Vanderbilt Economics Ph.D. Math Camp

Lecture Notes, Fall 2025

Maria Titova

Contents

1 Sets, Logic and Functions

1.1 Sets

Sets

Definition 1. A set is a collection of objects. These objects are called elements or members of the set.

- sets do not have to be finite
- you can define a set by:
  - specifying the elements directly (e.g. $\{Red, Yellow, Blue\}$)
  - specifying the elements indirectly (e.g. $\{x \in \mathbb{N} : x \text{ is prime}\}$)

Frequently Used Sets

- $\varnothing$ – empty set
- $\mathbb{N}$ – set of natural numbers
- $\mathbb{Q}$ – set of rational numbers
- $\mathbb{R}$ – set of real numbers (and $\mathbb{R}_+$, the set of weakly positive real numbers, and $\mathbb{R}_{++}$ – set of strictly positive real numbers)
- 2^S – set of all subsets of S

Notation

- let A be a set
  - $x \in A$ means that x is in A
  - $x \notin A$ means that x is not in A
- capital letters usually denote sets and lowercase letters denote elements
- x and $\{x\}$ are mathematically different: $\{x\}$ is the set whose only member is x

Subsets

Definition 2. A subset of B is a set A such that every element of A is an element of B. A proper subset of B is a subset of B that does not equal B.

- notation:
  - $A \subseteq B$ means A is a subset of B
  - $A \subset B$ means A is a proper subset of B
- examples:
  - $\mathbb{Q} \subset \mathbb{R}$
  - for any set S, $S \subseteq S$ and $\varnothing \subset S$

- if $x \in A$, then $\{x\} \subseteq A$
- if $A \subseteq B$ and $B \subseteq A$, then $A = B$

Complement

Definition 3. Suppose that all sets under consideration are subsets of a universal set U. Then, the complement of A is $A^c := \{x \in U : x \notin A\}$.

- **practice**: in $\mathbb{R}^2$, $\{(x, y) \in \mathbb{R}^2 : x + y > 1\}^c =$ ______________________________

Set Difference

Definition 4. The set difference of B and A is $B \setminus A := \{x : x \in B \text{ and } x \notin A\}$.

- fact: for any set A, $A^c = U \setminus A$

Union

Definition 5. The union of sets A and B is $A \cup B := \{x : x \in A \text{ or } x \in B\}$.

- we can take the union of more than two sets, even infinitely many
  - let $\mathcal{A}$ be a (possibly infinite) collection of sets
  - the union of the members of $\mathcal{A}$ is $\bigcup_{A \in \mathcal{A}} A := \{x : x \in A \text{ for some } A \in \mathcal{A}\}$

Intersection

Definition 6. The intersection of sets A and B is $A \cap B := \{x : x \in A \text{ and } x \in B\}$.

- we can take the intersection of more than two sets, even infinitely many
  - let $\mathcal{A}$ be a (possibly infinite) collection of sets
  - the intersection of the members of $\mathcal{A}$ is $\bigcap_{A \in \mathcal{A}} A := \{x : x \in A \text{ for all } A \in \mathcal{A}\}$

Cartesian Product

Definition 7. The Cartesian product of sets A and B is $A \times B := \{(a, b) \mid a \in A \text{ and } b \in B\}$.

- we denote the Cartesian product of A and itself as A^2
- **practice**:
  - if $A = \{a, b\}$ and $X = \{x, y, z\}$, then $A \times X =$ ______________________________
  - if $A = [0, 1]$, then $A^2 =$ ______________________________

1.2 Logic

Propositions

- for propositions P and Q, $P \Rightarrow Q$ means that if P is true, then Q is also true
- equivalently, $P \Rightarrow Q$ is true if one of the following holds:
  - P is true and Q is true
  - P is false and Q is true
  - P is false and Q is false
- **practice**: is the following proposition true or false?

$$(x \in \mathbb{N} \text{ and } x^2 < 0) \implies x = 1$$

Contrapositive

- contrapositive of "if P, then Q" is "if Q is not true, then P is not true"
- we use "$\neg Q$" to denote the negation of Q, so the contrapositive of $P \Rightarrow Q$ is $\neg Q \Rightarrow \neg P$
- **practice**: contrapositive of "if $x > 0$, then $x^3 > 0$" is ____________________________

Contrapositive

- **question**: What is the relationship between the truth value of $P \Rightarrow Q$ and its contrapositive?
- **answer**: They are equivalent, because
  - $P \Rightarrow Q$ is true if (1) P and Q are true, (2) $\neg P$ and Q are true, or (3) $\neg P$ and $\neg Q$ are true
  - $\neg Q \Rightarrow \neg P$ is true under exactly the same conditions
- useful because sometimes contrapositive is easier to prove

Converse

- converse of $P \Rightarrow Q$ is $Q \Rightarrow P$
- **question**: what is the relationship between the truth value of $P \Rightarrow Q$ and its converse?
- **answer**: there is none: a true statement could have a true or false converse, and a false statement could have a true or false converse
- **practice**:
  - converse of $x > 0 \Rightarrow x^3 > 0$ is ____________________________
  - converse of $x > 0 \Rightarrow x^2 > 0$ is ____________________________

If and Only If

- if $P \Rightarrow Q$ and its converse $Q \Rightarrow P$ are both true, we
  - write $P \Leftrightarrow Q$
  - say "P if and only if Q"
  - abbreviate "if and only if" as "iff"

Universal Quantifier

- universal, or "for all," quantifier denotes that a property holds for all elements of some set
- takes form "for all $a \in A$, a has a certain property"
- denoted $\forall$, e.g. $\forall x > 0 : x^2 > 0$

Existential Quantifier

- existential, or "there exists," quantifier denotes that a property holds for at least one element of some set
- takes the form "there exists $a \in A$ such that a has a certain property"
- denoted $\exists$, e.g. $\exists x > 0 : x - 100 > 0$

Negating Quantifiers

- negation of a universal quantifier involves an existential quantifier
  - negation of $\forall a \in A : P(a)$ is $\exists a \in A : \neg P(a)$
  - to disprove a universal statement, you only need to find a single counterexample
- negation of an existential quantifier involves a universal quantifier
  - negation of $\exists a \in A : P(a)$ is $\forall a \in A : \neg P(a)$
  - "there exists" statements are not very useful as falsifiable predictions of a theory

Negating Quantifiers

- order of multiple quantifiers of the same type does not matter:
  - $\forall x \in X, \forall y \in Y : P(x,y)$ is equivalent to $\forall (x,y) \in X \times Y : P(x,y)$
  - $\exists x \in X, \exists y \in Y : P(x,y)$ is equivalent to $\exists (x,y) \in X \times Y : P(x,y)$
- order of quantifiers of different types can matter:
  - $\forall n \in \mathbb{N}, \exists n' \in \mathbb{N} : n' > n$ is true
  - $\exists n' \in \mathbb{N}, \forall n \in \mathbb{N} : n' > n$ is false

Necessary/Sufficient

- if $P \Rightarrow Q$ is true, we say that Q is necessary for P and that P is sufficient for Q

  - Q is necessary for P because we cannot have P without Q
  - P is sufficient for Q because, as long as we have P, we definitely have Q
- if $P \Leftrightarrow Q$, we say that Q is necessary and sufficient for P, and vice versa

Necessary/Sufficient

- if a theory makes a particular prediction, that prediction is a *necessary* condition for the theory to be true
  - you can reject the theory by showing the prediction incorrect
  - showing the prediction correct does not confirm the theory, because other theories (incompatible with this one) might have the same prediction
- sometimes, if you are lucky, you can find *necessary and sufficient conditions* for a theory
  - in principle, you could confirm the theory by verifying these conditions
  - in practice, this is usually impossible (we never have that much data), but they still help us understand the model
- we often impose "regularity conditions" on objects in our models. We choose conditions that are *sufficient* (but maybe not necessary) for the model to yield "nice" predictions

1.3 Functions

Functions

- let A and B be nonempty sets. A function $f : A \to B$ is a rule that assigns to each element $x \in A$ a single element $f(x) \in B$.
- A is the domain of the function, and B is the codomain
- **example**: let $\mathcal{A}$ be a finite set of products, and let $\mathcal{K}(\mathcal{A})$ denote the set of nonempty subsets of $\mathcal{A}$. A choice function $c : \mathcal{K}(\mathcal{A}) \to \mathcal{A}$ tells you, for each menu the consumer might face, which product she will choose

Range/Graph

- range (or image) of a function $f : A \to B$ is

$$\{y \in B : y = f(x) \text{ for some } x \in a\}$$

- graph of a function $f : A \to B$ is

$$\{(x, y) \in A \times B : y = f(x)\}$$

One-to-One

Definition 8. Function f is one-to-one (injective) if $x \neq x'$ implies $f(x) \neq f(x')$.

- a function is one-to-one if each element in the domain is mapped to a unique element of the codomain

Onto

Definition 9. Function $f : A \to B$ is onto (surjective) if, for all $y \in B$, there exists $x \in A$ such that $y = f(x)$.

- depends on how you specify codomain: x^2 is onto if codomain is $\mathbb{R}_+$, but not if it is $\mathbb{R}$

Bijective

Definition 10. A function is bijective if it is both one-to-one and onto.

Composition

Definition 11. Let $f : A \to B$ and $g : B \to C$. The composition $g \circ f$ is the function from A to C defined by

$$(g \circ f)(\cdot) = g(f(\cdot)).$$

- **practice**: $f(x) = -x$ and $g(x) = |x|$; $g(f(x)) =$ ________________________________

Properties of Compositions

Theorem 12. *Let $f : A \to B$ and $g : B \to C$. Then,*

- *if f and g are one-to-one (injective), then so is $g \circ f$*
- *if f and g are onto (surjective), then so is $g \circ f$*
- *if f and g are bijective, then so is $g \circ f$*

Inverse

Definition 13. Let $f : A \to B$. If there exists a function $g : B \to A$ such that $(g \circ f)(x) = x$ for all $x \in A$ and $(f \circ g)(y) = y$ for all $y \in B$, then g is the inverse of f and is denoted f^{-1}.

- **practice**: do these f's have an inverse?
  - $f : [0,1] \to [0,1]$ where $f(x) = x^2$
  - $f : [-1,1] \to [0,1]$ where $f(x) = x^2$
  - $f : [0,1] \to [0,2]$ where $f(x) = x^2$

Existence of Inverse

Theorem 14. *For any function $f : A \to B$, f^{-1} exists if and only if f is bijective.*

Correspondences

- correspondence $f : A \rightrightarrows B$ is a rule that assigns to each element $x \in A$ a subset $f(x)$ of B
- **example**: consider a consumer who makes different decisions at different times; her choice correspondence $c : \mathcal{K}(\mathcal{A}) \rightrightarrows \mathcal{A}$ tells you, for each menu the consumer might face, which products she sometimes chooses
- technically, correspondence $f : A \rightrightarrows B$ is a function $f : A \to 2^B$
- but it's worth having separate notation for correspondences because they come up a lot in economics

1.4 Proofs

Proofs

- proofs are simply "a set of carefully crafted directions, which, when followed, should leave the reader absolutely convinced of the truth of the proposition in question"
- there are generally two types of proofs: direct and indirect (or proof by contradiction)
- even if you don't want to be a theorist, you will need to be comfortable reading and writing proofs

General Advice

- **don't be intimidated**. Most proofs economists write down, even theorists, do not require a flash of mathematical brilliance. They are intended to establish interesting facts about a model. Often they use standard tools
- if you don't know where to start (which, in research, is much of the time), just start somewhere. Try some examples and draw some illustrations. Then see if you can generalize
- sometimes, it is helpful to attempt to prove the opposite of the desired result (even if you know the desired result to be true). You may learn something when you get stuck
- an underrated skill: realizing what you *don't* know how to prove. It's better to be stuck than to think you've proven something you haven't. Check carefully for weak points and hand-waving

Practice: Direct Proof

Claim: If n is an even integer, then $-5n - 3$ is an odd integer.

Practice: Proof by Contradiction

Claim: $\sqrt{2}$ is not a rational number.

Induction

- you use induction when you want to show that something holds in a set of scenarios that can be indexed by $\mathbb{N}$
- start by showing that the result holds for an initial value (usually 0 or 1)
- then show that $(\text{result holds for } n-1) \implies (\text{result holds for } n)$

Practice: Proof by Induction

Claim: For any $n \in \mathbb{N}$,

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Practice: Proof by Cases/Steps

Claim: If n is an integer, then $n^2 + 3n + 5$ is an odd integer.

2 Metric Spaces and Sequences

Least Upper Bounds

Definition 15. $x \in \mathbb{R}$ is an upper bound for $A \subseteq \mathbb{R}$ if $x \geq a$ for all $a \in A$. An upper bound x is the *least* upper bound (supremum) if, for any upper bound b for A, $x \leq b$. If $x \in A$, then x is the maximum of A

- there can be many upper bounds, but at most one supremum/maximum
- we write $x = \sup A$, and $x = \max A$
- for definitions of lower bound, infimum, and minimum, replace $\geq$ with $\leq$, "least" with "greatest," and "upper" with "lower"

Cardinality

- cardinality of a set can be thought of as the "size" of a set (but there are other measures of size)
- cardinality of a finite set is just the number of elements
- cardinality of an infinite set should presumably be greater than that of any finite set – but what is it?

Cardinality

- Georg Cantor (1845—1918) came up with a way to do this
  - A has cardinality less than or equal to the cardinality of B if there is a one-to-one $f : A \to B$
  - then, A and B have the same cardinality if there is a bijection $f : A \to B$
  - or, A and B have the same cardinality if the elements in A and B can be paired up without any omissions or repetitions

Infinite Sets

Definition 16. Set is countable if it is finite or has the cardinality of natural numbers.

Theorem 17. *The rationals and integers are countable.*

Theorem 18.
- *Finite unions of countable sets are countable.*
- *Subsets of countable sets are countable.*
- *The Cartesian product of finitely many countable sets is countable.*

- *note: any infinite subset of a countably infinite set has the same cardinality as the original set*

Infinite Sets

Definition 19. Set is uncountable if its cardinality is strictly greater than that of the natural numbers.

Theorem 20. *The set of reals ($\mathbb{R}$) is uncountable. The power set of the natural numbers ($2^{\mathbb{N}}$) is uncountable.*

Continuum Hypothesis

Theorem 21. *The Continuum Hypothesis states that there is no set whose cardinality is strictly between that of the rationals and the reals.*

- due to the work of Kurt Godel in the 1930s and Paul Cohen in the 1960s, we now know that this question is undecidable. This means essentially that the standard axioms of set theory do not provide enough structure to determine the answer to the question. Both the Continuum Hypothesis and its negation are consistent with the working rules of mathematics

2.1 Metric Spaces

Metric spaces

Definition 22. A metric space is a pair (S, d), where S is a set and $d : S^2 \to \mathbb{R}_+$ satisfies the following properties:

- for all $x, y \in S$, $d(x, y) = 0$ iff $x = y$
- for all $x, y \in S$, $d(x, y) = d(y, x)$
- (triangle inequality) for all $x, y, z \in S$, $d(x, y) \leq d(x, z) + d(z, y)$

- concept of a metric space formalizes the intuitive idea of "distance"

Common Metrics: Euclidean

- when $S = \mathbb{R}$, we use the metric given by $d(x, y) = |x - y|$
- when $S = \mathbb{R}^n$, we usually use the Euclidean metric:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$

  which is equivalent to the above when $n = 1$

Common Metrics: Hausdorff

- when S is a collection of non-empty subsets of a metric space $(\widehat{S}, \widehat{d})$, we usually use the Hausdorff metric:

$$d(X, Y) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} \widehat{d}(x, y), \sup_{y \in Y} \inf_{x \in X} \widehat{d}(x, y) \right\}$$

- intuitively, the Hausdorff distance between two sets is the greatest distance between a point in one set and the closest point in the other set
- **practice**: $d([-2,0],[0,1]) =$ ________________________

- when S is a set of bounded functions mapping X to $\widehat{S}$, where $(\widehat{S}, \widehat{d})$ is a metric space, we usually use the supremum metric:

$$d(f,g) = \sup_{x \in X} \widehat{d}(f(x), g(x))$$

$\mathbb{R}^n$ and Euclidean Distance

- for rest of this section, we focus on $S = \mathbb{R}^n$ and d as the Euclidean metric
- concepts in this lecture are much more generally applicable

2.2 Convergence

Sequences

Definition 23. A <u>sequence</u> is a function whose domain is $\mathbb{N}$.

