Math Camp Lecture Notes
Autumn 2022

Chase Abram*

September 29, 2022

Contents

*abram@uchicago.edu: These notes are mine...but of course I referenced many math texts and web pages when writing. I think all the ideas (except where noted) are well-known enough math results that original authors are not always cited/known.

1 These Notes

Some thoughts to kick us off:

1. Technically, I think I am covering the "micro" part of the camp, and Camilla the "macro" part. In reality, I am macro-leaning, and that will no doubt show somewhat. Also, I am more focusing on refreshing/teaching pure math that will be useful in the first year, so you should use these concepts in micro, macro, and even metrics[1].

2. Many times, I will state a proposition, example, or even problem somewhat vaguely. Sometimes this is unintentional, but often the reason is that if I leave it up to you to interpret what is going on, you will learn more.

3. Sometimes examples in one topic will jump ahead or back to another topic. This is typically because the core of the current example does not rely on deeply understanding the topic to which we jump. Additionally, I assume everyone has at least a moderate familiarity with calculus and linear algebra, but ...

4. ...if you don't understand something, speak up and ask about it! There may be others who don't understand, and even if not, one purpose of math camp is get everyone on somewhat the same page, and I may have overly omitted certain topics.

5. Don't take what I am saying as gospel truth. If something seems off, or you believe could be stated better, look into it! These notes function more as a primer and review than a proper rigorous encyclopedia, so just like in real life and research, you will find gaps, and some you may want to explore and fill. Stay curious!

6. I (probably) won't be doing too much computational stuff with you all during the math camp, but I'm a big fan of `Julia`. Happy to chat about that more (and I encourage you to use `Julia` instead of `Matlab` in your courses this year).

7. The appendices were prepared because I covered the core material of the notes after the first two weeks, so used the last week for cover some examples. They use math from the core text, and also give a sampling of some basic economic ideas, in highly specific (mainly macro) contexts.

8. All of this being said, I make mistakes. If you find issues or have questions feel free to contact me via email: **abram@uchicago.edu**.

9. Last, but certainly not least: **Please enjoy your few weeks of math camp**. The expectations for you are basically nil, you have plenty of free time to do whatever you want, and are surrounded by people who you will soon be spending lots of time with. I suggest relaxing, exercising, exploring Chicago, and generally having a fun last hurrah. It will be much better for your mental health than trying to do math/econ in your free time outside of lectures, which you will soon do anyway.

[1] Especially metrics for linear algebra.

2 References

You will need math stuff outside of these notes for the first-year. Wikipedia is often a fantastic source for math concepts, largely because people edit the pages until they are clear and correct. Some of my personal favorite other math references (in no particular order):

1. (J. Hubbard and B. Hubbard, 2009): Probably the weirdest math textbook I've ever seen. I used it for two semesters of calculus, hated it the whole time, then later realized it taught me a ton that other books miss. It covers what the title says, but be warned the difficulty amps up quickly. Many of the exercises are tricky, and the proofs use strange notations. Worth a read if you like this math.

2. (Rudin, 1976): This little blue book is the absolute gold standard for learning real analysis. Proofs are clean and elegant, and you should be able to follow the book only knowing a little multivariable calculus. Only downside is that the differential forms section is a bit weak, and examples are bit lacking throughout. But the problems cover some of the "obvious" missing examples.

3. (Royden and Fitzpatrick, 2010): Excellent first-year math PhD analysis text. Notation is clean and the introduction to measure theory is beautiful, with lots of great exercises and examples.

4. (Spivak, 1965): Parts of it are basically a worse version of Rudin, but the material on manifolds and differential forms is better. Probably the easiest way to get to Stokes' theorem, if you care.

5. (Luenberger, 1969): Excellent for linear algebra and basic functional analysis. The approach of viewing optimization as a projection problem helps put some econometric ideas on a mathematical footing, which is appealing. Big plus is that lots of stuff is in one place (lin alg, analysis, optimization)

6. (Evans, 2010): This book is too smart for me, but I've used it a little. It starts from the foundations and builds up PDE theory (which is an order of magnitude harder than ODE theory, generally).

7. (Strogatz, 2018) A more approachable way to learn ODE ideas, but also goes deeper than some other ODE books. Big strength is examples galore, with pictures!

3 Assignments and Grading

If this has not already been made clear, allow me to clarify: you are not being assessed on your performance during Math Camp. Some programs include a course like this one as part of their first-year courses that must be passed, but we do not. We will use Canvas for communication and I might even post "Assignments" purely to provoke thinking, but entirely ignoring them will not be detrimental to you in terms of passing the Core.

4 Logic and Proofs

At the end of the day, most any discussion in economics (or most fields) comes down to what we assume, and what we are deducing from what we assume. This is probably somewhat more obvious when you think of micro theory results, wherein we assume nice mathematical properties, then use them to deduce results, but it also applies to empirical work. When you are in a seminar and an argument breaks out about identification, that is someone saying "I assume X about this economic problem, and in that case your estimation procedure does not do what you claim it does, because you assume Y". In the first year you will spend nearly all your time learning how to move from assumptions to conclusions, but don't lose sight of the fact that the assumptions are the core of whatever you are doing. If you assume people are rational, you may get agents that are able to solve problems which we cannot even solve with computers. If you assume people's preferences take into account habit formation, you should not be surprised when habits arise in your equilibrium.

4.1 Truth and Fiction

For these notes, and probably most of our careers, we will be working with first-order binary logic. So a statement P is either true or false, no in between.

Example 4.1. "Hummingbirds can fly" is a true statement (if an object is a hummingbird, then it can fly), but "birds can fly" is false (if an object is a bird, then it can fly), since there exist birds that cannot fly.

Note that it may be tempting to check examples of statement being true in certain cases, but for the statement to be true, we need it to be true in all cases.

Example 4.2. Consider the Borwein integral. If you only check the pattern for $n = 1, 3, \ldots, 13$ you will conclude the integral is $\frac{\pi}{2}$ always, and if you are sloppy you might convince yourself it holds at higher numbers, even though the pattern breaks at $n = 15$.

We now clarify some of the above ideas.

Definition 4.1. Let X be a set, and P be a statement about X.

1. We say $\forall x \in X, P(x)$ to mean that if $x \in X$, then $P(x)$
2. We say $\exists x \in X, P(x)$ to mean that there is at least one $x \in X$, such that $P(x)$

You might notice something bizarre: $\forall x \in X, P(x)$ when X is empty, but we never have $\exists x \in X, P(x)$ when X is empty.

4.2 Proof

A proof is a means of combining assumptions with logic to show whether a statement is true or false. There is sometimes discussion of "level of rigor" of proofs, but really a proof is either right or wrong, and when people discuss rigor, they are referring to the amount of assumptions allowed to be used in the proof. In the worst cases, however, someone will essentially assume the conclusion of the proof in making the proof. That's not bad rigor, that's just wrong.

There are many types of proofs, and sometimes one is easier/better than another, depending on the problem at hand. Additionally, we often use previous results to build new results (this is how modern mathematics is built).

4.2.1 Direct

Direct proofs are what you typically picture as a proof. They take some assumptions and concatenate them logically into a result directly.

Example 4.3. We want to prove $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. We may do the following calculation

$$\begin{aligned}\sum_{i=1}^{n} i &= \frac{1}{2}\left(2\sum_{i=1}^{n} i\right) \\ &= \frac{1}{2}\left(\sum_{i=1}^{n}(i + [n - i + 1])\right) && \text{(Reverse term order for second set)} \\ &= \frac{1}{2}n(n+1)\end{aligned}$$

Though not usually considered a separate technique, constructive proofs also fall under the category of direct proof. By construction, I mean a proof wherein we want to show that something exists, and do to so we just find an example of it existing. Note that this is special to an existence proof, and generally examples do not prove generalities.

Example 4.4. We want to prove there exists a function which is continuous everywhere and has zero derivative almost everywhere, but is not constant. Note that this sounds kinda impossible, since we are demanding that the function can basically never change, and when it can it can only do so by an infinitesimal amount. But consider the Devil's Staircase.

4.2.2 Contraposition

Contraposition is like direct proof, but negating a negative. The idea is that if we want to prove that P implies Q, we could instead prove that Q not true implies P not true. If this is proven, then P being true must imply Q is also true.

Example 4.5. We want to prove "if x^2 is even, then x is even". The contrapositive statement is "if x is not even, then x^2 is not even". Since the product of two even numbers is even, and two odd numbers is odd, if x is not even, then x is odd, so x^2 is odd, so x^2 is not even.

4.2.3 Contradiction

This is economists' favorite proof method, though I once had a maths prof tell me that almost all contradiction proofs can be made direct with a little effort, and I once heard an economist refer to this as the "weenie"-style proof. The basic idea is that you assume the conclusion is false, then you show that your assumptions are contradicted, therefore your assumption about the false conclusion was wrong.

Example 4.6. Here is a classic. We want to prove $\sqrt{2}$ is irrational. Suppose it is rational (this is the assumption we will show is wrong). Then $\sqrt{2} = \frac{p}{q}$ for some $p, q \in \mathbb{N}$, where p and q are coprime ($\frac{p}{q}$ is reduced as much as possible). Then square both sides and move q, so $2q^2 = p^2$. The left side is even, so the right side must also be even, hence p is even. Then p^2 must be divisible by 4, say $p^2 = 4r$, $r \in \mathbb{N}$.Then $q^2 = \frac{p^2}{2} = \frac{4r}{2} = 2r$ is also even. But then p and q were not in lowest common form, we have a contradiction. Hence $\sqrt{2}$ is not rational.

4.2.4 Counterexample

This type of proof is only useful for showing a general statement is false. The idea is that we can find one instance where a statement does not hold, therefore the whole statement is false.

Example 4.7. Euler had a conjecture. A computer helped find a counterexample.

4.2.5 Induction

This one is tricky if you have not seen it before. The idea is that we prove something for a base case (usually $n = 0$ or $n = 1$), then show that if it holds for case n, it also must hold for case $n+1$. This then proves the statement for all n. You can reason this out by considering that if you prove for $n = 1$, then the $n \Rightarrow n+1$ inductive step implies it holds for $n = 2$. But then the inductive step also shows it holds for $n = 3$. But then the inductive step...
I like to visualize this as dominoes falling. We push the first domino $n = 1$, then check that every other domino is close enough that, when the domino before it falls, it too will fall. Then all the dominoes will fall!

Example 4.8. We want to prove $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ (this is the same as above, yes, and we are showing there are multiple ways to prove the statement). Let's check this for $n = 1$: $\sum_{i=1}^{1} i = 1 = \frac{1(1+1)}{2}$. Great. Now suppose it holds for the case n, the $n+1$ case is then

$$\begin{aligned}
\sum_{i=1}^{n+1} i &= n + 1 + \sum_{i=1}^{n} i \\
&= n + 1 + \frac{n(n+1)}{2} \\
&= \frac{2(n+1) + n(n+1)}{2} \\
&= \frac{[n+1]([n+1]+1)}{2}
\end{aligned}$$

4.2.6 Inspection/"It's Obvious"

Sometimes you will look at a problem and it is just obviously true, right? No. If it is obvious, then there exists a method of proving it that relies on assumptions and logic. So this proof technique is not valid. Most "proofs" of the Riemann Hypothesis use this fallacy at some point.

5 Miscellaneous Prelims

First, we will lay out a handful of definitions and miscellaneous ideas. This section will feel disjointed, but the idea is that I am handing you some of the fundamental tools you will need throughout right now, and you can reference back when needed.

It will be quite helpful to have terms precisely describing how functions may map between spaces in nice ways.

5.1 Jections

Definition 5.1. A function $f : X \to Y$ is

(i) injective[2] if for all $x, x' \in X$, $x \neq x'$ implies $f(x) \neq f(x')$

(ii) surjective[3] if for all $y \in Y$, there exists $x \in X$ such that $f(x) = y$.

(iii) bijective[4] if both injective and surjective

Intuitively, we may think that if there exists an injection from X to Y, then Y is at least as big as X, and vice versa if there exists a surjection. For finite sets, this works exactly as we would like via the pigeonhole principle, which hilariously says that if you try to smash more pigeons through holes than the number of holes you have, you must smash at least two pigeons through the same hole[5]. Not so if the sets have infinitely many elements.

Example 5.1. The function $f(x) = x + 1$ is an injection from $\mathbb{N} \setminus \{0\}$ to $\mathbb{N}$.

Problem 5.1. Find a surjection from $\mathbb{N} \setminus \{0, 1, \ldots, 10^{10}\}$ to $\mathbb{N}$.

Example 5.2. Functions which are strictly monotonic (derivative strictly greater than 0) are injective $\mathbb{R} \to \mathbb{R}$.

Example 5.3. Functions which are continuous and have limits at $\pm\infty$ are surjective $\mathbb{R} \to \mathbb{R}$.

5.2 Cardinality

Definition 5.2. A set X is called

(i) Finite if there exists an n such that there exists a surjection from $\{1, \ldots, n\}$ to X

(ii) Countable if there exists a surjection from $\mathbb{N}$ to X[6]

(iii) Uncountable if not countable

Problem 5.2. Are the rationals countable? Find a surjection if so, or prove that one does not exist if not.

[2] one-to-one

[3] onto

[4] invertible

[5] Oh that's not how you think of the pigeonhole principle?

[6] Really this is the def for "at most countable". To be sure a set is exactly countable (not finite) we would require a bijection from $\mathbb{N}$ to X.

Problem 5.3. Are the reals countable? (Hint: look up Cantor's diagonalization argument and be careful.)

Example 5.4. Let $a < b$ and $c < d$. Then $f(x) = \frac{d-c}{b-a}(x-a)+c$ provides a surjection $[a,b] \to [c,d]$. It is also an injection, therefore a bijection.

Problem 5.4. Let $a_1 < b_1 < a_2 < b_2 < \ldots < a_n < b_n$ and $c < d$. Also let $X = \bigcup_i^n [a_i, b_i]$ and $Y = [c,d]$

(i) Provide an injection from X to Y.

(ii) Provide a surjection from X to Y.

(iii) Does there exist a bijection between X and Y?

5.3 Norms

Definition 5.3. A norm $\|\cdot\| : X \to \mathbb{R}$ satisfies the following properties:

(i) $\|x\| \geq 0$ for all $x \in X$, and $\|x\| = 0$ iff $x = 0$

(ii) $\|x + y\| \leq \|x\| + \|y\|$

(iii) $\|\alpha x\| = |\alpha| \cdot \|x\|$ (α is a scalar)

Problem 5.5. Are the following norms? Verify if so, and if not, identify which property is not satisfied.

(i) $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$

(ii) $\|x\|_\ell = |x_1|$ (x is still n-dimensional)

(iii) $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$

(iv) $\|x\|_{-\infty} = \min\{|x_1|, \ldots, |x_n|\}$

5.4 Inner Product

Definition 5.4. An inner product $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{F}$ satisfies the following properties, where $\mathbb{F}$ is a field.

(i) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$, $\lambda \in \mathbb{F}$

(ii) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$

(iii) $\langle x, y \rangle = \overline{\langle y, x \rangle}$

(iv) $\langle x, x \rangle \geq 0$, with equality only if $x = 0$

Theorem 5.1 (Cauchy-Schwarz)**.** On an inner product space, if we consider the norm as the one induced by the inner product, then for all $x, y \in X$, we have $|\langle x, y \rangle| \leq \|x\| \|y\|$

Definition 5.5. We say two elements are orthogonal if their inner product is zero.

A common inner product in $\mathbb{R}^n$ is the dot product $\langle x, y \rangle \equiv \sum_{i=1}^n x_i y_i$, and unless stated otherwise, this is the default inner product for $\mathbb{R}^n$.

Problem 5.6. Some orthogonality practice. (Don't just put the zero vector, you know that is lazy.)

(i) Find a vector orthogonal to $(1, 2, 3)$.

