$ for all $x, y \in S$.

**Definition 8.12.** A set $S \subseteq \mathbb{R}^n$ is **compact** iff it is closed and bounded (Definitions 8.10 and 8.11).

We are often interested in **convergence**. In $\mathbb{R}$, convergence of sequence $x_n$ to limit $x$ (i.e., $\lim_{n\to\infty} x_n = x$) is equivalent to the following condition: for any $\epsilon > 0$, there exists $N$ such that $|x_n - x| < \epsilon$ for all $n \geq N$. In $\mathbb{R}^d$, it's the same but with the absolute value $|x_n - x|$ replaced by $d(x_n, x)$. In $\mathbb{R}^d$, this is equivalent to the convergence of each element of $x_n$ to the corresponding element of $x$. In more general metric spaces, however, element-wise convergence is not necessarily equivalent to convergence of the vector.

**Definition 8.13.** A **subsequence** of sequence $\{x_i\}_{i=1}^{\infty}$ can be written as $\{x_{k_i}\}_{i=1}^{\infty}$ for a strictly increasing sequence of natural numbers $\{k_i\}_{i=1}^{\infty}$.

**Theorem 8.3** (Bolzano–Weierstrass)**.** *A subset $S \subseteq \mathbb{R}^n$ is compact if and only if every infinite sequence of points in $S$ has a convergent subsequence whose limit is also a point in $S$; that is, for every sequence $\{x_i\}_{i=1}^{\infty}$ with all $x_i \in S$, there exists strictly increasing sequence of natural numbers $\{k_i\}_{i=1}^{\infty}$ such that $\lim_{i\to\infty} x_{k_i} = x \in S$.*

## 8.3 Function properties: continuity

Here we define continuity of multivariate functions and then provide a result on continuity of a value function.

### 8.3.1 Multivariate function continuity

For $f\colon S \to \mathbb{R}$, we extend the definition of limit for $S \subseteq \mathbb{R}$ (Definition 1.18) to $S \subseteq \mathbb{R}^n$.

**Definition 8.14.** Given $f\colon S \to \mathbb{R}$ (for some $S \subseteq \mathbb{R}^n$), the **limit** $\lim_{x\to x_0} f(x) = y_0$ means that for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in S$, $0 < d(x, x_0) < \delta$ implies $|f(x) - y_0| < \epsilon$.

The definition of continuity remains identical to Definition 1.19: $f$ is continuous at $x_0$ if $\lim_{x\to x_0} f(x) = f(x_0)$.

**Example 8.1.** For $x \in \mathbb{R}^2$, using the Euclidean distance metric $d((x_1, x_2), (y_1, y_2)) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$, we show that $f(x) = x_1^2 + x_2^2$ is continuous at point $z = (0, 0)$, where $f(z) = 0$. Note $f(x) \geq 0$ for all $x \in \mathbb{R}^2$, so $|f(x) - 0| = f(x)$, and the condition $|f(x) - f(z)| < \epsilon$ simplifies to $f(x) < \epsilon$, or $x_1^2 + x_2^2 < \epsilon$. This is equivalent to $\sqrt{x_1^2 + x_2^2} < \sqrt{\epsilon}$, which is equivalent to $d(x, z) < \sqrt{\epsilon}$. That is, for any $\epsilon > 0$, choosing $\delta = \sqrt{\epsilon}$ ensures $|f(x) - f(z)| < \epsilon$ for all $d(x, z) < \delta$. ■

