Vanderbilt Economics Ph.D. Math Camp

Lecture Notes, Fall 2022

Maria Titova

Contents

1 Lecture 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 $B \setminus A := \{x : x \in B \text{ and } x \notin A\}$.

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

Union

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

- in math, "or" is inclusive unless otherwise specified—so if $x \in A$ and $x \in B$, then $x \in A \cup B$
- we can take the union of more than two sets, even infinitely many
  - to do this, let $\mathcal{A}$ be a (possibly infinite) collection of sets
  - then, the union of the members of A is $\bigcup_{\mathcal{A}} A := \{x : x \in A \text{ for some } A \in \mathcal{A}\}$

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 A is $\bigcap_{\mathcal{A}} A := \{x : x \in A \text{ for all } A \in \mathcal{A}\}$

Cartesian Product

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

definition

- 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 $f(x) = x^2$
  - And $f : [0,1] \to [0,2]$ 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

Theorem 15. *If n is an even integer, then $-5n-3$ is an odd integer.*

Practice: Proof by Contradiction

Theorem 16. *$\sqrt{2}$ 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

Theorem 17. *For any $n \in \mathbb{N}$,*

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

Practice: Proof by Cases/Steps

Theorem 18. *If n is an integer, then n^2+3n+5 is an odd integer.*

2 Lecture 2: Metric Spaces and Sequences

Least Upper Bounds

Definition 19. $x \in \mathbb{R}$ 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 20. Set is countable if it is finite or has the cardinality of natural numbers.

Theorem 21. *The rationals and integers are countable.*

Theorem 22.
- *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 23. Set is uncountable if its cardinality is strictly greater than that of the natural numbers.

Theorem 24. *The set of reals ($\mathbb{R}$) is uncountable. The power set of the natural numbers ($2^{\mathbb{N}}$) is uncountable.*

Continuum Hypothesis

Theorem 25. *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 26. 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

Theorem 27. *A sequence 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: $(a_1, a_2, \ldots)$ 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

Theorem 28. *Let* (S, d) *be a metric space. A sequence* (x_n) *in* S *converges 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 29. *If (x_n) converges to x and y, then $x = y$.*

- **Proof:** ____

A Simple Convergence Proof

Theorem 30.

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

- **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
- will state and prove this result for real numbers, but it can easily be extended to functions, sets, etc

Theorem 31. *Suppose $\forall n > N_0$, $x_n < z_n < y_n$. Then, $(x_n \to L \text{ and } y_n \to L) \Rightarrow z_n \to L$.*

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 32. 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 33. 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 34. *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 35. *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 36. *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

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

Definition 37. 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 38. *If a sequence in $\mathbb{R}$ is monotone and bounded, it converges.*

Monotone Convergence Theorem: Proof

- suppose that (a_n) is bounded and increasing
- since (a_n) is bounded, $\sup a_n$ exists. Let $a := \sup a_n$
- fix any $\varepsilon > 0$. If there is no N such that $a - \varepsilon < a_N$, then $a - \varepsilon < a$ is an upper bound for (a_n), which contradicts $a = \sup a_n$
- thus, for all $\varepsilon > 0$, there exists N such that $a - \varepsilon < a_N < a$
- since (a_n) is increasing and a is an upper bound, we have $a - \varepsilon < a_n < a$ for all $n > N$, which implies $|a_n - a| < \varepsilon$ for all $n > N$

Subsequences

Definition 39. 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 40. *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 41. *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 42. A sequence (x_n) is <u>Cauchy</u> if, for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $|a_n - a_m| < \varepsilon$ 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

Theorem 43. *$\mathbb{R}^n$ is <u>complete</u>, that is, a sequence in $\mathbb{R}^n$ is Cauchy if and only if it converges.*

3 Lecture 3: Topology

3.1 Topological Space

Topological Space

Definition 44. 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 45. 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 46. *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 $T = \bigcup_{(\varepsilon,s)\in S} B(s,\varepsilon)$ 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

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

- 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 47. 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 48. *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 49. Let (X, τ) be a topological space. If $T \in \tau$ and $x \in T$, then T is an open neighborhood of x.

Definition 50. 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 51. *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 52. *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 $A^c = \bigcup\limits_{x \in A^c} T(x)$. 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 53. Closure of A ($\mathrm{cl}(A)$ or $\overline{A}$) is the union of A and all its limit points.

Theorem 54. *$\overline{A}$ 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

3.2 Sequential Compactness

Sequential Compactness

Definition 55. Let (X, d) be a metric space. $S \subseteq X$ is <u>sequentially compact</u> 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 56. Let (x_n) be a sequence in a topological space. (x_n) <u>converges</u> to x if every open neighborhood of x contains all but finitely many elements of the sequence.

Definition 57. Let (X, τ) be a topological space $S \subseteq X$ is <u>sequentially compact</u> 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 58. *$S \subset \mathbb{R}^n$ 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, $s \in S$

3.3 Compactness

Open Covers

Definition 59. 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 60. 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

4 Lecture 4: Limits and Continuity

Functional Limits in Metric Spaces

Definition 61. 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, $\underline{\lim_{x \to \overline{x}} f(x) = \overline{y}}$ 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 62. *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 63. *Let X be a metric space, and let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ be functions. If $\lim_{x \to \overline{x}} f(x) = y$ and $\lim_{x \to \overline{x}} g(x) = z$, then*

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

4.1 Continuous Functions

Continuous Function

Definition 64. Let (X, d_X) and (Y, d_Y) be metric spaces. Function $f : X \to Y$ is continuous at $c \in X$ if $\forall \varepsilon > 0$, $\exists \delta > 0$ such that, for all $x \in X$,

$$d_X(c, x) < \delta \quad \Longrightarrow \quad d_Y(f(c), f(x)) < \varepsilon$$

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

Practice

- is $f(x) = \sqrt{x}$ continuous?

Continuity and Limits

- we can characterize continuity in terms of functional limits

Theorem 65. *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, f is continuous at $\overline{x}$ if and only if $\lim_{x \to \overline{x}} f(x) = f(\overline{x})$.*

- follows from definitions of continuity and functional limits

Continuity and Sequences

- we can characterize continuity in terms of sequences

Theorem 66. *Suppose that (X, d_X) and (Y, d_Y) are metric spaces; $f : X \to Y$ is a function. For any limit point $\overline{x}$ 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 67. *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 68. *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

Continuity and Compactness

Theorem 69. *Suppose that (X, d_X) and (Y, d_Y) are metric spaces and $f : X \to Y$ is a continuous function. If X is compact, then so is $f(X)$.*

- recall that $X \subset \mathbb{R}^n$ is compact $\iff$ X is closed and bounded

Extreme Value Theorem

Theorem 70. *Let X be a compact set, and let $f : X \to \mathbb{R}$ be a continuous function. There exists $\underline{x}$ and $\overline{x}$ in X such that $f(\underline{x}) \leq f(x) \leq f(\overline{x})$ for all $x \in X$.*

- closed and bounded subset of $\mathbb{R}$ has a max and a min
- the range of a real-valued function on a compact set is closed and bounded

Uniform Continuity

- recall that "f is continuous at x"means that there exists a function $\delta : \mathbb{R}_{++} \to \mathbb{R}_{++}$ such that $f(y)$ is within ε of $f(x)$ whenever y is within $\delta(\varepsilon)$ of x
- notice that δ implicitly depends on x – even if f is continuous at all x in its domain, there is no requirement that the same δ works for all x
- a stronger notion of continuity adds this requirement:

Definition 71. Suppose that (X, d_X) and (Y, d_Y) are metric spaces and $f : X \to Y$ is a function. f is uniformly continuous if, for each $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in X$ and $y \in Y$

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

Example

- is $f(x) = x^2$ uniformly continuous?

Uniform Continuity and Compactness

Theorem 72. *Suppose that (X, d_X) and (Y, d_Y) are metric spaces; $S \subseteq X$; $f : X \to Y$ is a function. If S is compact and f is continuous on S, then f is uniformly continuous on S.*

Intermediate Value Theorem

Theorem 73. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. For any $y \in \mathbb{R}$ such that*

$$\min\{f(a), f(b)\} < y < \max\{f(a), f(b)\},$$

there exists $x \in (a, b)$ such that $f(x) = y$.

- intuitively, says that a continuous function on a closed interval $[a, b]$ attains every value that falls between $f(a)$ and $f(b)$

IVT – Proof

- let $K := \{x \in [a,b] : f(x) \leq y\}$. This set is non-empty and bounded, so it has a finite supremum, s. We claim $f(s) = y$
- suppose $f(s) \neq y$. By continuity of f, we can find $\delta > 0$ such that $|x - s| < \delta$ implies $|f(x) - f(s)| < |y - f(s)|$
- if $f(s) < y$, then we can find $x > s$ such that $f(x) < y$, which contradicts the definition of s
- if $f(s) > y$, then we also have $f(x) > y$ for all $x < s$ such that $s - x < \delta$. Again, this contradicts the definition of s

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

Contraction

Definition 75. 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 $x, y \in X$.

- all contractions are continuous—in fact, they satisfy a property even stronger than uniform continuity:

Definition 76. Suppose that (X, d_X) and (Y, d_Y) are metric spaces and $f : X \to Y$ is a function. f is Lipschitz continuous if there exists $k \in \mathbb{R}_+$ such that, for all $x_1, x_2 \in X$,

$$d_Y(f(x_1), f(x_2)) \leq k d_X(x_1, x_2).$$

Contraction Mapping Theorem

Theorem 77. *Let (X, d) be a complete metric space. If $f : X \to X$ is a contraction, then f has a unique fixed point.*

Fixed Points in One Dimension

Theorem 78. *If $f : [a, b] \to [a, b]$ is a continuous, then f has a fixed point.*

Brouwer's Theorem

Theorem 79. *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.*

5 Lecture 5: Derivative

5.1 Basics of Derivatives

Derivative

Definition 80. Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}$ be a function.
The derivative of f at $\overline{x} \in A$ 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 $\forall \overline{x} \in A$, 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 $f'(1)$

Continuity vs. Differentiability

Theorem 81. *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 82. *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* $\left(\frac{f}{g}\right)(\overline{x}) = \frac{g(\overline{x})f'(\overline{x}) - f(\overline{x})g'(\overline{x})}{g(\overline{x})^2}$

Proof

- for (3),

$$\begin{aligned}\lim_{x\to\overline{x}} \frac{f(x)g(x) - f(\overline{x})g(\overline{x})}{x-\overline{x}} &= \lim_{x\to\overline{x}} \left(g(x)\frac{f(x)}{x-\overline{x}} - \frac{f(\overline{x})g(\overline{x})}{x-\overline{x}} \right) \\ &= \lim_{x\to\overline{x}} g(x)\frac{f(x)-f(\overline{x})}{x-\overline{x}} + \lim_{x\to\overline{x}} \frac{g(x)f(\overline{x}) - f(\overline{x})g(\overline{x})}{x-\overline{x}} \\ &= g(\overline{x})f'(\overline{x}) + f(\overline{x})g'(\overline{x})\end{aligned}$$

- for (4), the main step is

$$\lim_{x\to\overline{x}} \frac{\frac{1}{g(x)} - \frac{1}{g(\overline{x})}}{x-\overline{x}} = \lim_{x\to\overline{x}} \frac{1}{g(x)g(\overline{x})} \frac{g(\overline{x}) - g(x)}{x-\overline{x}} = -\frac{g'(\overline{x})}{g(\overline{x})^2}$$

- then apply (3) to get (4)

Chain Rule

- composition of two differentiable functions is also differentiable

Theorem 83. *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 $f(\overline{x})$, then*