(ii) Find two vectors orthogonal to $(1, 0, 1)$ which are also orthogonal to each other.

(iii) Consider the inner product $\langle x, y\rangle = \int_{-\pi}^{\pi} x(t)y(t)\mathrm{d}t$. Under what conditions for m (an integer) is $\cos(mx)$ orthogonal to $\sin(nx)$ (n also an integer)?

5.5 Homogeneity

Definition 5.6. A function f is homogeneous of degree α if $f(\lambda x) = \lambda^\alpha f(x)$ for all $\lambda \in \mathbb{R}$.

Example 5.5. Linear functions are homogeneous of degree 1.

Example 5.6. Constant functions are homogeneous of degree 0.

Problem 5.7. What can we say about the degree of homogeneity of production functions that are decreasing, constant, and increasing returns to scale (assuming they are homogeneous of some degree)?

Problem 5.8. Let $f(x) = \prod_{i=1}^{n} x_i^{\gamma_i}$ What is the degree of the homogeneity of f?

Problem 5.9. If a function f is homogeneous of degree r, show that its derivative is homogeneous of degree $r - 1$.

Theorem 5.2 (Euler)**.** If f is homogeneous of degree r, then $Df(x) \circ x = rf(x)$.

Proof.

$$\begin{aligned} f(\lambda x) &= \lambda^r f(x) && \text{(Def.)} \\ Df(\lambda x) \circ x &= r\lambda^{r-1} f(x) && (\partial\lambda) \\ Df(x) \circ x &= rf(x) && (\text{At } \lambda = 1) \end{aligned}$$

□

5.6 Convexity

Definition 5.7. A set C is convex if, for all $x, y \in C$, $\lambda \in (0, 1)$, we have $\lambda x + (1 - \lambda)y \in C$.

Intuitively, in convex sets you can draw lines between any of the points, and the lines stay inside the set. This property is quite nice for when we want to consider a "mixture" of set elements, but want to remain inside the set.

Example 5.7. Recently, I was interested in better understanding the set $\mathcal{A}(x) \equiv \{A : Ax = x, a_{ij} \geq 0, \sum_j a_{ij} = 1\}$ for some given x[7]. It turns out this is not the easiest set to imagine or ascribe simple properties. But it is convex, since if $A, B \in \mathcal{A}(x)$, then

$$\begin{aligned} (\lambda A + (1 - \lambda)B)\, x &= \lambda Ax + (1 - \lambda)Bx \\ &= \lambda x + (1 - \lambda)x && (A, B \in \mathcal{A}(x)) \\ &= x \end{aligned}$$

and non-negativity and summing to unity are also easy to check. Thus, $\lambda A + (1 - \lambda)B \in \mathcal{A}(x)$.

[7] It can be shown that $\mathcal{A}(x)$ is isomorphic to the set of transportation polytopes with marginals x in both dimensions.

Problem 5.10. Find an example of a (non-singleton) set which is convex, and a set which is not convex, in each of:

(i) $\mathbb{R}$

(ii) $\mathbb{R}^2$

(iii) The set of functions $f : \mathbb{R} \to \mathbb{R}$

Definition 5.8. A function f is quasi-convex (quasi-concave) if either of the following equivalent statements hold, for every $x, y \in X$, $\lambda \in (0, 1)$

(i) $f(\lambda x + (1 - \lambda y) \leq \max\{f(x), f(y)\}$ $(\geq \min\{f(x), f(y)\})$

(ii) Given arbitrary $\alpha \in \mathbb{R}$, the set $\{x \mid f(x) \leq \alpha\}$ $(\{x \mid f(x) \geq \alpha\})$ is convex

Definition 5.9. A function f is convex (concave) if, for every $x, y \in X$, $\lambda \in (0, 1)$, $f(\lambda x+(1-\lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ $(\geq \lambda f(x) + (1 - \lambda)f(y))$. We can add strictness when the inequality becomes strict.

Quasi-convexity disciplines what sublevel sets are allowed to exist for a function, and convexity discplines the rate of change of the rate of change of a function.

Problem 5.11. Show that convex functions are quasi-convex.

Theorem 5.3. Convex functions $f : \mathbb{R} \to \mathbb{R}$ are continuous.

The typical (only?) proof of this theorem requires the Chordal Slope Lemma, which you will encounter in fall metrics.

Problem 5.12. Find a quasi-convex function that is

(i) Convex

(ii) Quasi-concave[8]

(iii) Concave

(iv) Neither convex nor concave (over its whole domain, that is)

[8] Being both, this function will be quasi-linear.

6 Linear Algebra

Linear algebra is the tool that lets us work in higher dimensions and analyze the structure of systems.

6.1 Basics

Definition 6.1. A vector space X is a set of elements which is closed under the operations of addition and scalar multiplication. This means

(i) $x, y \in X$ implies $x + y \in X$

(ii) $x \in X$, $\alpha \in \mathbb{F}$[9] implies $\alpha x \in X$

Problem 6.1. Show that $\mathbb{R}^n$ is a vector space if $F = \mathbb{R}$ and we define $+$ by component-wise addition.

Example 6.1. The space of $n \times m$ matrices is a vector space with the standard component-wise definition of addition.

Example 6.2. The space of periodic function on a given interval is a vector space, since combining periodic functions gives another periodic function.

Problem 6.2.

(i) Is the space of continuous function such that $f(0) = f(1) = 0$ a vector space?

(ii) What about if we require $f(0) = f(1)$ instead?

(iii) What about $f(0) = 1$ instead?

Definition 6.2. A set of elements $\{x_i\}$ is linearly dependent if there exists a set of scalars $\{a_i\}$, not all zero, such that $0 = \sum_i a_i x_i$. If such scalars do not exist (i.e. the only way to have $0 = \sum_i a_i x_i$ is if $a_i = 0$ for all i), then the set is linearly independent.

Problem 6.3. Find a pair of linearly independent vectors in $\mathbb{R}^n$, for $n = 1, 2$.

Definition 6.3. A set $\{s_i\}$ spans the set S if, for any $s \in S$, there exists scalars $\{a_i\}$ such that $s = \sum_i a_i s_i$.

Problem 6.4. Find a set that spans $\mathbb{R}^n$, for $n = 1, 2$, but is not linearly independent.

Definition 6.4. A basis $\{b_i\}$ for a vector space X is a set of elements that are linearly independent and span X. We refer to $|\{b_i\}|$ as the dimension of the space X. If every element of the basis is orthogonal to all other basis elements and has norm 1, then the basis is orthonormal.

Problem 6.5. Find a basis for each of the following spaces:

(i) $\mathbb{R}^n$

(ii) The set of functions $f : \mathbb{R} \to \mathbb{R}$ which only take nonzero values at $\{x_1, \ldots, x_n\} \in \mathbb{R}$

(iii) The set of n-dimensional square matrices

(iv) The set of n-dimensional square symmetric matrices

[9] $\mathbb{F}$ is some field with which we are defining the vector space with respect to. In all of our applications $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

6.2 Transformations

Definition 6.5. Let X and Y be vector spaces. A transformation $T : X \to Y$ assigns each element $x \in X$ to some element $y \in Y$. If $X = Y$, then T is an operator.

Note there is no requirement about uniqueness.

Example 6.3. The function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined $f(x, y) = (x^2 + y^2, \sin(x) * pl(y)$, where $pl(y)$ is the number of primes less than y, is a transformation. I have made it intentionally gross to emphasize the current freedom, before we impose more restrictions.

Operators are ubiquitous, even though we often just call them "square matrices" or talk about them without using the word operator. I prefer starting with this super abstract definition so that we can encapsulate lots of different settings, including infinite dimensional spaces, which are increasingly important to understand, as the equilibrium to many economic models is a distribution or function, both of which live in infinite-dimensional spaces.

Example 6.4. (Allen and Arkolakis, 2014) has lots of neat math. On page 1094, the two equations introduced can be viewed as the following operators

$$\Phi(i)[x] = \int_S W(s)^{1-\sigma} T(i, s)^{1-\sigma} A(i)^{\sigma-1} u(s)^{\sigma-1} x(s) \mathrm{d}s$$
$$\Psi(i)[x] = \int_S W(i)^{1-\sigma} T(s, i)^{1-\sigma} A(s)^{\sigma-1} u(i)^{\sigma-1} x(s) \mathrm{d}s$$

These are adjoint, which basically means Ψ is the "transpose" of Φ (we won't really get into this). At this point in the notes this is probably all confusing, but the idea is that you have now seen a real application of how operators are used, and you can return after we talk about other stuff and you feel more comfortable.

Definition 6.6. The kernel of a transformation $f : X \to Y$ is the set $K \subseteq X$ such that $x \in K$ implies $f(x) = 0$.

Definition 6.7. The image of a transformation $f : X \to Y$ is the set $I \subseteq Y$ such that $y \in I$ implies there exists $x \in X$ such that $f(x) = y$.

Problem 6.6. Find the kernel and image of the following transformations. Let F be the set of functions from $[0, 1]$ to $\mathbb{R}$:

(i) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$

(ii) $f : F \to \mathbb{R}$, $f(x) = \int_0^1 x(t)\mathrm{d}t$

(iii) $f : F \to F$, $f(z)[x] = x(z)^2$

(iv) The above trade example (just kidding)

Definition 6.8. A transformation $T : X \to Y$ is linear if, for any $x_1, x_2 \in X$, $\alpha_1, \alpha_2 \in F$, we have $T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 T(x_1) + \alpha_2 T(x_2)$.

This really cuts down on what T can be.

Example 6.5. The mapping $T(x, y) = (5x + 3y, 7y)$ is linear.

Problem 6.7. Consider the space of all sequences in $\mathbb{R}$, and the shift operator $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ defined by $T((x_1, x_2, \ldots,)) = (x_2, x_3, \ldots)$. Is T linear?

6.3 Eigenstuff

Now we turn to eigenstuff. My feeling is that most people either learn this and conclude it is meaningless technicalites, or conclude that it is crucial to understanding the properties of a linear transformation. I'm gonna try to get everyone into the second camp.

Definition 6.9. An eigenvector of a linear transformation T is a non-zero element $x \in X$ for which there exists $\lambda \in \mathbb{F}$ such that $Tx = \lambda x$. The eigenvalue corresponding the eigenvector is λ.

Example 6.6. Consider the following matrix

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

The map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by Ax is a linear transformation. It is easy to check that if $x = (1, -1)$, then $Ax = (1, -1)$ so $(1, -1)$ is an eigenvector with eigenvalue 1.

We now want to understand why eigenvalues matter for linear transformations.

Definition 6.10. The determinant is the product of the eigenvalues of a linear transformation.

Definition 6.11. The trace is the sum of the eigenvalues of a linear transformation.

Maybe these are familiar definitions, or maybe you are currently saying "that's the wrong definition. I know the determinant is the thing where you take the product of each element going down a column of a matrix, but each element multiplies the submatrix..." Yes. That is one way to compute the determinant. But this definition makes it more obvious that the determinant is tied to the eigenvalues.

Let's consider how to find eigenvalues. Consider

$$\begin{aligned} Ax &= \lambda x \\ \Rightarrow (A - \lambda I)x &= 0 \end{aligned}$$

Therefore if λ is an eigenvalue of A with eigenvector x, then 0 is an eigenvalue of $A - \lambda I$ with eigenvector x. Therefore the determinant of $A - \lambda I$ will also be 0. Conversely, if 0 is an eigenvalue of $A - \lambda I$, there exists non-zero x such that $(A - \lambda I)x = 0$, meaning $Ax = \lambda x$. Thus, finding eigenvalues of A is equivalent to solving

$$\det(A - \lambda I) = 0$$

If we carry out the determinant calculations, the above equation will reduce to a polynomial in λ, motivating:

Definition 6.12. The characteristic polynomial of A is $p(\lambda) = \det(A - \lambda I)$. Its roots are the eigenvalues of A.

With small dimensions, this gives us a nice way to find roots with pen and paper. More generally, we can use a computer with a simple nonlinear solver to find the roots (polynomials are well-behaved).

Problem 6.8. Consider the following matrix (again)

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Use the previous determinant formula to find the characteristic polynomial, and all the eigenvalues of A.

Again, you may still be wondering "why do I care about eigenvalues?". Some motivation:

Example 6.7. Consider the VAR

$$x_{t+1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} x_t + \epsilon_{t+1}$$

where $\epsilon_{t+1} \sim \mathcal{N}(0, I)$. We might be curious about the long-term behavior of this process, and a priori it is not obvious how a, b, c, d should be restricted to guarantee, for example, stationarity. The classification is simple with eigenvalues, though: if all the eigenvalues are inside the unit circle, then the process is stationary.

Problem 6.9. What must be the restrictions on a, b, c, d so that the process is stationary? What are the restrictions so that the process does not oscillate (in response to a one time impulse)?

Example 6.8. The eigenstuff also gives us a way to think about the kernel and image of a linear operator. For simplicity, consider an operator that has n linearly independent eigenvectors[10]. In this case, we can decompose the operator into how it acts upon this eigenbasis: If $x = \sum_i a_i x_i$, then $Ax = \sum_i a_i A x_i = \sum_i a_i \lambda_i x_i$. Note that this means the eigenvalues quickly tell us the dimension of the kernel and image: The number of λ_i which are zero gives the dimension of the kernel, and the remainder give the dimension of the image.

This last example also connects to how we defined the determinant above. If even one eigenvalue is zero, then the operator cannot be bijective, since a whole subspace will map to zero, and one eigenvalue being zero will clearly make the determinant zero. On the other hand, if no eigenvalue is zero, then the determinant will not be zero.

6.4 Jordan

As might have become obvious from the above example, faced with a transformation A, we may wish to change its representation in order to better understand its properties.

Definition 6.13. The inverse of a matrix A, denoted A^{-1}, is the matrix satisfying $AA^{-1} = I$.

Example 6.9. The form of A^{-1} for a 2×2 can easily be checked.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\Rightarrow A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

[10] In some sense, "most" operators satisfy this.

Definition 6.14. Given a matrix A, its Jordan Normal Form is VJV^{-1}, where V has columns which are the (potentially generalized) set of eigenvectors, and J is diagonal except for possibly some 1s above the diag.

It is worth clarifying: there exist matrices which do not have a basis of eigenvectors, but there always exists a Jordan normal form as described above.

Problem 6.10. Find the eigenvalues and eigenvectors of

$$B = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$$

Problem 6.11. Look up generalized eigenvectors. Find the unit-length generalized eigenvector of B

Problem 6.12. Find the Jordan decomposition of

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Jordan form is particularly nice if you want to take powers of a matrix, as $A^n = VJ^nV^{-1}$.

Problem 6.13. Once again, consider

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

What is A^{100}? (Recall that you already know the eigenstuff from above.)

Problem 6.14. Recall $\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$(if this is new, now you know!). We define the matrix exponential $\exp(M) = \sum_{k=0}^{\infty} \frac{1}{k!} M^k$.[11]. Find $\exp(A)$, using the A from above.

The other nice use of Jordan form is that we can transform a system so that it becomes diagonal (or at least really close, and upper triangular).

Problem 6.15. Using the A we have been using all along, solve the following system of differential equations.

$$\dot{x} = Ax$$
$$x(0) = (1, 2)$$

6.5 Symmetry and Definiteness

Definition 6.15. A matrix A is symmetric if $a_{ij} = a_{ji}$ for all elements in the matrix, where i is the row, and j is the column. If we instead require $a_{ij} = \bar{a}_{ji}$, where the bar denotes complex conjugation, then A is Hermitian (or self-adjoint).

In the case of real matrices, the definitions of symmetric and Hermitian coincide. We also have the following useful property:

[11] If you are worried about existence, rest assured that the factorial keeps everything from exploding.

Theorem 6.1. It is always possible to find a set of orthonormal eigenvectors of a symmetric matrix, which therefore constitute a basis for X.

Now we turn to the question of definiteness.