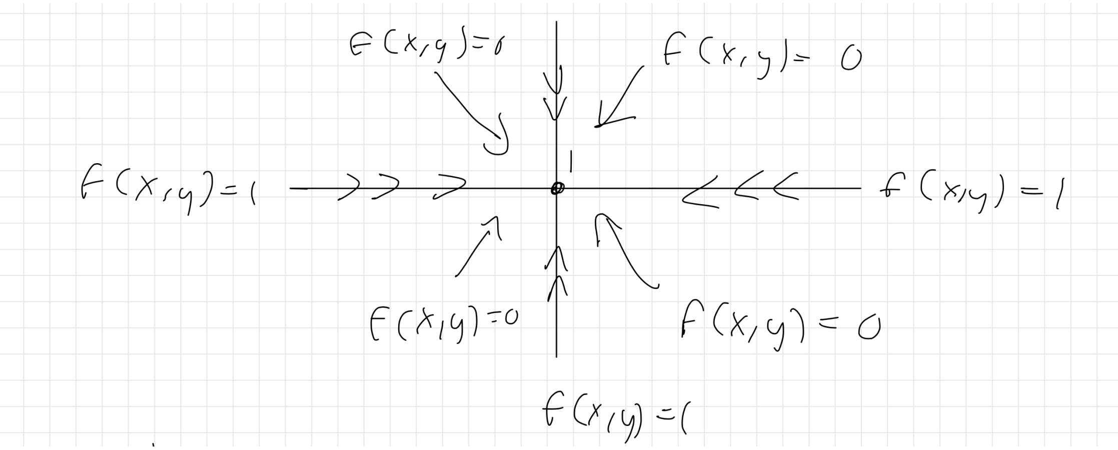



Figure 8.1: Function from Example 8.2 that is not continuous at the origin.

**Example 8.2.** We now look a function that is not continuous at $z = (0, 0)$: for $x = (x_1, x_2) \in \mathbb{R}^2$,

$$f(x) = \mathbb{1}\{x_1x_2 = 0\} = \begin{cases} 1 & \text{if } x_1x_2 = 0 \\ 0 & \text{otherwise.} \end{cases}$$

This function and its lack of continuity at the origin ($z$) are illustrated in Figure 8.1. If we approach $z$ along a path outside either axis, then $f(x) = 0$ along the path until jumping up discontinuously at $f(z) = 1$. That is, in terms of the definition, $|f(x) - f(z)| = 1$ even for points $x$ arbitrarily close to $z$, i.e., for arbitrarily small $d(x, z)$. (So for $\epsilon < 1$, we cannot find a corresponding $\delta$: no matter how small $\delta$, we can find $x$ with $d(x, z) < \delta$ but $|f(x) - f(z)| = 1 > \epsilon$.) ■

### 8.3.2 Value function continuity: maximum theorem

We now present a simplified version of a result about value function continuity known as the maximum theorem. (The full version is later in Theorem 9.1.) Economically, imagine we are maximizing objective function $f$ wrt choice variable vector $x$, given parameter vector $\theta$. Mathematically, the objective and value functions are

$$f\colon X \times \Theta \to \mathbb{R},\ X \subset \mathbb{R}^n,\ \Theta \subset \mathbb{R}^m,\ V(\theta) \equiv \max_{x \in X} f(x, \theta),\ x^*(\theta) \equiv \arg\max_{x \in X} f(x, \theta). \tag{8.3}$$

Without further assumptions, $x^*(\theta)$ may not exist, or it may be a set of multiple (even infinite) points.

**Theorem 8.4** (simplified maximum theorem)**.** *Given* (8.3)*, assume $X$ is a compact set and $f$ is continuous. Then,*

1. *$V(\cdot)$ is a continuous function;*
2. *if $x^*(\theta)$ is unique for all $\theta$, then $x^*(\cdot)$ is a continuous function.*

## Optional resources

Optional resources for this chapter

- Wikipedia articles like "Compact space" and "Metric space"

## Exercises

**Exercise E8.1.** For $x, y \in \mathbb{R}$, let $d(x, y) = |(x - y)/x|$ for $x, y > 0$. Assess each of the four properties of a distance function in Definition 8.1; if violated, provide a specific counterexample. XXX TO DO

**Exercise E8.2.** Consider the production possibilities of a firm that produces goods $x_1 \geq 0$ and $x_2 \geq 0$, such that $x_1^2 + x_2^2 \leq q$ for some $q > 0$. Let $X$ denote the set of feasible $(x_1, x_2)$.

a. Assess whether or not $X$ has each of the following properties: open; closed; bounded; compact. (Try to at least explain your judgment, even if you don't have a formal proof.)

b. Explicitly write the set of boundary points, $\partial X$. (Suggestion: write it as a union of three sets.)