- intuitively, a sequence is an ordered list of objects
- codomain could be $\mathbb{R}^n$, or it could be a set of sets, functions, preferences…
- can denote a sequence in several ways: $(a_1, a_2, \dots)$ or $(a_n)_{n=1}^{\infty}$ or (a_n)

Convergence

- intuitively, a sequence (x_n) converges to a limit x if, by going far enough out in the sequence, we can get as close to x as we wish

Definition 24. Let (S,d) be a metric space. A sequence (x_n) in S <u>converges</u> to $x \in S$ if, for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for all $n > N$.

- if (x_n) converges to x, we often write $x = \lim_{n \to \infty} x_n$ or $x_n \to x$

Uniqueness

Theorem 25. *If (x_n) converges to x and y, then $x = y$.*

- **Proof:** ________________________

A Simple Convergence Proof

Theorem 26.

$$\lim_{n\to\infty} \frac{1}{n} = 0$$

- **hint** *(Archimedean property)*: $\forall \varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $N > \frac{1}{\varepsilon}$
- **proof**: ______________________________

Showing Convergence

- pick an arbitrary $\varepsilon > 0$, and then you show that, at some point N in the sequence, all the terms beyond N are no more than ε away from the limit
- proof boils down to coming up with a formula that delivers the N required for each value of ε
- this formula is typically reverse engineered: you start with the inequality you need, and then back out a candidate N
- finally, you check that the candidate works: $|x_n - x| < \varepsilon$ actually holds for all $n > N$

Squeeze Lemma

- sometimes, it is hard to show convergence directly
- it may be easier to establish upper and lower bounds for the sequence of interest, and then show that they converge to the same object
- we state this result for real numbers, but it can easily be extended to functions, sets, etc

Theorem 27. *Suppose $\forall n > N_0$, $x_n < z_n < y_n$. Then, $(x_n \to L$ and $y_n \to L) \Rightarrow z_n \to L$.*

Negating Convergence

- how do we show that (x_n) does not converge to a given limit x?
- we show the following:

"For some $\varepsilon > 0$, for all $N \in \mathbb{N}$, there exists $n > N$ such that $|x_n - x| \geq \varepsilon$"

Practice

- consider (x_n) given by $x_n = (-1)^n$, show that x_n does not converge to 1

Example

- what if we want to show that (x_n) does not converge to anything, i.e. that (x_n) diverges?
- we suppose that $(x_n) \to x$ and derive a contradiction for every x

Bounded

Definition 28. Fix a metric space (S, d). $T \subseteq S$ is bounded if there exists $r \in \mathbb{R}$ such that $d(x, y) < r$ for all $x, y \in T$.

Definition 29. A sequence (x_n) is bounded if its range is bounded.

- sequence (x_n) in $\mathbb{R}$ is bounded iff $\exists r \in \mathbb{R}$ such that $|x_n| < r$ for all n

Bounded

Theorem 30. *Every convergent sequence is bounded.*

- converse does not hold: $(-1)^n$ is bounded, but does not converge

Algebraic Limit Theorem

- conveniently, convergent sequences of real numbers behave well with respect to addition, multiplication and division

Theorem 31. *If (a_n) and (b_n) are convergent sequences in $\mathbb{R}$, then*

- *for any $c \in \mathbb{R}$, $\lim(ca_n) = c \lim a_n$*
- $\lim(a_n + b_n) = \lim a_n + \lim b_n$
- $\lim(a_n b_n) = (\lim a_n)(\lim b_n)$
- *if $\lim b_n \neq 0$, then $\lim(a_n/b_n) = (\lim a_n)/(\lim b_n)$*

Order Limit Theorem

Theorem 32. *If (a_n) and (b_n) are convergent sequences in $\mathbb{R}$, and if there exists N such that $a_n \leq b_n$ for all $n > N$, then $\lim a_n \leq \lim b_n$.*

Monotone Sequences

- while not all bounded sequences converge, bounded increasing (and bounded decreasing) sequences in $\mathbb{R}$ always do

Definition 33. A sequence (a_n) in $\mathbb{R}$ is increasing if $a_n \leq a_{n+1}$ for all n, and decreasing if $a_n \geq a_{n+1}$ for all n. It is monotone if it is increasing or decreasing.

Monotone Convergence Theorem

Theorem 34. *If a sequence in $\mathbb{R}$ is monotone and bounded, it converges.*

Subsequences

Definition 35. If (x_n) is a sequence and I is an infinite subset of $\mathbb{N}$, then $(x_n)_{n \in I}$ is a subsequence of (x_n).

Convergence of Subsequences

Theorem 36. *If (x_n) converges to x, then every subsequence of (x_n) converges to x.*

- subsequence of a *divergent* sequence may or may not converge
- **practice**: find convergent and divergent subsequences of $x_n = (-1)^n$ ________________________

Bolzano–Weierstrass Theorem

- the fact that $(-1)^n$ contains a convergent subsequence is not a coincidence

Theorem 37. *If (x_n) is a bounded sequence in $\mathbb{R}^n$, then it has a convergent subsequence.*

Convergence of Subsequences

- there are generalizations of Bolzano-Weierstrass beyond $\mathbb{R}^n$
- however, the theorem *cannot* be generalized to the claim "every bounded sequence has a convergent subsequence"
  - consider the metric space $(\mathbb{Q}, d)$ where d is the usual metric
  - sequence $(3, 3.1, 3.14, \ldots)$ does not have a convergent subsequence because $\pi \notin \mathbb{Q}$
- we often choose to work in spaces with the property "every sequence has a convergent subsequence"—these are called "sequentially compact" spaces
- $\mathbb{R}^n$ not sequentially compact, but any closed and bounded subset of $\mathbb{R}^n$ is

Cauchy Sequences

- if (x_n) is a convergent sequence, then if you go far enough out, you can make all the remaining terms as close together as you like
- we can formally define this property *without* mentioning limits

Definition 38. A sequence (x_n) is Cauchy if, for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $|x_n - x_m| < \varepsilon$ for all $n, m > N$

Cauchy Sequences

- if a sequence is convergent, it is Cauchy—but what about the converse?

- Cauchy sequence need not be convergent:
  - in $\mathbb{Q}$, $(3, 3.1, 3.14, \ldots)$ doesn't have a limit, but it is Cauchy
- in many of the metric spaces we usually work with, Cauchy sequences are always convergent
  - a metric space that satisfies this property is called "complete"

Completeness

Definition 39. A metric space (X, d) is complete if every Cauchy sequence of points in X has a limit that is also in X.

Theorem 40. $\mathbb{R}^n$ *is complete.*

3 Limits, Continuity and Differentiation

Some Loose Definitions

Definition 41. Let (X, d) be a metric space. Then, $x \in X$ is a limit point of $A \subseteq X$ if $x = \lim x_n$ for some sequence (x_n) in A such that $x_n \neq x$ for all n.

Definition 42. A subset A of a metric space (X, d) is closed if it contains all its limit points.

Theorem 43. *In $\mathbb{R}^n$, a set is compact if and only if it is **closed** and **bounded**.*

Functional Limits in Metric Spaces

Definition 44. Suppose that (X, d_X) and (Y, d_Y) are metric spaces; $\overline{x}$ is a limit point of X; $f : X \to Y$ is a function. Then, $\lim_{x \to \overline{x}} f(x) = \overline{y}$ if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in X$,

$$d_X(\overline{x}, x) < \delta \quad \Longrightarrow \quad d_Y(\overline{y}, f(x)) < \varepsilon.$$

- by getting close enough to $\overline{x}$, we can get the value of the function to be arbitrarily close to $\overline{y}$
- we do not require $f(\overline{x}) = \overline{y}$ or $\overline{x} \in X$
- alt. notation: $f(x) \to \overline{y}$ as $x \to \overline{x}$

Practice

- let $f(x) = 2 + \frac{x}{3}$; show that $\lim\limits_{x \to 6} f(x) = 4$

Limits and Sequences

- we previously defined limits for sequences (Lecture 1)
- those are related to functional limits

Theorem 45. *Suppose that (X, d_X) and (Y, d_Y) are metric spaces and $f : X \to Y$ is a function. For any limit point $\overline{x}$ of X, the following are equivalent:*

- $\lim\limits_{x \to \overline{x}} f(x) = \overline{y}$
- *for any $(x_n) \to \overline{x}$, $(f(x_n)) \to \overline{y}$*

Algebra for Functional Limits

Theorem 46. *Let X be a metric space, and let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ be functions. If $\lim\limits_{x \to \overline{x}} f(x) = \overline{y}$ and $\lim\limits_{x \to \overline{x}} g(x) = \overline{z}$, then*

- *for any* $c \in \mathbb{R}$, $\lim_{x \to \bar{x}} cf(x) = c\bar{y}$
- $\lim_{x \to \bar{x}} (f(x) + g(x)) = \bar{y} + \bar{z}$
- $\lim_{x \to \bar{x}} f(x)g(x) = \overline{yz}$
- *if* $\bar{z} \neq 0$, $\lim_{x \to \bar{x}} \frac{f(x)}{g(x)} = \frac{\bar{y}}{\bar{z}}$

3.1 Continuous Functions

Continuous Function

Definition 47. Let (X, d_X) and (Y, d_Y) be metric spaces. Function $f : X \to Y$ is continuous at $x' \in X$ if $\forall \varepsilon > 0$, $\exists \delta > 0$ such that, for all $x \in X$,

$$d_X(x, x') < \delta \implies d_Y(f(x), f(x')) < \varepsilon$$

- a function is continuous if it is continuous $\forall x' \in X$

Practice

- is $f(x) = \sqrt{x}$ continuous at $x = 0$?
- **practice (harder)**: is $f(x) = \sqrt{x}$ continuous?

Continuity and Limits

- we can characterize continuity in terms of functional limits

Theorem 48. *Suppose that*

- (X, d_X) *and* (Y, d_Y) *are metric spaces*
- $\bar{x}$ *is a limit point of* X
- $f : X \to Y$ *is a function*

Then, f *is continuous at* $\bar{x}$ *if and only if* $\lim_{x \to \bar{x}} f(x) = f(\bar{x})$.

- follows from definitions of continuity and functional limits

Continuity and Sequences

- we can characterize continuity in terms of sequences

Theorem 49. *Suppose that* (X, d_X) *and* (Y, d_Y) *are metric spaces;* $f : X \to Y$ *is a function. For any limit point* $\bar{x}$ *of* X*, the following are equivalent:*

- *f is continuous at $\overline{x}$*
- *for any sequence (x_n) in X that converges to $\overline{x}$, the sequence $(f(x_n))$ converges to $f(\overline{x})$*

Showing Discontinuity

- when is limit point $\overline{x}$ of X discontinuous?

Theorem 50. *Suppose that*

- *(X, d_X) and (Y, d_Y) are metric spaces*
- *$\overline{x}$ is a limit point of X*
- *$f : X \to Y$ is a function*

If there exists a sequence (x_n) in X that converges to $\overline{x}$ such that $(f(x_n))$ does not converge to $f(\overline{x})$, then f is discontinuous at $\overline{x}$.

Algebra for Continuous Functions

Theorem 51. *Suppose that (X, d_X) is a metric space and $x \in X$. Let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ be functions that are continuous at x. Then, the following functions are continuous at x:*

- *cf for any $c \in \mathbb{R}$*
- *$f + g$*
- *fg*
- *$\frac{f}{g}$, provided $g(x) \neq 0$*
- so, all polynomials, i.e. functions of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

  are continuous

Extreme Value Theorem

Theorem 52. *Let (X, d) be a metric space and suppose that X is a compact set. If $f : X \to \mathbb{R}$ is continuous, then there exist points*

$$\underline{x}, \overline{x} \in X$$

such that

$$f(\underline{x}) \leq f(x) \leq f(\overline{x}) \quad \text{for all } x \in X.$$

In other words, f attains its minimum at $\underline{x}$ and its maximum at $\overline{x}$.

- in $\mathbb{R}$ (or $\mathbb{R}^n$), this means a continuous function on any **closed and bounded** set reaches its maximum and minimum values

Intermediate Value Theorem

Theorem 53. *Let $f : [a,b] \to \mathbb{R}$ be a continuous function. For every value*

$$y \in \left[\min\{f(a), f(b)\}, \max\{f(a), f(b)\}\right],$$

there exists

$$x \in [a,b]$$

such that

$$f(x) = y.$$

- a continuous function on a closed interval takes on *every* value between $f(a)$ and $f(b)$
- the values are inclusive: the function hits the endpoints $f(a)$ and $f(b)$ as well as all values in between

3.2 Fixed Points

Fixed Points

- fixed points are used to show that an equation, or system of equations, has a solution
- e.g., they are used to show that Nash equilibria exist, and to show that certain dynamic programming problems have a solution

Definition 54. A fixed point of a function $f : X \to Y$ is $x \in X$ such that $f(x) = x$.

Contraction

Definition 55. Let (X, d) be a metric space. A function $f : X \to X$ is called a contraction with modulus $\alpha \in [0, 1)$ if $d(f(x), f(y)) \leq \alpha d(x, y)$ for all $x, y \in X$.

- note: all contractions are continuous

Contraction Mapping Theorem

Theorem 56. *Let (X, d) be a complete metric space. If $f : X \to X$ is a contraction, then f has a unique fixed point.*

- we defined **complete** in Lecture 2
- $\mathbb{R}^n$ with Euclidean metric is complete

Fixed Points in One Dimension

Theorem 57. *If $f : [a, b] \to [a, b]$ is a continuous, then f has a fixed point.*

Brouwer's Theorem

Theorem 58. *If $X \subset \mathbb{R}^N$ is compact, convex and non-empty and $f : X \to X$ is continuous, then f has a fixed point.*

3.3 Differentiation in $\mathbb{R}$

Derivative

Definition 59. Let $X \subseteq \mathbb{R}$ and let $f : X \to \mathbb{R}$ be a function.
The derivative of f at $\overline{x} \in X$ is

$$f'(\overline{x}) := \lim_{\Delta x \to 0} \frac{f(\overline{x} + \Delta x) - f(\overline{x})}{\Delta x} \qquad \left(= \lim_{x \to \overline{x}} \frac{f(x) - f(\overline{x})}{x - \overline{x}} \right)$$

if this limit exists, in which case we call f differentiable at $\overline{x}$.
If $f'(\overline{x})$ exists $\forall \overline{x} \in X$, we say that f is differentiable.