Definition 6.16. Expressions of the form $x'Ax$ are quadratic forms. Note that, since we transpose x on the left, the expression is a scalar.

Quadratic forms are often helpful for identifying properties about operators.

Definition 6.17. Given a symmetric matrix A,

(i) If $x'Ax \geq 0$ for all $x \neq 0 \in X$, then A is positive semi-definite. If the inequality is strict for all i, then A is positive definite.

(ii) If $x'Ax \leq 0$ for all $x \neq 0 \in X$, then A is negative semi-definite. If the inequality is strict for all i, then A is negative definite.

Whether a matrix is positive (negative) (semi)-definite will determine whether a critical point is a maximum, minimum, or something else. We also use the concept of positive semi-definiteness to formalize what we mean by "best" in the Gauss-Markov Theorem (OLS is BLUE).

Yet again, properties of eigenvalues of a transformation are essential to understanding the transformation.

Problem 6.16. Prove that a symmetric A is positive definite if and only if all of its eigenvalues are positive. (Hint: the above theorem may be helpful).

The closest we can get to having a square root of a matrix is the Cholesky decomposition.

Definition 6.18. For a real positive definite matrix, A, its Cholesky decomposition is $A = LL^T$, where L is lower triangular and positive along the diagonal.

Again, it is worth hammering on this point: for any positive definite matrix, there exists a unique Cholesky decomposition. If the matrix is only positive semidefinite, then the decomp need not be unique.

Problem 6.17. Show why, if A has a Cholesky decomposition, then A must be positive semidefinite. (One line proof)

Example 6.10. If we consider a multivariate normal random variable defined by $Y = AX + BW$, where X is fixed, and $W \sim \mathcal{N}(0, I)$. Then $Y \sim \mathcal{N}(AX, BB')$. This is also nice in reverse, since if we are given $Y \sim \mathcal{N}(M, \Omega)$, but we know $\Omega = LL^T$, then we may consider $Y = M + LW$. The fact that L is lower triangular can prove handy for conditioning, when working with jointly normal random variables[12].

Definition 6.19. A linear operator T is

(i) Idempotent if $T^2 = T$

(ii) Nilpotent if there exists k such that $T^k = 0$

[12] You'll probably see this again in the winter quarter when messing with the Kalman filter. If you don't get satisfactory notes on this in the winter quarter, hit me up, because I think I have a note somewhere that is pretty clean on this point.

(iii) Orthogonal if $TT^* = I$

(iv) Involutory if $T^2 = I$

Problem 6.18. What can we say about the eigenvalues in each case?

(i) T is idempotent

(ii) T is nilpotent

(iii) T is involutory

Problem 6.19. Let X be an $n \times m$ matrix such that $X'X$ is nonsingular. Let $F(v) = X(X'X)^{-1}X'v$.

(i) Prove that F is a symmetric operator.

(ii) What can you say in terms of trying to construct an eigenbasis for the operator F.

(iii) Prove that F is an idempotent operator.

(iv) What can you say about the way in which OLS projects vectors into explained and unexplained components?

7 Real Analysis

7.1 Setup

Definition 7.1. A metric space (X, d) is a set X and metric $d : X \times X \to \mathbb{R}$ which satisfies

(i) $d(x, y) \geq 0$ for all $x, y \in X$, with equality only if $x = y$.

(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$

(iii) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$

Definition 7.2.

(i) $B_r(x) = \{y \mid d(x, y) < r\}$ is the open ball of radius r around x[13]

(ii) A neighborhood of x is a set N such that there exists $r > 0$ such that $B_r(x) \subseteq N$

(iii) An element x is a limit point of the set E if every neighborhood of x has nonempty intersection with $E \setminus \{x\}$

(iv) A set E is closed if it contains all its limit points

(v) An element x is in the interior of E if there exists a neighborhood N of x such that $N \subseteq E$

(vi) A set E is open if all of its elements are interior

(vii) A set E is bounded if there exists $M < \infty$ such that $d(x, y) < M$ for all $x, y \in E$

Note that open and closed are not opposites, nor are they exhaustive, nor are they exclusive.

Example 7.1. In $(\mathbb{R}, |\cdot|)$, sets (a, b) are open, $[a, b]$ are closed, and $(a, b]$ are neither open nor closed. The empty set is both open and closed (vacuously), so we can (no joke) call it clopen.

Problem 7.1. Show $\mathbb{R}$ is clopen.

Our favorite counterintuitive example will be the discrete metric space. For concreteness, when we refer to the discrete metric, we mean the set $\mathbb{N}$ with the metric

$$d(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

This should immediately be confusing, because somehow 5 has distance 1 from 100, 100 has distance 1 from 200, yet 5 has distance 1 from 200. This idea makes sense if you have three points in an equilateral triangle, or tetrahedron, so try to picture that in infinitely many dimensions.

Problem 7.2. Prove the discrete metric is a metric.

Problem 7.3. The open sets of the discrete metric space are interesting...so are the closed sets. What are they?

[13] Annoying technical point: we have not yet defined open, so after we do we would need to prove that the "open ball" is actually open. It is. If this confusing, rename the open ball in this part of the definition as the "special" ball or something, and you will see there is no circular logic.

The following problem gets at the more general reason why we care about these ideas (and maybe maths generally) at all in economics: they can help us understand the nature of solutions (and lack thereof).

Problem 7.4. Consider the "lowest number game", wherein 2 agents choose a number in $A \subset \mathbb{R}$, and the agent that chooses the lower number wins the reward. In the case of a tie, the reward is split. For each of the following A, determine whether it is open or closed (or both or neither) and find the optimal strategy and equilibrium outcome.

(i) $A = \{\frac{1}{2}, 1\}$

(ii) $A = \{\frac{1}{n} \mid n \in \{1, 2, \dots, 100\}\}$

(iii) $A = \{\frac{1}{n} \mid n \in \{1, 2, \dots\}\}$

(iv) $A = \{\frac{1}{n} \mid n \in \{1, 2, \dots\}\} \cup \{0\}$

(v) $A = [0, 1]$

(vi) $A = [0, 1)$

(vii) $A = (0, 1)$

Definition 7.3. An open cover of a set E is a collection of open sets $\{O_\alpha\}$ such that $E \subseteq \bigcup_\alpha O_\alpha$. A subcover is a subset of the sets $\{O_\alpha\}$ which also covers E.

Definition 7.4. A set K is compact if, for every open cover of K, there exists a finite subcover.

My intuition is that compactness is saying that K can be contained in a nice way. In terms of optimization, compactness is extraordinarily helpful for disciplining the domain. Note that there always exists an open cover of any set, since we can just put a ball around every element. The trick is whether or not we can remove all but a finite number of balls and still cover the whole set.

Problem 7.5. Show that $(0, 1)$ is not compact by finding an open cover for which no finite subcover exists. (Hint: Make it so that the edges are only covered in the limit of a countable sequence of sets.)

There is slightly more work to be done to prove the following theorem, but the result is what we really care about.

Theorem 7.1 (Heine-Borel)**.** In $\mathbb{R}^n$, a subset A is compact iff it is closed and bounded.

Problem 7.6. For each A in the "lowest number game" above, decide whether A is compact (this should be easy if your above answers are right).

Heine-Borel is great and basically tells us that we can think about compactness in terms of closedness and boundedness, which are a bit more intuitive than the open cover definition. Is there a reason why we restricted to $\mathbb{R}^n$?

Example 7.2. Consider the discrete metric space, and the open cover $\bigcup_{x \in X} B_{\frac{1}{2}}(q)$. In this case, every ball contains only x, so removing even one of them would mean we no longer cover X. So there does not exist a finite subcover, and X is not compact. Nonetheless, X is bounded (let $M = 2$, for example) and closed (if every neighborhood of y intersects X, then y must be in X).

7.2 Sequences and Notions of Convergence

Let's start with two properties that sequences may possess.

Definition 7.5. A sequence (x_n) converges to x if, for all $\epsilon > 0$, there exists N such that $n \geq N$ implies $d(x_n, x) < \epsilon$.

Definition 7.6. A sequence (x_n) is Cauchy if, for all $\epsilon > 0$, there exists N such that $n, m \geq N$ implies $d(x_n, x_m) < \epsilon$

What's the difference? In the first case, x_n is approaching some given x. In the second case, the x_n are getting closer together, which may mean they are approaching a given x, but it does not necessarily.

Example 7.3. Note that any convergent sequence is a Cauchy sequence. Let $\epsilon > 0$, and choose N such that $n \geq N$ implies $d(x_n, x^*) < \epsilon/2$. Then for $n, m \geq N$, $d(x_n, x_m) \leq d(x_n, x^*) + d(x_m, x^*) < \epsilon$.

Problem 7.7. It is quite tempting to think that Cauchy sequences must also always converge (the converse of above). Let C be the space of function $[0, 1] \to \mathbb{R}$ which are continuous. Consider the sequence

$$f_n(x) = \begin{cases} 0 & x \leq 1 - \frac{1}{n} \\ n(x - 1 + \frac{1}{n}) & x \geq 1 - \frac{1}{n} \end{cases}$$

(i) Let the metric be $d(f, g) = \int_0^1 |f(x) - g(x)| \mathrm{d}x$. What does f_n converge towards, and is it in C? Is f_n a Cauchy sequence?

(ii) Let the metric be $d(f, g) = \sup_x |f(x) - g(x)|$. What does f_n converge towards, and is it in C? Is f_n a Cauchy sequence?

We can classify metric spaces using Cauchy sequences in the following way.

Definition 7.7. A metric space X is (Cauchy) complete if every Cauchy sequence converges.

As seen in the above examples, it is generally possible to find a Cauchy sequence which does not converge. However, in the space that we most frequently care about, this problem does not exist.

Theorem 7.2. Every Cauchy sequence in $\mathbb{R}^n$ converges.

Some other notions of convergence are helpful to know.

Definition 7.8. Let (f_n) be a sequence of functions $f_n : X \subseteq \mathbb{R} \to \mathbb{R}$

(i) Say $f_n \to f$ pointwise if, for all $x \in X$, and for all $\epsilon > 0$, there exists $N(x)$ such that $n > N$ implies $|f_n(x) - f(x)| < \epsilon$

(ii) Say $f_n \to f$ uniformly if, for all $\epsilon > 0$, there exists N such that $n > N$ implies $|f_n(x) - f(x)| < \epsilon$ for all $x \in X$

The difference is key: in the first case the convergence occurs point-by-point, and the rate of convergence may vary by x, whereas in the second case everything has to converge together, so the N cannot depend on x. We are naturally interested in understanding when these notions of convergence differ.

Problem 7.8. Let

$$g_n(x) = \begin{cases} 0 & x \in [0, 1 - \frac{1}{n}] \\ nx - (n-1) & x \in (1 - \frac{1}{n}, 1] \end{cases}$$
$$x_n = \frac{n-1}{n}(1 - \frac{1}{n})$$
$$y_n = 1 - \frac{1}{n^2}$$

(i) What is $\lim_{n\to\infty} x_n$?

(ii) What is $\lim_{n\to\infty} y_n$?

(iii) What is $g(x) \equiv \lim_{n\to\infty} g_n(x)$ (for arbitrary $x \in [0,1]$?

(iv) What is $\lim_{n\to\infty} g_n(x_n)$?

(v) What is $\lim_{n\to\infty} g_n(y_n)$?

(vi) Either find the uniform limit of g_n or show it does not exist.

(vii) What can you conclude about the conjecture "$x_n \to x$ and $g_n \to g$ (pointwise) implies $g_n(x_n) \to g(x)$"?

Let's consider one more prescient example of how different notions of convergence may not agree. These notes have no measure theory nor rigorous probability, but I'd be remiss to mention notions of convergence and not mention the following concepts, which are key to econometrics.

Definition 7.9. A sequence of random variables X_n converges to random variable X...

(i) almost surely ($X_n \to^{a.s.} X$) if $P[\lim_{n\to\infty} X_n = X] = 1$

(ii) in probability ($X_n \to^p X$) if for all $\epsilon > 0$, $\lim_{n\to\infty} P[|X_n - X| > \epsilon] = 0$

(iii) in distribution ($X_n \to^d X$) if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x where F is continuous

Example 7.4. Let's check that these are not the same.

(i) Consider the sequence of sets such that for $n = 1$, we have $[0,1]$, for $n = 2, 3$, we have $[0, \frac{1}{2}]$ and $[0, \frac{1}{2}]$, for $n = 4, 5, 6$, we have $[0, \frac{1}{3}]$, $[\frac{1}{3}, \frac{2}{3}]$, $[\frac{2}{3}, 1]$, and so on. Let $\Omega = [0,1]$, P be uniform over Ω, $X_n(\omega) = 1$ for ω in the n-th set, $X_n(\omega) = 0$ otherwise, and $X(\omega) = 0$. Then

$$\lim_{n\to\infty} P[|X_n - X| > \epsilon] \leq \lim_{n\to\infty} P(C_n)$$
$$= 0$$

where C_n is the interval for n, and thus the measure of C_n is going to 0.

So $X_n \to^p X$. However, for every ω and every N, there exists $n > N$ such that $X_n(\omega) = 1$ and $X_{n+1}(\omega) = 0$, so $\lim_{n\to\infty} X_n(\omega)$ is not even defined for any ω, and $P[\lim_{n\to\infty} X_n(\omega) = X] = 0$. So $X_n \not\to^{a.s.} X$.

(ii) Let $\Omega = [0,1]$, P be uniform over Ω, $X_n(\omega) = \omega$ if n is odd, $X_n(\omega) = 1 - \omega$ if n is even, $X(w) = \omega$. Then

$$\begin{aligned} F_n(x) &= P[X_n \leq x] \\ &= \int_0^x dP(\omega) \\ &= x \end{aligned}$$

and similarly $F(x) = x$. So $F_n = F$ and $X_n \to^d X$. However, $P[|X_n - X| > \epsilon] = 0$ if n is odd, and $P[|X_n - X| > \epsilon] = 1 - \epsilon > 0$[14] if n is even, so $X_n \not\to^p X$.

7.3 Banach and Hilbert Spaces

We are now able to define a central structure for functional analysis.

Definition 7.10. A Banach space is a vector space equipped with a norm, such that it is a complete metric space with the metric induced by the norm.

Theorem 7.3. The space $\mathcal{C}[0,1]$ (the set of functions $[0,1] \to \mathbb{R}$ which are continuous) is a Banach space (under the metric induced by the sup norm).

Proof. I include this proof because I think it is instructive for mathematical thinking and tackling a proof. In this vein, the following is not a good example of how to write a "neat" proof, because I wander around and return to earlier expressions to explain that I am guessing at strategies. A better approach is to do all of this on your own, then write down the clean final result. I write all the thinkings and musings to be more pedagogical.

I will use $\|\cdot\|$ interchangeably with $\sup_x |\cdot|$.

First, we want to prove the space is a Banach space, which means we want to show that every Cauchy sequence converges. Then the proof approach is to consider an arbitrary Cauchy sequence in $\mathcal{C}[0,1]$ and show that it converges to some element of the space. Let $\{f_n\}$ be an arbitrary Cauchy sequence.

How do we consider what the limit might be? In this case, it seems intuitive that the limit should match the pointwise limit, i.e. $f_n \to f$, where f is defined by $\lim_{n\to\infty} f_n(x)$ for each x. Is this limit well-defined for each x? If it were a Cauchy sequence, it would be since, $\mathbb{R}$ is complete. To verify this, note that for any $\epsilon > 0$, there exists an N such that $n, m > N$ implies $\sup_x |f_n(x) - f_m(x)| < \epsilon$ by definition of f_n being a Cauchy sequence under the sup norm, and therefore for any given x we have $|f_n(x) - f_m(x)| \leq \sup_x |f_n(x) - f_m(x)|$, so $f_n(x)$ is a Cauchy sequence for any x, and $f(x) = \lim_{n\to\infty} f_n(x)$ is well-defined for each x.