**Exercise E8.3.** Let $x = (x_1, x_2) \in \mathbb{R}^2$ and $f\colon \mathbb{R}^2 \to \mathbb{R}$ with $f(x) = |x_1 - x_2|$. Is this function continuous? Try to prove your claim formally.

# Chapter 9

# Correspondences and Fixed-Point Theory

We introduce correspondences, which are like functions that return sets instead of single values, and fixed point theory, which can be helpful for finding equilibria in economics.

*Unit learning objectives for this chapter*

9.1. Define vocabulary words (in **bold**) related to correspondences and fixed point theory, both mathematically and intuitively [TLO 1]

9.2. Characterize the properties of a particular correspondence [TLO 3]

9.3. Assess whether a fixed point exists for a given correspondence or function [TLO 3]

## 9.1 Correspondences

In economics, sometimes we want a "function" that returns not just a single value but a set of values, called a **correspondence**, as motivated by the following example. Imagine a consumer's problem of choosing $x = (x_1, x_2)$ to maximize utility $u(x) = x_1 + x_2$ subject to non-negative $x$, prices $p = (p_1, p_2)$, and budget constraint $p \cdot x = p_1x_1 + p_2x_2 \leq M$. When $p_1 = p_2$, assuming all the money is spent so that $p \cdot x = M$, utility is $M/p_1 = M/p_2$ regardless of $x$. That is, rather than $x^*$ being a single point, $x^*$ is a set of points in $\mathbb{R}^2$: $x^* = \{x \in \mathbb{R}^2 : 0 \leq x_1 \leq M/p_1, x_2 = (M - p_1x_2)/p_2\}$. Thus, the relationship between $x^*$ and $p$ is not a function but a correspondence. Besides consumer (and producer) choice, we may also need a correspondence to describe multiple equilibria in macroeconomics, game theory (like industrial organization), and other areas.

### 9.1.1 Basics

Notationally, instead of a function $f\colon X \to Y$, a correspondence is written as $f\colon X \rightrightarrows Y$ or $f\colon X \twoheadrightarrow Y$. (Or, sometimes it is just written with $\to$.) The domain is $X$, and the codomain is all possible subsets of $Y$. In economics, we typically have $f\colon \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, oten with $n = m$.

Two caveats about having a set-valued output:

1. For any $Y$, $\emptyset \in Y$, so generally $f(x) = \emptyset$ is allowed; but sometimes this possibility is excluded.
2. Technically, $y$ the value differs from $\{y\}$ the set (that contains only one value, $y$); so even if $f(x) = \{y\}$ for all $x$, the correspondence $f$ is not technically a "function." However, in that case, we usually proceed by analyzing the function $\tilde{f}(x) = y$.