- alternative notation:

$$f_x(\overline{x}) \qquad \left.\frac{df}{dx}\right|_{x=\overline{x}} \qquad \frac{df}{dx}(\overline{x})$$

Practice

- let $f(x) = x^2$; find $f'(1)$ using the definition of derivative

Continuity vs. Differentiability

Theorem 60. *If f is differentiable at $\overline{x}$, then f is continuous at $\overline{x}$.*

- converse is not true: $f(x) = |x|$ is continuous but not differentiable at 0

Algebra for Derivatives

Theorem 61. *Let $f, g : \mathbb{R} \to \mathbb{R}$. If f and g are differentiable at $\overline{x}$, then*

(1) $(cf)'(\overline{x}) = cf'(\overline{x})$ *for any* $c \in \mathbb{R}$

(2) $(f + g)'(\overline{x}) = f'(\overline{x}) + g'(\overline{x})$

(3) $(fg)'(\overline{x}) = f'(\overline{x})g(\overline{x}) + g'(\overline{x})f(\overline{x})$

(4) *if* $g(\overline{x}) \neq 0$, *then* $\left(\frac{f}{g}\right)'(\overline{x}) = \frac{g(\overline{x})f'(\overline{x}) - f(\overline{x})g'(\overline{x})}{g(\overline{x})^2}$

Chain Rule

- composition of two differentiable functions is also differentiable

Theorem 62. *Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ be functions, where $A, B \subseteq \mathbb{R}$ and $g(B) \subseteq A$. If g is differentiable at $\overline{x}$ and f is differentiable at $g(\overline{x})$, then*

$$\left[f(g(x))\right]'_{\overline{x}} = f'(g(\overline{x})) \cdot g'(\overline{x})$$

Interior Extremum Theorem

Theorem 63. *Suppose f is differentiable on $(a,b) \subset \mathbb{R}$. If f attains a local maximum or minimum at $\overline{x} \in (a,b)$, then $f'(\overline{x}) = 0$.*

- $\overline{x} \in X$ is a global max of $f : X \to \mathbb{R}$ if $\forall x \in X$, $f(\overline{x}) \geq f(x)$
- $\overline{x} \in X$ is a local max of $f : X \to \mathbb{R}$ if $\exists \varepsilon > 0$ such that $\forall x \in X$ s.t. $|x - \overline{x}| < \varepsilon$, $f(\overline{x}) \geq f(x)$

First Order Condition

- "derivative equals zero" is called the first-order condition (FOC)
- first-order condition is necessary (for an interior extremum) but not sufficient
  - **example**: $f : [-1,1] \to \mathbb{R}$ given by $f(x) = x^3$, at $x = 0$

Darboux's Theorem

Theorem 64. *Let $f : [a,b] \to \mathbb{R}$ be a differentiable function. If*

$$f'(a) < \phi < f'(b),$$

then there exists some $\overline{x} \in (a,b)$ such that

$$f'(\overline{x}) = \phi.$$

Rolle's Theorem

Theorem 65. *Let $f : [a,b] \to \mathbb{R}$ be continuous on the closed interval $[a,b]$ and differentiable on the open interval (a,b). If*

$$f(a) = f(b),$$

then there exists some point

$$\overline{x} \in (a,b)$$

such that

$$f'(\overline{x}) = 0.$$

Mean Value Theorem (MVT)

- interesting consequence of Rolle's theorem: if $f : [a, b] \to \mathbb{R}$ is continuous and differentiable, then there is a point in $[a, b]$ at which the derivative equals the average slope of the function
- **intuition**: if you travel 50 miles in an hour, there was a point during your journey where your speed was exactly 50 mph

Theorem 66. *If f is continuous on $[a, b]$ and differentiable on (a, b), then there exists $\overline{x} \in (a, b)$ such that*

$$f'(\overline{x}) = \frac{f(b) - f(a)}{b - a}.$$

Applying MVT

- MVT can be used to prove useful results about derivatives

Theorem 67. *Suppose $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b). If*

$$f'(x) = 0 \quad \text{for every } x \in (a, b),$$

then f is a constant function on $[a, b]$.

- **proof:** fix any $c, d \in [a, b]$ such that $c < d$. By the MVT, there exists $\overline{x} \in (c, d)$ such that

  $$\ldots$$

  which implies $\ldots$

Sign of Derivative

Theorem 68. *Let X be an open interval, and let $f : X \to \mathbb{R}$ be differentiable.*

(a) If $f'(x) \geq 0$ for all $x \in X$, then f is non-decreasing:

$$x_1 < x_2 \implies f(x_1) \leq f(x_2).$$

If $f'(\overline{x}) > 0$ for some $\overline{x} \in (x_1, x_2)$, then $f(x_1) < f(x_2)$.

(b) *If $f'(x) \leq 0$ for all $x \in X$, then f is non-increasing:*

$$x_1 < x_2 \implies f(x_1) \geq f(x_2).$$

If $f'(\overline{x}) < 0$ for some $\overline{x} \in (x_1, x_2)$, then $f(x_1) > f(x_2)$.

Second Derivative

Definition 69. Let X be an interval, and let $f : X \to \mathbb{R}$ be a differentiable function.
The second-order derivative of f at $\overline{x} \in X$, denoted $f''(\overline{x})$, is the derivative of the function $g : X \to \mathbb{R}$ given by $g(x) = f'(x)$.
If $f''(\overline{x})$ exists $\forall \overline{x} \in X$, we say that f is twice differentiable.

- alternative notation:

$$f_{xx}(\overline{x}) \qquad \left.\frac{d^2 f}{dx^2}\right|_{x=\overline{x}} \qquad \frac{d^2 f}{dx^2}(\overline{x})$$

- f is twice continuously differentiable if f is twice differentiable and $f''(x)$ is a continuous function (common requirement in optimization)

Existence of Derivative

- does a continuous function on an interval in $\mathbb{R}$ have to have a derivative at some point in its domain?
  - **no**: *Weierstrass function* is continuous everywhere in $\mathbb{R}$, but differentiable nowhere
  - also, differentiable nowhere functions are used in economics (especially financial economics)
  - **example**: asset prices are often modeled using a stochastic process called *Brownian motion*; with probability 1, the sample paths of a Brownian motion are continuous everywhere but differentiable nowhere

3.4 L'Hopital

Limits of Quotients

- (from Lecture 4) for any functions $f, g : (a, b) \to \mathbb{R}$

$$\lim_{x \to \overline{x}} f(x) = a, \quad \lim_{x \to \overline{x}} g(x) = b \neq 0 \quad \implies \quad \lim_{x \to \overline{x}} \frac{f(x)}{g(x)} = \frac{a}{b}$$

- what happens when $b = 0$?
- what happens when a and b are ∞?

Convergence to Infinity

Definition 70. Suppose that (X, d_X) is a metric space, $D \subseteq X$, $\overline{x}$ is a limit point of D, and $f : D \to \mathbb{R}$ is a function. Then, $\underline{\lim_{x \to \overline{x}} f(x) = +\infty}$ if, for all $M > 0$, there exists δ such that

$$d_X(x, \overline{x}) < \delta \quad \Longrightarrow \quad f(x) > M.$$

L'Hopital's Rule

Theorem 71. *Suppose f and g are continuous on (a, b) and let $\overline{x} \in (a, b)$. If all of the following conditions*

- $\lim\limits_{x \to \overline{x}} f(x) = \lim\limits_{x \to \overline{x}} g(x) = 0$ *or* $\pm\infty$,
- $g(x) \neq 0$ *for all* $x \in (a, b) \setminus \{\overline{x}\}$,
- f *and* g *are differentiable on* $(a, b) \setminus \{\overline{x}\}$
- *the limit* $\lim\limits_{x \to a} \frac{f'(x)}{g'(x)}$ *exists,*

hold, then

$$\lim_{x \to \overline{x}} \frac{f'(x)}{g'(x)} \quad = \quad \lim_{x \to \overline{x}} \frac{f(x)}{g(x)}.$$

Scope of Application

- **caution**: l'Hopital does not imply that

  $$\lim_{x \to \overline{x}} \frac{f'(x)}{g'(x)}$$

  exists; it only says what happens *when* this limit exists
- **example**: $f(x) = x + \sin(x)$ and $g(x) = x$,

  $$\frac{f'(x)}{g'(x)} = \cos(x),$$

  which does not have a limit (finite or infinite) as $x \to \infty$
- failure of l'Hopital's rule does not imply that the limit doesn't exist; we just cannot use l'Hopital to find it
- we can, however, calculate the limit directly:

  $$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1$$

4 Linear Algebra

4.1 Vector Spaces

Vector Space

A vector space over $\mathbb{R}$ is a non-empty set V of objects (called *vectors*), on which are defined two operations, $+$ (called *addition*) and $\cdot$ (called *scalar multiplication*), satisfying the following axioms for all $\vec{u}, \vec{v}, \vec{w} \in V$ and $a, b \in \mathbb{R}$

A0 (closed under $+$): $\vec{u} + \vec{v} \in V$

A1 (commutativity of $+$): $\vec{u} + \vec{v} = \vec{v} + \vec{u}$

A2 (associativity of $+$): $\vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w}$

A3 (existence of additive identity): $\exists \vec{0} \in V$ such that $\vec{0} + \vec{v} = \vec{v}$

A4 (existence of additive inverse): $\forall \vec{v} \in V$, $\exists (-\vec{v}) \in V$ s.t. $\vec{v} + (-\vec{v}) = \vec{0}$

[-0.5cm]

M0 (closed under $\cdot$): $a \cdot \vec{v} \in V$

M1 ($\cdot$ by 1): $1 \cdot \vec{v} = \vec{v}$

M2 (associativity of $\cdot$): $a \cdot (b \cdot \vec{v}) = (a \cdot b) \cdot \vec{v}$

M3 (distributivity-1 of $\cdot$): $a \cdot (\vec{u} + \vec{v}) = a \cdot \vec{u} + a \cdot \vec{v}$

M4 (distributivity-2 of $\cdot$): $(a + b) \cdot \vec{v} = a \cdot \vec{v} + b \cdot \vec{v}$

Standard Example

- $\mathbb{R}^n$ is a vector space
- addition is defined as
$$\vec{u} + \vec{v} = (u_1 + v_1, \ldots, u_n + v_n)$$
- multiplication is defined as
$$a \cdot \vec{v} = (av_1, \ldots, av_n)$$

Other Examples

- set of all $m \times n$ matrices $M_{m,n}$
  - element-wise addition and scalar multiplication
- set of all sequences $\{x_n\}$ of real numbers
  - element-wise addition and scalar multiplication
- set of all all real functions $f : \mathbb{R} \to \mathbb{R}$
  - $(f + g)(x) := f(x) + g(x)$ and $(af)(x) = af(x)$

- set of all solutions to a system of equations

Vector Subspace

Definition 72. A subspace of a vector space V is a subset of V which is itself a vector space, under the same operations as in V.

Theorem 73. *W is a subspace of vector space V if and only if $\forall a, b \in \mathbb{R}$ and all $\vec{u}, \vec{v} \in W$, $a\vec{u} + b\vec{v} \in W$.*

- need to check that W is closed under $+$ and $\cdot$ (conditions **A0** and **M0**)

Span

- what set of vectors can we construct from a fixed set of vectors?

Definition 74. A linear combination of vectors $\{\vec{v}^1, \ldots, \vec{v}^n\}$ is a vector of the form $a_1\vec{v}^1 + \ldots + a_n\vec{v}^n$, where $a_i \in \mathbb{R}$ for all $i \in \{1, \ldots, n\}$.

Definition 75. Let U be a set of vectors.

- The span of U, denoted $\text{span}(U)$, is the set of all linear combinations of finite subsets of U.
- We say that U spans a vector space V if $V \subseteq \text{span}(U)$.

Linear Independence

Definition 76. Set of vectors $\{\vec{v}^{\,1}, \ldots, \vec{v}^{\,n}\}$ is linearly independent if

$$\sum_{i=1}^{n} a_i \vec{v}^{\,i} = \vec{0} \quad \Longrightarrow \quad a_i = 0 \text{ for all } i \in \{1, \ldots, n\}.$$

Otherwise, it is linearly dependent.

- note that zero on the right-hand side is the zero *vector*
- **example**: are vectors $(2, 1)$ and $(1, 3)$ linearly independent?

Basis

Definition 77. Let V be a vector space. $U \subseteq V$ is a basis for V if (1) U spans V and (2) every finite subset of U is linearly independent.

- **example**: $\{e^i\}_{i=1}^n$, where e^i is the real vector that has 1 in the i-th position and 0 elsewhere, is a basis for $\mathbb{R}^n$
- every basis of V has the same number of (linearly independent) vectors which we call dimension and denote $\dim(V)$
  - in an n-dimensional space, every set of more than n vectors is dependent

4.2 Matrices

Matrices

- an $m \times n$ real matrix is an array

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

  where $a_{ij} \in \mathbb{R}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$
- here,

$$\begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \text{ is the } i\text{-th row of } A \qquad \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} \text{ is the } j\text{-th column of } A$$

Matrix Operations

- matrix addition

$$\underbrace{A}_{m\times n} + \underbrace{B}_{m\times n} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$$

- multiplication of matrix by a scalar $\lambda \in \mathbb{R}$

$$\lambda \cdot \underbrace{A}_{m\times n} = \begin{bmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{bmatrix}$$

Matrix Multiplication

- To multiply matrix A by matrix B, the number of columns of A must equal the number of rows of B:

$$\underbrace{A}_{m\times n} \cdot \underbrace{B}_{n\times l} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}b_{j1} & \sum_{j=1}^{n} a_{1j}b_{j2} & \cdots & \sum_{j=1}^{n} a_{1j}b_{jl} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} a_{mj}b_{j1} & \sum_{j=1}^{n} a_{mj}b_{j2} & \cdots & \sum_{j=1}^{n} a_{mj}b_{jl} \end{bmatrix}$$

Properties of Matrix Multiplication

- in general, $AB \neq BA$
- $(AB)C = A(BC)$
- $A(B+C) = (AB) + (AC)$
- $(A+B)C = (AC) + (BC)$

Matrix Multiplication: Notable Case

- **inner product** aka. dot product of two vectors of the same length n

$$\vec{u} \cdot \vec{v} = \underbrace{\begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}}_{1\times n} \cdot \underbrace{\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}}_{n\times 1} = u_1v_1 + \ldots + u_nv_n = \sum_{i=1}^{n} u_iv_i$$

  - also denoted by $\langle \vec{u}, \vec{v} \rangle$
  - note that when we do matrix multiplication to find $C = AB$, c_{ij} is the inner product of the i-th row of A and the j-th column of B

Rank

Definition 78. Let A be an $m \times n$ matrix. Then,

- column rank of A is the number of linearly independent columns of A.
- row rank of A is the number of linearly independent rows of A.
- column rank $=$ row rank $=$ rank (A) for any matrix of any size.

Transpose

Definition 79. The transpose of an $m \times n$ matrix A is the $n \times m$ matrix B such that $b_{ij} = a_{ji}$. The transpose of A is denoted A' or A^T.

Linear Transformation

Definition 80. $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if

$$f(x) = \begin{pmatrix} \sum_{i=1}^{n} a_{1i}x_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi}x_i \end{pmatrix}.$$

- f is a linear transformation $\iff$ $f(x) = Ax$, where A is an $m \times n$ matrix:

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

- here, x is an $n \times 1$ vector which in this lecture we denote by $\vec{x}$

Using Matrices to Solve Systems of Linear Equations

- system with m equations and n unknowns:

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases}$$

- matrix form:

$$A\vec{x} = \vec{b}, \quad A \in \mathbb{R}^{m \times n},$$

  where $\vec{x} \in \mathbb{R}^n$ is the variable vector and $\vec{b} \in \mathbb{R}^m$ the constants vector.
- solution methods:
  - Gaussian elimination
  - matrix inversion (if square and invertible)
  - Cramer's rule (small systems)

When Does a Solution Exist?