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

Proof

- let $\gamma : g(B) \to \mathbb{R}$ be the function given by

$$\gamma(z) := \frac{f(z) - f(g(\overline{x}))}{z - g(\overline{x})}$$

- then, $\lim_{z \to g(\overline{x})} \gamma(z) = \dots$
- hence,

$$\lim_{x \to \overline{x}} \frac{f(g(x)) - f(g(\overline{x}))}{x - \overline{x}} = \dots$$

Interior Extremum Theorem

Theorem 84. *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. $|x - \overline{x}| < \varepsilon$, $f(\overline{x}) \geq f(x)$

Proof

- suppose that $\overline{x}$ is a local max; the argument for a local min is similar
- fix any (x_n) such that $x_n \leq \overline{x}$ for all n. Since $\overline{x}$ is a local max, $f(x_n) \leq f(\overline{x})$ for all n sufficiently large. We have

$$\frac{f(x_n) - f(\overline{x})}{x_n - \overline{x}} \geq 0$$

  for all n sufficiently large, which implies $f'(\overline{x}) \geq 0$
- now fix any (x_n) such that $x_n \geq \overline{x}$ for all n. Since $\overline{x}$ is a local max, we have

$$\frac{f(x_n) - f(\overline{x})}{x_n - \overline{x}} \leq 0$$

  for all n sufficiently large, which implies $f'(\overline{x}) \leq 0$
- conclude that $f'(\overline{x}) = 0$

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 85. *If f is differentiable on $[a, b]$, and if ϕ satisfies $f'(a) < \phi < f'(b)$, then there exists $\overline{x}$ such that $f'(\overline{x}) = \phi$.*

Proof

- define $g : [a, b] \to \mathbb{R}$ by $g(x) = f(x) - \phi x$; since f is continuous on $[a, b]$, so is g; hence, g achieves a global min on $[a, b]$
- suppose that a is a min, so $f(a) - \phi a \leq f(x) - \phi x$ for all $x \in [a, b]$; this implies

$$\frac{f(x) - f(a)}{x - a} \geq \phi$$

  for all $x \geq a$; then, $f'(a) \geq \phi$ and that contradicts $f'(a) < \phi$
- use a similar argument to show that b is not a min; conclude that g has a min $\overline{x} \in (a, b)$
- since f is differentiable on $[a, b]$, so is g; since $\overline{x}$ is interior, g satisfies first-order condition at $\overline{x}$; we have $g'(\overline{x}) = f'(\overline{x}) - \phi = 0$

Rolle's Theorem

Theorem 86. *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists a point $\overline{x} \in (a, b)$ such that $f'(\overline{x}) = 0$.*

Proof

- since f is continuous on $[a, b]$, it achieves a max and min on $[a, b]$
- if max and min are equal, f is constant, so it has zero derivative everywhere on (a, b)
- now suppose $\max f > \min f$; since $f(a) = f(b)$, at least one of the extrema must be interior; call it $\overline{x}$
- since f is differentiable on (a, b) and $\overline{x}$ is an extremum imply, $f'(\overline{x}) = 0$

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

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

Proof

- let

$$g(x) = f(x) - \left(f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right)$$

- we have $g(a) = g(b) = 0$; g is continuous and differentiable; by Rolle's theorem, there exists $\overline{x} \in (a, b)$ such that

$$g'(\overline{x}) = f'(x) - \frac{f(b) - f(a)}{b - a} = 0$$

Applying MVT

- MVT can be used to prove other results about the derivative

Theorem 88. *Let A be an interval in $\mathbb{R}$, and let $f : S \to \mathbb{R}$ be a continuous function. If the derivative at every interior point of S exists and is 0, then f is a constant function.*

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

$$f'(\overline{x}) = \frac{f(b) - f(a)}{b - a} = 0,$$

which implies $f(b) = f(a)$

Sign of Derivative

Theorem 89. *Let A be an open interval, and let $f : A \to \mathbb{R}$ be a differentiable function.*

- *If $f'(x) \geq 0$ for all $x \in A$, then $f(x) \geq f(y)$ whenever $x > y$. If in addition $f'(\bar{x}) > 0$ for some $x \leq \bar{x} \leq y$, then $f(x) > f(y)$.*
- *If $f'(x) \leq 0$ for all $x \in A$, then $f(x) \leq f(y)$ whenever $x > y$. If in addition $f'(\bar{x}) > 0$ for some $x \leq \bar{x} \leq y$, then $f(x) < f(y)$.*

Second Derivative

Definition 90. Let A be an interval, and let $f : A \to \mathbb{R}$ be a differentiable function.
The second-order derivative of f at $\bar{x} \in A$, denoted $f''(\bar{x})$, is the derivative of the function $g : A \to \mathbb{R}$ given by $g(x) = f'(x)$.
If $f''(\bar{x})$ exists $\forall \bar{x} \in A$, we say that f is twice differentiable.

- alternative notation:

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

- 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

5.2 L'Hopital

Limits of Quotients

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

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

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

Convergence to Infinity

Definition 91. Suppose that (X, d_X) is a metric space, $D \subseteq X$, $\bar{x}$ is a limit point of D, and $f : D \to \mathbb{R}$ is a function. Then, $\underline{\lim_{x \to \bar{x}} f(x) = +\infty}$ if, for all $M > 0$, there exists δ such that

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

L'Hopital's Rule

Theorem 92. *Suppose f and g are continuous on (a, b) and let $\bar{x} \in (a, b)$. If all of the following conditions*

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

hold, then

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

Extended Mean Value Theorem

- L'Hopital's Rule follows from the extended version of the MVT

Theorem 93 (MVT)**.** *If f and g are continuous on $[a, b]$ and differentiable on (a, b), then there exists $\bar{x} \in (a, b)$ such that*

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

Theorem 94 (Extended MVT)**.** *If f and g are continuous on $[a, b]$ and differentiable on (a, b), then there exists $\bar{x} \in (a, b)$ such that*

$$(f(b) - f(a)) \cdot g'(\bar{x}) = (g(b) - g(a)) \cdot f'(\bar{x}).$$

Scope of Application

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

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

  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:

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

6 Lecture 6: Linear Algebra

6.1 Vector Spaces

Overview

- in Lecture 2, we saw a class of spaces (metric spaces) that generalizes $\mathbb{R}^n$ but keeps the notion of "distance" between two points

- we will now study another way to generalize $\mathbb{R}^n$, this time keeping the notions of addition (of vectors) and multiplication (of a vector by a number)

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 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $a, b \in \mathbb{R}$

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

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

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

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

A3 (existence of additive inverse): $\forall \boldsymbol{v} \in V$, $\exists (-\boldsymbol{v}) \in V$ s.t. $\boldsymbol{v} + (-\boldsymbol{v}) = \boldsymbol{0}$

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

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

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

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

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

Standard Example

- $\mathbb{R}^n$ is a vector space
- addition is defined as

$$\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, \ldots, u_n + v_n)$$

- multiplication is defined as

$$a \cdot \boldsymbol{v} = (av_1, \ldots, av_n)$$

Other Examples

- set of all $m \times n$ matrices $M_{m,n}$
  - element-wise addition and scalar multiplication
- set of all sequences $\{x_n\}$ if 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 95. 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 96. *W is a subspace of vector space V if and only if $\forall a, b \in \mathbb{R}$ and all $\boldsymbol{u}, \boldsymbol{v} \in W$, $a\boldsymbol{u} + b\boldsymbol{v} \in W$.*

- 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 97. A linear combination of vectors $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}\}$ is a vector of the form $a_1\boldsymbol{v_1} + \ldots + a_n\boldsymbol{v_n}$, where each $a_i \in \mathbb{R}$.

Definition 98. Let U be a set of vectors. The span of U, denoted $\mathrm{span}(U)$, is the set of all linear combinations of finite subsets of U. We say that U spans a vector space V if $V \subseteq \mathrm{span}(U)$.

Linear Independence

Definition 99. Set of vectors $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}\}$ is linearly independent if

$$\sum_{i=1}^{n} a_i\boldsymbol{v_i} = \boldsymbol{0} \quad \implies \quad a_i = 0 \text{ for all } i.$$

Otherwise, it is linearly dependent.

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

Basis

Definition 100. Let V be a vector space. $U \subset V$ is a basis for V if U spans V, and 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 $\dim(V)$

6.2 Linear Algebra

Matrices

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

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

  where each $a_{ij} \in \mathbb{R}$
- we sometimes write $A = (a_{ij})$ so that

$$(a_{i1} \cdots a_{im}) \text{ is the } i\text{th row of } A \qquad \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} \text{ is the } j\text{th column of } A$$

Determinant

Definition 101. The determinant of $A \in M_{n,n}$ is inductively defined by

$$\det(A) := \sum_{i=1}^{n} (-1)^{i+1} a_{1i} \det(A_{-1i}),$$

where $A_{-ij} \in M_{n-1,n-1}$ is the matrix formed from A by deleting the ith row and jth column.

- **example**: if $A \in M_{2,2}$, $\det(A) = a_{11}a_{22} - a_{12}a_{21}$
- alternative notation: $|A|$

- knowing properties is more important than remembering the formula

Trace and Rank

Definition 102. The trace of $A \in M_{n,n}$ is the sum of its diagonal elements: $\operatorname{tr}(A) := \sum_{i=1}^{n} a_{ii}$.

Definition 103. The column rank of $A \in M_{m,n}$ is number of linearly independent columns of A. The 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

Matrix Addition / Multiplication

- matrix addition and scalar multiplication are done element-wise
- we can also multiply a matrix by a matrix

Definition 104. Let $A \in M_{n,m}$, $B \in M_{m,l}$, $C \in M_{n,l}$. C is the matrix product of A and B if

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}.$$

- $(A \times B)_{ij}$ is the dot product of the ith row of A and the jth column of B:

$$c_{ij} \quad = \quad (a_{i1} \cdots a_{im}) \cdot \begin{pmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{pmatrix}$$

Matrix Addition / Multiplication Properties

Theorem 105.
- $A + B = B + A$;
- $(A + B) + C = A + (B + C)$;
- $(AB)C = A(BC)$;
- $A(B + C) = AB + AC$.

- note that matrix multiplication often does not satisfy $AB = BA$

Transpose

Definition 106. The transpose of an $n \times m$ matrix A is the $m \times n$ matrix B such that $b_{ij} = a_{ji}$. The transpose of A is denoted A' or A^T.

Inverse Matrix

- $n \times n$ matrices (called square) may have a (multiplicative) inverse

Definition 107.
- Let $A \in M_{n,n}$. A is invertible if there exists $B \in M_{n,n}$ such that

$$AB = BA = I_n,$$

  where $I_n \in M_{n,n}$ is the identity matrix with ones on the main diagonal and zeros elsewhere.
- If A is invertible, then B is unique, is called the inverse of A, and denoted by A^{-1}.

Theorem 108. *$A \in M_{n,n}$ is invertible* $\iff \det(A) \neq 0$.

Properties of Inverses

Theorem 109. *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

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

Eigenvalue

Definition 110. Let $A \in M_{n,n}$. If $A\boldsymbol{v} = \lambda\boldsymbol{v}$ for some $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and some real or complex number λ, then $\boldsymbol{v}$ an eigenvector of A, and λ is the corresponding eigenvalue.