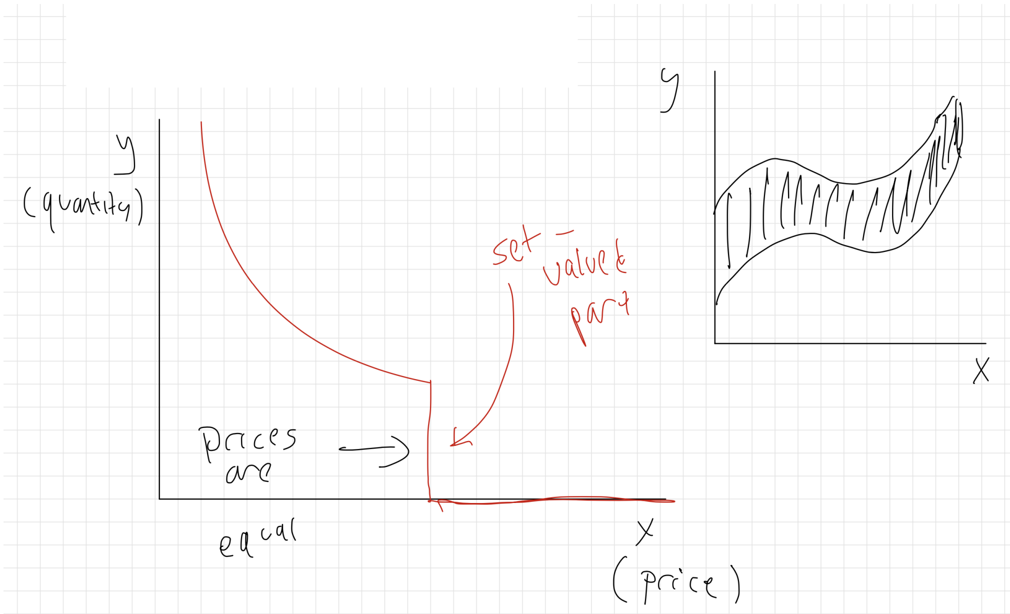



Figure 9.1: Examples of correspondences.

**Example 9.1.** Figure 9.1 shows two examples of a correspondence $f\colon \mathbb{R} \rightrightarrows \mathbb{R}$. The example on the left (demand curve) is almost a function, but as in the motivating example above, there is one specific price where any quantity between zero and some upper bound is possible. The example on the right maps each $x$ to a (non-trivial) interval of values $f(x) \subseteq Y$. ■

### 9.1.2 Properties

This section characterizes three properties that not all correspondences have, but that are required for certain theorems to apply, as we will see.

The **closed graph property** requires certain limits to be "included" in $f$. As with a relation (Definition 1.14), we can consider correspondence $f$ to describe a subset of $X \times Y$: the "graph" part of "closed graph property" refers to the subset

$$G = \{(x, y) \in X \times Y : y \in f(x)\}. \tag{9.1}$$

The closed graph property requires that if we have a sequence of points $(x_n, y_n)$ that are all inside the graph so $(x_n, y_n) \in G$, then (if it exists) the limit $\lim_{n\to\infty}(x_n, y_n) = (x^*, y^*)$ must also be in the graph. This is formally stated in Definition 9.1. Figure 9.2 shows a violation: there is a "hole" missing at a point that can be the limit of a sequence of points $(x_n, y_n) \in G$.

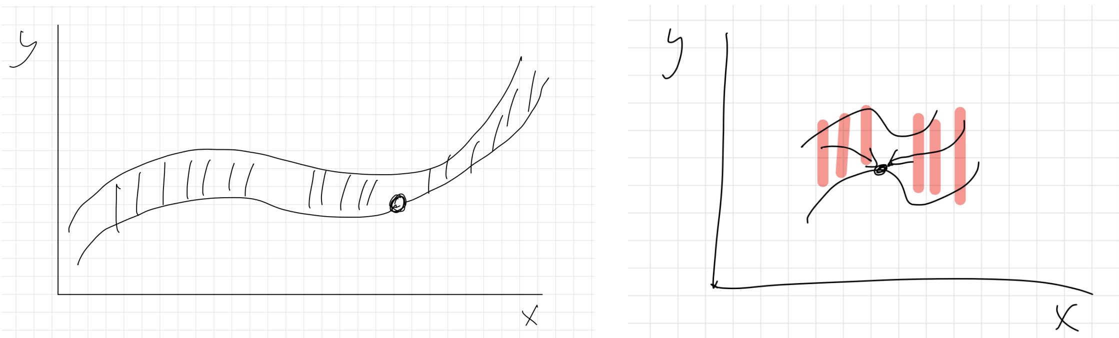



Figure 9.2: Examples of closed graph property violated (left) and satisfied (right).

**Definition 9.1** (closed graph property)**.** Correspondence $f\colon X \rightrightarrows Y$ has the closed graph property when the following implication holds: if a) $x_n \in X$ for all $n \in \mathbb{N}$ with $\lim_{n\to\infty} x_n = x^* \in X$, and b) $y_n \in f(x_n)$ for all $n \in \mathbb{N}$ with $\lim_{n\to\infty} y_n = y^*$, then $y^* \in f(x^*)$.

There are two main types of continuity for correspondences. One builds on the closed graph property, and the other essentially requires that any point in the graph $G$ is the limit point of a sequence in $G$. As often is true, there are other equivalent definitions of each. Sometimes a correspondence is called **continuous** if it satisfies both of the following definitions.