For the system $A\vec{x} = \vec{b}$, where A is an $m \times n$ matrix:

- **At least one solution** exists if

$$\text{rank}(A) = \text{rank}([A \mid \vec{b}]),$$

  where $[A \mid \vec{b}]$ is the augmented matrix ($\vec{b}$ added to A as an extra column)

- **Solution is unique** if consistent and

$$\text{rank}(A) = n.$$

- Otherwise,
  - no solution if $\text{rank}(A) \neq \text{rank}([A \mid \vec{b}])$
  - infinitely many solutions if consistent but $\text{rank}(A) < n$

Farkas' Lemma

Theorem 81. *Let A be a real $m \times n$ matrix and $\vec{b} \in \mathbb{R}^m$. Exactly one of the following is true:*

- *there exists $\vec{x} \in \mathbb{R}^n_+$ such that $\underbrace{A}_{m\times n} \underbrace{\vec{x}}_{n\times 1} = \underbrace{\vec{b}}_{m\times 1}$;*
- *there exists $\vec{y} \in \mathbb{R}^m$ such that $\underbrace{A^T}_{n\times m} \underbrace{\vec{y}}_{m\times 1} \geq \underbrace{\vec{0}}_{n\times 1}$ and $\vec{b} \cdot \vec{y} < 0$.*

- useful for proving that a system of equations has (or lacks) a positive solution
- comes in a few slightly different variations—pick the one that works for the problem at hand

Noteworthy Matrices

- **square matrix** has the same number of rows as columns ($n \times n$)
- **identity matrix** I_n is a square $n \times n$ matrix that has ones on the diagonal and zeros everywhere else
  - it is called that because $AI_n = I_n A = A$ for every $n \times n$ matrix A
  - often just called I without the subscript
- a square matrix A is **symmetric** if $a_{ij} = a_{ji}$ for all i and j
- a square matrix A is **diagonal** if it is all zeros except along the diagonal
  - every diagonal matrix is symmetric
- a square matrix A is (upper) **triangular** if all entries below the diagonal are zero

4.3 Square Matrices

Square Matrix: Determinant

Definition 82. The determinant of an $n \times n$ matrix A is recursively defined by

$$\det(A) = a_{11} \text{ if } n = 1,$$

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} \cdot a_{1i} \cdot \det(A_{-1i}), \text{ if } n \geq 2.$$

where A_{-ij} is the $(n-1) \times (n-1)$ matrix formed from A by deleting the i-th row and j-th column.

- alternative notation: $|A|$
- **example**: if $n = 2$, $\det(A) =$ ______________________________

Square Matrix: Trace

Definition 83. The trace of an $n \times n$ matrix A is the sum of its diagonal elements:

$$\operatorname{tr}(A) := \sum_{i=1}^{n} a_{ii}.$$

Square Matrix: Inverse

Definition 84.
- An $n \times n$ matrix A is invertible if there exists an $n \times n$ matrix B such that

$$AB = BA = I_n.$$

- If A is invertible, then B is unique, is called the inverse of A, and denoted by A^{-1}.

Theorem 85. *A is invertible $\iff$ $\det(A) \neq 0$.*

Properties of Inverses

Theorem 86. *If A is invertible, then*

- $(A^{-1})^{-1} = A$;
- $(kA)^{-1} = \frac{1}{k} A^{-1}$;
- $(A^T)^{-1} = (A^{-1})^T$;
- $\det(A) = \frac{1}{\det(A^{-1})} = \det(A^T)$.

If $A_1, \ldots, A_m$ are all invertible, then

$$(A_1 \cdot \ldots \cdot A_m)^{-1} = A_m^{-1} \cdot \ldots \cdot A_1^{-1}.$$

If A is an invertible 2×2 matrix, then its inverse is

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Eigenvalue

Definition 87. Let A be an $n \times n$ matrix. If $A\vec{v} = \lambda\vec{v}$ for some $\vec{v} \neq \vec{0}$ and some real or complex number λ, then

- $\vec{v}$ an eigenvector of A;
- λ is the corresponding (to $\vec{v}$) eigenvalue.
- $A\vec{v} = \lambda\vec{v} \iff (A - \lambda I_n)\vec{v} = \vec{0} \iff \det(A - \lambda I_n) = 0$
  - $\det(A - \lambda I)$ is an n-degree polynomial in λ
  - there is at least one and at most n solutions to an n-degree polynomial; eigenvalues are often complex
  - once you have eigenvalues, return to $(A - \lambda I)\vec{v} = 0$ to compute the corresponding eigenvectors (they will not be unique)

Properties of Eigenvalues

Theorem 88. *$\det(A)$ is the product and $\operatorname{tr}(A)$ is the sum of its eigenvalues.*

Theorem 89. *A is invertible $\iff$ A has no zero eigenvalues $\iff$ $\det(A) \neq 0$.*

- 0 is an eigenvalue of $A \iff A\vec{v} = 0$ for some $\vec{v} \in \mathbb{R}^n$ so that
  - $\operatorname{rank}(A) < n$
  - some rows (and columns) of A are linearly dependent

More on Eigenvalues and Eigenvectors

- the following are true for an invertible square matrix A:
  - if A is triangular, then diagonal elements of A are its eigenvalues
  - λ is eigenvalue of A with eigenvector $\vec{v} \iff \frac{1}{\lambda}$ is eigenvalue of A^{-1} with eigenvector $\vec{v}$
  - λ is eigenvalue of $A \implies \lambda$ is eigenvalue of A^T
  - λ is eigenvalue of A with eigenvector $\vec{v} \implies \forall a \neq 0$, $a\lambda$ is eigenvalue of aA with eigenvector $\vec{v}$

Diagonalization

Theorem 90. *Let A by an $n \times n$ matrix. If A has n linearly independent eigenvectors $\vec{v}_1, \dots, \vec{v}_n$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, then*

$$A = Q\Lambda Q^{-1},$$

where

- *Q is an $n \times n$ matrix with $\vec{v}_j$ as its j-th column*
- *Λ is an $n \times n$ matrix with λ_i in position (i, i) and zeros everywhere else*

- this theorem applies to square matrices with "full rank" of n
- this result is most useful for symmetric $n \times n$ matrices, which always have n linearly independent eigenvectors

4.4 Square Symmetric Matrices: Definiteness

Matrix Definiteness

Definition 91. Let A be an $n \times n$ real symmetric matrix. A is

- positive definite ($A > 0$) if $\vec{x}^T A \vec{x} > 0$ for all $\vec{x} \in \mathbb{R}^n \setminus \{\vec{0}\}$;
- positive semidefinite ($A \geq 0$) if $\vec{x}^T A \vec{x} \geq 0$ for all $\vec{x} \in \mathbb{R}^n$;
- negative definite ($A < 0$) if $\vec{x}^T A \vec{x} < 0$ for all $\vec{x} \in \mathbb{R}^n \setminus \{\vec{0}\}$;
- negative semidefinite ($A \leq 0$) if $\vec{x}^T A \vec{x} \leq 0$ for all $\vec{x} \in \mathbb{R}^n$;
- indefinite otherwise.

Definiteness via Eigenvalues

- eigenvalues of a matrix provide a neat way to check definiteness

Theorem 92. *Let A be an $n \times n$ real symmetric matrix. Then,*

- *A is positive (negative) definite if and only if all its eigenvalues are strictly positive (negative);*
- *A is positive (negative) semidefinite if and only if all its eigenvalues are weakly positive (negative).*

Definiteness via Principal Minors

Definition 93. Let A be an $n \times n$ real symmetric matrix, and let A_k be the matrix constructed from A by taking only the first k rows and columns. A_k is called the k-th leading principal minor of A.

Theorem 94. *An $n \times n$ real symmetric matrix A is*

- *negative definite if and only if $(-1)^k \det(A_k) > 0$ for all $k \in \{1, \ldots, n\}$;*
- *positive definite if and only if $\det(A_k) > 0$ for all $k \in \{1, \ldots, n\}$.*

- unfortunately, there is no analogous conditions for positive/negative semidefiniteness by weakening the inequalities

5 Multivariate Calculus

Multivariate Functions

- if we have m functions from $\mathbb{R}^n$ to $\mathbb{R}$, we can stack them together into a single function from $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

- note that in this lecture, $x \in \mathbb{R}^n$ is a vector; to simplify notation, we no longer use arrows

Differentiation

Definition 95. Let $X \subseteq \mathbb{R}^n$ be an open set and $\bar{x} \in X$. Then, the function $f : X \to \mathbb{R}^m$ is differentiable at $\bar{x}$ if there exists an $m \times n$ matrix $\mathcal{D}$ such that

$$\lim_{x \to \bar{x}} \frac{\|f(x) - f(\bar{x}) - \mathcal{D} \cdot (x - \bar{x})\|}{\|x - \bar{x}\|} = 0.$$

The (total) derivative of f at $\bar{x}$ is matrix $\mathcal{D}$, denoted as $f'(\bar{x})$ or $Df(\bar{x})$.
f is said to be differentiable if f is differentiable at any $\bar{x} \in X$.

- $||y||$ is Euclidean norm of the vector $y \in \mathbb{R}^n$: $||y|| = \sqrt{\sum_{i=1}^{n} y_i^2}$
- $\mathcal{D}$ is the derivative of f if linear function $f(\bar{x}) + \mathcal{D} \cdot (x - \bar{x})$ of x approximates $f(x)$ well when $x \to \bar{x}$

Total Derivative is Unique if Exists

Theorem 96. *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\bar{x} \in \mathbb{R}^n$, then there exists a unique $m \times n$ matrix $\mathcal{D}$ such that*

$$\lim_{x \to \bar{x}} \frac{\|f(x) - f(\bar{x}) - \mathcal{D} \cdot (x - \bar{x})\|}{\|x - \bar{x}\|} = 0.$$

Properties of Total Derivative

- If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are differentiable at $\bar{x} \in \mathbb{R}^n$, then
  - f and g are continuous at $\bar{x}$
  - $f + g$ is differentiable at $\bar{x}$, and $(f + g)'(\bar{x}) = f'(\bar{x}) + g'(\bar{x})$
  - $\forall \lambda \in \mathbb{R}$, $(\lambda f)'(\bar{x})$ is differentiable, and $(\lambda f)'(\bar{x}) = \lambda f'(\bar{x})$

Chain rule

Theorem 97. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^k \to \mathbb{R}^n$ be functions. If g is differentiable at $\overline{x} \in \mathbb{R}^k$, and if f is differentiable at $g(\overline{x})$, then $f\big(g(x)\big)$ is differentiable at $\overline{x}$ and*

$$(f \circ g)'(\overline{x}) = f'(g(\overline{x})) \cdot g'(\overline{x}).$$

- the derivative of g at $\overline{x}$ is an $n \times k$ matrix and the derivative of f at $g(\overline{x})$ is an $m \times n$ matrix
- composing f and g means applying their derivatives in sequence (multiplying a $m \times n$ matrix by a $n \times k$ one):

$$(f \circ g)'(\overline{x}) = f'(g(\overline{x})) \cdot g'(\overline{x})$$

- resulting $m \times k$ matrix is the derivative of the composition from $\mathbb{R}^k$ to $\mathbb{R}^m$

5.1 Functions $f : \mathbb{R}^n \to \mathbb{R}$ (scalar-valued multivariate functions)

Functions to $\mathbb{R}$: Partial Derivative

Definition 98. Let $f : \mathbb{R}^n \to \mathbb{R}$, $\overline{x} \in \mathbb{R}^n$, and $i \in \{1, \dots, n\}$. The partial derivative of f with respect to x_i at $\overline{x}$ is

$$\frac{\partial f}{\partial x_i}(\overline{x}) := \lim_{\Delta x \to 0} \frac{f(\overline{x} + \Delta x \cdot e^i) - f(\overline{x})}{\Delta x}$$

if this limit exists.

The vector of partial derivatives of f at $\overline{x}$ is called the gradient of f and denoted $\nabla f(\overline{x})$:

$$\nabla f(\overline{x}) := \begin{bmatrix} \frac{\partial f}{\partial x_1}(\overline{x}) & \cdots & \frac{\partial f}{\partial x_n}(\overline{x}) \end{bmatrix}.$$

- $e^i \in \mathbb{R}^n$ is i-th canonical basis vector of $\mathbb{R}^n$ (1 at i-th position, zeros o/w)
- $\frac{\partial f}{\partial x_i}(\overline{x})$ measures sensitivity of f w.r.t. x_i at $\overline{x}$

Example

Given the function

$$f(x,y) = \begin{cases} \frac{x^2 y}{x^4 + y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0), \end{cases}$$

find the partial derivatives $\frac{\partial f}{\partial x}(0,0)$ and $\frac{\partial f}{\partial y}(0,0)$.

Example: Solution

Partial derivative with respect to x at $(0,0)$:

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h \to 0} \frac{\frac{h^2 \cdot 0}{h^4 + 0^2} - f(0,0)}{h} = \lim_{h \to 0} \frac{0 - 0}{h} = 0.$$

Since $f(h,0) = 0$ for all h, this limit equals zero.

Partial derivative with respect to y at $(0,0)$:

$$\frac{\partial f}{\partial y}(0,0) = \lim_{h \to 0} \frac{f(0,h) - f(0,0)}{h} = \lim_{h \to 0} \frac{0 - 0}{h} = 0.$$

Since $f(0,h) = 0$ for all h, this limit also equals zero.

Hence,

$$\frac{\partial f}{\partial x}(0,0) = 0 \quad \text{and} \quad \frac{\partial f}{\partial y}(0,0) = 0.$$

Total vs. Partial Derivative

Definition 99. Let $f : \mathbb{R}^n \to \mathbb{R}$, $\overline{x} \in \mathbb{R}^n$, and $i \in \{1, \ldots, n\}$. If f is differentiable at $\overline{x}$, then all partial derivatives $\frac{\partial f}{\partial x_i}(\overline{x})$ exist and

$$f'(\overline{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\overline{x}) & \cdots & \frac{\partial f}{\partial x_n}(\overline{x}) \end{bmatrix} = \nabla f(\overline{x}).$$

- the total derivative $f'(\overline{x})$ is the <u>vector</u> of partial derivatives, i.e. the gradient $\nabla f(\overline{x})$
- total derivative and gradient both represent the best linear approximation of f at $\overline{x}$
- **for scalar-valued functions $f : \mathbb{R}^n \to \mathbb{R}$ only**

Cross Partials

- if the i-th partial derivative of f exists everywhere, we get a function $\frac{\partial f}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}$
- the j-th partial derivative of $\frac{\partial f}{\partial x_i}$ at $\overline{x}$ is the second-order partial $\frac{\partial^2 f}{\partial x_i \partial x_j}(\overline{x})$

Theorem 100. *If $\frac{\partial^2 f}{\partial x_i \partial x_j}$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}$ are continuous on an open set containing $\overline{x}$, then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\overline{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\overline{x}).$$

First-Order Condition

Theorem 101. *Let $f : X \to \mathbb{R}$ be a function, where $X \subseteq \mathbb{R}^n$. If f has a local extremum at $x \in int(X)$, and if $\nabla f(x)$ exists, then $\nabla f(x) = 0$.*

- if f has a local extremum at x inside its domain (not on the boundary), and if the gradient $\nabla f(x)$ exists, then $\nabla f(x)$ must be zero
- the gradient $\nabla f(x)$ points toward steepest increase, so at a local extremum it is zero since no immediate increase or decrease is possible
- this gives candidate points for local extrema called critical or stationary points (where the gradient $\nabla f(x)$ is zero) – not all such points are minima or maxima, some may be *saddle points*

Mean Value Theorem

Theorem 102. *Let $f : X \to \mathbb{R}$ be a differentiable function, where $X \subset \mathbb{R}^n$ is open and convex. For any $x, y \in X$, there exists $\lambda \in (0, 1)$ such that*

$$f(y) - f(x) = \langle \nabla f(\lambda x + (1 - \lambda)y), y - x \rangle.$$

- MVT: between two points on a curve, there is at least one point where the tangent (derivative) matches the average rate of change over the interval
- one-dimensional: $f(b) - f(a) = f'(c)(b - a)$ for some $c \in (a, b)$
- multivariate: derivative $\to$ gradient, and multiplication $\to$ inner product

5.2 Back to Functions $f : \mathbb{R}^n \to \mathbb{R}^m$

Jacobian

Definition 103. Suppose all of the partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}^m$ exist everywhere. The Jacobian of f is the $m \times n$ matrix given by

$$J = \begin{bmatrix} \nabla f_1 \\ \vdots \\ \nabla f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

Hessian

Definition 104. Suppose all of the partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}$ exist everywhere. The Hessian of f is the $n \times n$ matrix given by

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- if cross-partials are continuous, the order in which the partial derivatives are taken does not matter, so H is symmetric