- $A\boldsymbol{v} = \lambda\boldsymbol{v} \iff (A - \lambda I_n)\boldsymbol{v} = \mathbf{0} \iff \det(A - \lambda I_n) = 0$

- $\det(A - \lambda I)$ is an n-degree polynomial in λ
- there is at least one and at most n solutions to an n-degree polynomial; in general, eigenvalues could be complex However, all the eigenvalues of a symmetric matrix are real
- once you have the eigenvalues, return to $(A - \lambda I)\boldsymbol{v} = 0$ to compute the corresponding eigenvectors (they will not be unique)

Properties of Eigenvalues

Theorem 111. *$\det(A)$ is the product and $\operatorname{tr}(A)$ is the sum of its eigenvalues.*

- immediate consequence is

Theorem 112. *A is invertible $\iff$ A has no zero eigenvalues $\iff$ $\det(A) \neq 0$.*

- 0 is an eigenvalue of A $\iff$ $A\boldsymbol{v} = 0$ for some $\boldsymbol{v} \in \mathbb{R}^n$ 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 $A \in M_{n,n}$:
  - if A is triangular, then diagonal elements of A are its eigenvalues
  - λ is eigenvalue of A with eigenvector $\boldsymbol{v}$ $\iff$ $\frac{1}{\lambda}$ is eigenvalue of A^{-1} with eigenvector $\boldsymbol{v}$
  - λ is eigenvalue of A $\implies$ λ is eigenvalue of A^T
  - λ is eigenvalue of A with eigenvector $\boldsymbol{v}$ $\implies$ $\forall a \in \mathbb{R} \setminus \{0\}$, $a\lambda$ is eigenvalue of aA with eigenvector $\boldsymbol{v}$

Diagonalization

Theorem 113. *Let $A \in M_{n,n}$. If A has n linearly independent eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, then*

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

where $Q \in M_{n,n}$ has $\boldsymbol{v}_i$ as its ith column, and $\Lambda \in M_{n,n}$ has $\lambda_{ii} = \lambda_i$ and $\lambda_{ij} = 0$ for $i \neq j$. We call this decomposition diagonalization of A.

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

6.3 Other Spaces

Inner Product Space

- vector space only allows to add vectors and multiply them by a scalar
- inner product space is a vector space that also allows to "multiply" vectors

Definition 114. An inner product space over $\mathbb{R}$ is a vector space V together with an operation $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{R}$ (inner product) that satisfy the following properties:

- for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and all $a \in \mathbb{R}$, $\langle a\boldsymbol{u}, \boldsymbol{v} \rangle = a\langle \boldsymbol{u}, \boldsymbol{v} \rangle$;
- for all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$;
- for all $\boldsymbol{v} \in V \setminus \{\mathbf{0}\}$, $\langle \boldsymbol{v}, \boldsymbol{v} \rangle > 0$.

- standard inner product in $\mathbb{R}^n$ is $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^{n} u_i \cdot v_i$ (dot product)

Normed Vector Space

- normed vector space also has a "length" of each vector

Definition 115. A normed vector space over $\mathbb{R}$ is a vector space V together with an operation $\|\cdot\| : V \to \mathbb{R}$ (norm) that satisfy the following properties:

- for all $\boldsymbol{v} \in V$, $\|\boldsymbol{v}\| \geq 0$, with $\|\boldsymbol{v}\| = 0$ if and only if $\boldsymbol{v} = 0$;
- for all $a \in \mathbb{R}$ and all $\boldsymbol{v} \in V$, $\|a\boldsymbol{v}\| = |a| \cdot \|\boldsymbol{v}\|$;
- (triangle inequality) for all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$.

- can turn inner product space $\to$ normed vector space via $\|\boldsymbol{v}\| := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$
- can turn normed vector space $\to$ metric space via $d(\boldsymbol{u}, \boldsymbol{v}) := \|\boldsymbol{u} - \boldsymbol{v}\|$

Cauchy-Schwarz Inequality

Theorem 116. *For all vectors $\boldsymbol{u}$, $\boldsymbol{v}$ of an inner product space,*

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2 \leq \langle \boldsymbol{u}, \boldsymbol{u} \rangle \langle \boldsymbol{v}, \boldsymbol{v} \rangle.$$

The weak inequality holds with equality iff $\boldsymbol{u} = a \cdot \boldsymbol{v}$ for some $a \in \mathbb{R}$.

- if we define the induced norm $\|\boldsymbol{v}\| := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$, we get

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\|$$

- this inequality has many applications, e.g.

$$\mathrm{var}(X)\mathrm{var}(Y) \geq \mathrm{cov}(X, Y)$$

6.4 Hyperplanes

Hyperplane

Definition 117. Let $\boldsymbol{p} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$, and $c \in \mathbb{R}$. The set

$$H(\boldsymbol{p}, c) := \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{p} \rangle = c\}$$

is a hyperplane in $\mathbb{R}^n$.

- note that $H(\boldsymbol{p}, c)$ is a vector space $\iff c = 0$

Separating Hyperplane

Definition 118. Two sets A, $B \in \mathbb{R}^n$ are separated by hyperplane $H(\boldsymbol{p}, c)$ if and only if $\langle \boldsymbol{p}, \boldsymbol{a} \rangle \geq c$ for all $\boldsymbol{a} \in A$ and $\langle \boldsymbol{p}, \boldsymbol{b} \rangle \leq c$ for all $\boldsymbol{b} \in B$.

- A and B lie on opposite sides of the hyperplane $H(\boldsymbol{p}, c)$

Convex Combination

Definition 119. Given any finite collection of points $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$ in $\mathbb{R}^n$, $\boldsymbol{v}$ is a convex combination of these points if there exists $\boldsymbol{\lambda} \in \mathbb{R}^m_+$ such that $\sum_{i=1}^{m} \lambda_i = 1$ and

$$\boldsymbol{v} = \sum_{i=1}^{m} \lambda_i \boldsymbol{v_i}.$$

Definition 120. A set is convex if any convex combination of any two points in the set is also in the set.

Separating Hyperplane Theorem

Theorem 121. *Let A and B be nonempty, convex, disjoint sets in $\mathbb{R}^n$. Then, there exists a hyperplane in $\mathbb{R}^n$ that separates A and B.*

- frequently used in economics, e.g. for finding a set of market-clearing prices, or a utility function that predicts certain choices

7 Lecture 7: Multivariate Calculus and Correspondences

7.1 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 $\mathbb{R}^m$:

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

Linear Transformation

Definition 122. $A : \mathbb{R}^n \to \mathbb{R}^m$ is a <u>linear transformation</u> if

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

- if A is a linear transformation $\iff$ $A \in M_{m,n}$:

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

Differentiation

Definition 123. Let $A \subset \mathbb{R}^n$ and $x \in \text{int}(A)$. Function $f : A \to \mathbb{R}^m$ is differentiable at x if $\exists A \in M_{m,n}$ such that

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

The (total) derivative of f at x is matrix A, denoted as $f'(x)$ or $Df(x)$.
f is said to be differentiable if A is open and f is differentiable at any $x \in A$.
Let $A_1 \subset \text{int}(A)$ be the set of points at which f is differentiable. Function $f' : A_1 \to \mathbb{R}^{mn}$ is the derivative (function) of f.

- $||h||$ is Euclidean norm of the vector h: $||h|| = \sqrt{\sum_{i=1}^{n} h_i^2}$

Understanding the Definition

- let's denote vectors in bold like we did in the previous lecture
- $f : A \to \mathbb{R}^m$, and f is differentiable at $\boldsymbol{x} \in \text{int}(A)$, where $A \subset \mathbb{R}^n$ if

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\|f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x}) - A\boldsymbol{h}\|}{\|\boldsymbol{h}\|} = 0$$

- A is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$
  - **important**: codomain of $f' = A$ is $\mathbb{R}^{mn}$
- A is the derivative of f if linear function $f(\boldsymbol{x}) + A\boldsymbol{h}$ of $\boldsymbol{h}$ approximates $f(\boldsymbol{x}+\boldsymbol{h})$ well when $\boldsymbol{h} \to \boldsymbol{0}$
- **question**: how does this compare to definition of derivative of $f : \mathbb{R} \to \mathbb{R}$?

Total Derivative is Unique if Exists

Theorem 124. *Let $A \subset \mathbb{R}^n$ and $x \in \text{int}(A)$. If $f : A \to \mathbb{R}^m$ is differentiable at x, then there exists a unique $A \in M_{m,n}$ such that*

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

Proof

- suppose that A and B are total derivatives of f at x
- letting $d(h) := f(x+h) - f(x)$, we have

$$\begin{aligned}\lim_{h\to 0}\frac{\|Ah - Bh\|}{\|h\|} &= \lim_{h\to 0}\frac{\|d(h) - Bh - (d(h) - Ah)\|}{\|h\|}\\ &\leq \lim_{h\to 0}\left(\frac{\|d(h) - Bh\|}{\|h\|} + \frac{\|d(h) - Ah\|}{\|h\|}\right) = 0,\end{aligned}$$

  where the second line uses the triangle inequality and $\|-x\| = \|x\|$
- fix any $x \neq 0$. Since $ax \to 0$ as $a \to 0$, the last step implies

$$\begin{aligned}0 &= \lim_{a\to 0}\frac{\|A(ax) - B(ax)\|}{\|ax\|} = \lim_{a\to 0}\frac{\|a\,(Ax - Bx)\|}{|a|\|x\|}\\ &= \lim_{a\to 0}\frac{\|Ax - Bx\|}{\|x\|} = \frac{\|Ax - Bx\|}{\|x\|},\end{aligned}$$

  where the second and third lines use the fact that $\|ax\| = |a|\|x\|$
- since $\|Ax - Bx\| = 0$ for all x, and since $\|x\| = 0$ implies that x is the zero vector, we have $Ax = Bx$ for all x

Properties of Total Derivative

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

Chain rule

Theorem 125. *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 $x \in \mathbb{R}^k$, and if f is differentiable at $g(x)$, then $f \circ g$ is differentiable at x, and*

$$D(f \circ g)(x) = Df(g(x))Dg(x).$$

Functions to $\mathbb{R}^1$

- consider f from $A \subset \mathbb{R}^n$ to $\mathbb{R}^1$ ($m = 1$)
- derivative at $x \in int(A)$ reduces to $1 \times n$ vector
- we call it **gradient**, and denote it $\nabla f(x)$
- for these f, $\nabla f(x) = Df(x)$

Partial Derivative

Definition 126. Let $A \subset \mathbb{R}^n$ and $x \in \text{int}(A)$. For function $f : A \to \mathbb{R}^m$, its partial derivative of i-th coordinate w.r.t. j-th argument at $x \in A$ is

$$\frac{\partial f_i}{\partial x_j}(x) := \frac{d}{dt} f_i(x + te_j)\bigg|_{t=0}$$

if the right-hand side derivative exists.