**Definition 9.2** (upper hemicontinuity)**.** An **upper hemicontinuous** correspondence $f\colon X \rightrightarrows Y$ has a closed graph and bounded set $f(S) = \{y \in f(x) : x \in S\}$ for every compact $S \subset X$.

**Definition 9.3** (lower hemicontinuity)**.** Correspondence $f\colon X \rightrightarrows Y$ is **lower hemicontinuous** at point $x^*$ iff for every sequence with $x_n \in X$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n = x^* \in X$, and for every $y \in f(x^*)$, there exists a sequence with $y = \lim_{n\to\infty} y_n$ and $y_n \in f(x_n)$ for all $n > N$ for some $N$. (Some sources replace $y_n$ with $y_k \in f(x_{m_k})$ for subsequence $x_{m_k}$, i.e., sequence $m_k \in \mathbb{N}$ is strictly increasing.) A lower hemicontinuous correspondence is lower hemicontinuous at all $x^* \in X$.

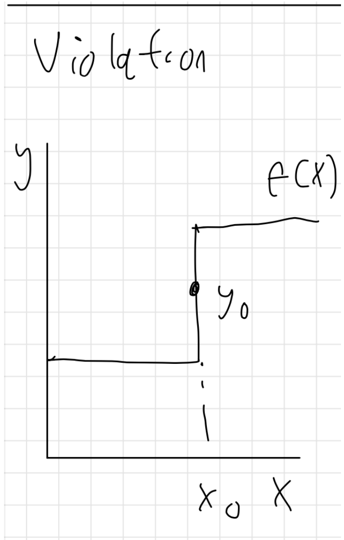



Figure 9.3: Violation of lower hemicontinuity.

Figure 9.3 shows a violation of lower hemicontinuity. To make it concrete, imagine $f$ assigns $f(x_0) = [a, b]$ with $a < y_0 < b$, with $f(x) = \{a\}$ for $x < x_0$ and $f(x) = \{b\}$ for $x > x_0$. Consider the sequence $x_n = x_0 - (1/n)$. Then $\lim_{n\to\infty} x_n = x_0$, but $f(x_n) = a$ for all $n \in \mathbb{N}$, so $\lim_{n\to\infty} f(x_n) = a \neq y_0$.

### 9.1.3 General maximum theorem

Using correspondences, the simplified maximum theorem of Section 8.3.2 can now be generalized. The setup is similar to (9.2) but allows the choice set to depend on the parameter $\theta$; that is, instead of maximizing over $x \in X$ regardless of $\theta$, maximization is over $x \in C(\theta)$ for a correspondence $C$. Additionally, there may be a set of multiple maximizers. The new setup is

$$\begin{aligned} &f\colon X \times \Theta \to \mathbb{R}, && C\colon \Theta \rightrightarrows X, \\ &V(\theta) \equiv \sup_{x\in C(\theta)} f(x,\theta), && C^*(\theta) \equiv \{x \in C(\theta) : f(x,\theta) = V(\theta)\}. \end{aligned} \tag{9.2}$$