Inverse Function Theorem

Theorem 105. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on an open set containing $\overline{x}$, and suppose*

$$\det\left(J(\overline{x})\right) \neq 0.$$

Then there exist open sets $U \ni \overline{x}$ and $V \ni f(\overline{x})$ such that $f : U \to V$ has a continuous differentiable inverse $f^{-1} : V \to U$. Moreover, for all $v \in V$,

$$J(f^{-1})(v) = \left(Jf(f^{-1}(v))\right)^{-1}.$$

- INVERSE FUNCTION THEOREM guarantees a local inverse near $\overline{x}$ if the jacobian there is invertible ($\det\left(J(\overline{x})\right) \neq 0$)
- this inverse allows us to “undo” the function locally
- the jacobian of the inverse is the matrix inverse of the original jacobian

Inverse Function Theorem: Example

Consider the function

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \quad f(x, y) = (2x + y,\ x - y)$$

Step 1: jacobian matrix

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}$$

Step 2: check determinant

$$\det(J) = 2 \cdot (-1) - 1 \cdot 1 = -2 - 1 = -3 \neq 0$$

since the jacobian is invertible, f has a differentiable inverse near every point by the INVERSE FUNCTION THEOREM

Inverse Function Theorem: Example, Continued

Step 3: solve for (x, y) given $(u, v) = f(x, y)$:

$$\begin{cases} u = 2x + y \\ v = x - y \end{cases}$$

The solution is

$$f^{-1}(u, v) = \left(\frac{u + v}{3}, \frac{u - 2v}{3} \right).$$

Step 4: verify jacobian of inverse

$$J^{-1} = \frac{1}{-3} \begin{bmatrix} -1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{bmatrix}$$

which matches the partial derivatives of f^{-1}

Implicit Function Theorem

Theorem 106. *Let $f : X \to \mathbb{R}^n$ be a continuously differentiable function, where $X \subset \mathbb{R}^{m+n}$ is open. If $f(x^*, y^*) = 0$ and*

$$J_y(x^*, y^*) := \begin{bmatrix} \frac{\partial f_1}{\partial y_1}(x^*, y^*) & \cdots & \frac{\partial f_1}{\partial y_n}(x^*, y^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_1}(x^*, y^*) & \cdots & \frac{\partial f_n}{\partial y_n}(x^*, y^*) \end{bmatrix}$$

is invertible, then there exists an open set $U \subset \mathbb{R}^m$ containing x^* and a unique continuously differentiable $g : U \to \mathbb{R}^n$ such that $g(x^*) = y^*$ and $f(x, g(x)) = 0$ for all $x \in U$. Also, for all $x \in U$ and $i = 1, \ldots, n$

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_i}(x) \\ \vdots \\ \frac{\partial g_n}{\partial x_i}(x) \end{bmatrix} = -\left[J_y(x, g(x)) \right]^{-1} \begin{bmatrix} \frac{\partial f_1}{\partial x_i}(x, g(x)) \\ \vdots \\ \frac{\partial f_n}{\partial x_i}(x, g(x)) \end{bmatrix}.$$

Implicit Function Theorem in Two Variables

- Suppose $f : X \to \mathbb{R}$ is continuously differentiable on an open set $X \subseteq \mathbb{R}^2$ with coordinates (x, y)
- Assume $f(x^*, y^*) = 0$ and $\frac{\partial f}{\partial y}(x^*, y^*) \neq 0$
- Then there exists an open interval U containing x^* and a differentiable function $g : U \to \mathbb{R}$ such that:
  - $g(x^*) = y^*$
  - For all $x \in U$, $f(x, g(x)) = 0$
  - For all $x \in U$,
$$g'(x) = -\frac{\frac{\partial f}{\partial x}(x, g(x))}{\frac{\partial f}{\partial y}(x, g(x))}$$
- **Example:** If $f(x, y) = x^2 + y = 0$, then $g'(x) = -2x$

Using the Implicit Function Theorem

- Let x be a vector of parameters and y a vector of choices or outcomes (e.g., demand and supply)
- Suppose the system $f(x, y) = 0$ characterizes optimal or equilibrium conditions for y given x
- If (x^*, y^*) solves the system and the IFT conditions hold at this point, then locally around x^* the solutions y can be expressed as a function $g(x)$
- The derivatives of g describe how optimal choices change in response to parameter variations near x^*

5.3 Miscellaneous, for $f : \mathbb{R}^n \to \mathbb{R}$ functions only

Classes of Functions

- <u>differentiability class</u> indicates the highest order of continuous derivatives a function possesses

Definition 107. Let $f : X \to \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ open. We say f is of class C^k if all derivatives up to order k, i.e., $f', f'', \dots, f^{(k)}$, exist and are continuous on X.

For example:

- $f \in C^0$ means f is continuous
- $f \in C^2$ means f is twice continuously differentiable

Directional Derivative

- recall: the i-th partial derivative of f measures how f changes when we move along the i-th coordinate axis
- question: how does f change when we move in an *arbitrary* direction $\vec{v}$, not just along coordinate axes?

Definition 108. Let $f : X \to \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ open. The directional derivative of f at $\vec{x} \in X$ along $\vec{v} \in \mathbb{R}^n$ is

$$\nabla_{\vec{v}} f(\vec{x}) := \lim_{t \to 0} \frac{f(\vec{x} + t\vec{v}) - f(\vec{x})}{t},$$

if this limit exists.

Directional vs. Partial Derivative

Theorem 109. *If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{x} \in \mathbb{R}^n$, then the directional derivative of f at $\vec{x}$ along any vector $\vec{v}$ exists and is given by*

$$\nabla_{\vec{v}} f(\vec{x}) = \langle \nabla f(\vec{x}), \vec{v} \rangle,$$

where $\nabla f(\vec{x})$ is the gradient of f at $\vec{x}$.

- recall the inner product or dot product of two vectors $\vec{a}, \vec{b} \in \mathbb{R}^n$ is

$$\langle \vec{a}, \vec{b} \rangle = \sum_{i=1}^{n} a_i b_i.$$

- in particular, the directional derivative along the i-th coordinate axis is the i-th partial derivative

Example: Directional Derivative

- Find the directional derivative of

$$f(x_1, x_2) = x_1^2 x_2$$

  in the direction $\vec{v} = (1, 2)$ at the point $(3, 2)$.

Example: Directional Derivative, Solution

Step 1: Compute the gradient of f

$$\nabla f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) = (2x_1 x_2, x_1^2)$$

Step 2: Evaluate the gradient at $(3,2)$

$$\nabla f(3,2) = (2 \times 3 \times 2, 3^2) = (12, 9)$$

Step 3: Normalize the direction vector $\vec{v}$

$$\|\vec{v}\| = \sqrt{1^2 + 2^2} = \sqrt{5}, \quad \hat{v} = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)$$

Step 4: Compute the directional derivative

$$D_{\vec{v}} f(3,2) = \nabla f(3,2) \cdot \hat{v} = (12, 9) \cdot \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) = \frac{12}{\sqrt{5}} + \frac{18}{\sqrt{5}} = \frac{30}{\sqrt{5}} = 6\sqrt{5}$$

So, the directional derivative is $\boxed{6\sqrt{5}}$.

Total Derivative

- For a small increment $\Delta\vec{x}$,

$$f(\vec{x} + \Delta\vec{x}) - f(\vec{x}) \approx \langle \nabla f(\vec{x}), \Delta\vec{x} \rangle = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{x}) \Delta x_i$$

- The linear map $Df(\vec{x})(\vec{v}) := \langle \nabla f(\vec{x}), \vec{v} \rangle$ is called the total derivative of f at $\vec{x}$

For your thought: How does this relate to our previous definition of the total derivative as the gradient vector?

Connecting Total Derivative Definitions

- Previously, the total derivative $f'(\vec{x})$ was introduced as the gradient vector $\nabla f(\vec{x})$
- Now, the total derivative $Df(\vec{x})$ is seen as a *linear map* acting on $\vec{v}$ via

$$Df(\vec{x})(\vec{v}) = \langle \nabla f(\vec{x}), \vec{v} \rangle$$

- The gradient vector is the coordinate representation of this linear map under the standard dot product; thus, both perspectives describe the same object

6 Unconstrained Optimization

6.1 Introduction & Motivation

Optimization Problem

Definition. An *optimization problem* can be mathematically stated as:

$$\max_{\vec{x}\in D} f(\vec{x}) \quad \text{or} \quad \min_{\vec{x}\in D} f(\vec{x})$$

where

- $f : X \to \mathbb{R}$ is the objective function (typically and in this course, $X \subseteq \mathbb{R}^n$)
- $D \subseteq X$ is the constraint (feasible) set
- the goal is to find $\vec{x}^* \in D$ such that $f(\vec{x}^*)$ attains the maximum (or minimum) value

Modern Economics $\approx$ Optimization

- **Micro (consumer/producer theory):** agents maximize utility/profit subject to budget/technology constraints
- **Game theory:** players choose best responses
- **Mechanism / Market / Institutional design:** a *designer* chooses rules/institutions to optimize objectives (welfare, revenue) subject to incentive-compatibility and feasibility constraints
- **Public economics & Political economy:** voters, politicians, governments optimize social, political, and personal objectives (e.g., redistribution, electoral success) subject to resource, institutional, and strategic constraints.
- **Econometrics / Empirical methods:** estimation is optimization – OLS / MLE / GMM minimize loss or maximize likelihood
- **Macro:** micro-founded models start from intertemporal optimization (agents maximize lifetime utility)

6.2 Unconstrained Optimization

Unconstrained Optimization Problem

- **Unconstrained Optimization Problem** has the form

  $$\max_{\vec{x}\in X} f(\vec{x}),$$

  where $f : X \to \mathbb{R}$ and $X \subseteq \mathbb{R}^n$
- we aim to answer the following questions:
  - is there a solution?
  - if there is a solution, is it unique?

- how do we find the solution(s)?

Existence

- we have already answered the existence question in Lecture 3

Theorem 110 (Extreme Value Theorem). *Let $X \subseteq \mathbb{R}^n$ be nonempty and compact. If $f : X \to \mathbb{R}$ is continuous, then there exist points*

$$\underline{x}, \overline{x} \in X$$

such that

$$f(\underline{x}) \leq f(x) \leq f(\overline{x}) \quad \textit{for all } x \in X$$

That is, f attains its minimum at $\underline{x}$ and its maximum at $\overline{x}$

- key ingredients:
  - compact domain X (closed and bounded)
  - continuous objective f

First-Order Condition

- if $\vec{x}^*$ is a local extremum and $\nabla f(\vec{x}^*)$ exists, then

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

- such $\vec{x}^*$ are called stationary points
- stationary points include local maxima, local minima, and saddle points

Exercise: Stationary Points

- **Problem**: find all stationary points of the function

$$f(x, y) = x^4 + y^4 - 4x^2 - 2y^2$$

- **Solution**:
  - compute gradient

$$\nabla f(x, y) = (4x^3 - 8x,\ 4y^3 - 4y)$$

  - set gradient equal to zero

$$4x(x^2 - 2) = 0, \quad 4y(y^2 - 1) = 0$$

- stationary points are:

$$(0,0),\ (0,1),\ (0,-1),\ (\sqrt{2},0),\ (\sqrt{2},1),$$
$$(\sqrt{2},-1),\ (-\sqrt{2},0),\ (-\sqrt{2},1),\ (-\sqrt{2},-1)$$

Second-Order Condition

Theorem 111 (Second-Order Condition)**.** *Let f be twice continuously differentiable and $\nabla f(\vec{x}^*) = 0$. Then:*

- *if the Hessian $\nabla^2 f(\vec{x}^*)$ is negative definite, $\vec{x}^*$ is a local maximum*
- *if the Hessian is positive definite, $\vec{x}^*$ is a local minimum*

- also:
  - if the Hessian is indefinite (has both positive and negative eigenvalues), $\vec{x}^*$ is a saddle point
  - **careful**: if the Hessian is semidefinite, the test is inconclusive

Checking the SOC – Eigenvalues

Input: Hessian $\nabla^2 f(\vec{x}^*)$ evaluated at a stationary point

- Determine definiteness by checking eigenvalues of the Hessian:

$$\begin{array}{ll}
\text{all eigenvalues } > 0 & \Longrightarrow \text{ local min} \\
\text{all eigenvalues } < 0 & \Longrightarrow \text{ local max} \\
\text{mix of } + \text{ and } - \text{ eigenvalues} & \Longrightarrow \text{ saddle point} \\
\text{at least one zero eigenvalue, rest } \geq \text{ or } \leq 0 & \Longrightarrow \text{ inconclusive}
\end{array}$$

Checking the SOC – Principal Minors

Input: Hessian $\nabla^2 f(\vec{x}^*)$ evaluated at a stationary point

- Determine definiteness using leading principal minors:

$$
\begin{array}{llll}
\text{all minors } > 0 & \Longrightarrow & \text{positive definite} \Longrightarrow & \text{local min} \\
\text{minors alternate signs } -, +, \ldots & \Longrightarrow & \text{negative definite} \Longrightarrow & \text{local max} \\
\text{positive and negative minors} & \Longrightarrow & \text{saddle point} & \\
\text{zero in a minor} & \Longrightarrow & \text{inconclusive} &
\end{array}
$$

Exercise: Hessian Classification

- **Problem**: classify the stationary points

$$(0,0), \quad (\sqrt{2},0), \quad (\sqrt{2},1)$$

of $f(x,y) = x^4 + y^4 - 4x^2 - 2y^2$ using the Second-Order Condition

- **Solution**:

$$H(x,y) = \begin{pmatrix} 12x^2 - 8 & 0 \\ 0 & 12y^2 - 4 \end{pmatrix}$$

$$H(0,0) = \begin{pmatrix} -8 & 0 \\ 0 & -4 \end{pmatrix} \quad \Longrightarrow \text{local max}$$

$$H(\sqrt{2},0) = \begin{pmatrix} 16 & 0 \\ 0 & -4 \end{pmatrix} \quad \Longrightarrow \text{saddle}$$

$$H(\sqrt{2},1) = \begin{pmatrix} 16 & 0 \\ 0 & 8 \end{pmatrix} \quad \Longrightarrow \text{local min}$$

General Procedure Summary

- **Step 1**: find all stationary points $\vec{x}^*$ by solving $\nabla f(\vec{x}^*) = 0$
- **Step 2**: classify each $\vec{x}^*$ using the Second-Order Condition:
  - via eigenvalues or principal minors – see previous slides

- **Step 3**: <u>for global extrema</u>, compare f at all local extrema and at boundaries

6.3 Convexity and Concavity

Definitions Using Inequality Form

Definition 112. Given a convex set $X \subseteq \mathbb{R}^n$, a function $f : X \to \mathbb{R}$ is <u>convex</u> (<u>concave</u>) if for all distinct $\vec{x}, \vec{y} \in X$ and all $\alpha \in (0,1)$

$$f(\alpha\vec{x} + (1-\alpha)\vec{y}) \;\leq\; (\geq)\; \alpha f(\vec{x}) + (1-\alpha) f(\vec{y}).$$

- for strict convexity (concavity), replace $\leq$ ($\geq$) with $<$ ($>$)
- inequality direction captures curvature: convex functions lie below chords, concave above
- definition requires domain to be convex

Hessian Criteria for Convexity/Concavity

Theorem 113. *Let $X \subseteq \mathbb{R}^n$ be convex and open. Let $f : X \to \mathbb{R}$ be C^2. Then,*

- *f is convex $\iff$ $\nabla^2 f(\vec{x})$ is positive semidefinite $\forall \vec{x} \in X$*
- *f is concave $\iff$ $\nabla^2 f(\vec{x})$ is negative semidefinite $\forall \vec{x} \in X$*
- *$\nabla^2 f(\vec{x})$ is positive definite $\forall \vec{x} \in X \implies f$ is strictly convex*
- *$\nabla^2 f(\vec{x})$ is negative definite $\forall \vec{x} \in X \implies f$ is strictly concave*