- $e_j \in \mathbb{R}^n$ is j-th canonical basis vector of $\mathbb{R}^n$ (1 at j-th position, zeros o/w)
- **idea**:

$$\text{fix } x \implies f_i(x + te_j) \text{ a single variable function in } t$$

$$\text{take derivative w.r.t. } t \implies \text{evaluate at } t = 0$$

- $\frac{\partial f_i}{\partial x_j}(x)$ measures sensitivity of f_i w.r.t. x_j

Example

- let $f(x, y,) = \frac{x^2 y}{x^4 + y^2}$ if $(x, y) \neq (0, 0)$, and $f(0, 0) = 0$
- calculate partial derivatives at $(0, 0)$

Total vs. Partial Derivative

Definition 127. Let $A \subset \mathbb{R}^n$ and $x \in \text{int}(A)$. If $f : A \to \mathbb{R}^m$ is differentiable at x, then $\frac{\partial f_i}{\partial x_j}(x)$ exists for any $(i, j) \in \{1, \ldots, m\} \times \{1, \ldots, n\}$, and

$$f'(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x). \end{bmatrix}.$$

Cross Partials

- if the i-th partial derivative of f exists for all $x \in \mathbb{R}^n$, then we can use it to define a function $\frac{\partial f}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}$
- j-th partial derivative of the function above at x is denoted $\frac{\partial^2 f}{\partial x_j \partial x_i}(x)$

Theorem 128. *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 x, then*

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

- we then refer to "cross-partial of f with respect to i and j" without specifying order

First-Order Condition

Theorem 129. *Let $f : A \to \mathbb{R}$ be a function, where $A \subseteq \mathbb{R}^n$. If f has a local extremum at $x \in int(A)$, and if $\nabla f(x)$ exists, then $\nabla f(x) = 0$.*

Mean Value Theorem

- in $\mathbb{R}^n$, we denote the inner product using a dot, so

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

Theorem 130. *Let $f : A \to \mathbb{R}$ be a differentiable function, where $A \subset \mathbb{R}^n$ is open and convex. For any $x, y \in A$, there exists $\lambda \in (0, 1)$ such that*

$$f(y) - f(x) = \nabla f(\lambda x + (1 - \lambda) y) \cdot (y - x).$$

Jacobian

Definition 131. 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 132. 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 133. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on some open set containing x, and suppose $\det(J(x)) \neq 0$. Then, there is an open set U containing x and an open set V containing $f(x)$ such that $f : U \to V$ has a continuous inverse $f^{-1} : V \to U$ which is differentiable for all $v \in V$. Also, the Jacobian of f^{-1} is the inverse of the Jacobian of f.*

Implicit Function Theorem

Theorem 134. *Let $f : A \to \mathbb{R}^n$ be a continuously differentiable function, where $A \subset \mathbb{R}^{m+n}$ is open and has coordinates (x, y). 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 $i = 1, \ldots, n$

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

IFT in Two Variables

- suppose that $f : A \to \mathbb{R}$ is a continuously differentiable function, where $A \subseteq \mathbb{R}^2$ and has coordinates (x, y)
- suppose that $f(x^*, y^*) = 0$ and $\frac{\partial f}{\partial y}(x^*, y^*) \neq 0$
- then, there is a function g defined on an open interval I of x^* such that:
  - $g(x^*) = y^*$
  - $\forall x \in I$, $f(x, g(x)) = 0$
  - $\forall x \in I$

$$g'(x) = -\frac{\frac{\partial f}{\partial x}(x, g(x))}{\frac{\partial f}{\partial y}(x, g(x))}$$

Using the IFT

- suppose we have a model where x is a vector of parameters and y is a vector of choices (e.g. demand and supply)
- suppose that $f(x, y) = 0$ if and only if y is optimal given x. (That is, the solution of the model can be written as a system of equations with 0 on the right-hand side)
- suppose we know that the model has a solution (x^*, y^*). Then, if the conditions of the theorem are satisfied, the solutions of the model for parameters near x^* are given by a function g
- partial derivatives of g tell us how optimal behavior changes as the parameters change (provided they remain close to x^*)

7.2 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 135. 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 136. *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 137. *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 138. 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 139. *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 140. *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 141. *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

8 Lecture 8: Intro to Optimization

8.1 Preliminaries

Classes of Functions

- differentiability class reflects the highest order of derivative that exists and is continuous

Definition 142. Let $f : X \to \mathbb{R}$ be a function, where $X \subseteq \mathbb{R}^n$ is open. f is said to be of differentiability class C^k if the derivatives f', f'', ..., $f^{(k)}$ exist and are continuous on X.

- for example, $f \in C^2$ is twice continuously differentiable

Directional Derivative

- recall that the i-th partial derivative of f reflects how f changes in i-th direction
- how does f changes when we move in an arbitrary direction (not necessarily along one of the axes)?

Definition 143. Let $f : X \to \mathbb{R}$ be a function, where $X \subseteq \mathbb{R}^n$ is open. The directional derivative of f at $\boldsymbol{x} \in X$ along $\boldsymbol{v} \in \mathbb{R}^n$ is

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

if this limit exists.

Directional vs. Partial

Theorem 144. *If f is differentiable at $\boldsymbol{x}$, then the directional derivative always exists and is given by*

$$\nabla_{\boldsymbol{v}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{v}.$$

- recall that $\boldsymbol{u} \cdot \boldsymbol{v} := \sum_{i=1}^{n} u_i v_i$ is the dot product of two vectors in $\mathbb{R}^n$

Example

- derivative of $f(x, y) = x^2 y$ in the direction $\boldsymbol{v} = (1, 2)$ at point $(3, 2)$

Direction of Steepest Ascent

- suppose we want to know the direction in which a differentiable function f ascends most steeply
  - for a given $\boldsymbol{x}$, we maximize $\nabla_{\boldsymbol{v}} f(\boldsymbol{x})$ over vectors $\boldsymbol{v}$
  - this is equivalent to maximizing $\nabla f(\boldsymbol{x}) \cdot \boldsymbol{v}$
  - we can restrict attention to $v \in \mathbb{R}^n$ such that $||v|| = 1$ (unit vectors only)
- by Cauchy-Schwarz inequality,

$$|\nabla f(\boldsymbol{x}) \cdot \boldsymbol{v}| \leq \|\nabla f(\boldsymbol{x})\| \|\boldsymbol{v}\|,$$

  with equality iff $\boldsymbol{v} = \lambda \nabla f(\boldsymbol{x})$ for some $\lambda \in \mathbb{R}$
- direction of steepest ascent is then the (normalized) gradient itself:

$$\frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|}$$

Total Derivative

- suppose that $\boldsymbol{v} = \Delta \boldsymbol{x}$ is a "small" vector
- then,

$$f(\boldsymbol{x} + \Delta \boldsymbol{x}) - f(\boldsymbol{x}) \approx \nabla f(\boldsymbol{x}) \cdot \Delta \boldsymbol{x} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x}) \cdot \Delta x_i$$

- we call $Df(\boldsymbol{x})$ the total derivative of f at $\boldsymbol{x}$

Matrix Definiteness

Definition 145. Let $A \in M_{n,n}$ be a real-valued symmetric matrix. A is

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

Definiteness via Principal Minors

Definition 146. 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 147. *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

Definiteness via Eigenvalues

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

Theorem 148. *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).*

8.2 Optimization Problem

Optimization Problem

- we maximize a function $f : X \to \mathbb{R}$ over all $\boldsymbol{x}$ in set $D \subseteq X$
- f is called the objective function, and D is called the constraint set
- the problem is often written

$$\max_{\boldsymbol{x} \in X} f(\boldsymbol{x}) \text{ subject to } \boldsymbol{x} \in D \qquad \text{or just} \qquad \max_{\boldsymbol{x} \in D} f(\boldsymbol{x})$$

- $\boldsymbol{x}^*$ is a solution to the maximization problem if $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in D$
  - we refer to $\boldsymbol{x}^*$ as maximizer and $f(\boldsymbol{x}^*)$ as maximum or value function

Parameters

- objective function and/or the constraint set often depend on a parameter
- let Θ be the set of parameters; $\boldsymbol{\theta} \in \Theta$ is a vector of parameters
- we then write

$$\max_{\boldsymbol{x} \in X} f(\boldsymbol{x}, \boldsymbol{\theta}) \text{ subject to } \boldsymbol{x} \in D(\boldsymbol{\theta})$$

- solutions to the maximization problem define a correspondence (or function, if the solution is always unique) on Θ
  - that is, $\boldsymbol{x}^*$ depends on $\boldsymbol{\theta}$, and so does $f(\boldsymbol{x}^*)$

Example: Consumer Problem

- consider an agent who consumes n commodities in nonnegative quantities
- agent's utility from consumption is given by $u : \mathbb{R}^n_+ \to \mathbb{R}$
- agent has income $I \geq 0$ and faces price vector $\boldsymbol{p} \in \mathbb{R}^n_+$
- agent maximizes $u(\boldsymbol{x})$ subject to

$$\boldsymbol{x} \in B(\boldsymbol{p}, I) := \{\boldsymbol{x} \in \mathbb{R}^n_+ : \boldsymbol{x} \cdot \boldsymbol{p} \leq I\}$$

- we call $B(\boldsymbol{p}, I)$ the budget set

Example: Political Economy

- policy space is $[0,1]^2$, status quo policy is $\boldsymbol{q} \in [0,1]^2$
- majority party is deciding what policy to pass
  - their utility is by $u_M : [0,1]^2 \to \mathbb{R}$
- president's utility is given by $u_P : [0,1]^2 \to \mathbb{R}$
  - president vetoes policy $\boldsymbol{x}$ if and only if $u_P(\boldsymbol{x}) < u_P(\boldsymbol{q})$
- then, majority party maximizes $u_M(\boldsymbol{x})$ subject to $\boldsymbol{x} \in D(\boldsymbol{q})$, where

$$D(\boldsymbol{q}) := \{\boldsymbol{x} \in [0,1]^2 : u_P(\boldsymbol{x}) \geq u_P(\boldsymbol{q})\}$$

8.3 Unconstrained Optimization

Unconstrained Optimization Problem

- today we consider optimization problems of the form $\max\limits_{\boldsymbol{x} \in X} f(\boldsymbol{x})$, 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 existential question
- recall from Lecture 4:

Theorem 149 (Extreme Value Theorem)**.** *Let X be a compact set, and let $f : X \to \mathbb{R}$ be a continuous function. There exists $\underline{\boldsymbol{x}}$ and $\overline{\boldsymbol{x}}$ in X such that $f(\underline{\boldsymbol{x}}) \leq f(\boldsymbol{x}) \leq f(\overline{\boldsymbol{x}})$ for all $\boldsymbol{x} \in X$.*

First-Order Condition

- recall that if $\boldsymbol{x}^*$ is a local extremum and $\nabla f(\boldsymbol{x}^*)$ exists, then $\nabla f(\boldsymbol{x}^*) = 0$
  - we call all such $\boldsymbol{x}^*$ <u>stationary points</u>

Example

- let $X = \mathbb{R}^2_+$ and $f(x, y) = xy - 2x^4 - y^2$

Second-Order Condition

- denote by $\nabla^2 f(\boldsymbol{x})$ the Hessian (matrix) of f evaluated at $\boldsymbol{x}$
- FOC is satisfied at points that is a max, a min, or neither
- second derivative often resolves this uncertainty

Theorem 150. *Let f be a twice continuously differentiable function and $\nabla f(\boldsymbol{x}^*) = 0$. Then,*

- *$\boldsymbol{x}^*$ is a local maximum (minumum) $\implies \nabla^2 f(\boldsymbol{x}^*) \leq (\geq)\ 0$;*
- *$\nabla^2 f(\boldsymbol{x}^*) > (<)\ 0 \implies \boldsymbol{x}^*$ is a local minimum (maximum).*

Example

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

General Procedure

- find all stationary points $\boldsymbol{x}^*$ of f such that $\nabla f(\boldsymbol{x}^*) = 0$
- for every $\boldsymbol{x}^*$, evaluate $\nabla^2 f(\boldsymbol{x}^*)$
  - if $\nabla^2 f(\boldsymbol{x}^*) > 0$, $\boldsymbol{x}^*$ is a local min
  - if $\nabla^2 f(\boldsymbol{x}^*) < 0$, $\boldsymbol{x}^*$ is a local max
  - discard all other $\boldsymbol{x}^*$
- for global max (min), compare values of f at all local maxima (minima) and boundary points
  - more on global extrema below