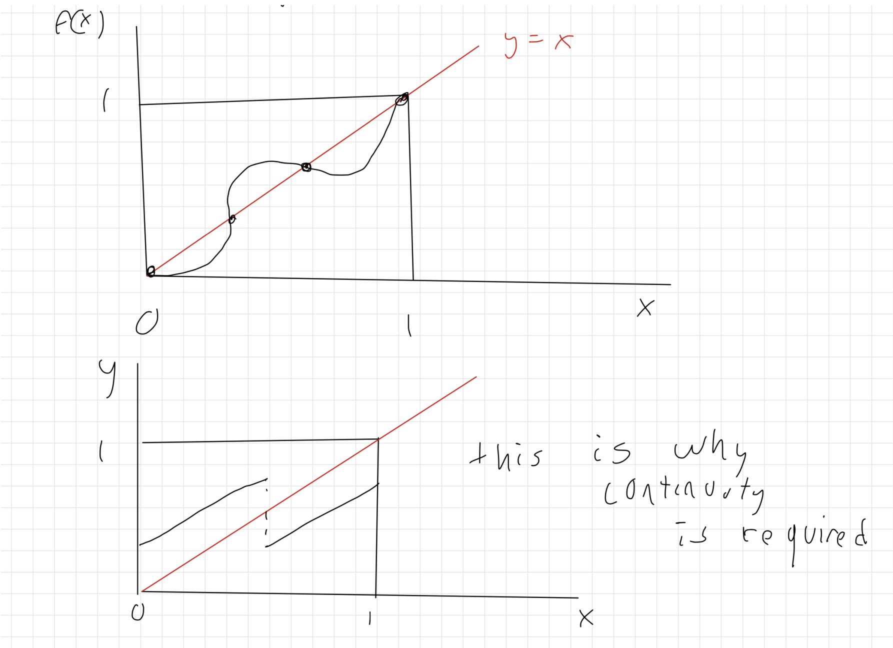
**Theorem 9.1** (maximum theorem)**.** *Given (9.2), assume that for all $\theta \in \Theta$, $C(\theta)$ is a compact set and $C(\theta) \neq \emptyset$. If $C$ is continuous (i.e., both upper and lower hemicontinuous) and $f$ is continuous, then*

1. *$V(\cdot)$ is a continuous function;*
2. *$C^*(\cdot)$ is an upper hemicontinuous correspondence with $C^*(\theta)$ compact and nonempty for all $\theta \in \Theta$.*

## 9.2 Fixed-point theory

Economically, a **fixed point** represents an equilibrium, mathematically represented by $f(x) = x$. Initially, we consider a function $f\colon A \to A$. (Remember that this notation says $A$ is the codomain, which is more general than requiring $A$ to be the range; that is, the fixed-point theorem still applies even if $\{f(x) : x \in A\}$ is not all of $A$ but rather a strict subset of $A$.) Later, we consider a correspondence $f\colon A \rightrightarrows A$, in which case the fixed point satisfies $x \in f(x)$.

**Theorem 9.2** (Brouwer's fixed-point theorem)**.** *If $S \subset \mathbb{R}^n$ is nonempty, compact, and convex, and if $f\colon S \to S$ is a continuous function, then $f$ has at least one fixed point $x^*$ such that $x^* = f(x^*)$.*




Section 9.2 shows two functions $f\colon [0,1] \to [0,1]$, one with a fixed point and one without. The top example satisfies the conditions of Theorem 9.2: $S = [0,1]$ is nonempty, compact, and convex, and visually we can see $f$ is continuous. The fixed points are the values of $x$ corresponding to where $y = f(x)$ intersects the line $y = x$. There are actually four fixed points, including $x^* = 0$ and $x^* = 1$. More generally, you can imagine three cases (that cover all the possibilities): 1) $f(0) = 0$, so $x^* = 0$ is a fixed point; 2) $f(1) = 1$, so $x^* = 1$ is a fixed point; 3) $f(0) > 0$ and $f(1) < 1$, so $y = f(x)$ must cross $y = x$ at some $0 < x < 1$. The bottom example in Section 9.2 shows why continuity of $f$ is required: otherwise, in the third case with $f(0) > 0$ and $f(1) < 1$, the function can jump over the $y = x$ line, so there is no fixed point.

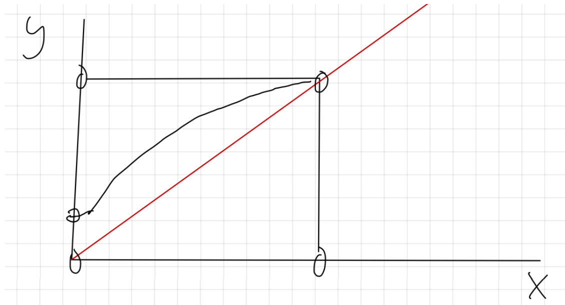



Figure 9.4: Illustration of Brouwer's fixed-point theorem.