- note that this criterion is for f being convex (concave) everywhere on its domain
- the SOC discussed earlier applies at a single point $\vec{x}^*$

Example: Cobb–Douglas

- **Problem**: find $\alpha, \beta \geq 0$ for which

$$f(x, y) = x^\alpha y^\beta, \quad x, y > 0$$

  is concave

- **Solution**:
  - hessian

$$\nabla^2 f = \begin{bmatrix} \alpha(\alpha-1)x^{\alpha-2}y^\beta & \alpha\beta x^{\alpha-1}y^{\beta-1} \\ \alpha\beta x^{\alpha-1}y^{\beta-1} & \beta(\beta-1)x^\alpha y^{\beta-2} \end{bmatrix}$$

  - negative semidefinite if

$$\alpha(\alpha-1) \leq 0, \quad \det\left(\nabla^2 f\right) \geq 0$$

$$\Rightarrow \quad 0 \leq \alpha \leq 1, \quad \alpha + \beta \leq 1$$

  - answer:

$$\boxed{\alpha, \beta \geq 0, \quad \alpha \leq 1, \quad \alpha + \beta \leq 1}$$

Uniqueness Implications of Strict Concavity

Theorem 114. *Let $X \subseteq \mathbb{R}^n$ be convex and $f : X \to \mathbb{R}$ concave*

- *any local maximizer of f is a global maximizer*
- *if the set of maximizers is nonempty, it is convex*

If f is strictly concave, there is at most one maximizer

- two points are maximizers $\implies$ all points between them are maximizers
- strict concavity $\implies$ no flat parts $\implies$ at most one maximizer exists

Global Optimality and Concave Functions

- immediate implication: if f is a concave function defined on a convex set and $\vec{x}$ satisfies the FOC, then $\vec{x}$ is a global maximizer

Theorem 115. *Let $X \subseteq \mathbb{R}^n$ be convex, $\vec{x} \in \text{int}(X)$, and $f : X \to \mathbb{R}$ concave (convex). Then, $\vec{x}$ is a global maximizer (minimizer) of f if and only if $\nabla f(\vec{x}) = \vec{0}$.*

- for concave f, the FOC is necessary and sufficient for an *interior* maximum
- boundary optima may not satisfy the FOC

6.4 Quasi-Concavity — A Weaker Notion

Motivation

- concavity is strong and excludes many useful functions, e.g. $f(x_1, x_2) = x_1^\alpha x_2^\beta$ when $\alpha + \beta > 1$
- in economics, we care mainly about preserving orderings rather than exact functional form
- all positive monotonic transformations preserve order but might destroy concavity
- **quasi-concavity** weakens concavity to allow flexibility while preserving preference ordering

Definition: Quasi-Concavity

Definition 116. Let $X \subseteq \mathbb{R}^n$ be convex. A function $f : X \to \mathbb{R}$ is **quasi-concave** if, for every $a \in \mathbb{R}$,

$$\{\vec{x} \in X : f(\vec{x}) \geq a\} \quad \text{is convex.}$$

Equivalently, for all $\vec{x}, \vec{y} \in X$ and $\alpha \in (0, 1)$,

$$f(\alpha\vec{x} + (1-\alpha)\vec{y}) \;\geq\; \min\{f(\vec{x}), f(\vec{y})\}.$$

- *intuition:* function does not dip below the lower of two values when moving between any two points, so its 'high-value' regions are convex and hill-shaped
- strict quasi-concavity: replace "$\geq$" with "$>$"

Bordered Hessians

- Another characterization of quasi-concavity uses the bordered Hessian, defined as

$$B_f := \begin{bmatrix} 0 & \nabla f \\ (\nabla f)^T & \nabla^2 f \end{bmatrix}$$

- The first row and column form the *border* of the matrix
- The k-th leading principal minor of the bordered Hessian is the determinant of its top-left $(k+1) \times (k+1)$ submatrix

Bordered Hessian Characterization of Quasi-Concavity

Theorem 117. *Let $X \subseteq \mathbb{R}^n_+$ be convex and open, and let $f : X \to \mathbb{R}$ be C^2. Denote by D_k the k-th leading principal minor of the bordered Hessian $B_f(x)$ at any $x \in X$. Then:*

$$f \text{ is quasi-concave on } X \iff D_1 \leq 0, D_2 \geq 0, D_3 \leq 0, \ldots$$

- similar to characterization of concavity (Hessian $\to$ bordered Hessian)
- for strict quasi-concavity, replace weak inequalities with strict ones

Example: Quasi-Concavity Check

- **Problem**: check if
$$f(x) = -x^2, \quad x \in \mathbb{R}$$
is quasi-concave using the bordered Hessian
- **Solution**:
  - bordered Hessian matrix:
$$B_f(x) = \begin{bmatrix} 0 & \nabla f(x) \\ \nabla f(x) & \nabla^2 f(x) \end{bmatrix} = \begin{bmatrix} 0 & -2x \\ -2x & -2 \end{bmatrix}$$
  - leading principal minors:
$$\text{0th principal minor: } 0 \quad \text{(not checked in the test)}$$

$$\text{1st principal minor: det } (B_f) = (-2)(0) - (-2x)^2 = -4x^2 \leq 0$$

- conclusion:

$$f \text{ is quasi-concave}$$

Uniqueness Implications of Strict Quasi-Concavity

Theorem 118. *Let $X \subseteq \mathbb{R}^n$ be convex and $f : X \to \mathbb{R}$ quasi-concave*

- *any local maximizer of f is a global maximizer*
- *if the set of maximizers is nonempty, it is convex*

If f is strictly quasi-concave, there is at most one maximizer

- exactly the same result we had for concavity

6.5 Envelope Theorems

Parameters in Optimization

- **Definition** A *parameterized maximization problem* is

$$\max_{\vec{x} \in D(\vec{\theta})} f(\vec{x}; \theta)$$

  where θ is a parameter (often a vector)
- Both the solution $\vec{x}^*(\theta)$ and the value $f(\vec{x}^*(\theta); \theta)$ depend on parameters
  - *comparative statics*: how do solutions/predictions change with θ?

Envelope Theorems

- Changing θ usually means re-solving the maximization problem
- **Envelope theorems:** for the value function $v(\theta) = f(x^*(\theta); \theta)$,

$$\frac{dv}{d\theta} = \frac{\partial f}{\partial \theta}(x^*(\theta), \theta)$$

  — no need to account for $\frac{dx^*}{d\theta}$
- Key insight: optimality conditions "cancel out" the effect of $x^*(\theta)$'s change

Unconstrained Envelope Theorem

Theorem 119. *Let $f : \mathbb{R}^{m+n} \to \mathbb{R}$ be C^1, with $x \in \mathbb{R}^m$ and $\theta \in \mathbb{R}^n$. Suppose there exists a C^1 function $x^* : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$x^*(\theta) = \arg\max_{x \in \mathbb{R}^m} f(x, \theta),$$

and define $V(\theta) := f(x^*(\theta), \theta)$. Then

$$D_\theta V(\theta) = \left(\frac{\partial f}{\partial \theta_1}, \ldots, \frac{\partial f}{\partial \theta_n} \right) \bigg|_{x = x^*(\theta)}$$

- at the optimum, the effect of θ on V comes *only* from the direct partial derivatives – we ignore $dx^*/d\theta$

Example: Unconstrained Envelope Theorem

- **Problem**: maximize

$$f(x, \theta) = \theta x - x^2, \quad x \in \mathbb{R}$$

  and find $\dfrac{\partial f}{\partial \theta}$ at $x^*(\theta)$
- **Solution**:
  - First-order condition: $\frac{\partial f}{\partial x} = \theta - 2x = 0 \;\Rightarrow\; x^*(\theta) = \theta/2$
  - Envelope theorem: $\frac{dV}{d\theta} = \frac{\partial f}{\partial \theta}\big|_{x = x^*(\theta)}$
  - Compute partial derivative and substitute:

$$\frac{\partial f}{\partial \theta} = x \quad \Rightarrow \quad \frac{dV}{d\theta} = \frac{\theta}{2}$$

  - Answer:

$$\boxed{\frac{dV}{d\theta} = \frac{\theta}{2}}$$

7 Equality-Constrained Optimization

Equality-Constrained Problem

- as before, objective function is $f : X \to \mathbb{R}$, where $X \subseteq \mathbb{R}^n$ is open
- **[new!]** m constraints of the form $g_j(\vec{x}) = 0$ for $j = 1, \ldots, m$
  - alternative representation: $g(\vec{x}) = \begin{bmatrix} g_1(\vec{x}) \\ g_2(\vec{x}) \\ \vdots \\ g_m(\vec{x}) \end{bmatrix} = \vec{0} \in \mathbb{R}^m$

Simple Consumer Problem

- maximize $\log x_1 + \log x_2$ over $(x_1, x_2) \in \mathbb{R}^2_+$ subject to

$$p_1 x_1 + p_2 x_2 = I$$

Motivation

$$\max_{(x_1,x_2)\in\mathbb{R}^2} f(x_1, x_2) \quad \text{s.t.} \quad g(x_1, x_2) = 0$$

one approach:

- solve $g(x_1, x_2) = 0$ for $x_2(x_1)$, plug into f:

$$\max_{x_1\in\mathbb{R}} f\big(x_1, x_2(x_1)\big)$$

- FOC for this unconstrained problem:

$$f_{x_1} + f_{x_2} \cdot x_2'(x_1) \;=\; f_{x_1} - f_{x_2} \cdot \frac{g_{x_1}}{g_{x_2}} = 0$$

- equivalently, $\exists\, \lambda \in \mathbb{R}$ such that

$$\frac{f_{x_1}}{g_{x_1}} = \lambda = \frac{f_{x_2}}{g_{x_2}}$$

Lagrangian

$$\max_{(x_1,x_2)\in\mathbb{R}^2} f(x_1,x_2) \quad \text{s.t.} \quad g(x_1,x_2)=0$$

another approach:

- let
$$L(x_1,x_2,\lambda) = f(x_1,x_2) - \lambda\, g(x_1,x_2)$$
- FOCs w.r.t. x_1 and x_2:
$$\frac{f_{x_1}}{g_{x_1}} = \lambda = \frac{f_{x_2}}{g_{x_2}}$$
- FOC w.r.t. λ:
$$g(x_1,x_2)=0$$

Lagrangian: General Formulation

$$\max_{\vec{x}\in\mathbb{R}^n} f(\vec{x}) \quad \text{s.t.} \quad g_j(\vec{x})=0,\ j=1,\dots,m$$

general formulation:

- define
$$L(\vec{x},\vec{\lambda}) = f(\vec{x}) - \sum_{j=1}^{m} \lambda_j\, g_j(\vec{x})$$
- FOCs:
$$\frac{\partial f}{\partial x_i}(\vec{x}) = \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(\vec{x}), \quad i=1,\dots,n$$
$$g(\vec{x}) = \vec{0} \in \mathbb{R}^m$$
- L is the Lagrangian, $\vec{\lambda}\in\mathbb{R}^m$ are Lagrange multipliers

Exercise: FOCs with a Constraint

- **problem**: write down FOCs for maximization of

$$f(x_1, x_2) = x_1^2 - x_2^2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$

- **solution**:
  - lagrangian:

$$L(x_1, x_2, \lambda) = x_1^2 - x_2^2 - \lambda(x_1^2 + x_2^2 - 1)$$

  - FOCs:

$$\frac{\partial L}{\partial x_1} = 2x_1 - 2\lambda x_1 = 0$$

$$\frac{\partial L}{\partial x_2} = -2x_2 - 2\lambda x_2 = 0$$

$$\frac{\partial L}{\partial \lambda} = -(x_1^2 + x_2^2 - 1) = 0$$

Interpretation of Multiplier

$$\text{consider constraint } g_k(\vec{x}) = c, \quad c > 0$$

- Lagrangian: $L(\vec{x}, \vec{\lambda}, c) = f(\vec{x}) - \sum_{j=1}^{m} \lambda_j g_j(\vec{x}) + \lambda_k c$
  - value function: $V(c) = L(\vec{x}^*(c), \vec{\lambda}^*(c))$
  - envelope theorem: $V'(c) = \lambda_k^*$
- **interpretation**: λ_k^* is value of relaxing k-th constraint
  - relaxing $g_k(\vec{x}) = 0$ to $g_k(\vec{x}) = c$ increases the objective by $\lambda_k^* c$
  - if $\lambda_k^* = 0$, relaxing the constraint has no effect

Lagrange Theorem

Theorem 120. *Suppose the following conditions hold:*

- $f(\vec{x})$ *and* $g_j(\vec{x})$, $j = 1, \ldots, m$, *are all continuously differentiable* (C^1) *over an open set* $X \subseteq \mathbb{R}^n$
- $\vec{x}^*$ *is an interior extremum of* $f(\vec{x})$ *subject to* m *equality constraints*
- *the gradients* $\nabla g_1(\vec{x}^*), \ldots, \nabla g_m(\vec{x}^*)$ *are linearly independent*

Then, there exists $\vec{\lambda}^* \in \mathbb{R}^m$ such that

$$\nabla f(\vec{x}^*) = \sum_{j=1}^{m} \lambda_j^* \, \nabla g_j(\vec{x}^*).$$

Lagrange Theorem: Limitations

- core structure: if $\vec{x}^*$ is an interior optimum, then the Lagrangian First-Order Conditions will find it
  - still need to check whether each candidate from the FOCs is a minimum, a maximum, or neither
- objective function f must be differentiable
- only finds interior optima — does not detect boundary solutions
- applies only to equality constraints

Constraint Qualification

- requirement that $\nabla g_1(\vec{x}^*), \ldots, \nabla g_m(\vec{x}^*)$ are linearly independent is called the <u>Constraint Qualification</u> (CQ)
  - equivalent to $\text{rank}\; Dg(\vec{x}^*) = m$, where

$$Dg(\vec{x}^*) := [\nabla g_1^T(\vec{x}^*), \ldots, \nabla g_m^T(\vec{x}^*)]$$

- if $\vec{x}^*$ is a local optimum but $\text{rank}\; Dg(\vec{x}^*) < m$, there may not exist $\vec{\lambda}$ such that $(\vec{x}^*, \vec{\lambda})$ solves the FOC
  - any point where CQ fails must be checked separately
- special case: if $m = 1$, CQ becomes $\nabla g(\vec{x}^*) \neq \vec{0}$

Exercise: Constraint Qualification

- **problem**: $\max f(x_1, x_2) = -x_2$ s.t. $g(x_1, x_2) = x_2^3 - x_1^2 = 0$
- **solution**:
  - Lagrangian: $L = -x_2 - \lambda(x_2^3 - x_1^2)$
  - FOCs:

$$\begin{cases} \frac{\partial L}{\partial x_1} = 2\lambda x_1 = 0 \\ \frac{\partial L}{\partial x_2} = -1 - 3\lambda x_2^2 = 0 \\ x_2^3 - x_1^2 = 0 \end{cases}$$

- Check Constraint Qualification (CQ):
  - ▷ gradient of constraint: $\nabla g = \left(-2x_1,\, 3x_2^2\right)$
  - ▷ CQ requires $\nabla g(\vec{x}^*) \neq \vec{0}$
- CQ fails at points: $(0, 0)$

Second-Order Conditions: Setup

- unconstrained SOC: check Hessian sign at a stationary point
- with equality constraints, we only check feasible directions — those that keep us on the Constraint surface
- Hessian of the Lagrangian (w.r.t. $\vec{x}$ variables):

$$H_f(\vec{x}) - \sum_{j=1}^{m} \lambda_j \, H_{g_j}(\vec{x})$$

- tangent space at $\vec{x}^*$:

$$Z^* = \{\vec{z} \in \mathbb{R}^n \mid Dg(\vec{x}^*)\, \vec{z} = 0\}$$

- idea: test curvature of f along $\vec{z} \in Z^*$ only

Second-Order Conditions

Theorem 121. *Let all f and g_j be C^2 and suppose that*

- *$\vec{x}^*$ and $\vec{\lambda}^*$ satisfy the Lagrangian First-Order Conditions,*
- *the Constraint Qualification holds,*
- *let $H^* = H_f(\vec{x}^*) - \sum_{j=1}^{m} \lambda_j^* \, H_{g_j}(\vec{x}^*), \quad Z^* = \{\vec{z} \in \mathbb{R}^n \mid Dg(\vec{x}^*)\, \vec{z} = \vec{0}\}$.*