Convexity and Concavity: Definition

Definition 151. Given a convex set $X \subseteq \mathbb{R}^n$, we say $f : X \to \mathbb{R}$ is <u>convex</u> (<u>concave</u>) if, for all distinct $\boldsymbol{x}, \boldsymbol{y} \in X$ and all $\alpha \in (0, 1)$,

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

- for strict concavity, replace $\geq$ ($\leq$) with $>$ ($<$)

Convexity and Concavity: Criteria

Theorem 152. *Let $X \subseteq \mathbb{R}^n$ be convex and open. Let $f : X \to \mathbb{R}$ be C^2.*

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

- note that this definition is for f being convex (concave) *everywhere* on its domain; the SOC a few slides above is about just point $\boldsymbol{x}^*$

Standard Example: Cobb-Douglas

- function $f : \mathbb{R}^n \to \mathbb{R}$ given by
$$f(\boldsymbol{x}) = \prod_{i=1}^{n} x_i^{\alpha_i},$$
where each $\alpha_i > 0$, is called the Cobb-Douglas function. It appears in many applications
- f is concave if $\sum\limits_{i=1}^{n} \alpha_i \leq 1$; for $n = 2$, we have

$$\nabla f(\boldsymbol{x}) = f(\boldsymbol{x}) \begin{pmatrix} \frac{\alpha_1}{x_1} \\ \frac{\alpha_2}{x_2} \end{pmatrix} \qquad \nabla^2 f(\boldsymbol{x}) = f(\boldsymbol{x}) \begin{bmatrix} \frac{\alpha_1(\alpha_1-1)}{x_1^2} & \frac{\alpha_1\alpha_2}{x_1x_2} \\ \frac{\alpha_1\alpha_2}{x_1x_2} & \frac{\alpha_2(\alpha_2-1)}{x_2^2} \end{bmatrix}$$

- positive semidefiniteness requires $\alpha_1 \leq 1$ and
$$\alpha_1\alpha_2(\alpha_1 - 1)(\alpha_2 - 1) - \alpha_1^2\alpha_2^2 = \alpha_1\alpha_1\left(1 - \alpha_1 - \alpha_2\right) \geq 0,$$
which is equivalent to $\alpha_1 + \alpha_2 \leq 1$

Uniqueness

Theorem 153. *Let $X \subseteq \mathbb{R}^n$ be convex. Let $f : X \to \mathbb{R}$ be a concave function.*

- *any strict local maximizer of f is a global maximizer;*

- *if the set of maximizers is nonempty, it is a convex set.*

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

Global Maximizers

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

Theorem 154. *Let $X \subseteq \mathbb{R}^n$ be convex, $\boldsymbol{x} \in int(X)$, and $f : X \to \mathbb{R}$ concave (convex). $\boldsymbol{x}$ is a global maximizer (minimizer) of f if and only if $\nabla f(\boldsymbol{x}) = 0$.*

- under concavity, FOC is necessary and sufficient for an *interior* optimum; optima on the boundary still don't have to satisfy the first-order condition

Weakening Concavity

- concavity is a strong requirement not satisfied by many functions that are otherwise nice, e.g. $f(x, y) = x^\alpha y^\beta$, where $\alpha + \beta > 1$
- in economics, we often use functions purely to capture an ordering over the domain; all positive monotonic transformations preserve order, and we don't want to take a stand on which one is "right." However, positive monotonic transformations can break concavity

Quasi-Concavity

Definition 155. Given a convex $X \subseteq \mathbb{R}^n$, we say $f : X \to \mathbb{R}$ is quasi-concave if, for all distinct $\boldsymbol{x}, \boldsymbol{y} \in X$ and all $\alpha \in (0, 1)$,

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

f is quasi-convex if

$$f(\alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) \leq \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}.$$

- for strict quasi-concavity, replace $\geq$ with $>$
- intuitively, quasi-concavity says that a function defined on $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot dip below both $f(\boldsymbol{x})$ and $f(\boldsymbol{y})$ on the segment connecting $\boldsymbol{x}$ and $\boldsymbol{y}$

Derivative Characterization

Theorem 156. *Let $X \subseteq \mathbb{R}^n$ be convex and open, and let $f : X \to \mathbb{R}$ be C^1. f is quasi-concave if and only if, for all x and $y \in X$,*

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) \quad \Longrightarrow \quad \nabla f(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) \geq 0.$$

- recall that $\nabla f(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) = \nabla_{(\boldsymbol{y}-\boldsymbol{x})} f$
- if f increases in total as we move from $\boldsymbol{x}$ to $\boldsymbol{y}$, then predicted change in the function given by directional derivative at $\boldsymbol{x}$ must be positive
- intuitively, if directional derivative at $\boldsymbol{x}$ is negative, then the function dips down before rising again, so it cannot be quasi-concave

Bordered Hessians

- another characterization of quasi-concavity uses the bordered Hessian, which is given by
$$B_f := \begin{bmatrix} 0 & (\nabla f)^T \\ \nabla f & \nabla^2 f \end{bmatrix}$$
- first row and column are called the "border" of the matrix
- k-th leading principal minor of a *bordered* matrix is the determinant of the top left-hand $(k+1) \times (k+1)$ matrix

Another Derivative Characterization

- for concavity, we checked leading principal minors of the Hessian
- change concavity $\to$ quasi-concavity and Hessian $\to$ bordered Hessian

Theorem 157. *Let $X \subseteq \mathbb{R}^n_+$ be convex and open, and $f : X \to \mathbb{R}$ be C^2.*

- *f is strictly quasi-concave if, for each $x \in X$, the 1-st through n-th leading principal minors of $B_f(x)$ are alternately strictly negative and strictly positive;*
- *if f is quasi-concave, then, for each $x \in X$, the 1-st through n-th leading principal minors of $B_f(x)$ are alternately weakly negative and weakly positive.*

Example

- is $f(x) = x^2$ concave? quasi-concave?

Uniqueness

Theorem 158. *Let $X \subseteq \mathbb{R}^n$ be convex, and $f : X \to \mathbb{R}$ be quasi-concave.*

- *Any strict local maximizer of f is a global maximizer.*
- *If the set of maximizers is nonempty, it is a convex set.*

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

- notice that this is exactly the same result we had for concavity

8.4 Envelope Theorems

Envelope Theorems

- suppose an economic agent chooses an action to maximize her utility, which depends on parameter $\boldsymbol{\theta}$
- for a given parameter value $\boldsymbol{\theta}$, she finds a solution (maximizer) $\boldsymbol{x}^*$
- she would like to know how $\boldsymbol{x}^*$ (and $f(\boldsymbol{x}^*)$) changes with $\boldsymbol{\theta}$
- **issue**: when $\boldsymbol{\theta}$ changes, she needs to re-solve the maximization problem
- presumably, taking partial derivative of $\boldsymbol{x}^*$ wrt $\boldsymbol{\theta}$ would not work ... OR WOULD IT?
- **envelope theorems** state that naively calculating $\frac{\partial \boldsymbol{x}^*}{\partial \boldsymbol{\theta}}$ DOES work and agent need not take reoptimization into account

Unconstrained Envelope Theorem

Theorem 159. *Let $f : \mathbb{R}^{n+m} \to \mathbb{R}$ be a C^1 function, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{\theta} \in \mathbb{R}^m$. Suppose that there exists a C^1 function $\boldsymbol{x}^* : \mathbb{R}^m \to \mathbb{R}^n$ given by*

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

and let $V(\boldsymbol{\theta}) := f(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{\theta})$. Then,

$$DV(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial f}{\partial \theta_1} f(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{\theta}) & \cdots & \frac{\partial f}{\partial \theta_m} f(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{\theta}) \end{pmatrix}.$$

- we can ignore changes in $\boldsymbol{x}^*$ when computing the effect of a small change in $\boldsymbol{\theta}$ on the objective function f

9 Lecture 9: 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 $g_j(\boldsymbol{x}) = 0$ for $m = 1, \ldots, M$
  - alternative representation:

$$g(\boldsymbol{x}) = \begin{bmatrix} g_1(\boldsymbol{x}) \\ g_2(\boldsymbol{x}) \\ \vdots \\ g_M(\boldsymbol{x}) \end{bmatrix} = \mathbf{0} \in \mathbb{R}^M \tag{1}$$

Simple Consumer Problem

- $\max\limits_{(x_1,x_2)\in\mathbb{R}^2_+} \log x_1 + \log x_2$ subject to $p_1 x_1 + p_2 x_2 = I$

Motivation

- let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be functions; we solve

$$\max_{(x,y)\in\mathbb{R}^2} f(x,y) \quad \text{subject to} \quad g(x,y) = 0$$

one approach

- use $g(x,y) = 0$ to solve for $y(x)$, then plug back into f and solve

$$\max_{x\in\mathbb{R}} f(x, y(x))$$

- FOC to the unconstrained problem is

- then, for some $\lambda \in \mathbb{R}$, we have

Lagrangian

- let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be functions; we solve

$$\max_{(x,y)\in\mathbb{R}^2} f(x,y) \quad \text{subject to} \quad g(x,y) = 0$$

another approach

- let

$$L(x,y,\lambda) = f(x,y) - \lambda g(x,y)$$

- FOCs wrt x and y are the same:

- FOC wrt λ is

Lagrangian Generalized

- for any $f : \mathbb{R}^n \to \mathbb{R}$ and set of constraints $\{g_m(\boldsymbol{x}) = 0\}_{m=1}^M$, we can define $L(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) - \sum_{m=1}^{M} \lambda_m \cdot g_m(\boldsymbol{x})$. The FOCs are

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \sum_{m=1}^{M} \lambda_m \frac{\partial g_m}{\partial x_i}(\boldsymbol{x}) \quad \text{for all } i = 1, \dots, n$$
$$g(\boldsymbol{x}) = \mathbf{0} \in \mathbb{R}^M$$

- L is called the Lagrangian, and $\boldsymbol{\lambda} \in \mathbb{R}^M$ the vector of Lagrange multipliers

Example

- maximize $f(x,y) = x^2 - 2xy + y^2$ subject to $x^2 + y = 0$

Interpretation of Multiplier

- to interpret λ_m^*, consider constraint $g_m(\boldsymbol{x}) = c$, where $c > 0$
- the Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{m=1}^{M} \lambda_m \cdot g_m(\boldsymbol{x}) + \lambda_m \cdot c$$

- let $V(c) = L\big(\boldsymbol{x}^*(c), \boldsymbol{\lambda}^*(c)\big)$ be the value function
- apply the unconstrained envelope theorem to get $V'(c) = \lambda_m^*$
- then, λ_m^* can be thought of as the *value of relaxing m-th constraint*:
  - if we change the constraint from $g_m(\boldsymbol{x}) = 0$ to $g_m(\boldsymbol{x}) = c$, the maximized objective increases by $\lambda_m^* \cdot c$
  - if $\lambda_m^* = 0$, then relaxing the constraint has no effect

Lagrange

Theorem 160. *Let $X \subseteq \mathbb{R}^n$ be open, $f : X \to \mathbb{R}^n$ and $g_1, \ldots, g_M$ (all $X \to \mathbb{R}^n$) be C^1. Suppose $\boldsymbol{x}^*$ is a local maximum (or minimum) of f on the set*

$$\mathcal{D} := \{\boldsymbol{x} \in X \mid g_m(\boldsymbol{x}) = 0 \textit{ for all } m = 1, \ldots, M\}.$$

Suppose also that vectors $Dg_1(\boldsymbol{x}^*), \ldots, Dg_M(\boldsymbol{x}^*)$ are linearly independent. Then, there exists $\boldsymbol{\lambda} \in \mathbb{R}^M$ such that

$$Df(\boldsymbol{x}^*) = \sum_{m=1}^{M} \lambda_m \cdot Dg_m(\boldsymbol{x}^*).$$