We have our candidate limit f, but it remains to verify that $f_n \to f$ under the sup norm (we above only verified pointwise convergence), and that $f \in \mathcal{C}[0,1]$. To verify the limit statement, let $\epsilon > 0$, and N such that $n, m > N$ implies $\|f_n - f_m\| < \frac{\epsilon}{2}$, which can be done since f is a Cauchy sequence. Now let $n > N$, and use the triangle inequality that holds for norms

[14] For ϵ sufficiently small, i.e. $0 < \epsilon < 1$.

$$\begin{aligned}\|f_n - f\| &= \|f_n - f_{N+1} + f_{N+1} - f\| \\ &\leq \|f_n - f_{N+1}\| + \|f_{N+1} - f\|\end{aligned}$$

By definition of Cauchy and our N choice, $\|f_n - f_{N+1}\| < \frac{\epsilon}{2}$. Additionally, since $\|f_{N+1} - f_m\| < \frac{\epsilon}{2}$ for all $m > N$,

$$\begin{aligned}\|f - f_{N+1}\| &= \|\lim_{m\to\infty} f_m - f_{N+1}\| \\ &= \lim_{m\to\infty} \|f_m - f_{N+1}\| \\ &< \frac{\epsilon}{2}\end{aligned}$$

The first line is just the way we defined f, the second line uses that the norm is a continuous function, and the third line uses that the sequence is bounded by $\frac{\epsilon}{2}$. There is potentially the issue of $\lim_{m\to\infty} \|f_m - f_{N+1}\|$ existing at all (consider the sequence which bobs between $\pm\frac{\epsilon}{4}$), but since the first term on the left is well-defined the interchanging of limit with $\|\cdot\|$ since the norm is continuous guarantees us it is well-defined. Then we return to our above unfinished calculation

$$\begin{aligned}\|f_n - f\| &= \|f_n - f_{N+1} + f_{N+1} - f\| \\ &\leq \|f_n - f_{N+1}\| + \|f_{N+1} - f\| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon\end{aligned}$$

Great, so $f_n \to f$ in the sense of the sup norm. Last bit: is $f \in \mathcal{C}[0,1]$? We might guess that somehow we need to use that each f_n is continuous at each x (where else would the continuity come from?), and probably again that the sequence is Cauchy under the sup norm. The strategy is then to break $|f(x) - f(y)|$ into terms which use the fact that f is close to f_n, and that for any given f_n, if $|x - y|$ is sufficiently small, then $|f_n(x) - f_n(y)| < \epsilon$. First, let $\epsilon > 0$, and note that the triangle inequality gives

$$|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)|$$

We need a δ to bound $|x - y|$, so our idea is to first pick an N big enough that for $n > N$ *both* $|f_n(x) - f(x)| < \frac{\epsilon}{3}$ and $|f_n(y) - f(y)| < \frac{\epsilon}{3}$, which is possible since $f_n \to f$, so we just pick N so that this holds for any x, y. Once we have this, we can just pick δ such that $|x - y| < \delta$ implies $|f_n(x) - f_n(y)| < \frac{\epsilon}{3}$, and

$$\begin{aligned}|f(x) - f(y)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon\end{aligned}$$

So f is continuous, and we are done. Note that in the last step the δ choice depended on the n choice, which depended on the N choice, which depended on the Cauchy assumption, so we did use the limit sequence definition of f to show its continuity.

□

The brother of Banach is also quite helpful.

Definition 7.11. A Hilbert space is a vector space equipped with an inner product, such that it is a Banach space with the norm induced by the inner product.

You will likely see little to no discussion of Banach and Hilbert spaces in your first year, but it is helpful to at least see them once and get a flavor for what they are.

Example 7.5. The space of real-valued Borel measurable random variables, with finite variance, is a Hilbert space. Its inner product is $\langle x, y\rangle = \int_\Omega x(\omega)y(\omega)\mathrm{d}\mu(\omega) = \mathbb{E}[xy]$. Why care? You can dig up a vector space or functional analysis textbook and find lots of nice theorems regarding Hilbert spaces (including theorems regarding projections!). Therefore working in Hilbert space potentially gives us nice properties, and in fact you'll see this exact Hilbert space in Azeem's course (though I don't think he points out that it is a Hilbert space).

Problem 7.9. What exactly do we gain from the additional structure of a Hilbert space, compared to a Banach space? Can you think of problems or ideas that cannot be addressed in a Banach space, but can be in a Hilbert space?

Example 7.6. The core ideas of generalized method of moments (GMM) rest on the idea of random variables living in a Hilbert space, and us wanting to find parameter values such that our assumptions about orthogonality are met.

8 Continuity

I present you with a definition of continuity which looks unhelpful, at first.

8.1 Basics

Definition 8.1. A function $f : X \to Y$ is continuous if $f^{-1}(V)$[15] is open for every open $V \subseteq Y$.

I give this definition because it is more general than the standard $\epsilon - \delta$ definition, in that it can work outside of metric spaces. However, in metric spaces they are equivalent:

Theorem 8.1. A function $f : X \to Y$ is continuous iff[16] for every $\epsilon > 0$, there exists a $\delta > 0$ such that if $d(x, x') < \delta$, then $d(f(x), f(x')) < \epsilon$.

Example 8.1. Cobb-Douglas production, $F(K, L) = K^\alpha L^{1-\alpha}$ is continuous. CRRA utility, $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$, is as well. Not really sure why I felt the need to give these, since most everything that we use as standard is continuous.

Example 8.2. Let X be a metric space with finitely many elements. Then every subset of X is open, since we may take a small enough neighborhood so that x is the only point in the neighborhood. So the pre-image of any open set for any function will be open. So *any* function on X will be continuous.

If you have not messed around with stuff much, or have not spent much time around pedantic mathematicians, the old "you can draw without lifting your pencil" seems a good enough intuition that you may think you'll never need the formal definitions. Here are some trickier ones:

Problem 8.1. Determine where each of the following functions is continuous (e.g. is the function continuous everywhere, nowhere, or at some subset, and what is that subset?).

(i) $f(x) = \sin(\frac{1}{x})$, $f(0) = 0$.

(ii) $f(x) = x \sin(\frac{1}{x})$, $f(0) = 0$

(iii) $f(x) = x^2 \sin(\frac{1}{x})$, $f(0) = 0$

(iv) $f(x) = \mathbf{1}_{\mathbb{Q}}(x)$

(v) Stars over Babylon function: If $x \in \mathbb{Q}$, then express $x = \frac{p}{q}$ in lowest possible terms (no integer divides both p and q) and $f(x) = \frac{1}{q}$. Otherwise $f(x) = 0$.

Sometimes we want a stronger notion of continuity. A few exist (with varying strengths), but I think this one is worth knowing.

Definition 8.2. A function $f : X \to Y$ is Lipschitz continuous if there exists a constant K such that $d(f(x), f(y)) \leq K \cdot d(x, y)$

Problem 8.2. Prove that if a function is Lipschitz continuous, then it is continuous.

[15] $f^{-1}(V) = \{x \in X \mid f(x) \in V\}$

[16] This double f means "if and only if".

8.2 Intermediate Value Theorem

Theorem 8.2 (Intermediate Value)**.** Let $x \leq y$, and $f : \mathbb{R} \to \mathbb{R}$ continuous. If $f(x) = a$, $f(y) = b$, then $[a, b] \subseteq f([x, y])$ (assuming without loss of generality $a \leq b$, otherwise we have $[b, a]$). If f is strictly monotonic[17], there exists an inverse mapping $f^{-1} : [a, b] \to [x, y]$.

I have slightly generalized the standard statement, mainly so that people already familiar with the theorem have to read it more closely. But the basic idea is intuitive: if I go from value a to value b as I move from x to y, and I know f is continuous, it has to hit all the values between a and b along the way.

Example 8.3. Most often, we use the intermediate value theorem to show existence of an equilibrium. For example, we might have the goods market clearing condition

$$\sum_i c_i(r) = \sum_i y_i(r)$$

where consumption c and output y both depend on the interest rate r. We can reframe this as

$$F(r) = \sum_i c_i(r) - \sum_i y_i(r)$$

If we can show that F is continuous (often the case), and that there exist $\underline{r}$ and $\overline{r}$ such that $F((\underline{r}) < 0 < F(\overline{r})$, then we may conclude there exists $r^* \in [\underline{r}, \overline{r}]$ such that the market clears. If we are also able to conclude F is strictly monotonic (also sometimes the case), then we may also conclude the equilibrium is unique.

Problem 8.3 (Baby Brouwer)**.** Consider the set of functions $F = \{f : [0, 1] \to [0, 1] \mid f \text{ is continuous}\}$.

(i) What are the possible values that may be taken by $f(x) - x$ at $x = 0$ (what is $\{y \mid f(0) - 0 = y, f \in F\}$)?

(ii) What are the possible values of $f(x) - x$ at $x = 1$ (what is $\{y \mid f(1) - 1 = y, f \in F\}$)?

(iii) What can you conclude about the values that *must* be take by $f(x) - x$ in $[0, 1]$ for any $f \in F$?

(iv) What can you conclude about the existence and uniqueness of fixed points for operators in F?

8.3 Extreme Value Theorem

One of the most important theorems we have connects continuity and compactness to optimization (which we will get to later).

Theorem 8.3 (Extreme Value Theorem)**.** If $f : K \to Y$ is continuous, where K is compact, then f has a maximum and minimum.

This might seem obvious or unhelpful, because extrema are always achieved, right? No. Therefore having a condition for existence is quite helpful.

[17] Here meaning $x > y$ implies $f(x) > f(y)$ (no equality allowed)

Problem 8.4. Find an example of $f : K \to Y$ which breaks the Extreme Value Theorem by satisfying the assumptions except:

(i) f is not continuous

(ii) K is not compact

Example 8.4. A simple application is that if an agent has a compact set of choices (defined by a budget constraint), and we consider the implied utility of each choice as a continuous function, then there exists a solution to the agent's problem. Usually we have a simple budget set, like a n-dimensional simplex, but note that the Extreme Value Theorem says we can accommodate weirder stuff, like prices endogenously changing, provided the budget set remains compact.

8.4 Hemicontinuity

I am intentionally omitting this section. It is worth covering if you want to be a micro-theorist, but it came up maybe only once during the first year, and spending time on it in math camp does not seem prudent to me.

9 Differentiation

9.1 General

Functions are generally not linear, which is quite unfortunate since linear objects are comparatively easy to deal with. Additionally, we often care about how a function changes as we vary its arguments. If the function is nice enough that we can make this "change behavior" precise, we may consider the linear[18] approximation of the function. This is all the derivative is: a locally linear approximation to a function. Below we will see Taylor's theorem, which I believe is often taught in an unmotivating context, so I will try to motivate a little differently.

Much of economics (and the physical sciences, for that matter) requires nice (linear) approximations to nonlinear functions in order to solve problems. It is worth keeping in mind that this is really all the derivative is.

I'll start with a really general looking definition that you almost certainly did not see in calc I, and maybe have not seen at all (I don't think I saw it in undergrad), but first, we need to clarify what we mean by "bounded", for a transformation.

Definition 9.1. A transformation $f : X \to Y$ is bounded if, for all bounded subsets $X' \subseteq X$, we have that $f(X') \subseteq Y$ is bounded.

Now, to the good stuff:

Definition 9.2. Let $x \in X$, $f : X \to Y$. If there exists a bounded linear transformation $A_x : X \to Y$ such that

$$\lim_{\|h\|_X \to 0} \frac{\|f(x+h) - f(x) - A_x(h)\|_Y}{\|h\|_X} = 0$$

Then A_x is the Fréchet derivative of f at x, and we say f is Fréchet differentiable at x. If such an A_x exists for each $x \in U \subseteq X$ (possibly different across x), then we may define $Df(x)$ as the A_x for each given $x \in U$, and say f is Fréchet differentiable over U.[19]

This is the generalization of the standard derivative you see in basic calculus. If you apply the definition to a simple space (like $\mathbb{R}^n$) you will see your old friends from calc I, II, and III pop out. Additionally, consider

Definition 9.3. Let $x \in X$, $f : X \to Y$. If there exists a transformation $A_x : X\Big|_{\|x\|=1\|} \to Y$ such that

$$\lim_{t \to 0} \frac{\|f(x+th) - f(x) - tA_x(h)\|_Y}{\|th\|_X} = 0$$

for all $h \in X\Big|_{\|x\|=1}$, then $A_x(h)$ is the Gateaux derivative of f at x. It depends on the direction, h, chosen.

[18] More generally affine

[19] Note that, while we require A_x to be linear, we do not have any requirements for how A_x may vary with x. Hence, as a function of x, $Df(x)$ may be nonlinear.

This is the generalization of the directional derivative you likely saw in multivariable calc.

Note the difference: the Fréchet derivative requires that the limit exists for any means of the norm of h approaching 0, whereas the Gateaux derivative only requires linear approaches (draw picture here). So if a function is Fréchet differentiable, it has a Gateaux derivative.

If we are working in a single dimension, $f : \mathbb{R} \to \mathbb{R}$, then Fréchet and Gateaux are the same, and we can see

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

$$\lim_{h\to 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = 0 \qquad (A \text{ linear, norm is scalar})$$

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

This last expression might look like the typical derivative definition you have seen. The reason we did not start with it is that we generally cannot divide by vectors, so require the general def, but we can divide when we have a scalar.

Example 9.1. We can easily find the derivative of $f(x) = x^2$ for $x \in \mathbb{R}$:

$$\begin{aligned}
\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} &= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} \\
&= \lim_{h\to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\
&= \lim_{h\to 0} \frac{2xh + h^2}{h} \\
&= \lim_{h\to 0} 2x + h \\
&= 2x
\end{aligned}$$

Problem 9.1. Find the derivative of $f(x) = x^n$.

Problem 9.2. Consider $f(x, y) = \frac{x^3}{x^2+y^2}$, except $f(0, 0) = 0$.

(i) Find the Gateaux derivative of f.

(ii) Either find the Fréchet derivative, or show it does not exist.

Definition 9.4. Some derivatives have special names.

(i) For $f : \mathbb{R}^n \to \mathbb{R}$, Df is called the gradient, and $D^2 f \equiv Hf$ is called the Hessian

(ii) For $f : \mathbb{R}^n \to \mathbb{R}^m$, Df is called the Jacobian (matrix)[20]

[20] I once lost a crap ton of points in a maths class for referring to this matrix as the Jacobian, because apparently mathematicians reserve "Jacobian" (without the word matrix after) to mean the determinant of the Jacobian matrix.

These have special names because they are the most common, and easily represented. Note that for $f : \mathbb{R}^n \to \mathbb{R}^n$, $D^2 f$ is a tensor-like object. Since, in this case, Df is a square matrix, $D^2 f$ can be visualized as a cube of matrix entries. And before you think I've wondered away from anything you'll ever use, I can tell you we had to deal with this type of cube object in an IO field course.

(Draw matrices with dimensions to illustrate that Df is linear operator. Good mnemonic for remembering the dimensions of Df is that it must operate on the same space f operates on.)

Theorem 9.1 ((i))**.** 1. A function is concave if Hf is negative semidefinite. A function is strictly concave if Hf is negative definite.

2. A function is convex if Hf is positive semidefinite. A function is strictly convex if Hf is positive definite.

Theorem 9.2. Let $C^\infty(\mathbb{R})$ be the space of functions $f : \mathbb{R} \to \mathbb{R}$ which are infinitely differentiable. Then the operator $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ which maps functions to their derivatives is a linear operator.

A tricky little point: be sure to keep track of your operators and the underlying space, or you will venture into sloppyland. For a given function, the Fréchet derivative is $\mathbb{R} \to \mathbb{R}$. The operator D which takes a function and returns its Fréchet derivative is $C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$.