Then,

- **necessary:** *if $\vec{x}^*$ is a local max, then $\vec{z}^T H^* \vec{z} \leq 0$ for all $\vec{z} \in Z^*$,*
- **sufficient:** *if $\vec{z}^T H^* \vec{z} < 0$ for all $\vec{z} \in Z^* \setminus \{\vec{0}\}$, then $\vec{x}^*$ is a strict local max*

SOC: Explanation

- unconstrained SOC: check the Hessian of f at $\vec{x}^*$ in *all* directions
- constrained SOC: same idea, but only check curvature in directions that keep us on the constraint surface
- feasible directions: $\vec{z} \in Z^*$ (tangent to the constraint surface)
- test H^* only on $\vec{z} \in Z^*$:
  - negative semidefinite on Z^* $\Rightarrow$ necessary condition for local max

– negative definite on Z^* $\Rightarrow$ sufficient condition for strict local max

Exercise: SOC with Equality Constraint (1)

- **Problem**: maximize $f(x_1, x_2) = -x_1^2$ s.t. $x_1^2 + x_2^2 - 4 = 0$
- **Solution**:
  – Lagrangian:

$$L = -x_1^2 - \lambda\,(x_1^2 + x_2^2 - 4)$$

  – FOCs:

$$\begin{aligned} -2x_1^* - 2\lambda^* x_1^* &= 0, \\ -2\lambda^* x_2^* &= 0, \\ x_1^{*2} + x_2^{*2} - 4 &= 0 \end{aligned}$$

  – Candidates:

$$(\pm 2, 0, -1), \quad (0, \pm 2, 0)$$

  – CQ fails at: nowhere

Exercise: SOC with Equality Constraint (2)

- H^* matrices:

$$(\pm 2, 0, -1): \quad H^* = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}, \quad (0, \pm 2, 0): \quad H^* = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}$$

- Tangent space Z^* at each candidate (from $Dg(\vec{x}^*)\vec{z} = 0$):

$$(\pm 2, 0): \; z_1 = 0, \quad (0, \pm 2): \; z_2 = 0$$

- SOC test: check $\vec{z}^\top H^* \vec{z}$ on Z^*:
  – for $(\pm 2, 0)$, directions $\vec{z} = (0, z_2)$:

$$\vec{z}^\top H^* \vec{z} = 2z_2^2 > 0 \quad \Rightarrow \quad \text{strict local minimum}$$

  – for $(0, \pm 2)$, directions $\vec{z} = (z_1, 0)$:

$$\vec{z}^\top H^* \vec{z} = -2z_1^2 < 0 \quad \Rightarrow \quad \text{strict local maximum}$$

General Procedure

Equality-Constrained Optimization Problems: Step 1

- **Problem**:
$$\max_{\vec{x} \in X} f(\vec{x}) \quad \text{s.t.} \quad \vec{x} \in \mathcal{D} := \{\vec{x} \in X \mid g_j(\vec{x}) = 0 \text{for all } j = 1, \dots, m\}$$
where $X \subseteq \mathbb{R}^n$ is open, $f : X \to \mathbb{R}$ and all $g_j : X \to \mathbb{R}$ are C^2
- **Step 1**: argue that a solution exists
  - e.g., show that f is continuous and the Constraint Set $\mathcal{D}$ is compact

Step 2: Constraint Qualification

- check whether the Constraint Qualification (CQ),
$$\text{rank } Dg(\vec{x}) = m,$$
holds at any $\vec{x} \in \mathcal{D}$
  - if CQ fails, compute
$$\overline{y} := \sup_{\vec{x} \in \mathcal{D} \cap \{\text{CQ fails}\}} f(\vec{x})$$
over the set where CQ fails
    - ▷ if CQ fails at some feasible point, Lagranges theorem might miss it
    - ▷ we check those points separately ($\overline{y}$ is the best they can deliver)

Step 3: First-Order Condition

- set up the Lagrangian
$$L(\vec{x}, \vec{\lambda}) := f(\vec{x}) - \sum_{j=1}^{m} \lambda_j g_j(\vec{x})$$
- find all points $(\vec{x}^*, \vec{\lambda}^*)$ that satisfy the FOCs
$$\frac{\partial f}{\partial x_i}(\vec{x}^*) = \sum_{j=1}^{m} \lambda_j^* \frac{\partial g_j}{\partial x_i}(\vec{x}^*) \qquad \text{for all } i \in \{1, \dots, n\},$$
$$g_j(\vec{x}^*) = 0 \qquad \text{for all } j \in \{1, \dots, m\}.$$

Step 4: Identify Global Maximizer

- compute $f(\vec{x}^*)$ at each of the candidate points $(\vec{x}^*, \vec{\lambda}^*)$ from Step 3
- if $f(\vec{x}^*)$ at one candidate is weakly greater than f at all the other candidates *and* greater than or equal to $\overline{y}$ (from Step 2), then that $\vec{x}^*$ is a global maximizer
- if $\overline{y}$ is weakly greater than f at all the Step 3 candidates, and there is some $\vec{x} \in \mathcal{D}$ with CQ failing such that $f(\vec{x}) = \overline{y}$, then that $\vec{x}$ is a global maximizer

7.1 Envelope Theorem

Equality-Constrained Envelope Theorem

Theorem 122. *Let $f : U \times \Theta \to \mathbb{R}$ and $g_j : U \times \Theta \to \mathbb{R}$ ($j = 1, \dots, m$) be C^1, where $U \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^m$ are open. For each $\vec{\theta} \in \Theta$, let $\vec{x}^*(\vec{\theta})$ solve*

$$\max_{\vec{x} \in U} f(\vec{x}, \vec{\theta}) \quad s.t. \quad g_j(\vec{x}, \vec{\theta}) = 0 \ \ \forall j,$$

and let $\lambda_1^*(\vec{\theta}), \dots, \lambda_m^*(\vec{\theta})$ satisfy Lagrange FOCs. Let $V(\theta) := f(x^*(\theta), \theta)$. Then,

$$D_\theta V(\theta) = \left(\frac{\partial L}{\partial \theta_1}, \ \dots, \ \frac{\partial L}{\partial \theta_m} \right) \Bigg|_{x = x^*(\theta), \ \lambda = \lambda^*(\theta)}$$

- follows from the unconstrained envelope theorem because maximizing the Lagrangian is an unconstrained problem

8 Inequality-Constrained Optimization

8.1 Inequality-Constrained Problem

Inequality-Constrained Problem

- as before, objective function is $f : X \to \mathbb{R}$, where $X \subseteq \mathbb{R}^n$ is open
- [**new!**] constraint set is $\mathcal{D} := \{\vec{x} \in X : g(\vec{x}) \leq 0\}$, where $g : X \to \mathbb{R}^m$
- notes:
  - if (some of the) constraints are ≥ 0, multiply them by -1
  - if (some of the) constraints are $= 0$, write them as ≥ 0 and ≤ 0
    - ▷ this problem generalizes equality-constrained optimization

Overview

- we will see the **Karush-Kuhn-Tucker (KKT)** conditions
  - necessary conditions for local optima
- we say
  - j-th constraint is slack at $\vec{x} \in \mathbb{R}^n$ if $g_j(\vec{x}) < 0$
  - j-th constraint is binding at $\vec{x} \in \mathbb{R}^n$ if $g_j(\vec{x}) = 0$
- when applying KKT, we have to make a conjecture (or argument) about which constraints are binding

KKT Theorem

Theorem 123. *Suppose the following conditions hold:*

- $f(\vec{x})$ *and* $g_j(\vec{x})$, $j = 1, \ldots, m$ *are all* C^1 *functions over open set* $X \subseteq \mathbb{R}^n$
- $\vec{x}^*$ *is an interior extremum of* $f(\vec{x})$ *subject to* m *(inequality) constraints*
- *vectors* $\nabla g_1(\vec{x}^*), \ldots, \nabla g_m(\vec{x}^*)$ *are linearly independent*

Then, there exists $\vec{\lambda}^* \in \mathbb{R}^m$ *such that*

$$
\begin{aligned}
Df(\vec{x}^*) &= \sum_{j=1}^{m} \lambda_j \cdot \nabla g_j(\vec{x}^*) \\
\lambda_j &\geq 0 \textit{ for all } j = 1, \ldots, m \\
g_j(\vec{x}^*) &\leq 0 \textit{ for all } j = 1, \ldots, m \\
\lambda_j \cdot g_j(\vec{x}^*) &= 0 \textit{ for all } j = 1, \ldots, m.
\end{aligned}
$$

The latter equation is called the complementary slackness condition.

Importance of Constraint Qualification

- as in equality-constrained optimization, KKT conditions may fail to find a local optimum if the CQ is not satisfied at the optimum
- consider example

$$\max_{(x_1,x_2)\in\mathbb{R}^2} -(x_1^2+x_2^2) \quad \text{subject to} \quad -(x_1-1)^3+x_2^2 \le 0$$

- **claim**: $(1,0)$ is the optimum, but is not picked up by the KKT conditions

Interpretation of Multiplier

- interpretation of λ_j is the same as in the equality-constrained case
- notice that $\lambda_j = 0$ if $g_j(\vec{x}) < 0$, and $g_j(\vec{x}) = 0$ if $\lambda_j > 0$
- consider optimum $\vec{x}^*$. If $g_j(\vec{x}^*) < 0$, then the j-th constraint is not binding, so there is no benefit to relaxing it. If $\lambda_j > 0$, then there is a benefit to relaxing the constraint, so the constraint must be binding
- can also think about λ as a way of penalizing high values of $g_j(\vec{x}^*)$. If a particular constraint is not binding at the optimum, then values of $\vec{x}$ that violate the constraint are not desirable anyway, so there is no need to penalize them
- **caution**: complementary slackness does not rule out $\lambda_j = 0$ and $g_j(\vec{x}^*) = 0$. This will occur if the optimum *without* the j-th constraint happens to satisfy the j-th constraint with equality

Procedure

- the procedure is very similar to one for the equality-constrained problem:

Step 1 argue that an optimum exists

Step 2 check whether the constraint qualification fails at any points that satisfy the constraint

Step 3 find all solutions to KKT conditions, i.e. all $(\vec{x},\vec{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ such that

$$\begin{aligned}
\frac{\partial f}{\partial x_i}(\vec{x}) &= \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i}(\vec{x}) \text{ for all } i \in \{1,\dots,n\} \\
\lambda_j &\ge 0 \text{ for all } j = 1,\dots,m \\
g_j(\vec{x}) &\le 0 \text{ for all } j = 1,\dots,m \\
\lambda_j g_j(\vec{x}) &= 0 \text{ for all } j = 1,\dots,m
\end{aligned}$$

Step 4 take all the $\vec{x}$'s from step (3) and any $\vec{x}$'s from step (2); points that maximize f over this set are optima

Practice: Importance of Step 1

- $\max_{x \in \mathbb{R}} x^2 - x$ subject to $-x \leq 0$

Practice: Importance of Step 2

- $\max_{x \in \mathbb{R}} 2x^3 - 3x^2$ subject to $-(3-x)^3 \leq 0$

Practice

- $\max_{(x_1,x_2) \in \mathbb{R}^2} x_1^2 - x_2$ subject to $x_1^2 + x_2^2 - 1 \leq 0$

Practice

- $\max_{(x_1,x_2) \in \mathbb{R}^2_{++}} \log x_1 + \log x_2$ subject to $x_1^2 + x_2^2 - 2 \leq 0$

Practice (Hard)

- $\max_{(x_1,x_2) \in \mathbb{R}^2} x_1 x_2$ subject to $(x_1 - 1)^2 + (x_2 - 1)^2 \leq 9$ and $x_1 \geq 1$

Special Types of Optimization Problems

Convex Optimization

- sometimes, we can get around Step 1 (proving a solution exists)
  - when f is concave and $\mathcal{D}$ is convex

Concavity and KKT

Theorem 124. *Suppose that [-0.75cm]*

- *X is a convex open set*
- *$f : X \to \mathbb{R}$ is a concave C^1 function*
- *$g_j : X \to \mathbb{R}$ is a convex C^1 function for all $j \in \{1, \ldots, m\}$*
- *there exists at least one point $\vec{x} \in X$ such that $g_j(\vec{x}) < 0$ for all j*

[-0.75cm]Then, $\vec{x}^*$ maximizes $f(\vec{x})$ subject to $g_j(\vec{x}) \leq 0$ for all j if and only if there exists $\vec{\lambda}^* \in \mathbb{R}^m$ such that $(\vec{x}^*, \vec{\lambda}^*)$ satisfies the KKT conditions.

- notice there is no CQ here

- **useful fact**: if the g_j are all convex, then so is the constraint set;
  - if f is strictly concave on the constraint set, there is at most one solution

Slater's Condition

- the requirement that there exists some point at which none of the constraints bind is called Slater's condition
- this condition is needed to prove that the KKT conditions are necessary for an optimum
- Slater's condition + convexity of the constraint functions replace the constraint qualification

Practice: Slater's Condition

- $\max_{x \in \mathbb{R}} x$ subject to $x^2 \leq 0$

B Basic Topology

B.1 Topological Space

Topological Space

Definition 125. Let X be a set, and let τ be a collection of subsets of X. (X, τ) is a topological space if it satisfies the following:

- $X \in \tau$ and $\varnothing \in \tau$;
- (closure under unions) any union of members of τ belongs to τ;
- (closure under intersections) any finite intersection of members of τ belongs to τ.

From Metric to Topology

- if we have a metric space (X, d), we can use it to define a topological space (X, τ)
- first, for each $\varepsilon > 0$ and each $x \in X$, we define the ε-neighborhood of x:

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

- then, let τ be the collection of unions of the $B(x, \varepsilon)$
- can show that this τ is indeed closed under finite intersections

Open Sets

- the members of τ are called "open sets"
- you may be more familiar with a different definition:

Definition 126. Let (X, d) be a metric space. $T \subseteq X$ is open if, for all $t \in T$, there exists $\varepsilon > 0$ such that $B(t, \varepsilon) \subset T$.

- there is no contradiction here:

Theorem 127. *Let (X, d) be a metric space. $T \subseteq X$ is a union of ε-neighborhoods of members of X if and only if, for all $t \in T$, there exists $\varepsilon > 0$ such that $B(t, \varepsilon) \subset T$.*

Equivalence of Definitions

- suppose that $T = \bigcup\limits_{(\varepsilon, s) \in S} B(s, \varepsilon)$ for some $S \subset \mathbb{R}_{++} \times X$
- fix any $t \in T$. We have $t \in B(\widetilde{s}, \widetilde{\varepsilon})$ for some $(\widetilde{\varepsilon}, \widetilde{s}) \in S$

- let $\delta := \widetilde{\varepsilon} - d(t, \widetilde{s})$. The above inequality implies $B(t, \delta) \subseteq B(\widetilde{s}, \widetilde{\varepsilon})$ because

$$d(y,t) < \delta \implies d(t, \widetilde{s}) + d(y,t) < \widetilde{\varepsilon} \implies d(y, \widetilde{s}) < \widetilde{\varepsilon}$$

- conclude that $B(t, \delta) \in T$
- now suppose that, for all $t \in T$, there exists $\varepsilon(t) > 0$ such that $B(t, \varepsilon(t)) \subset T$
- it is easy to see that $T = \bigcup_{t \in T} B(t, \varepsilon(t))$, so T is a union of ε-neighborhoods of members of X

Things to Remember

- confused? here are some takeaways:
  - arbitrary unions of open sets are open, and so are finite intersections of open sets (we can use this property to *define* the collection of open sets)
  - if X is an open subset of $\mathbb{R}^n$, then for every x in X, there is an open ball around x that is a subset of X
  - we can build up every open set in $\mathbb{R}^n$ by taking unions of open balls

Closed Sets

Definition 128. A set is closed if its complement is open.