Constraint Qualification

- requirement that $Dg_1(\boldsymbol{x}^*), \ldots, Dg_M(\boldsymbol{x}^*)$ are linearly independent is called the constraint qualification (CQ)
  - same as $\operatorname{rank} Dg(\boldsymbol{x}^*) = M$, where $Dg(\boldsymbol{x}^*) := [Dg_1(\boldsymbol{x}^*), \ldots, Dg_M(\boldsymbol{x}^*)]^T$
- if $\boldsymbol{x}^*$ is a local optimum but $\operatorname{rank} Dg(\boldsymbol{x}^*) < M$, then there may not be any $\boldsymbol{\lambda}$ such that $(\boldsymbol{x}^*, \boldsymbol{\lambda})$ solves the FOC
  - any point where CQ fails needs to be checked separately
- if $M = 1$ (only one constraint), CQ becomes $Dg(\boldsymbol{x}^*) \neq 0$

Example

- maximize $f(x,y) = -y$ subject to $g(x,y) = y^3 - x^2 = 0$

Second-Order Conditions

Theorem 161. *Let $X \subseteq \mathbb{R}^n$ be open, $f : X \to \mathbb{R}^n$ and $g_1, \ldots, g_M$ (all $X \to \mathbb{R}^n$) be C^2.*
Let $\mathcal{D} := \{\boldsymbol{x} \in X \mid g_m(\boldsymbol{x}) = 0 \text{ for all } m = 1, \ldots, M\}$,
suppose that $\boldsymbol{x}^* \in \mathcal{D}$ and $\operatorname{rank} Dg(\boldsymbol{x}^*) = M$.
Let $\boldsymbol{\lambda} \in \mathbb{R}^M$ be such that $Df(\boldsymbol{x}^*) = \sum_{m=1}^{M} \lambda_m \cdot Dg_m(\boldsymbol{x}^*)$.
Let $Z(\boldsymbol{x}^*) := \{z \in \mathbb{R}^n \mid Dg_m(\boldsymbol{x}^*) \cdot z = 0, \text{ for all } m = 1, \ldots, M\}$.

Then,

- *if $z^T[D^2_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda})]z < 0$ for all $z \in Z(\boldsymbol{x}^*) \setminus \{\boldsymbol{0}\}$, then f achieves a local maximum at $\boldsymbol{x}^*$ on $\mathcal{D}$;*
- *if $z^T[D^2_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda})]z > 0$ for all $z \in Z(\boldsymbol{x}^*) \setminus \{\boldsymbol{0}\}$, then f achieves a local minimum at $\boldsymbol{x}^*$ on $\mathcal{D}$*

- xxx

Arguing Uniqueness

- if there is a unique $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ that satisfies the FOC <u>and</u> CQ is satisfied $\forall \boldsymbol{x} \in X$, then $\boldsymbol{x}^*$ is a global optimum
  - use local SOC to figure out whether minimum or maximum
- if $\boldsymbol{x}^*$ is a strict local maximum <u>and</u> $\mathcal{D}$ is a convex set <u>and</u> f is strictly quasi-concave on $\mathcal{D}$, then $\boldsymbol{x}^*$ is a global maximum

Equality-Constrained Optimization Problems: Step 1

- problem:

$$\max_{\boldsymbol{x} \in X} f(\boldsymbol{x}) \text{ subject to } \boldsymbol{x} \in \mathcal{D} := \{\boldsymbol{x} \in X \mid g_m(\boldsymbol{x}) = 0 \text{ for all } m = 1, \ldots, M\},$$

  where $X \subseteq \mathbb{R}^n$ be open, $f : X \to \mathbb{R}^n$ and $g_1, \ldots, g_M$ (all $X \to \mathbb{R}^n$) be C^2
- **step 1**: argue that a solution exists

- e.g. by showing that the objective function is continuous and the constraint set is compact

Step 2: Constraint Qualification

- check whether CQ, rank $Dg(\boldsymbol{x}) = M$, holds at any $\boldsymbol{x} \in \mathcal{D}$
  - if it fails, define $\overline{y} \in \mathbb{R} \cup \{\infty\}$ by

$$\overline{y} := \sup\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{D} \text{ and rank } Dg(\boldsymbol{x}) < m\}$$

Step 3: First Order Condition

- set up the Lagrangian

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) - \sum_{m=1}^{M} \lambda_m \cdot g_m(\boldsymbol{x})$$

- find all points $(\boldsymbol{x}^*, \boldsymbol{\lambda}) \in X \times \mathbb{R}^M$ at which $DL(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$
  - each $(\boldsymbol{x}^*, \boldsymbol{\lambda})$ solves

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}^*) = \sum_{m=1}^{M} \lambda_m \cdot \frac{\partial g_m}{\partial x_i}(\boldsymbol{x}^*) \text{ for all } i \in \{1, \ldots, n\}$$
$$g_m(\boldsymbol{x}^*) = 0 \text{ for all } m \in \{1, \ldots, M\}$$

Step 4

- compute the value of f at each of the points from Step 3
- if the value of f at one of these points is weakly greater than the value of f at all the others *and* $\overline{y}$, then that point is a global maximizer
- if $\overline{y}$ is weakly greater than the value of f at all the points from step 3, and if there is some $\overline{\boldsymbol{x}} \in \mathcal{D}$ at which the CQ fails and $f(\overline{\boldsymbol{x}}) = \overline{y}$, then $\overline{\boldsymbol{x}}$ is a global maximizer

Summing Up

Theorem 162. *Suppose the following conditions hold for a given equality-constrained optimization problem.*

- *A global optimum $\boldsymbol{x}^*$, exists.*
- *The constraint qualification is met at $\boldsymbol{x}^*$.*

Then, there exists $\boldsymbol{\lambda}^*$ such that $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a critical point of the Lagrangian.

- follows directly from Lagrange theorem—a way of remembering what Lagrangian does

Example

- maximize $f : \mathbb{R}^2 \to \mathbb{R}$ given by

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

over the constraint set $x^2 + y^2 = 1$

Equality-Constrained Envelope Theorem

Theorem 163. *Let $X \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^K$ be open, and let $f : X \times \Theta \to \mathbb{R}$ and $g : X \times \Theta \to \mathbb{R}^M$ be C^1 functions. Let $\boldsymbol{x}^* : \Theta \to \mathbb{R}^n$ be a C^1 function such that, for all $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{x}^*(\boldsymbol{\theta})$ maximizes $f(\boldsymbol{x}, \boldsymbol{\theta})$ over $\boldsymbol{x} \in X$ subject to $g(\boldsymbol{x}, \boldsymbol{\theta}) = 0$. Let $\boldsymbol{\lambda}^* : \Theta \to \mathbb{R}^M$ be a C^1 function such that, for all $\boldsymbol{\theta} \in \Theta$, $(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{\lambda}^*(\boldsymbol{\theta}))$ satisfies the first-order conditions of the Lagrangian. Let $V^*(\boldsymbol{\theta}) := f(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{\theta})$. Then,*

$$DV^*(\theta) = \begin{pmatrix} \frac{\partial L}{\partial \theta_1}(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{\lambda}^*(\boldsymbol{\theta}), \boldsymbol{\theta}) & \cdots & \frac{\partial L}{\partial \theta_K}(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{\lambda}^*(\boldsymbol{\theta}), \boldsymbol{\theta}) \end{pmatrix}.$$

- apply envelope theorem for the unconstrained optimization problem of maximizing the Lagrangian
- **additional step**: observing that $V^*(\boldsymbol{\theta}) = L(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{\lambda}^*(\boldsymbol{\theta}))$

10 Lecture 10: Inequality-Constrained Optimization

10.1 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 $\mathcal{D} := \{\boldsymbol{x} \in X : g(\boldsymbol{x}) \leq 0\}$, 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
  - m-th constraint is slack at $\boldsymbol{x} \in \mathbb{R}^n$ if $g_m(\boldsymbol{x}) < 0$
  - m-th constraint is binding at $\boldsymbol{x} \in \mathbb{R}^n$ if $g_m(\boldsymbol{x}) = 0$
- when applying KKT, we have to make a conjecture (or argument) about which constraints are binding

KKT Theorem

Theorem 164. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^M$ be C^1. Suppose that the first k constraints are binding at $\boldsymbol{x} \in \mathbb{R}^n$, and that the gradients of the binding constraints are linearly independent. If $\boldsymbol{x}^*$ maximizes f subject to $g(\boldsymbol{x}) \leq 0$, then there is a unique $\boldsymbol{\lambda} \in \mathbb{R}^k_+$ such that*

$$Df(\boldsymbol{x}^*) = \sum_{m=1}^{k} \lambda_m \cdot Dg_m(\boldsymbol{x}^*),$$
$$\lambda_m \cdot g_m(\boldsymbol{x}^*) = 0 \textit{ for all } m = 1, \ldots, k.$$

The latter equation is called the complementary slackness condition.

Necessary vs. Sufficient

- as with FOCs of the Lagrangian, KKT conditions are necessary but not sufficient for a local optimum
- consider example
  - $\max_{x \in \mathbb{R}} = x^3$ subject to $g(x) = -(x+1) \leq 0$
  - let $x^* = \lambda^* = 0$

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

$$\max_{(x,y)\in\mathbb{R}^2} -(x^2+y^2) \quad \text{subject to} \quad -(x-1)^3+y^2 \leq 0$$

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

Interpretation of Multiplier

- interpretation of λ_m is the same as in the equality-constrained case
- notice that $\lambda_m = 0$ if $g_m(\boldsymbol{x}) < 0$, and $g_m(\boldsymbol{x}) = 0$ if $\lambda_m > 0$
- consider optimum $\boldsymbol{x}^*$. If $g_m(\boldsymbol{x}^*) < 0$, then the m-th constraint is not binding, so there is no benefit to relaxing it. If $\lambda_m > 0$, 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 $g_m(\boldsymbol{x}^*)$. If a particular constraint is not binding at the optimum, then values of $\boldsymbol{x}$ that violate the constraint are not desirable anyway, so there is no need to penalize them
- **caution**: complementary slackness does not rule out $\lambda_m = 0$ and $g_m(\boldsymbol{x}^*) = 0$. This will occur if the optimum *without* the m-th constraint happens to satisfy the m-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. If so, compute $\overline{y} := \sup f$ over this set of points

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

$$\begin{aligned}
\frac{\partial f}{\partial x_i}(\boldsymbol{x}) &= \sum_{m=1}^{M} \lambda_m \frac{\partial g_m}{\partial x_i}(\boldsymbol{x}) \text{ for all } i \in \{1,\ldots,n\} \\
g_m(\boldsymbol{x}) &\leq 0 \text{ for all } m \in \{1,\ldots,M\} \\
\lambda_m g_m(\boldsymbol{x}) &= 0 \text{ for all } m \in \{1,\ldots,M\}
\end{aligned}$$

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

Example: Importance of Step 1

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

Example: Importance of Step 2

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

Summing Up

Theorem 165. *Suppose the following conditions hold for a given inequality-constrained optimization problem:*

- *a global optimum, $\boldsymbol{x}^*$, exists;*
- *the constraint qualification is satisfied at $\boldsymbol{x}^*$.*

Then, there exists $\boldsymbol{\lambda}^*$ such that $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ satisfies the KKT conditions.