Problem 9.3. The Fréchet derivative is a bounded linear operator by definition, and we above said that D is linear. Is it bounded (we endow $C^\infty(\mathbb{R})$ with the sup norm)?

Problem 9.4. What are the eigenvectors and eigenvalues of D? (Hint: Don't overthink. What was the simpest derivative to remember in calc I?)

9.2 L'Hospital

This theorem may feel one-off, but it is surprisingly handy when dealing with lots of limits.

Theorem 9.3 (L'Hospital)**.** Let f and g be differentiable and $g'(x) \neq 0$ in (a, b). Suppose $\frac{f'(x)}{g'(x)} \to A$ as $x \to a$. If $f(x) \to 0$ and $g(x) \to 0$ or if $g(x) \to \infty$ (as $x \to a$), then $\frac{f(x)}{g(x)} \to A$.

The basic idea is that sometimes we have a fraction which approaches an indeterminate form, but if we know that the ratio of the rates of change of the numerator and denominator approach a finite limit, then we can say the same for the original fraction.

Problem 9.5. Find $\lim_{x \to -1} \frac{x^2-1}{x+1}$.

Problem 9.6. Find the following limits

(i) $\lim_{\gamma \to 1} \frac{c^{1-\gamma}-1}{1-\gamma}$

(ii) $\lim_{x \to 0} x \ln x$

9.3 Chain Rule

The following theorem is crazy useful, and I guarantee you will use it a ton in your first year. It details how the derivative of a composition of functions relates to the derivatives of each individual function.

Theorem 9.4 (Chain Rule)**.** Suppose $T : X \to Y$ and $S : Y \to Z$ are differentiable, and define $P : X \to Z$ as $P(x) = S(T(x))$ (composition). Then P is differentiable, and $DP(x) = DS(T(x)) \circ DT(x)$.

Once again, this statement is overkill. The simpler version is that if $h(x) = g(f(x))$, then $h'(x) = g'(f(x)) \cdot f'(x)$, so the derivatives form a "chain" of composition.

Problem 9.7. Consider the function $h(x) = \sqrt{x + \sqrt{x}}$. In terms of the chain rule, what are f and g, and what is h'?

Problem 9.8. Use the chain rule to find the derivative of the following functions

(i) $f(x) = (\sum_{j=1}^{n} x^{\alpha_j})^{\frac{1}{\sum_{j=1}^{n} \alpha_j}}$

(ii) $f(x) = (\sum_{j=1}^{n} x_j^{\alpha_j})^{\frac{1}{\sum_{j=1}^{n} \alpha_j}}$ (note that x is now n-dimensional)

(iii) Look up Adrien Bilal's FAME framework. His approach of using the master equation to represent the equilibrium with one object leans heavily on recognizing how to apply the chain rule in more difficult spaces.

9.4 Taylor's Theorem

The following theorem may be stated more generally than I have it here, but will require more cumbersome notation since there will be multi-indices, etc. The basic idea of Taylor's approximation is that we may use the derivatives of a function to approximate that function. It is a generalization of the mean value theorem to higher orders of approximation

Theorem 9.5 (Taylor)**.** Suppose f is a smooth real function on $[a, b]$, and let $\alpha < \beta \in [a, b]$. Then there exists $\gamma \in [\alpha, \beta]$ such that

$$f(\beta) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(\beta - \alpha)^k + \frac{f^{(n)}(\gamma)}{n!}(\beta - \alpha)^n$$

The intuition here is that each term "corrects" some of the error of the previous term. I put "corrects", because if we truncate the final correcting term (the one with γ), there is no guarantee that increasing the number of terms increases the degree of approximation[21].

Example 9.2. The mean value theorem is just Taylor's theorem where $n = 1$: We can find a point b between two points (a, c) such that the slope at b matches the average slope from a to c: $f'(b) = \frac{f(c)-f(a)}{c-a}$.

Problem 9.9. Find an example where f is continuous, but not differentiable, that breaks the mean value theorem.

Problem 9.10. Can you extend the mean value theorem (or Taylor's theorem) to higher dimensions? What is the extension?

Problem 9.11. Find the n-th order Taylor approximation for each function about z, where by approximation, we mean dropping the correcting term with the γ.

[21] This can actually be quite subtle. The bounds on the remaining error shrink as we add more terms, but there is no guarantee that less terms will not perform better, even though their bounds are worse

(i) $f(x) = x^3$, $n = 1$, $z = 1$

(ii) $f(x) = x^2$, $n = 2$, $z = 1$

(iii) $f(x) = x^2$, $n = 3$, $z = 1$

(iv) $f(x) = e^x$, $n = 100$, $z = 0$

In math courses my experience is that we often fail to emphasize what this theorem is good for. After all, if I know the function, why am I using derivatives to approximate the function? We don't always know the function, or have an analytical derivative available.

Example 9.3. Consider calibrating a model. For concreteness, you have a consumption-savings model, and want to match some level of aggregate assets. Within this setting, holding everything else in the model constant, we can consider the assets level as a function of, say, the interest rate, $a(r)$. The rub is that we can find $a(r)$ for any given r, but we do not know the analytical form. Gradient descent takes advantage of the fact that we can approximate $a(r)$ by a first-order Taylor expansion, and we can computer numerical (approximate) derivatives via $\frac{a(r+h)-a(r)}{h}$. The shiny new way to do this is with automatic differentiation (look it up if interested).

9.5 Inverse Function Theorem

This next one is a bit strange, and not always the most useful, but it can often be helpful when we need an inverse for part of a proof (say existence or uniqueness, for example). Basically, if things are sufficiently "non-constant", we can locally invert a function.

Theorem 9.6 (Inverse Function)**.** Suppose $f : X \to X$ is smooth, $a \in X$, and $Df(a) : X \to X$ is bijective. Then

(i) There exist open sets $U, V \subset X$ such that $a \in U$, $f(a) \in V$, and $f : U \to V$ is a bijection

(ii) $f^{-1} : V \to U$ is smooth, and $D[f^{-1}(x)] = Df(f^{-1}(x))^{-1}$

I'm guessing if you saw this in a real analysis course at first it blew your mind a little bit, but my hope is that my phrasing makes things not too bad. The gist of it is that if the derivative at a point is invertible, then we can find a little area around that same point where the function itself is invertible.

Example 9.4. By far the clearest way to see this theorem in action (and how it breaks) is with a baby example. Consider $f(x) = -(x-1)^2 + 1$, and let $0 < a < 1$. Then f is smooth, and $Df(a)(x) \mapsto -2(a-1)x$ is bijective (aka $Df(a) \neq 0$). This case is simple enough we can even construct U and V without much hassle. Let $U = (0, \frac{1}{2}(a+1))$, $V = (0, f(\frac{1}{2}(a+1)))$, and we have $f^{-1}(x) = 1 - \sqrt{1-x}$. If we instead consider $a = 1$, then $Df(a) = 0$, and indeed intuitively there is no way to build an inversion, since f will fail the "horizontal line test" over any open neighborhood containing $a = 1$.

Problem 9.12. Consider the inverse demand curve $f(Q) = \frac{1}{Q} - \frac{1}{1-Q}$ for $Q \in (0, \frac{1}{2})$, and the (elastic) inverse supply curve $g(Q) = P$. Use the inverse function theorem to say that equilibrium Q is a smooth function of P.

9.6 Implicit Function Theorem

The Inverse Function Theorem, Implicit Function Theorem, and Rank Theorem are tightly linked, and with the proper formulation any one can be used to prove the other two. I start with Inverse because it is my favorite and what I find the most useful, but now we can quickly go through Implicit. I'm skiping Rank, but you can look it up yourself (say in (Rudin, 1976)) if interested. Whereas the Inverse FT discussed smooth mappings in the context of directly mapping from a space to itself, The Implicit FT instead uses the context of an *implicit* manifold which we may want to say things about.

Theorem 9.7 (Implicit Function)**.** Suppose $f : X \times Y \to X$ is smooth, $(a, b) \in X \times Y$ is such that $f(a, b) = 0$, and $D_X^b f : X \to X$ is bijective (D_X^b denotes only taking the derivatives with respect to the X coordinates, holding the Y coordinates at b). Then

(i) There exists an open set $U \subseteq Y$ and a smooth mapping $g : U \to X$ such that $b \in U$, $g(b) = a$, and $f(g(y), y) = 0$ for all $y \in U$.

(ii) g is smooth and satisfies $Dg = -[D_X^y f(g(y), y)]^{-1} D_Y^{g(y)} f(g(y), y)$

Problem 9.13. Suppose the equilibrium of a system of interest is expressed as

$$\log y + cy = k$$

Let f map from exogenous parameters (c, k) to the solution y of the system. How does y depend on c and k (what are$\frac{\partial f}{\partial c}$ and $\frac{\partial f}{\partial k}$)?

10 Fixed Point Stuffs

There's a whole area of math devoted to understanding fixed points. As economists, we are interested in having these results in our pockets for proving existence, uniqueness, and even construction or search for equilibria. First the definition:

Definition 10.1. A fixed point of $f : X \to X$ is any point satisfying $f(x) = x$.

So at a fixed point, all the "forces" that the function takes into account balance out and x is unmoved.

10.1 Contraction Mapping Principle

We start with, I'm not gonna mince words, the best fixed point theorem. It gives existence, uniqueness, a method to find the fixed point, the proof is easy, and the intuition is easy.

Definition 10.2. An operator $\varphi : X \to X$ on a metric space (X, d) is a contraction if there exists $0 \leq \rho < 1$ such that $d(\varphi(x), \varphi(y)) \leq \rho d(x, y)$ for all $x, y \in X$. The minimum ρ for which this is true is called the modulus of φ.

Problem 10.1. Prove that a contraction is Lipschitz continuous.

Problem 10.2. We need to refresh geometric series real quick. Let $|a| < 1$, and find $\sum_{i=0}^{\infty} a^i$, taking as given that it is finite. (Hint: figure out how to write the series as an equation depending on itself, then solve the equation, assuming the sum is finite.)

Theorem 10.1 (Banach Contraction Mapping)**.** Let (X, d) be a complete metric space, and $\varphi : X \to X$ a contraction with modulus ρ. Then there exists a unique fixed point of φ, and furthermore the sequence $\varphi^n(x)$ converges to the fixed point exponentially quickly for any $x \in X$.

Proof. We start with existence, and get the convergent sequence result along the way. Let $x \in X$ be arbitrary, and consider the sequence $x_{n+1} = \varphi(x_n)$. We want to show this is a Cauchy sequence, because then we can say it converges. Let $m > n$, and note

$$\begin{aligned}
d(x_n, x_m) &\leq \sum_{i=0}^{m-n-1} d(x_{n+i}, x_{n+i+1}) \\
&\leq \sum_{i=0}^{m-n-1} d(x_{n+i}, \varphi(x_{n+i})) \\
&\leq \sum_{i=0}^{m-n-1} \rho^i d(x_n, \varphi(x_n)) \\
&\leq d(x, \varphi(x))\rho^n \sum_{i=0}^{m-n-1} \rho^i \\
&\leq d(x, \varphi(x))\rho^n \sum_{i=0}^{\infty} \rho^i \\
&= d(x, \varphi(x)) \frac{\rho^n}{1-\rho}
\end{aligned}$$

Then we can see that for any $\epsilon > 0$, there exists N such that $n, m > N$ implied $d(x_n, x_m) < \epsilon$, so our sequence is Cauchy. Therefore its limit exists, and we will call it x^*.

Now we note

$$\begin{aligned}
\varphi(x^*) &= \varphi(\lim_{n\to\infty} x_n) \\
&= \lim_{n\to\infty} \varphi(x_n) && (\varphi \text{ (Lipschitz) continuous}) \\
&= \lim_{n\to\infty} x_{n+1} && \text{(By def of sequence)} \\
&= x^*
\end{aligned}$$

Therefore we have shown that $\varphi(x^*) = x^*$, and the exponential convergence can be shown by considering the sequence we constructed.

Lastly, we suppose x^* and y^* are fixed points. Then $d(x^*, y^*) = d(\varphi(x^*), \varphi(y^*)) \leq \rho d(x^*, y^*)$. Since $\rho < 1$, this can only be the case if $d(x^*, y^*) = 0$, in which case $x^* = y^*$.[22]

□

So the proof gives us a way to find the fixed point, and it could not be simpler: iterate on the contraction until you are happy with the convergence[23]. The intuition should also be clear by now: the contraction pushes elements strictly closer together, so we find the points that everything is getting pushed towards.

Example 10.1. Probably the simplest contraction is $f(x) = ax$, where $|a| < 1$, as it has modulus $|a|$, and fixed point 0.

Example 10.2. One of my favorites: Take a to-scale map of a country and lay it on the ground of the country. The mapping which takes a point in the country and maps it to its position on the ground under the corresponding point on the map is a contraction. There exists a unique point of the map which is *exactly* on top of its physical counterpart!

Problem 10.3. Find a modulus and fixed point of the following contractions

(i) $f(x) = k + ax$, $k \in \mathbb{R}$, $|a| < 1$

(ii) The map that takes a point (x, y), converts it to polar coordinates, scales the radius by $|a| < 1$, and rotates the angle by k radians.

Problem 10.4. Consider $f(x) = Ax$, where A is a symmetric $n \times n$ matrix. What is a condition for this to be a contraction? Prove it. Can you generalize this to operators more generally?

If a function $\mathbb{R} \to \mathbb{R}$ is differentiable, then we can use the derivative to show a mapping is a contraction, in some cases.

[22] This uniqueness bit is a good example of a proof which is often done by contradiction, but a direct proof is just as easy.

[23] This is known as Picard iteration. It underlies many proofs for the existence of solutions to ODEs.

Theorem 10.2. If $f : \mathbb{R} \to \mathbb{R}$ is differentiable, and there exists an interval $[a, b]$ such that $x \in [a, b]$ implies $|f'(x)| \leq c$ for some $c < 1$, then f is a contraction over $[a, b]$

Proof. For $x < y \in [a, b]$, the mean value theorem implies there exists $z \in [x, y]$ such that $f'(z)(y - x) = f(y) - f(x)$. Then

$$\begin{aligned} |f(y) - f(x)| &= |f'(z)(y - x)| \\ &\leq |f'(z)||y - x| && \text{(Cauchy-Schwartz)} \\ &\leq c|y - x| \end{aligned}$$

where the last line is by assumption since $z \in [a, b]$. □

Example 10.3. Consider a simple growth model

$$\begin{aligned} k_{t+1} &= f(k_t) \\ &= Ak_t^{\alpha} \end{aligned}$$

Then $f'(x) = A\alpha x^{\alpha-1}$, so for $x > (A\alpha)^{\frac{1}{1-\alpha}}$ we have $f'(x) < 1$ and f is a contraction for $[(A\alpha)^{\frac{1}{1-\alpha}} + \epsilon, M + (A\alpha)^{\frac{1}{1-\alpha}} + \epsilon]$ for any $\epsilon, M > 0$.

Example 10.4. Bellman equations. You'll see them soon enough, so no reason for me to go into detail now.

10.2 Brouwer

Sometimes no contraction is available, but we at least know the function of interest is continuous, and we can consider it in a sufficiently "nice" space. Then we can sometimes still get existence of a fixed point.

Theorem 10.3 (Brouwer)**.** Every continuous function from a closed ball of a $\mathbb{R}^n$ into itself has a fixed point.

This should look familiar from a problem above, where we proved this in a super particular and easy setting.

Theorem 10.4 (Schauder)**.** Every continuous function $f : K \to K$, where K is a convex compact (nonempty) subset of a Banach space, has a fixed point.

The proofs for these theorems are somewhat more involved, though they do provide some intuition for what is going on. In some sense, proving them relies on an "iterative trapping" argument, wherein we show that some point must get mapped to a smaller and smaller set of points near it, until we show that at the limit this collapses to being the point itself. The continuity, compactness, and convexity are all crucial.