- a set can be both open and closed (clopen)
- recall that, for any topological space, (X, τ), we have $X \in \tau$ and $\varnothing \in \tau$
- thus, X and $\varnothing$ are clopen
- $\mathbb{R}$ and $\varnothing$ are the only clopen subsets of $\mathbb{R}$

Closed Sets

Theorem 129. *Arbitrary intersections of closed sets are closed, and finite unions of closed sets are closed.*

- **prove** using De Morgan's Laws
  - let $(A_i)_{i \in I}$ be a collection of sets, where I is a possibly uncountable index set . Then,

$$\left(\bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} (A_i)^c \text{ and } \left(\bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} (A_i)^c$$

Closed Sets

- as with open sets, we often use an alternative definition of closed sets

- this definition uses the concepts of “open neighborhood” and “limit point”

Definition 130. Let (X, τ) be a topological space. If $T \in \tau$ and $x \in T$, then T is an open neighborhood of x.

Definition 131. Let (X, τ) be a topological space. $x \in X$ is a limit point of $A \subseteq X$ if every open neighborhood of x contains at least one point in A other than x. A point x is an isolated point of A if $x \in A$ and x is not a limit point of A.

- note: the definition of limit point does not require $x \in A$

Limit vs, Boundary

- **caution**: don’t confuse limit points with boundary points
  - x is a boundary point of A if every open neighborhood of x contains a point in A and a point in A^c
  - 1 is a boundary point of $\{1, 2\}$ but not a limit point (it’s isolated)
  - 0 is a limit point of $(-1, 1)$ but not a boundary point

Limits and Convergence

- the term “limit point” suggests that this concept is related to convergence
- in metric spaces, we can indeed define limit points in terms of convergence

Theorem 132. *Let (X, d) be a metric space $x \in X$ is a limit point of A if and only if $x = \lim a_n$ for some sequence (a_n) in A such that $a_n \neq x$ for all n.*

- this result can actually be extended to a class of non-metric spaces

Closed Sets and Limit Points

Theorem 133. *Let (X, τ) be a topological space. $A \subseteq X$ is closed if and only if A contains all its limit points.*

Proof

- suppose A and suppose $x \notin A$ is a limit point of A
  - A is closed $\implies$ A^c is open $\implies$ $A^c \in \tau$ $\implies$ $\exists \tilde{\varepsilon} > 0$ s.t. $B_{\tilde{\varepsilon}} \subseteq A^c$ $\implies$ $B_{\tilde{\varepsilon}}$ does not intersect A so that x is not A’s limit point
- now suppose that A contains all its limit points. Fix $x \in A^c$
  - since x is not a limit point of A, there exists $T(x) \in \tau$ such that $x \in T(x)$ and $T(x) \cap A = \varnothing$
  - since $T(x) \subseteq A^c$ for all $x \in A^c$, we have $A^c = \bigcup_{x \in A^c} T(x)$. Since $T(x) \in \tau$ for all $x \in A$, we have $A^c \in \tau$. Since A^c is open, A is closed

Practice

- consider metric space $((0,1), d)$ where d is the absolute value metric
  - is set $(0,1)$ open or closed?
- consider the metric space $(\mathbb{R}, d)$ where d is the absolute value metric
  - is set $\{1/n : n \in \mathbb{N}\}$ open or closed?

Closure

- last example suggests that we can get a closed set from an arbitrary set by adding limit points
- indeed we can, even in a general topological space

Definition 134. Closure of A ($\mathrm{cl}(A)$ or $\overline{A}$) is the union of A and all its limit points.

Theorem 135. *$\overline{A}$ is the smallest closed set containing A.*

Proof

- we show that $\overline{A}$ is closed. Suppose that x is a limit point of $\overline{A}$ since $x \in A$ immediately implies $x \in \overline{A}$, assume $x \notin A$
- fix any open neighborhood T of x. Since x is a limit point of $\overline{A}$, there exists $y \in \overline{A} \cap T$ such that $y \neq x$
- since $T \in \tau$ and $y \in T$, T is an open neighborhood of y. Also, $y \in A$ or y is a limit point of A. Either way, there exists $z \in A \cap T$. Since $x \notin A$, we have $z \neq x$
- conclude that, for any open neighborhood T of x, there exists $z \in A \cap T$ such that $z \neq x$, so x is a limit point of A
- since any closed set containing A must contain all of A's limit points, $\overline{A}$ is included in any other closed set containing A

B.2 Sequential Compactness

Sequential Compactness

Definition 136. Let (X, d) be a metric space. $S \subseteq X$ is sequentially compact if every sequence in S has a subsequence that converges to a point in S.

- in fact, we don't need to limit ourselves to metric space

Sequential Compactness in General

Definition 137. Let (x_n) be a sequence in a topological space. (x_n) converges to x if every open neighborhood of x contains all but finitely many elements of the sequence.

Definition 138. Let (X, τ) be a topological space $S \subseteq X$ is sequentially compact if every sequence in S has a subsequence that converges to a point in S.

Compact vs Closed/Bounded

- fix a metric space (X, d) and $S \subset X$. Fix some limit point s of S
- recall that $s = \lim s_n$ for some sequence (s_n) in $S \setminus \{s\}$
- if S is sequentially compact, then (s_n) must have a subsequence that converges to a point in S. Since $s_n \to s$, that point can only be s, so $s \in S$
- so, a subset S of a metric space is sequentially compact only if it is closed
- also, S is sequentially compact only if it is bounded

Heine-Borel Theorem

- you might have seen "compact" defined as "closed and bounded"
- in $\mathbb{R}^n$, the two are equivalent:

Theorem 139. *$S \subset \mathbb{R}^n$ is sequentially compact if and only if it is closed and bounded.*

Heine-Borel: Proof

- suppose that $S \subset \mathbb{R}^n$ is closed and bounded
- fix any (s_n) in S. Since S is bounded, (s_n) has a convergent subsequence let $s = \lim s_n$
- if s appears in (s_n), then $s \in S$
- if s does not appear in (s_n), then s is a limit point of S. Since S is closed, $s \in S$

B.3 Compactness

Open Covers

Definition 140. Let (X, τ) be a topological space, and let $\{T_i\}_{i \in I}$ be a (possibly uncountable) collection of members of τ. $\{T_i\}_{i \in I}$ is an open cover of $S \subseteq X$ if

$$S \subseteq \bigcup_{i \in I} T_i.$$

A finite subcover is a finite subcollection of the original open cover that is itself an open cover

Compactness and Open Covers

Definition 141. Let (X, τ) be a topological space. $S \subseteq X$ is compact if every open cover of S has a finite subcover.

Examples of Compactness

- consider the interval $(0,1)$; the collection of sets $\{(i,1)\}_{i\in(0,1)}$ is an open cover, but it doesn't have a finite subcover (to see this, note that the union of any finite subcollection of $\{(i,1)\}_{i\in(0,1)}$ is equal to a member of the subcollection—but no member of $\{(i,1\}_{i\in(0,1)}$ is a superset of $(0,1)$); thus, $(0,1)$ is not compact according to our new definition
- now consider a finite set $\{x_1,\ldots,x_n\}$; take any open cover; for each $i\in\{1,\ldots,n\}$, pick one set in the open cover that includes x_i; the union of these sets is a finite subcover; thus, any finite set is compact

C Correspondences

C.1 Correspondences

Correspondences

- recall the definition of a correspondence from Lecture 1:
- correspondence $\Phi: X \rightrightarrows Y$ is a rule that assigns to each element $x\in X$ a subset $f(x)$ of Y
- we will look at some properties of correspondences, working up to a fixed point theorem for correspondences that is used in game theory
- for more on correspondences (in lots of generality), a good reference is Chapter 17 of *Infinite Dimensional Analysis: A Hitchhiker's Guide* by Aliprantis & Border

Basic Properties

- $\Phi: X \rightrightarrows Y$ is non-empty-valued if $\Phi(x)\neq\varnothing$ for all $x\in X$
- Φ is closed-valued if $\Phi(x)$ is closed for all $x\in X$
- Φ is compact-valued if $\Phi(x)$ is compact for all $x\in X$
- Φ is convex-valued if $\Phi(x)$ is convex for all $x\in X$

Upper Hemicontinuity

- it is useful to break correspondence continuity into two separate conditions

Definition 142. Let (X,τ_X) and (Y,τ_Y) be topological spaces. $\Phi: X\rightrightarrows Y$ is upper hemicontinuous at $x\in X$ if, for each $V\in\tau_Y$ such that $\Phi(x)\subseteq V$, there exists an open neighborhood U of x such that $\Phi(u)\subseteq V$ for all $u\in U$.

- intuitively, if $\Phi(x)$ is contained in the open set V, it will still be contained in V as long as you stay within a small enough neighborhood of x

Sequential Version

- as with continuity of a function, can restate UHC in terms of sequences
- will restrict attention to metric spaces, although this can be generalized

Theorem 143. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let $\Phi : X \rightrightarrows Y$ be a closed-valued correspondence. The following are equivalent:*

- *Φ is upper hemicontinuous at x, and $\Phi(x)$ is compact.*
- *For all (x_n) in X such that $x_n \to x \in X$, for all (y_n) in Y such that $y_n \in \Phi(x_n)$ for all n, (y_n) has a subsequence that converges to a point in $\Phi(x)$.*

Closed Graph Theorem

- the following result provides the easiest way to think about UHC

Theorem 144. *Let (X, d_X) and (Y, d_Y) be metric spaces, where Y is compact. $\Phi : X \rightrightarrows Y$ is closed-valued and upper hemicontinuous if and only if it has a closed graph, i.e. the set*

$$\{(x, y) \in X \times Y : y \in \Phi(x)\}$$

is closed.

Lower Hemicontinuity

Definition 145. Let (X, τ_X) and (Y, τ_Y) be topological spaces. $\Phi : X \rightrightarrows Y$ is lower hemicontinuous at $x \in X$ if, for each $V \in \tau_Y$ intersecting $\Phi(x)$, there exists an open neighborhood U of x such that $\Phi(u)$ intersects V for all $u \in U$.

- LHC says that $\Phi(x)$ cannot suddenly collapse when you perturb x

Sequential Version

Theorem 146. *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $\Phi : X \rightrightarrows Y$ be a correspondence. The following are equivalent:*

- *Φ is lower hemicontinuous at $x \in X$.*
- *If $x_n \to x$, then for each $y \in \Phi(x)$, there exists a subsequence (x_{n_k}) and elements $y_k \in \Phi(x_{n_k})$ such that $y_k \to y$.*

Continuity

- correspondence is continuous if it is both upper hemicontinuous and lower hemicontinuous

- if Φ has singleton values, there is a function such that $\Phi(x) = \{f(x)\}$ for all x in the domain of Φ
- in that case, UHC, LHC and continuity of Φ are all equivalent, and are equivalent to continuity of f

Preserving UHC

Theorem 147. *Let $\{\Phi_i : X \rightrightarrows Y\}_{i \in I}$ be a set of UHC correspondences.*

- *If I is finite, then the correspondence $\Phi : X \rightrightarrows Y$ given by $\Phi(x) := \bigcup_{i \in I} \Phi_i(x)$ is UHC.*
- *If each Φ_i is compact-valued, then the correspondence $\Phi : X^I \rightrightarrows Y$ given by $\Phi(x) := \prod_{i \in I} \Phi_i(x)$ is UHC.*

Kakutani's Fixed Point Theorem

- we say that x is a fixed point of Φ if $x \in \Phi(x)$

Theorem 148. *If $X \subseteq \mathbb{R}^n$ is nonempty, compact, and convex, and if $\Phi : X \rightrightarrows X$ is an upper hemicontinuous correspondence with nonempty, closed, convex values, then Φ has a fixed point.*

- this result is used to prove the existence of Nash equilibrium

D Dynamic Programming with Finite Horizon

What is Dynamic Programming?

- **Dynamic programming** (DP): A method for solving optimization problems over time.
- Find decisions today accounting for impacts on future outcomes.
- Used in economics, finance, engineering, etc.

Problem Statement

Maximize:

$$\max_{\{a_t\}_{t=0}^{T-1}} \sum_{t=0}^{T-1} u(x_t, a_t)$$

subject to:

$$x_{t+1} = f(x_t, a_t), \quad x_0 \text{ given}$$

- decision variable: a_t (action)
- state variable: x_t
- horizon: T periods (finite)

Bellman Equation

Value Function

$$V_t(x_t) = \max_{a_t} \{u(x_t, a_t) + V_{t+1}(x_{t+1})\}$$

with $x_{t+1} = f(x_t, a_t)$

- **boundary condition:** $V_T(x_T)$ given
-
  - often 0 if there are no rewards after final period

First-Order Conditions (FOCs)

Assume u and f are differentiable. The FOC for a_t^*:

$$u_a(x_t, a_t^*) + V'_{t+1}(x_{t+1}) f_a(x_t, a_t^*) = 0$$

Backward Induction

- Start from the last period $T-1$:
  - Compute $V_{T-1}(x_{T-1})$ using V_T.
  - Move backward: $V_{T-2}, V_{T-3}, \ldots, V_0$.
- Each value function summarizes optimal future payoffs.

Problem Setup: Two-Period Consumption-Savings

- Maximize lifetime utility:
$$\max_{c_0, c_1} \log(c_0) + \beta \log(c_1)$$
- Subject to budget constraints:
$$I_1 = I_0 - c_0, \quad c_1 = I_1, \quad I_0 \text{ given}$$
- Goal: find c_0^*, c_1^* to maximize utility given I_0.

Problem Setup: T-Period Consumption-Savings

- Maximize lifetime utility over T periods:

$$\max_{\{c_t\}_{t=0}^{T-1}} \sum_{t=0}^{T-1} \beta^t \log(c_t)$$

- Subject to sequential budget constraints:

$$I_{t+1} = I_t - c_t, \quad t = 0, 1, \ldots, T-1, \quad I_0 \text{ given}$$

- Goal: Find consumption plan $\{c_t^*\}$ maximizing utility given initial assets.

Bellman Equation Setup for T-Period Problem

Value function:

$$V_t(I_t) = \max_{c_t} \{\log(c_t) + \beta V_{t+1}(I_t - c_t)\}$$

Subject to:

$$0 \leq c_t \leq I_t$$

Terminal condition:

$$V_T(I_T) = 0$$

First-Order Condition (FOC) for Optimal Consumption

Bellman equation:

$$V_t(I_t) = \max_{c_t} \{\log(c_t) + \beta V_{t+1}(I_t - c_t)\}$$

derivative w.r.t. c_t:

$$\frac{\partial}{\partial c_t} [\log(c_t) + \beta V_{t+1}(I_t - c_t)] = 0$$

FOC:

$$\frac{1}{c_t^*} - \beta V'_{t+1}(I_t - c_t^*) = 0 \quad \Longrightarrow \quad \frac{1}{c_t^*} = \beta V'_{t+1}(I_t - c_t^*)$$

Backward Induction and Solution Strategy

- **Step 1: Terminal condition**
  Define the value function at final period:
  $$V_T(I_T) = 0$$
- **Step 2: Solve last period problem**
  At $t = T - 1$:
  $$V_{T-1}(I_{T-1}) = \max_{c_{T-1}} \log(c_{T-1}) \quad \Rightarrow \quad c^*_{T-1} = I_{T-1}$$
- **Step 3: Recursively solve for earlier periods**
  For $t = T - 2, T - 3, \ldots, 0$, use FOC and known V_{t+1} to find:
  $$c^*_t \text{ and } V_t(I_t)$$
- **Outcome:** Obtain optimal consumption policy $\{c^*_t\}$