Figure 9.4 illustrates why compactness of $S$ is required in Theorem 9.2. Instead of $S = [0,1]$, let $S = (0,1)$, the open unit interval (which is not compact because it is not closed). The graph shows a function with $f(x) > x$ for $x < 1$ and $f(1) = 1$, but because $S$ excludes 1, there is no fixed point (no solution to $f(x) = x$ on $S = (0,1)$). The set $S$ could also violate compactness by being unbounded even if it is closed, like $S = \mathbb{R}$. The function $f(x) = x + 1$ is continuous, and $S = \mathbb{R}$ is nonempty and convex, but not compact, so Brouwer's does not apply (and indeed this $f$ clearly has no fixed point). However, be careful with the fallacy of the inverse: non-compactness of $S$ does not imply that $f$ has no fixed point, but rather it merely implies that we cannot apply Theorem 9.2. For example, $f\colon \mathbb{R} \to \mathbb{R}$ with $f(x) = 0$ still has a fixed point at $x^* = 0$.

**Example 9.2.** Let $S = [-\pi, \pi]$ and $f(x) = \cos(x)$ with $f\colon S \to S$. Although we cannot solve for the value satisfying $x^* = \cos(x^*)$ analytically, this example satisfies the conditions of Theorem 9.2, so a fixed point must exist. ■

**Example 9.3.** Consider the following game theory example. Players 1 and 2 respectively choose $x_1 \in [0,1]$ and $x_2 \in [0,1]$. The payoffs are respectively $\pi_1 = 1 - x_1 + f(x_1, x_2)$ and $\pi_2 = 1 - x_2 + f(x_1, x_2)$ for a continuous, strictly concave function $f\colon [0,1]^2 \to \mathbb{R}$. The best response functions are

$$x_1^*(x_2) = \arg\max_{x_1 \in [0,1]} 1 - x_1 + f(x_1, x_2), \quad x_2^*(x_1) = \arg\max_{x_2 \in [0,1]} 1 - x_2 + f(x_1, x_2).$$

Because of the compactness of $[0,1]$ and properties of $f$, the best response functions are also continuous functions from $[0,1]$ to $[0,1]$. The function $G\colon [0,1]^2 \to [0,1]^2$ for $G(x_1, x_2) = (x_1^*(x_2), x_2^*(x_1))$ is thus also continuous, and $[0,1]^2$ is compact and convex, so the conditions of Theorem 9.2 are satisfied and a fixed point exists with $G(x^*) = x^*$. Economically, that means this game has a Nash equilibrium. ■

Brouwer's fixed-point theorem is extended to correspondences in Kakutani's fixed-point theorem.

**Theorem 9.3** (Kakutani's fixed-point theorem)**.** *If $S \subset \mathbb{R}^n$ is nonempty, compact, and convex, and if $f\colon S \rightrightarrows S$ is an upper hemicontinuous correspondence with nonempty convex $f(x)$ for all $x \in S$, then $f$ has at least one fixed point $x^*$ such that $x^* \in f(x^*)$.*

## Optional resources

Optional resources for this chapter

- Appendix M.I ("Fixed Point Theorems") of Mas-Colell, Whinston, and Green (1995)

## Exercises

**Exercise E9.1.** Recall the motivating example where the consumer's demand for good $x_1$ in terms of prices $(p_1, p_2)$ is not a function but a correspondence, due to the maximizing consumption having any $0 \leq x_1 \leq M/p_1$ when $p_1 = p_2$. Imagine $p_2$ is fixed, and let the relationship between the demanded $x_1$ and $p_1$ be described by $f\colon \mathbb{R}_+ \rightrightarrows \mathbb{R}_+$, where $\mathbb{R}_+ = \{r \in \mathbb{R} : r \geq 0\}$. For each of the following properties, assess whether or not $f$ has the property.

a. Closed graph property
b. Upper hemicontinuity
c. Lower hemicontinuity

**Exercise E9.2.** Let $f(x) = 1 + (x-1)^3$ over $0 \leq x \leq 2$, and consider a sequence with $x_{t+1} = f(x_t)$. Does an equilibrium point (in the sense of Section 6.2) exist? Why?

# Bibliography

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. Oxford University Press. URL https://worldcat.org/en/title/32430901. [99]