Problem 10.5. Construct examples where all of the conditions of Schauder's fixed point theorem hold, except for the following, and show that no fixed point exists in each example.

(i) f is not continuous

(ii) K is not compact

(iii) K is not convex

Note that these theorems say nothing about uniqueness. This is actually quite annoying, because we may have contexts where we can show that Brouwer does what we want, but we desire uniqueness. In this case we need to rely on other arguments, such as showing that if two equilibria were to exist, then a better one between them would exist, and getting a contradiction.

Example 10.5. Consider again the map example from above. Now you are allowed to also stretch and compress the map before putting it on the ground. Sometimes we will still have a contraction, but if you stretch map carefully enough we will not. However, the mapping is still continuous, so Brouwer lets us say there exists at least one fixed point.

Problem 10.6. Explain how to stretch the map so that there is more than one fixed point. Can you get exactly two? What if I let you tear the map? What does tearing the map do to our mapping?

Problem 10.7 (Cournot)**.** Consider a set of n firms producing the same product. Each firm i faces market price $P(q) = a - b\sum_{i=1}^{n} q_i$, and controls their own quantity of production, q_i. They face unit costs c.

(i) What the profit function $\pi_i(q)$ for firm i?

(ii) Holding fixed q_j for $j \neq i$, what is the FOC and optimal q_i for firm i?

(iii) If all firms play this optimal q_i and this is common knowledge, what is the optimal q^* that all firms choose?

(iv) **Show the symmetric equilibrium is the unique one, i.e. there does not exist an asymmetric q^* which is also an equilibrium.

(v) What is the implied price?

(vi) What are the implied firm and aggregate profits?

(vii) As $n \to \infty$

  (a) What does q^* go to?
  (b) What does $\sum_i q^*$ go to?
  (c) What do firm and aggregate profits go to?

The above problem is standard firm theory stuff. Now, for fun, we'll put a good old "Chicago Price Theory" spin on it. The following question is somewhat open-ended, and there are no "right" or "wrong" answers, but you should use the above model and results to reason through your responses. Channel your inner Friedman/Becker/Murphy.

Problem 10.8. You are now an economic policymaker. The industry of interest has a tendency to coalesce into a small number of firms. You have legislative tools that will allow you to "trust-bust", but you can also choose to not use your tools, which will lead to firms combining and/or forming cartels. You also (later in the problem) have the ability to tax the market, and this tool is independent of your trust-busting hammer.

(i) If you believe that (outside of the market within the model thus far) these goods provide a positive externality, what policies do you pursue?

(ii) Same as above, but negative externality.

(iii) A friend makes the statement "We need to break up the big firms, since their market power is allowing them to pollute more than if there were more competition." Does this (highly stylized toy) model support or refute this claim?

(iv) Suppose this industry, for whatever reason, is particularly easy to tax, but the industry also produces a massive negative externality. You have a big new budget bill that needs funded. What do you do?

(v) Does your above answer change if this industry is comparatively harder to tax?

(vi) Evaluate the following statement: Effective and simple tax policy will decrease pollution.

(vii) Evaluate the following statement: Monopolies are bad. Externalities are bad. A monopoly market with externalities must be worse than that same market with only one or the other!

Lest you think I have forgotten: the above is a toy model! Be careful with taking the "answers" to the above question to the real world! Do "innocuous" changes to the above model flip the results, or at least change the quantitative outcomes? This is a rhetorical question, but one we should always ask if we want to take price theory implications and apply them to the real world.

10.3 Kakutani

We skipped hemicontinuity and correspondences, but Kakutani's fixed point theorem is an extensions of Brouwer to correspondences, which map not to values, but to sets[24].

Definition 10.3. A correspondence $f : X \to 2^Y$ is a mapping which maps elements to sets.

Example 10.6. $f(x) = [-x, x]$ fans out as x increases.

Example 10.7. All functions (mappings to a single point) are correspondences.

Correspondences come with their own terminology. I think the terms are fairly intuitive, though, so hopefully the following definitions are not too surprising.

Definition 10.4. A property P of the graph of a correspondence refers to P holding under the product topology of $X \times Y$[25] for the set $\{(x, y) \mid y \in f(x)\}$.

Definition 10.5. A correspondence is P-valued, for some property P, if for all $x \in X$, $f(x)$ satisfies P.

Example 10.8. We will need these specifics in Kakutani's fixed point theorem.

(i) f has a closed graph if $\{(x, y) \mid y \in f(x)\}$ is closed under the product topology of $X \times Y$

(ii) f is non-empty valued if $f(x) \neq \varnothing$ for all $x \in X$.

(iii) f is convex-valued if $f(x)$ is a convex set for all $x \in X$.

Finally we need to amend our definition of "fixed point" slightly. The following definition matches our above definition when we restrict correspondences to functions.

[24] Though, of course, a set is just an element of the power set.

[25] "Product topology" is not defined in these notes, but it is intuitively what you think and want it to be for our contexts. If you are curious, do some further reading/Googling.

Definition 10.6. We say $x \in X$ is a fixed point of a correspondence if $x \in f(x)$.

Finally, we have all the ingredients to extend Brouwer's fixed point theorem to the case of correspondences.

Theorem 10.5 (Kakutani)**.** Let X be compact, convex, and non-empty. If $f : X \to 2^X$ is a correspondence which has a closed graph, is non-empty valued and convex valued, then f has a fixed point.

Problem 10.9. The function $f(x) = [0, -4(x - \frac{1}{2})^2 + 1]$ satisfies the conditions for Kakutani on $[0, 1]$. What are the fixed points?

Problem 10.10. Find an example of f and X that satisfy the conditions of Kakutani except for the following, and thus there is not a fixed point in each example.

(i) f does not have a closed graph

(ii) f is not non-empty valued

(iii) f is not convex-valued

The above theorem was integral to the development of game theory. The reason correspondences were needed (rather than just functions) is for the cases where the optimal strategy is not unique (e.g. sometimes all strategies are optimal), which turn out to be important, at least in terms of theory, for showing that equilibria exist.

Problem 10.11. Consider a 2×2 game (like Prisoner's Dilemma, Stag Hunt, Matching Pennies, etc.). Consider the mapping $BR(a_1, a_2)$, which takes an action by each player, then maps to the optimal action set for each player, *fixing the other player's option at the given value.*

(i) If we restrict to pure strategies, can we guarantee a fixed point exists? Use one of the above theorems to prove so, or find a counterexample.

(ii) Can we say anything about uniqueness with only pure strategies?

(iii) If we allow mixed strategies, can we guarantee a fixed point exists? Use one of the above theorems to prove so, or find a counterexample.

(iv) Can we say anything about uniqueness, when we allow mixed strategies?

(v) Economically, what do we call a fixed point in this context (think Russell Crowe)?

10.4 General Note

Hopefully, it is now somewhat clear why fixed point theorems are so often discussed for economists: we are obsessed with equilibria, and FPT often give conditions for us to know when equilibria exist.

Intuitively, Econ 101 probably taught you how to think about equilibria forming by saying that supply has to equal demand, and if one or the other move (say demand increases) then there is a mismatch at the current price (quantity demanded exceeds quantity supplied), so a new price comes about (a higher one) such that supply equals demand. In numerically solving models, this is almost exactly what we do. We'll try a price, see if markets clear, and adjust the price until they do. More generally, a model takes a set of agent choices, and maps to a new set of implied choices (think NE). The fixed point of this operator is the equilibrium.

Example 10.9. An economy has n agents which (heterogeneously) produce and consume m goods types. For any given m-dimensional price vector p, each agent will be willing to supply a certain set of goods $s_i(p)$, and will demand a certain quantity of goods, $d_i(p)$. The excess demand of the economy is then $z(p) = \sum_{i=1}^{n} d_i(p) - \sum_{i=1}^{n} s_i(p)$, and the economy is in equilibrium only when $z(p) = 0$. The fixed point application is that a given supply of goods will imply a price that gives the implied demand. So we can consider this mapping from supplies of goods to demands of goods (through the price mechanism!) and its fixed point is the equilibrium.

Example 10.10. A more complicated example would be to think of a dynamic heterogeneous-agent model (i.e. HANK), wherein the distribution of agents over the state space maps to a price (through supply side), which then maps to an implied distribution of agents. The fixed point of this model is an entire distribution. Sometimes, however, the model may be reduced instead to pairing of a value functions and prices which imply the distribution, so we need only find the prices that imply the value function that imply the distribution that imply the original prices.

Example 10.11. Yet another example from my idiosyncratic preferences for topics: spatial equilibria. In urban economics, or more generally spatial economics, we often search for the price vector and/or spatial distribution that "balances" agglomeration and congestion forces so that no individual wants to move from where they live. We can make this dynamic by either allowing a constant transition in the distribution to be taking place (stochastic steady-state) and/or by considering each agent atomistic (in which case the model becomes a mean-field game, and the "spatial shuffling" averages out to have the distribution remain stable).

Problem 10.12. Find a paper with a structural model, and check that the equilibria match our idea of a fixed point in these notes (figure out what variables make up x and what equations make f). Do the authors prove existence and/or uniqueness, and if so, with what theorem(s)? If they don't prove, what intuition do you have for existence/uniqueness (which may come from the above theorems applying or almost applying)?

Clearly, the point is that we are looking for way for everything to be "fixed" (but time may be part of the state space!), and there may be multiple ways of viewing which object we start with and want to consider as being mapped to itself.

11 Optimization

One aspect of economics that is quite special is that we deal with humans, and in particular agents that have choices and desires. Therefore, in almost any economic model (even reduced-form), in the background there is some degree of optimization taking place. In order to model this, we spend a lot of time thinking about how to model and solve optimization problems.

11.1 Unconstrained

An unconstrained problem is of the form

$$\max_x f(x)$$

If f is not smooth, and we don't have additional info about it, we basically have no theorems about how to find the max, and have to manually search. If f is smooth, then we can consider the derivative, Df, and move in the direction in which the function is increasing[26]. If we are ever at a point where f is not increasing in any direction, then we have a candidate for an extremum. If we can further say that the derivative of the derivative $D^2 f \equiv Hf$ is "positive" or "negative", then we can say the extremum is a local minimum or local maximum. This is not always possible. We now clarify these ideas.

Theorem 11.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be smooth[27]. If $Df(x) = 0$ and $Hf(x)$ is positive (negative) definite, then x is a local minimum (maximum).

Requiring that the gradient $Df(x) = 0$ is the first-order necessary condition (FONC), and requiring that the Hessian $Hf(x)$ is definite is the second-order sufficient condition (SOSC). Note that in one-dimension the SOSC is just checking the sign of $f''(x)$, and that if $Hf(x)$ is indefinite, we cannot say anything about x (without further investigation).

Problem 11.1. Find all the critical points of the below functions, and identify them as minima, maxima, or neither. Be sure to explain how you checked (you may need to be more creative than just FONC and SOSC).

(i) $f(x) = (x-5)^2$

(ii) $f(x) = (x-3)^4$

(iii) $f(x) = \sin(x)$

(iv) $f(x) = \sin(x) + \frac{x}{2}$

(v) $f(x) = \sin(x) + x$

(vi) $f(x) = \sqrt{|x|}$

(vii) $f(x,y) = (x-2)^2 + (y+6)^2$

(viii) $f(x_1, \ldots, x_n) = \prod_{i=1}^n x_i^{\alpha_i} - \sum_{i=1}^n w_i x_i$, where $w_i > 0$, $\alpha_i \in (0,1)$, and $\sum_{i=1}^n \alpha_i < 1$.

(ix) $f(x_1, \ldots, x_n) = \prod_{i=1}^n x_i^{\alpha_i} - \sum_{i=1}^n w_i x_i$, where $w_i > 0$, $\alpha_i \in (0,1)$, and $\sum_{i=1}^n \alpha_i = 1$.

[26] On a computer this is gradient ascent.

[27] Smooth generally means C^∞, but really C^2 is fine here.

(x) $f(x) = \int_0^1 \|x(t) - k(t)\|^2 \mathrm{d}t$ where k is a continuous function and the underlying space we are optimizing over is continuous functions over $[0, 1]$.

Theorem 11.2. Suppose $f : X \to \mathbb{R}$ obtains extrema at $\{x\}$ and $g : \mathbb{R} \to \mathbb{R}$ is strictly increasing[28]. Then the extrema of $h(x) = g(f(x))$ are $\{x\}$.

Example 11.1. As shown by the Cobb-Douglas examples above, firm production problems may be considered unconstrained optimization problems. Kevin Murphy likes to make the point that this is special and different from a consumer, precisely because a consumer faces a budget and maximizes utility, whereas a firm is producing and may use production profits to finance factor inputs.

11.2 Equality Constrained

An equality constrained problem is of the form

$$\max_x \; f(x)$$
$$\text{such that } g(x) = 0$$

As we might expect, these types of problems are, in general, even harder than unconstrained problems. However, if both f and g are smooth, we can use Lagrangian optimization to find the extrema.

Formally proving that Lagrangian optimization works is not particularly illustrative (though it is worth going through once on your own). The intuition behind what is going on is helpful. Let's first lay out the ideas. We define the Lagrangian $\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)$, where λ matches the dimension of g. We then find the critical points (x, λ) where $\mathcal{L}_x = 0$ and $\mathcal{L}_\lambda = 0$. These points are our candidates for extrema, which we can then test with the SOSC or via other methods if SOSC does not work. In reality, we typically just check the FONCs and make sure the problem is sufficiently well-behaved beforehand.

Why does this work? Again consider gradient descent, except note that, by the construction of the Lagrangian, $g(x) = 0$ at the critical points, since we require $\mathcal{L}_\lambda = 0$. Thus, by adding $\lambda g(x)$ to our $f(x)$, we have created an unconstrained maximization problem which can only have solutions where $g(x) = 0$. But anytime $g(x) = 0$, $\mathcal{L} = f$. So it must be that we are finding the maximum value of f, such that $g(x) = 0$.

Finally, is there anything to be taken away from the Lagrange multipliers? Yes! Note that $\mathcal{L}_g$ (the derivative of the Lagrangian with respect to the constraint) is λ, so λ tells us the rate of change of $\mathcal{L}$ (and by the envelope theorem[29] f) when we relax the constraint.

Problem 11.2. For each of the following constrained optimization problems, find the maximum. In all cases, $p_i > 0$ for all i.

(i) $f(x) = \prod_{j=1}^n x_j^{\alpha_j}$, $g(x) = w - \sum_{j=1}^n p_j x_j$

(ii) $f(x) = \sum_{j=1}^n \alpha_j \ln(x_j)$, $g(x) = w - \sum_{j=1}^n p_j x_j$

[28] While g may not be differentiable, the infimum of its set of subderivatives is positive.
[29] See below in the notes for more on the envelope theorem.

(iii) $f(x) = \max\{x_1, \ldots, x_n\}$, $g(x) = \sum_{j=1}^n p_j x_j - w$, $x_i \geq 0$

(iv) $f(x) = \min\{x_1, \ldots, x_n\}$, $g(x) = \sum_{j=1}^n p_j x_j - w$

Problem 11.3. You are given k meters of fencing to make a rectangular fence. How do you build the fence to house the maximum number of cattle?

Example 11.2. Consider the following problem: You are given a finite amount of cake today, and know the cake will expire in one week, and you have to decide how to consume that cake. We won't go over this here, because you will see such problems in Nancy Stokey's course in just a few weeks. In brief, however, the problem will feel the same as above: maximizing a utility criterion subject to a material (or budgetary) constraint.

11.3 Inequality Constrained

Not much is left to be said here. An inequality constrained problem takes the form

$$\max_x f(x)$$
$$\text{such that } g(x) \geq 0$$

These types of problems are interesting and often require lots of case-checking, but usually in econ, even thought we might phrase a problem as inequality constrained, we quickly find that the problem is equality constrained.