- follows directly from the KKT theorem

Example

- $\max\limits_{(x,y) \in \mathbb{R}^2} x^2 - y$ subject to $x^2 + y^2 - 1 \leq 0$

Example

- $\max\limits_{(x,y) \in \mathbb{R}^2_{++}} \log x + \log y$ subject to $x^2 + y^2 - 2 \leq 0$

Convex Optimization

- sometimes, we can get around Step 1 (proving a solution exists)
  - when f is quasi-concave and $\mathcal{D}$ is convex
  - we start with concavity, then move on to quasi-concavity

Concavity and KKT

Theorem 166. *Let $f : X \to \mathbb{R}$ be a concave C^1 function, where $X \subseteq \mathbb{R}^n$ is convex and open. For each $m \in \{1, \ldots, M\}$, let $g_m : X \to \mathbb{R}$ be a convex C^1 function. Suppose there exists $\overline{x} \in X$ such that $g_m(\overline{x}) < 0$ for all m. Then, $\boldsymbol{x}^*$ maximizes $f(\boldsymbol{x})$ subject to $g_m(\boldsymbol{x}) \leq 0$ for all m if and only if there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ satisfies the KKT conditions.*

- notice there is no CQ here
- **useful fact**: if the g_m 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 exist 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

Example: Slater's Condition

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

Quasiconcavity and KKT

Theorem 167. *Let $f : X \to \mathbb{R}$ be a quasiconcave C^1 function, where $X \subseteq \mathbb{R}^n$ is open and convex. For $m \in \{1, \ldots, M\}$, let $g_m : X \to \mathbb{R}$ be a quasiconvex C^1 function. Suppose that $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ satisfy the KKT conditions. Then, $\boldsymbol{x}^*$ is an optimum if (1) f is concave, or (2) $Df(\boldsymbol{x}^*) \neq 0$.*

- **caution**: this is just a *sufficient* condition for an optimum

10.2 Maximum Theorem

Continuity of Maxima

- now consider a more general problem:

$$\max_{\boldsymbol{x} \in \mathcal{D}(\boldsymbol{\theta})} f(\boldsymbol{x}, \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} \in \Theta$ is vector of parameters

- let
  - $f^*(\boldsymbol{\theta}) := \max_{\boldsymbol{x} \in \mathcal{D}(\boldsymbol{\theta})} f(\boldsymbol{x}, \boldsymbol{\theta})$ be the maximized objective
  - $\mathcal{D}^*(\boldsymbol{\theta}) := \{\boldsymbol{x} \in \mathcal{D}(\boldsymbol{\theta}) : f(\boldsymbol{x}, \boldsymbol{\theta}) = f^*(\boldsymbol{\theta})\}$ be the set of maximizers
- under which conditions are $f^*(\boldsymbol{\theta})$ and $\mathcal{D}^*(\boldsymbol{\theta})$ continuous?
  - note that $\mathcal{D}(\boldsymbol{\theta})$ and $\mathcal{D}^*(\boldsymbol{\theta})$ are correspondences

Berge's Theorem of the Maximum

Theorem 168. *Let (X, d_X) and (Θ, d_Θ) be metric spaces, and let $f : X \times \Theta \to \mathbb{R}$ be a continuous function. Let $\mathcal{D} : \Theta \rightrightarrows X$ be a correspondence with nonempty and compact values. Let*

$$f^*(\boldsymbol{\theta}) := \max_{\boldsymbol{x} \in \mathcal{D}(\boldsymbol{\theta})} f(\boldsymbol{x}, \boldsymbol{\theta})$$

$$\mathcal{D}^*(\boldsymbol{\theta}) := \{\boldsymbol{x} \in \mathcal{D}(\boldsymbol{\theta}) : f(\boldsymbol{x}, \boldsymbol{\theta}) = f^*(\boldsymbol{\theta})\}.$$

If $\mathcal{D}$ is continuous, then f^* is continuous and $\mathcal{D}^*$ is upper hemicontinuous with nonempty and compact values.

- recall that continuity of $\mathcal{D}$ is equivalent to UHC + LHC of $\mathcal{D}$

11 Lecture 11: Dynamic Programming with Finite Horizon

Motivation: Consumption over Two Periods

- suppose that you live for 2 periods, today and tomorrow
- today, you have wealth w_1
- at each period $t = 1, 2$ you choose
  - how much to consume, $c_t \in [0, w_t]$
  - how much to save, $w_t - c_t$
    - ▷ ROI is $R > 1$ (invest \$$1$ today to get \$$R$ tomorrow)
- there's no extra income at $t = 2$
- how do you maximize your lifetime consumption $\log c_1 + \beta \log c_2$?
  - $\beta \in (0, 1]$ is your discount factor

Overview

- in many economic decision problems, today's action affects what happens tomorrow
  - if investor sells stock today, she can't sell it in the future at a possibly higher price
  - if consumer spends more today, he will have to restrict his spending in the future
  - if firm engages in R&D today, it may discover some profitable products
- dynamic programming is used to determine the optimal course of actions in these types of problems
- classic reference: *Recursive Methods in Economic Dynamics* (1989) by Stokey and Lucas

Finite Horizon Problem

Definition 169. A finite-horizon Markov dynamic programming problem is given by

$$\{S, A, T, (r_t, f_t, \Phi_t)_{t=1}^{T}\}, \text{ where}$$

- S is the state space;
- A is the action space;
- $T \in \mathbb{N}$ is the time horizon;
- $r_t : S \times A \to \mathbb{R}$ is the period-t reward function;
- $f_t : S \times A \to S$ is the period-t transition function;
- $\Phi_t : S \rightrightarrows A$ is the period-t feasible action correspondence.

- this is the deterministic version; can add uncertainty

Interpretation

- agent must choose an action in each of T periods:
  - at $t = 1$, he chooses $a_1 \in \Phi(s_1)$ and gets reward $r_1(a_1, s_1)$
  - at $t = 2$, state is $s_2 = f_1(a_1, s_1)$ and $a_2 \in \Phi(s_2)$ and gets reward $r(a_2, s_2)$
  - ... and so on
- notice that the state space is a way of capturing both (1) which actions are feasible and (2) the rewards from each action.
- conditional on state, today's feasible actions and rewards are independent of past states and actions; this is what makes the problem "Markov"
- this *does not* mean that today's feasible actions and rewards are unconditionally independent of past states and actions; yesterday's states and actions still influence what state it is today

Problem

- agent's problem is to maximize
$$\sum_{t=1}^{T} r_t(s_t, a_t)$$
subject to
$$\begin{aligned} s_1 &= s \in S \\ s_{t+1} &= f_t(s_t, a_t) \text{ for all } t \in \{1, \dots, T-1\} \\ a_t &\in \Phi_t(s_t) \text{ for all } t \in \{1, \dots, T\} \end{aligned}$$

Example: Optimal Consumption Problem

- suppose that you live for $T \geq 2$ periods
- your initial wealth is $w_1 > 0$
- at each period $t = 1, \dots, T$, you choose
  - how much to consume, $c_t \in [0, w_t]$
  - how much to save, $w_t - c_t$
    - ▷ ROI is $R > 1$ (invest \$1 at t to get \$$R$ at $t+1$)
- there is no extra income apart from returns on investments
- how do you maximize your lifetime consumption $\log c_1 + \beta \log c_2 + \cdots + \beta^{T-1} \log c_T$, where $\beta \in (0, 1]$ is the discount factor?

History

- to describe the solution, will need a way of describing what has already happened up to a particular period
- for any $t \in \{1, \dots, T\}$, a t-history is a member of $S^t \times A^{t-1}$:
$$\eta_t = (s_1, a_1, \dots, s_{t-1}, a_{t-1}, s_t)$$

- we will restrict attention to possible t-histories, i.e. t-histories such that, for all $\tau \in$

$\{1, \ldots, t-1\}$,

$$a_\tau \in \Phi_\tau(s_\tau)$$
$$s_{\tau+1} = f_t(s_t, a_t)$$

- for each $t \in \{1, \ldots, T\}$, let H_t denote the set of possible t-histories
  - notice that $H_1 = S$
- it is convenient to write $s_t[\eta_t]$ to denote period-t state when history is η_t

Strategy

- a strategy σ is a sequence $\{\sigma_t\}_{t=1}^T$, where $\sigma_t : H_t \to A$ is such that

$$\sigma_t(\eta_t) \in \Phi_t(s_t[\eta_t])$$

- for each possible history, a strategy prescribes a feasible action for the agent to take
- notice that a strategy prescribes actions even at histories that will not be reached if the strategy is followed
  - this isn't necessary for computing the rewards from the strategy, it is useful for showing optimality
- let Σ (capital σ) denote the set of all possible strategies

Strategy

- starting from any initial state s, a strategy σ yields a unique sequence of states and actions

$$\{s_t(\sigma, s), a_t(\sigma, s)\}_{t=1}^T$$

  and a unique sequence of histories

$$\{\eta_t(\sigma, s)\}_{t=1}^T$$

- these are recursively defined by $\eta_1(\sigma, s) = s$ and

$$\begin{aligned} a_t(\sigma, s) &= \sigma_t(\eta_t(\sigma, s)) & \forall t \in \{1, \ldots, T\} \\ s_{t+1}(\sigma, s) &= f_t(s_t(\sigma, s), a_t(\sigma, s)) & \forall t \in \{1, \ldots, T-1\} \\ \eta_{t+1}(\sigma, s) &= (\eta_t(\sigma, s), a_t(\sigma, s), s_t(\sigma, s)) & \forall t \in \{1, \ldots, T-1\} \end{aligned}$$

Reward and Value

- reward delivered by σ in period t, given initial state s, is

$$r_t(\sigma, s) := r_t(s_t(\sigma, s), a_t(\sigma, s))$$

- the total value of σ, given initial state s, is

$$W(\sigma, s) := \sum_{t=1}^{T} r_t(\sigma, s)$$

- let

$$V(s) := \sup_{\sigma \in \Sigma} W(\sigma, s)$$

  be the value function

Optimality

- strategy $\sigma^* \in \Sigma$ is optimal if it maximizes $W(\sigma, s)$ **for any initial state** s
- that is, for all $s \in S$,

$$W(\sigma^*, s) = \sup_{\sigma \in \Sigma} W(\sigma, s) = V(s)$$

- solving the agent's problem is equivalent to finding an optimal strategy

Continuation Problem

- following any τ-history η_t in a Markov problem, the agent faces a Markov problem on the remaining $T-\tau+1$ periods

- this problem is $\{S, A, T-\tau+1, (r_t^\tau, f_t^\tau, \Phi_t^\tau)_{t=1}^{T-\tau+1})\}$, where

$$\begin{aligned} r_t^\tau &= r_{t+\tau-1} \\ f_t^\tau &= f_{t+\tau-1} \\ \Phi_t^\tau &= \Phi_{t+\tau-1} \end{aligned}$$

- we call this the $(T-\tau+1)$-period continuation problem or the continuation problem at τ
- to avoid re-indexing r, f, and Φ, we simply denote it $\{S, A, T-\tau+1, (r_t, f_t, \Phi_t)_{t=\tau}^{T}\}$

Markov Strategy

- all t-histories have the same continuation problem, so the value the agent can achieve cannot depend on past states or actions
- thus, it is reasonable to think (and turns out to be true) that we won't lose anything by searching for optimal strategies within the class of strategies that don't depend on past states or actions; these are called Markov strategies
- while there may be optimal strategies that are not Markov, they won't do any better than the best Markov strategy
- intuitively, current state captures all past information that is relevant for the future—so there is nothing to be gained by conditioning on additional past information

Markov Strategy

- formally, $\sigma \in \Sigma$ is Markov if, for all $t \in \{1, \ldots, T\}$, there exists $g_t : S \to A$ such that

$$\sigma_t(\eta_t) = g_t(s_t[\eta_t])$$

- thus, we often denote a Markov strategy by $\{g_t\}_{t=1}^{T}$, where $g_t : S \to A$ satisfies $g_t(s) \in \Phi_t(s)$ for all $s \in S$
- a Markov optimal strategy is a Markov strategy that is optimal
  - σ^* is optimal only if it does at least as well as *every* other strategy

From Optimal to Markov Optimal

Theorem 170. *Let σ be an optimal strategy for the problem $\{S, A, T, (r_t, f_t, \Phi_t)_{t=1}^{T}\}$. Suppose that, for some $\tau \in \{1, \ldots, T\}$, the $(T - \tau + 1)$-period continuation problem has a Markov optimal strategy $\{g_t\}_{t=\tau}^{T}$. Then, the strategy $\hat{\sigma}$ given by*

$$\hat{\sigma}_t(\eta_t) = \begin{cases} g_t(s_t[\eta_t]) \text{ if } t \in \{\tau, \ldots, T\} \\ \sigma_t(\eta_t) \text{ if } t \in \{1, \ldots, \tau - 1\} \end{cases}$$

is optimal for the original problem.