A general cookbook for how to view this most general type of problem:

1. Find the unconstrained extrema of f. If any satisfy $g(x) \geq 0$, keep them, but throw the rest away.
2. Solve the constrained maximization problem.
   (I) If the constraint is piecewise smooth, break into pieces and check each one.
   (II) You may need to view a subset where $g(x) = 0$ as its own sub-inequality constrained problem, and this process can be iterative.

I really included this section only for completeness, so do not want to dwell on it.

Problem 11.4. Minimize $f(x) = x^2+y^2+z^2$ such that $g(x)$ parameterizes the union of the ellipsoid $a(x-x_0)^2+b(y-y_0)^2+c(z-z_0)^2-1$ and the box with vertices at (e_i, f_i) where $i = 1, \ldots, 8$. Note that you will have to find the max on the ellipsoid, then check each face of the box, then check each edge of the box, then check each corner of the box, and finally compare all the extrema.

11.4 Envelope Theorem

The envelope theorem, or envelope condition, is one of those terms that gets thrown around a ton, but I feel like no one ever fully explains it, and consequently most people don't really know what it is. One reason is that when people say "envelope ..." they are usually not citing it the way you would cite a formal mathematical theorem, but instead meaning the general idea of an envelope, which is what we discuss here. If you want the technical details, go read Wikipedia, but I don't find it as enlightening.

The basic idea of the envelope condition that is mostly used is that, at an optimum, the marginal value of increasing one choice variable, but *not* readjusting the others, is the same as if you *did* adjust the others. (Draw picture)

You can open the example Simple Production in Desmos to play with parameters and see the enveloping in action (I suggest starting by clicking the play button by the K parameter). Note that the derivative of V_R matches the derivative of V, at the point where they coincide. This is what the envelope theorem tells us should happen.

The parameters are

- w: labor wage
- r: capital rental
- α: capital intensity
- β: labor intensity
- A: total factor productivity
- K: fixed capital level for V_R
- K_L: outputs optimal capital for a given labor level
- V_R: value for varying levels of labor, given fixed capital K
- V: value for varying levels of labor, given that capital also adjusts optimally

Can you interpret what is going on? Do you know where K_L came from? I've left this intentionally vague so you can think a bit.

11.5 KKT

The following theorem generalizes the Lagrange multiplier result to inequality constraints.

Theorem 11.3 (Karush-Kuhn-Tucker)**.** Let $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^m$. If x is a (local) maximum of f, such that $g(x) \geq 0$, and $Dg(x)$ has full row rank, then there exists $\lambda \in \mathbb{R}^m$ such that[30]

$$\begin{aligned} \lambda &\geq 0 \\ g(x) &\geq 0 \\ \lambda \odot g(x) &= 0 \\ Df(x) + \lambda \cdot Dg(x) &= 0 \end{aligned}$$

[30]The symbol $\odot$ stands for the Hadamard product. This is element-wise multiplication, so if x and y are n-vectors, then so is $x \odot y$, and it is $(x_1 y_1, \ldots, x_n y_n)$.

The first three expressions above are the complementary slackness conditions, named so because if one of the first two conditions is slack (> 0) the other must bind. The requirement $Dg(x)$ has full row rank is the constraint qualification.

Let's again think through the logic, as we did with Lagrangians. We may again consider the Lagrangian $\mathcal{L}(x, \lambda) = f(x) + \lambda \cdot g(x)$, which has the property that $\mathcal{L}(x, \lambda) = f(x)$ when $\lambda \cdot g(x) = 0$. With Lagrangians, this constraint was always satisfied because $g(x) = 0$ was a requirement. The amendment here is that if $g(x) > 0$ for some indices, then we require $\lambda = 0$ instead.

The interpretation of the Lagrange multipliers has barely changed: For positive Lagrange multipliers, they still answer "what would be the marginal increase in $\mathcal{L}(x, \lambda) = f(x)$ if $g(x) = 0$ were relaxed?". For zero Lagrange multipliers, the interpretation should now be clear: the marginal increase is zero because $g(x) = 0$ is not a binding constraint!

Again, these conditions are necessary but not sufficient.

Problem 11.5. Consider the problem of maximizing $f(x, y) = x^2 + y^2$ subject to $x \geq 0, y \geq 0$. Show that all the conditions of KKT are satisfied at some x, but that it is not a maximum.

A general cookbook approach is then:

1. Setup $\mathcal{L}(x, \lambda) = f(x) + \lambda \cdot g(x)$.
2. Take the first order conditions.
3. See if the nature of the problem allows for eliminating solutions, or narrowing such that λ is necessarily positive or zero. Inada conditions often imply $\lambda > 0$ for some indices, or sometimes all.

This last step is clearly vague and tricky. If would be nice if we had some flavor of SOSC to narrow our scope of solutions, and maybe even obtain uniqueness. We do.

Theorem 11.4. Suppose $f : X \to \mathbb{R}$ is (strictly) concave, X is convex, $g : X \to \mathbb{R}^m$ is (strictly) concave component-wise, and there exists $x \in \text{int}(X)$ such that $g(x) > 0$ (all indices). Then there exists (unique) (λ^*, x^*) satisfying the KKT conditions iff x^* is a local (global) maximum.

The only new weirdness is the $g(x) > 0$ in the interior condition, called Slater's condition. Basically, this says that we must be optimizing over a space which has an interior relevant to the problem: if $g(x) = 0$ everywhere in the interior, we are on an "edge" of the constraint set.

The concavity conditions are just a way of making sure that we find local maxima instead of minima. If they are strict, then the local maximum must also be global. We may relax the concavity assumptions somewhat.

Theorem 11.5. Take all the conditions of the above theorem except only require that f be quasi-concave, g be quasi-concave component-wise, and don't require Slater's condition. Suppose (x^*, λ^*) satisfy the KKT conditions. Then x^* is a global maximum provided either f is concave, or $Df(x^*) \neq 0$.

Note the difference in this statement: the conditions are no longer necessary. If satisfied, we have a maximum, but it is possible to not satisfy them and still have a maximum. Strengthening this theorem to make the FOCs into FONCs recovers one of the above two theorems.

11.6 Duality

Another way to think about constrained optimization is via a min-max game. Instead of considering an optimizer that must satisfy some constraints, we instead consider either an unconstrained or more simply constrained pair of optimizers which have different controls, and are optimizing in opposite directions. To be concrete, consider again the original problem,

$$\max_{x} \; f(x)$$
$$\text{such that } g(x) \geq 0$$

We will call solutions O. Also recall the Lagrangian

$$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)$$

Above, we considered that candidates satisfy $\mathcal{L}_x = \mathcal{L}_\lambda = 0$. I now claim that we can take this FOC idea a step further, and say that the original problem is equal to the following primal problem,

$$\max_{x} \min_{\lambda \geq 0} \mathcal{L}(x, \lambda)$$

We will call solutions to this problem P.

Problem 11.6. Show the above equivalence statement in the following steps:

(i) Argue that when $g(x) \geq 0$, then $f(x) = \min_{\lambda \geq 0} \mathcal{L}(x, \lambda)$

(ii) Argue that when $g_i(x) < 0$ for any i, then $\min_{\lambda \geq 0} \mathcal{L}(x, \lambda)$ is unbounded below (so the min does not exist).

(iii) Use the above statements to show that $O = P$

Thus far we have only had an alternative way to think of the (same) optimization problem. Now we consider a different problem, which will in some cases be the same problem. The dual problem is

$$\min_{\lambda \geq 0} \max_{x} \mathcal{L}(x, \lambda)$$

We will call solutions D.
I see your immediate confusion. That is the same problem, no? Switching the min and max should not matter since the (x, λ) are picked jointly? Your intuition is good for many (most?) economic problems, but can fail. Let's start with a fail to motivate why we need to be careful

Problem 11.7. Show $P \neq D$ for the following function.

$$\mathcal{L}(x, \lambda) = \begin{cases} 0 & \{x = 0, \lambda = 0\} \cup \{x \neq 0, \lambda \neq 0\} \\ 1 & \{x = 0, \lambda \neq 0\} \cup \{x \neq 0, \lambda = 0\} \end{cases}$$

Now we want to consider the way in which $P \neq D$, and when $P = D$, and why you would ever care. First, we may establish the following

$$P = \max_x \min_{\lambda \geq 0} \mathcal{L}(x, \lambda) = \min_{\lambda \geq 0} \mathcal{L}(x^P, \lambda) \leq \mathcal{L}(x^P, \lambda^D) \leq \max_x \mathcal{L}(x, \lambda^D) = \min_{\lambda \geq 0} \max_x \mathcal{L}(x, \lambda) = D$$

The result $P \leq D$ is known as weak duality, and $P = D$ is strong duality. The conditions for strong duality are, uh, strong, but simple to state. We just want a structure so that the competing optimizers don't care what order they play in. This solution (x^*, λ^*) then needs to be a saddle, defined as

$$\mathcal{L}(x, \lambda^*) \leq \mathcal{L}(x^*, \lambda^*) \leq \mathcal{L}(x^*, \lambda)$$

Intuitively, at (x^*, λ^*), neither the minimizer nor the maximizer will want to take the solution elsewhere.

Theorem 11.6. Suppose $\mathcal{L}$ has a saddle point. Then strong duality holds.

Problem 11.8. Prove the above theorem

Okay, so now we have like 2 or 3 ways to think about solving constrained optimization problems. Why should you care? A few reasons:

(i) Sometimes it will be easier to approach the dual than the primal problem, but we want to be sure that the solutions actually match. This may apply theoretically, or computationally (see below).

(ii) Sometimes two problems of interest come to us in the form of the primal and dual, and we can show they lead to the same outcome.

Example 11.3. Consider a consumer with a budget b for buying goods. Their problem may be stated as

$$\begin{aligned} & \max_x u(x) \\ \text{such that} \quad & c(x) \leq b \end{aligned}$$

The relevant Lagrangian is

$$\mathcal{L}(x, \lambda) = u(x) + \lambda(b - c(x))$$

Suppose $\mathcal{L}$ satisfies strong duality. To tackle this problem computationally, we need to answer a few questions

1. Given any x, how hard is it to compute $\lambda^*(x)$ for the minimizer?

2. Given any λ, how hard is it to compute $x^*(\lambda)$ for the maximizer?

The comparison above will help us decided whether to use the primal or dual approach to solve the problem. Suppose first that computing $\lambda^*(x)$ is super easy, but $x^*(\lambda)$ is super hard. Then the primal problem is likely a better choice, because we may note

$$P = \max_x \min_{\lambda \geq 0} \mathcal{L}(x, \lambda) = \max_x \mathcal{L}(x, \lambda^*(x))$$

where the second equality is from strong duality. Then we can search over the space x lives in to find $x^*(\lambda)$, but we will only need to do this once, since along the way we will maintain $\lambda^*(x)$, which is easy to compute for any x we test.

Now instead consider that $\lambda^*(x)$ is a real chore, but that $x^*(\lambda)$ is piece of cake. Then dual is better suited, since just search over the space λ lives in, maintaining $x^*(\lambda)$ along the way.

$$D = \min_{\lambda \geq 0} \max_x \mathcal{L}(x, \lambda) = \min_{\lambda \geq 0} \mathcal{L}(x^*(\lambda), \lambda)$$

Problem 11.9. The following problem draws heavily from (Fajgelbaum and Schaal, 2020), a beautiful paper.

Consider a world that consists of a network of locations, any of which may be linked or unlinked. At each location L_j agents reside, and have utility for consumption U. There is also production technology at each location that produces F_j goods. The interesting aspect is that we may ship goods between location in the network, but as we ship goods, some goods are lost. In particular, in order for 1 good to be directly shipped between connected locations k and l, $\tau_{kl}(Q_{kl}, I_{kl})$ goods must be shipped, where the "iceberg cost" τ depends on the flows through the link, Q_{kl}, and the investment in the transport network on the link.

We consider the social planner that faces a finite budget κ to finance the investment over all links. The planner also places Pareto weights ω_j on the utility of the agents at each location (agents cannot move). The planner problem is then

$$\max_{C,Q,I} \sum_j \omega_j L_j U(C_j)$$

$$F_j + \sum_k Q_{kj} = C_j + \sum_l Q_{jl}\tau_{jl}$$

$$\sum_{k,l} \delta_{kl} I_{kl} \leq \kappa$$

The objective simply weights the utilities across locations. The second constraint is the flow constraint, which says that the goods produced at any location, plus those shipped in, must equal the goods consumed, plus those shipped out (if this did not hold, either there would be wasted goods or more goods consumed than produced, so we rule these out immediately). The last constraint is the budget of the planner.

We will tackle this problem slowly, and hopefully it will help emphasize the potential power of duality.

(i) Suppose there are n locations. What are the dimensionalities of C, Q, and I?

(ii) How many Lagrange multipliers will we have?

(iii) Set up the Lagrangian. Use P_j for the multiplier on flow constraint j and P_κ for the budget constraint.

(iv) Find the FONCs.

(v) Specialize to

$$\tau_{kl}(Q_{kl}, I_{kl}) = 1 + \bar{t}_{kl}\frac{Q_{kl}^{\beta}}{I_{kl}^{\gamma}}$$
$$U(C_j) = C_j^{\alpha}$$

Solve for C_j and Q_{kl} as functions of the P_j and I_{kl}. (Note: Q_{kl} must not be negative)

(vi) Solve for I_{kl} as a function of the P_j, Q_{kl} and P_κ.

(vii) Solve for I_{kl} as a function of the P_j and P_κ alone.

(viii) In the language from above, is it easier to find $x^*(\lambda)$, or $\lambda^*(x)$?

(ix) Suggest a way to search for an equilibrium, given that you have a computer which can solve a non-linear system of equations.

(x) I skirted past the details of actually checking that strong duality holds, but it does for the given functional forms. But notice something odd: For any given link, holding Q_{kl} fixed, there are increasing returns to investment I_{kl}. So it should be the case that the problem is not properly concave, right? Explain what force mitigate the increasing returns to investment, mathematically and economically. Can you conjecture a condition on the underlying parameters that will make strong duality hold?

12 Conclusion

Now you should be all ready for your first year. Based on my recollection of the math required for the first-year courses, we (including the other instructor's concurrent material) have covered basically all the math ideas you will need to be familiar with to do the first year. To proactively address some questions:

- **Are you saying all the math needed for the first year is contained in the math camp notes?**

  No. First, I err, and therefore I am sure I have missed ideas, or at least not covered them as thoroughly as needed for the whole Core. Second, some math will be learned in the courses as you go, and we have omitted some of those parts. The hope is that we have given/reviewed the foundation so that the new stuff can be learned more easily.

- **Will we use all the math in these notes in the first year?**

  Nope. I never saw Fréchet or Gateaux mentioned, and in fact the majority of the functional generalizations were (intentional) overkill. I have no doubt that some parts of these notes were needless generalization, but my goal was to help clarify or re-frame thinking on the topics, so hopefully even the non-applications stretched your brain in a good way.

- **Returning to a topic in these notes later, I found something unclear, or maybe even wrong...?**

  Oh dang. Please let me know so I can look into it and fix it!

Have a great first year! I'll be around and I try to be friendly, so don't be a stranger!

(If you stumbled onto these notes and you were not a member of the Autumn 2022 classes taking Econ Math Camp at the University of Chicago, I hope the notes were helpful. If not... well at least you got them for free and therefore free disposal applies.)

These appendices are additional topics or applications that are outside of what I considered to be the necessary math for the Core. However, my hope is that these extensions increase your ability to problem-solve and see the economics.

A Substitution and Income Effects

We start here with a classic problem, but one which was maybe only graphically discussed in previous courses. We start by reviewing the graphical explanations, then actually check the math.

Setup: An agent has income y, utility function $u(c_1, c_2) = \alpha \ln c_1 + (1 - \alpha) \ln c_2$, and faces a price of 1 for c_1, and a price of p_2 for c_2.