- intuitively, once the agent gets to τ, he can't do any better by following σ than by following $\{g_t\}_{t=\tau}^{T}$ so he might as well play the Markov strategy starting that time

Existence

- can we guarantee that a problem and all of its continuation problems have an optimal strategy?
- if A and S are finite, then Σ is finite, so an optimal strategy exists
- more generally, we will need the following assumptions:
  - **A1**: each r_t is continuous and bounded
  - **A2**: each f_t is continuous
  - **A3**: each Φ_t is continuous and compact-valued

Backward Induction Solution

- if we assume **A1—A3**, we can construct a Markov strategy as follows:
  - for each $t < T$, let

$$g_t(s_t) \in \arg \max_{a_t \in \Phi_t(s_t)} \left(r_t(a_t, s_t) + W\left(\{g_\tau\}_{\tau=t+1}^{T}, f_t(a_t, s_t) \right) \right)$$

  and let

$$g_T(s_T) \in \arg \max_{a_T \in \Phi_T(s_T)} r_T(a_T, s_T)$$

- any such strategy is called a backward induction solution; next we verify that it is actually a solution

Proof

- clearly, g_T is optimal for the last-period continuation problem
- suppose that $\{g_\tau\}_{\tau=t+1}^T$ is optimal for the $(T-t)$-period continuation problem, and that $\{\sigma_\tau\}_{\tau=t}^T$ is optimal for the $(T-t+1)$-period problem
- then, the strategy for the $(T-t+1)$-period problem formed by replacing σ_τ with g_τ for all $\tau > t$ is also optimal for that problem. A strategy that agrees with g_τ for all $\tau > t$ cannot do any better than

$$\max_{a_t \in \Phi_t(s_t)} \left(r_t(a_t, s_t) + W(\{g_\tau\}_{\tau=t+1}^T, f_t(a_t, s_t)\right),$$

  which is the value achieved by $\{g_\tau\}_{\tau=t}^T$

Bellman Equation

Theorem 171. *Under* **A1—A3**, *a finite-horizon Markov dynamic programming problem admits a Markov optimal strategy. For each t, the value function V_t satisfies*

$$V_t(s_t) = \max_{a_t \in \Phi_t(s_t)} \left(r_t(s_t, a_t) + V_{t+1}(s_{t+1})\right)$$

for all $s_t \in S$.

- even in infinite-horizon problems, where we can no longer do backward induction, the Bellman equation still holds

12 Lecture 12: Dynamic Programming with Infinite Horizon

Infinite Horizon

- many problems in economics have infinite horizon, because:
  - we want to study the behavior of agents over time, and some problems don't have fixed horizons
  - even in situations that have to terminate eventually (e.g. a lab experiment or a person's life), the agent may never be able to say for sure whether she has reached the last period
- to make an infinite-horizon problem tractable, need to assume some degree of stationarity (meaning the environment does not change too much or too unpredictably with

time)

Infinite Horizon Problem

Definition 172. A stationary discounted dynamic programming problem is $\{S, A, \Phi, r, f, \delta\}$, where

- $S \subseteq \mathbb{R}^n$ is the state space,
- $A \subseteq \mathbb{R}^m$ is the action space,
- $\Phi : S \rightrightarrows A$ is the feasible action correspondence,
- $r : S \times A \to \mathbb{R}$ is the reward function,
- $f : S \times A \to S$ is the transition function, and
- $\delta \in (0, 1)$ is the discount factor.

Problem

- the agent's problem is to maximize

  $$\sum_{t=0}^{\infty} \delta^t r(s_t, a_t)$$

  subject to

  $$\begin{aligned} s_0 &= s \in S \\ s_{t+1} &= f(s_t, a_t) \text{ for all } t \in \{0, 1, \ldots\} \\ a_t &\in \Phi(s_t) \text{ for all } t \in \{0, 1, \ldots\} \end{aligned}$$

Strategies

- we can define histories and strategies exactly as in the finite horizon problem
- as before, from any given initial state, a strategy yields a unique sequence of states and actions and a unique sequence of histories; the sequence of states and actions pins down the sequence of rewards

- also as before, we can define

$$W(\sigma, s) := \sum_{t=0}^{\infty} \delta^t r(s, \sigma)$$
$$V(s) := \sup_{\sigma \in \Sigma} W(s, \sigma)$$

- again, a strategy σ is optimal if, for all $s \in S$,

$$W(s, \sigma) = V(s)$$

Stationary Strategy

- recall: for a finite-horizon Markov dynamic programming problem, we look for optimal strategies that depend only on the period and the current state
- for a stationary discounted dynamic programming problem, we will also drop dependence on the period; stationarity ensures that the continuation problem is identical to the original problem, so there is nothing to be gained by conditioning on the period
- thus, we will look for an optimal strategy $\pi^* : S \to A$; this is called a stationary strategy
- caution: a stationary strategy does not require the agent to do the same thing in every period

Bellman Equation

- recall: for a finite-horizon Markov dynamic programming problem, the value function satisfies the Bellman equation

$$V_t(s_t) = \max_{a_t \in \Phi_t(s_t)} \left(r_t(s_t, a_t) + V_{t+1}(f_t(s_t, a_t)) \right).$$

- for a stationary discounted dynamic programming problem, the value function will satisfy the **Bellman equation**

$$V(s) = \max_{a \in \Phi(s)} \left(r(s, a) + \delta V(f(s, a)) \right).$$

- notice that the value function is not indexed by the period in the infinite-horizon case; intuitively, this is to be expected because the environment looks exactly the same in every period

Existence

Theorem 173. *Suppose the stationary discounted dynamic programming problem satisfies the following conditions:*

- $r : S \times A \to \mathbb{R}$ *is continuous and bounded.*
- $f : S \times A \to S$ *is continuous.*
- $\Phi : S \rightrightarrows A$ *is a compact-valued, continuous correspondence.*

Then, there exists a stationary optimal strategy $\pi^* : S \to A$. *Moreover, the value function* $V(s) = W(\pi^*, s)$ *is continuous on* S, *and is the unique bounded function that satisfies the Bellman equation at each* $s \in S$.

12.1 Optimal Growth Problem

Optimal Growth Problem

- for the rest of today, will restrict attention to a special case of a stationary discounted dynamic programming problem, called the optimal growth problem
- working through this problem is useful because:
  - s similar approach can work with other dynamic programming problems
  - it demonstrates how much can be done *without* assuming anything about functional forms

Optimal Growth Problem

- consider an agent who can both produce and consume a single good
- the state variable is the current stock of the good, $y_t \in \mathbb{R}_+$
  - agent begins with endowment $y_0 > 0$
- each period $t = 0, \ldots$, the agent picks a consumption level $c_t \in [0, y_t]$
  - receives utility $u(c_t)$
- the rest of the stock, $x_t = y_t - c_t$, is invested

  - next period's stock becomes $y_{t+1} = f(y_t - c_t)$
- the agent discounts the future by $\delta \in (0,1)$

Boundedness

- we would like to apply our existence theorem, but we may not want to assume that u is bounded above on $\mathbb{R}_+$
- instead, will make the following assumptions about the production process, collectively denoted **A1a**:
  - $f(0) = 0$
  - f is continuous and nondecreasing on $\mathbb{R}_+$
  - there is $\overline{x} > 0$ such that $f(x) \leq x$ for all $x \geq \overline{x}$, and $y_0 \leq \overline{x}$
- also assume that $y_0 \in [0, \overline{y}]$, where $\overline{y} < \infty$ (**A1b**)
- together, these conditions ensure that y_t (and, by extension, c_t) cannot be outside $[0, \max\{\overline{x}, \overline{y}\}]$; we set $S = A = [0, \max\{\overline{x}, \overline{y}\}]$.
- now, continuity of u (**A2**) implies that u is bounded on A

Existence

- we now apply the existence result and conclude that there is a stationary optimal strategy
- the result also implies that $V(y)$ is continuous, and is the unique bounded function that satisfies

$$V(y) = \max_{c \in [0,y]} \left(u(c) + \delta V(f(y-c)) \right)$$

  for each $y \in S$

Concavity

- to say more, will add the following assumptions:
  - **A3**: u is strictly increasing
  - **A4**: u is strictly concave
  - **A5**: f is concave

Theorem 174. *Under* **A1—A5**, $V : S \to \mathbb{R}$ *is strictly concave and strictly increasing.*

Uniqueness

- we already know that the value function V is unique
- recall that a stationary optimal strategy solves

$$\max_{c\in[0,y]} (u(c) + \delta V(f(y-c)))$$

  for all $y \in S$
- since V is strictly concave, the solution to the above problem is unique for each $y \in S$, so the stationary optimal strategy $c^* : S \to A$ is unique
- by the maximum theorem, c^* is a continuous function

Inada Conditions

- often, it is desirable to avoid corner solutions by adding assumptions about what happens at the boundary of the feasible set; these are called Inada conditions
- we will impose:
  - **A6**: u is C^1, and $\lim_{c\to 0} u'(c) = +\infty$
  - **A7**: f is C^1, and $\lim_{x\to 0} f'(x) > \frac{1}{\delta}$

Theorem 175. *Under* **A1—A7***, for all $y > 0$, we have $c^*(y) \in (0, y)$.*

Euler Equation

- now we have a neat characterization for the optimal policy

Theorem 176. *Suppose that* **A1—A7** *hold. Let $y > 0$, and let $\{c_t\}_{t=0}^{\infty}$ and $\{y_t\}_{t=0}^{\infty}$ denote the production and consumption sequences generated by the optimal strategy. Then,*

$$u'(c_t) = \delta u'(c_{t+1}) f'(y_t - c_t).$$

- this is called the Euler equation
- Euler equation relates current costs to future benefits and when the agent is using the optimal strategy

Monotonicity

- we can use the Ramsey-Euler equation to show that the optimal solution is strictly increasing

Theorem 177. *Under* **A1—A7**, $y > z$ *implies* $c^*(y) > c^*(z)$.

What Next?

- sometimes (not too often) you can get explicit formulas for the value function/optimal strategy by imposing nice functional forms for utility, production, etc
- if you are working with a problem that *doesn't* admit a tidy solution, you can:
  - as we did today, try to prove that the value function and optimal strategy exist and have nice properties, without actually computing them
  - see whether the problem is more tractable in continuous time; graduate textbooks on asset pricing, e.g. *Dynamic Asset Pricing Theory* by Duffie, will cover this material
  - have a computer solve the problem; you won't get a theorem this way, but you can demonstrate that your model makes reasonable predictions; this is often done in macro