(1) Sketching

- (i) Draw the budget constraint (call in B_1) in (c_1, c_2) space, and sketch the indifference curve intersecting the optimal choice.
- (ii) Sketch the effect of a decrease in p_2 (say to $\hat{p}_2$) on the budget constraint (call it B_2), and sketch the new curve which intersects the optimal choice.
- (iii) Hicksian
  - (a) Sketch the budget curve which has a slope equal to that of B_2, but is tangent to the optimal indifference curve under B_1. Call it B_H.
  - (b) (SE) On the axes, show the effect on the optimal goods choice of moving from B_1 to B_H.
  - (c) (IE) On the axes, show the effect on the optimal goods choice of moving from B_H to B_2.
- (iv) Slutsky (You may want to do this on a different diagram than the Hicksian one, just for organization)
  - (a) Sketch the budget constraint which has slope equal to that of B_2, but intersects the optimal choice under B_1. Call it B_S.
  - (b) (SE) On the axes, show the effect on the optimal goods choice of moving from B_1 to B_S.
  - (c) (IE) On the axes, show the effect on the optimal goods choice of moving from B_S to B_2.

(2) Math

- (i) Find the optimal choice under B_1.
- (ii) Find the optimal choice under B_2.
- (iii) Hicksian
  - (a) Find the optimal goods choice under B_H
  - (b) Calculate the substitution effect (the change in goods choice when moving from B_1 to B_H).
  - (c) Calculate the income effect (the change in goods choice when moving from B_H to B_2)
- (iv) Slutsky
  - (a) Find the optimal goods choice under B_S

(b) Calculate the substitution effect (the change in goods choice when moving from B_1 to B_S).

(c) Calculate the income effect (the change in goods choice when moving from B_S to B_2)

(3) Basic Analysis

(i) Compare the substitution and income effects of the Hicksian and Slutsky decompositions

(ii) What is the income elasticity of both goods?

(iii) What is the price elasticity of demand for good i with respect to price j ($i, j \in \{1, 2\}$)?

(iv) Are the goods substitutes or complements? Answer in terms of both net and gross substitutability.

(v) Evaluate the following statement: "The Hicksian decomposition shows the true substitution effect, whereas the Slutsky decompsition cheats by changing the choice set".

(vi) Which decomposition seems empirically easier to work with?

(vii) Classify the goods as normal, inferior, or Giffen inferior.

(4) Policy Analysis: Reinterpret the decrease in p_2 as a subsidy to purchasing c_2

(a) If c_2 is a "green" good and your platform is environmentally focused was this a good policy?

(b) If c_2 is green, would taxing c_2 be better?

(c) If c_1 is a "green" good and your platform is environmentally focused was this a good policy?

(d) If c_1 is green, would taxing c_2 be better?

(e) What can you conclude about the efficacy of policies that only allowing taxing/subsidizing one good?

(f) Consider if the government runs a balanced budget, so that the tax/subsidy is offset by a lump-sum transfer/tax T.

i. What will be the goods choice made under this policy?

ii. Will utility be higher or lower at this point than on the original choice under B_1?

iii. What can you conclude about a social planner's ability to correct "inefficiencies" in this context?

(5) Check: What theorems or ideas did we use to analyze this problem?

(6) Challenge: How do the above answers change if we move away from utility having a log specification? What was the log assumption imposing on preferences?

B CES Magic

This section inspired by these notes (but no peeking until you work through the following problem yourself!), which contain references to papers that originate some of these ideas.

The problem that faces us is that it is not immediately obvious how to formulate a demand system (or market more generally) for many firms, but also maintain tractability. The Cournot example from earlier seems like an obvious candidate, but the "frictions" in that case come from strategic interaction among firms; the products themselves remain identical, and there is one price. We would like a setup where each firm can charge a different price, and lose some, but not all, of their market share. The following is the standard way of accomplishing this task, and indeed almost any modern paper using a New Keynesian model probably has this structure hidden somewhere in the model[31].

Consumer preferences drive basically all the results we will find. The budget constraint is standard, except that now we have a continuum of goods, each possibly with its own price, and we let I stand for household income, which is comprised of labor income and lump-sum profits.

$$U \equiv Q = \left(\int_0^1 q(\omega)^{\frac{\sigma-1}{\sigma}} \mathrm{d}\omega\right)^{\frac{\sigma}{\sigma-1}} \tag{Preferences}$$

$$LW + \Pi = I = \int_0^1 p(\omega) q(\omega) \mathrm{d}\omega \tag{Budget}$$

(i) Set up the constrained optimization problem. How many choice variables are there? How many constraints?

(ii) Take the first order conditions.[32] (Hint: Ignore the $LW + \Pi$ part, and just use the I term. The $LW + \Pi$ will be relevant later.)

(iii) Rewrite the $q(\omega)$ FOC so that it only depends on $p(\omega)$, $q(\omega)$, Q, and λ. This will involve noting that an integral in your original FOC can be written as Q to some power.

(iv) Write the demand relationship between two arbitrary ω and ω'. This should be in the form $q(\omega) = \ldots$, where the right side depends on $p(\omega)$, $p(\omega')$, and $q(\omega')$.

(v) Multiply both sides by $p(\omega)$, and integrate over ω. You should be left with an expression such that the left side is equal to I (from the budget constraint), and the right side depends on $p(\omega')$, $q(\omega')$, and some integral over prices.

(vi) Rearrange so that you have $q(\omega')$ in terms of I, $p(\omega')$, and a price index $P \equiv \left(\int p(\omega)^{1-\sigma} \mathrm{d}\omega\right)^{\frac{1}{1-\sigma}}$.

(vii) Calculate the price elasticity of demand between any good and price ($\epsilon_{\omega,\omega'} \equiv \frac{\mathrm{d} \ln q(\omega)}{\mathrm{d} \ln p(\omega')} = \frac{\mathrm{d}q(\omega)}{\mathrm{d}p(\omega')} \frac{p(\omega')}{q(\omega)}$):

(viii) As $\sigma \to \infty$, what is the economic interpretation, and what happens to the demand system?

[31] We may soon see it arising less on the goods side and more from a "labor union" channel in newer models.

[32] You may protest that attacking this problem with a Lagrangian is sloppy, since we need to consider measure-theoretic concerns. I'm sympathetic, but for now rest-assured that a naive attack will work for us.

(ix) As $\sigma \to 1$, what is the economic interpretation, and what happens to the demand system?

(x) If $\sigma < 1$, what is the economic issue that breaks things?

(xi) Suppose all firms' have production function $q(\omega) = L(\omega)$, where $L(\omega)$ is labor hired, and the wage is W.

  (a) What is firm ω's profit function and maximization problem?
  (b) What is the optimal price (in terms of W)?
  (c) What is the optimal quantity for ω (in terms of W, P, I)?
  (d) What are the profits per unit?
  (e) Suppose that all $\omega' \neq \omega$ are already producing their optimal quantities. What is the optimal quantity (in terms of L, W, Π)?
  (f) What happens to profits as $\sigma \to \infty$?

(xii) What happens to the results if we change the exogenous level of labor, L, available? What happens to the results if we change the wage level, W?

(xiii) Do you think this is a good way to model monopolistic competition, and generate profits? What are the advantages and disadvantages?

(xiv) The difference between the goods price and wage is often referred to as a "wedge", because it is inefficient, in some sense. What tax/subsidy policy would entirely neutralize the wedge?

(xv) Does the policy improve welfare? Does it matter how you finance it?

C Firm Size and Supply Structure

Many (most?) economic models model production as something bare bones simple, such as $Y = K^{\alpha}L^{1-\alpha}$. Why do we do this? The obvious answer is mathematical simplification, but that answer is not particularly satisfying, especially when we consider how far we have evolved in terms of modeling the demand side[33], so why has the supply side not progressed similarly?

Of course I am being somewhat facetious, and the answer is that it *has* progressed, but in fact this is really only true for the past decade or so. Let's dive into a few influential papers to see the progression. We first deal with the question of firm size and "washing out" of shocks, then move to the structure of the supply-side of the economy. In all cases, I am simplifying the original results.

(1) Lucas (allegedly): This part of the problem is concerned with the idea that idiosyncratic shocks get "washed out" due to a law of large numbers argument. It is generally attributed to Lucas, though it is unclear whether he ever actually put forward this idea (or believed it), or if someone else's interpretation of one of his papers led to this idea.

Suppose there are N firms in the economy, where firm i produces $S_{i,t}$ in year t. Firms receive idiosyncratic shocks such that their individual growth rate is

$$\frac{S_{i,t+1} - S_{i,t}}{S_{i,t}} = \frac{\Delta S_{i,t+1}}{S_{i,t}} = \sigma_i \epsilon_{i,t+1}$$
$$\epsilon_{i,t+1} \sim N(0,1)$$

(a) What is the mean growth rate for firm i? What is the standard deviation for firm i (these are easy, just checking that you understand the setup)?

(b) Define the aggregate output as $Y_t \equiv \sum_{i=1}^{n} S_{i,t}$. What is the aggregate growth rate? Express it in terms of the $\sigma_i, S_{i,t}, Y_t$, and $\epsilon_{i,t+1}$.

(c) What is the mean aggregate growth rate? What is the standard deviation (call it σ_Y)?

(d) Assume $\sigma_i \equiv \sigma$ is constant across i, and $\frac{S_{i,0}}{Y_0} = \frac{1}{N}$. What is σ_Y, for growth from time 0 to 1?

(e) What is the economic interpretation of the assumptions in the previous part?

(f) One (outdated but whatever) estimate of the number of firms in the U.S. is just shy of 6 million. In this model, should the U.S. have business cycles? Why?

(g) If your above answer does not match what we empirically observe, can you identify what assumption(s) might be causing the issue(s)?

(2) (Gabaix, 2011) puts forward the idea that the above "washing out" argument hinges critically on an implicit assumption that the firm size distribution is sufficiently thin-tailed. Consider the same setup as above, except now $\frac{S_{i,0}}{Y_0} = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{N+1}} \frac{1}{N^i}$. I am not even sure if this follows Gabaix's power law argument exactly, but it is close, and works to see the economics.

(a) Find σ_Y.

(b) Can the U.S. have business cycles now?

[33] Just Google "heterogeneous agents" or "non-homethetic preferences" to get a slew of results.

(c) What was the crucial change (economically) that led to this result? For concreteness, consider an economy with only chains of Walmarts and McDonald's, and all other firms are local shops. What would we need to assume about the firm shocks in each case, and which setup is more plausible?

(d) Consider if σ_i are again heterogeneous. How do you expect σ_Y to relate to σ_i, in the above case and this case?

(e) What do you conclude about how aggregate fluctuations depend on the firm size distribution (this is the "granular origins" hypothesis)?

(3) (Hulten, 1978) is a paper that answered the queries up top with a flavor of "it doesn't matter". Arguably, the influence of this paper led to widespread ignorance of the problem, not because it was boring or hard (or unimportant), but because we believed that the "big questions" were not affected by the supply structure.

Let $F(Y(t), J(t), t) = 0$ denote the production possibilities set, where $J(t)$ a set of inputs (m-dimensional) and $Y(t)$ is a set of outputs (n-dimensional). For example, in the above appendix, $F(Y(t), J(t), t) = F_{CES}(Y, L, t) = Y - L$.

(a) Assume F is homogeneous of degree zero in (Y, J). What does this imply about $D_{(Y,J)}F(Y, J, t) \circ (Y, J)$, for any Y?

(b) (First-Order) Take the derivative of $F(Y(t), J(t), t)$ with respect to time, and set it equal to 0, since $F = 0$ always. (Note that you will need to account for t's effect on Y and J, as well as the direct partial derivative of F with respect to t).

(c) Divide the answer from the second part of this question by the answer from the first part (in a neat way) so that the new expression can be understood as a sum of weighted sums and $\frac{\partial F}{\partial t} \equiv \dot{F}$

(d) (No Wedges) Define

$$p_i \equiv -\frac{\frac{\partial F}{\partial Y_i}}{\frac{\partial F}{\partial Y_1}}$$

$$w_j \equiv \frac{\frac{\partial F}{\partial J_j}}{\frac{\partial F}{\partial Y_1}}$$

Rewrite your previous expression using these terms, so that you have an expression for $T \equiv \frac{\dot{F}}{\sum \frac{\partial F}{\partial Y_i} Y_i}$

(e) What can we say about what is needed to measure TFP, empirically?

(f) Consider the following statement: "The supply chain in the U.S. is too interconnected, and it is hurting TFP. If we could get industries less connected, while preserving the same output and input shares, TFP would grow much more quickly." Is this statement supported or refuted by this model?

(g) What potential issues can you see by always assuming what was needed to get this result?

(4) (Acemoglu et al., 2012) approaches the issue slightly differently from (Gabaix, 2011), and instead assumes that the "size" differences are due to the network structure of production. This

provides a sense of microfoundations for the size result, and allows firm interactions to matter for shock propagation (this is the "network origins" hypothesis). Though this paper was certainly not the first to take the production network structure seriously[34], it is probably the paper that kicked off the current production networks literature (one lesson here might be that writing a paper really clearly and cleanly can have dividends). Also see (Carvalho and Tahbaz-Salehi, 2019) for excellent exposition (and if you get stuck on this problem).

Households are endowed with one unit of labor, and preferences are

$$u(c) = \sum_{i=1}^{n} \beta_i \ln(c_i)$$

The production structure can be summarized by

$$x_i = A z_i \ell_i^{\alpha} \prod_{j=1}^{n} x_{ij}^{w_{ij}}$$

where x_{ij} is the amount of output j used in production of output i. We assume $\sum_{j=1}^{n} w_{ij} = (1-\alpha)$ $(w_{ij} \geq 0)$[35] and $\sum_{i=1}^{n} \beta_i = 1$, so production is CRS, and preferences are homogeneous of degree 1. The material constraints are that total labor is unitary, and the sum of consumption goods and intermediate inputs from i must equal output from i:

$$1 = \sum_{i=1}^{n} \ell_i$$

$$x_i = \sum_{j=1}^{n} x_{ji} + c_i$$

(a) Solve the household's problem, given prices p_i and wage h. (This should be starting to feel more natural and familiar.)

(b) What are the first-order conditions for the firms, given prices p_i and wage h?

(c) Take logs of the production function, and plug in your FOCs from the firm. Once you simplify (a lot) you should find an affine matrix equation in the vector $\hat{p} = \ln(\frac{p_i}{h})$. You can choose A to be whatever you want, in particular you may choose it to help with cancelling some constants, and you will want to use the notation $\epsilon_i \equiv \log(z_i)$. (This will be hard, but give it a try!)

(d) Multiply the material constraint for x_i by p_i, then plug in the FOCs from the firm *and* household. Divide by h.

(e) You should now have a matrix expression in terms of $\lambda_i = \frac{p_i x_i}{h}$. Solve it. Can you interpret this?

[34] The first might have been Leontief himself!

[35] This will imply that the spectral radius of W is within the unit circle.

(f) Combine the matrix expressions to get a solution for $\ln(x_i)$. Your answer should be in the form of a matrix times the ϵ, plus a constant vector.

(g) Return to the pricing equation from part (c), and rearrange so that $\log(h)$ is the inner product (dot product) of two vectors, one being ϵ (You will need to choose the numeraire cleverly to do some cancelling).

For the remainder of the question, assume $\beta_i = \frac{1}{n}$.

(h) Consider W such that every entry is $\frac{1-\alpha}{n}$.

i. How does a shock from an arbitrary firm's z_i affect the other firms?

ii. Can you explain what this network is, economically?

(i) Consider W such that $w_{i,1} = 1 - \alpha$ for all i.

i. How does a shock from an arbitrary firm's z_i affect the other firms? (Explain the propagation mechanism)

ii. Can you explain what this network is, economically?

(j) Consider W such that $w_{11} = 1 - \alpha$ and $w_{i+1,i} = 1 - \alpha$ for $i < n$.

i. How does a shock from an arbitrary firm's z_i affect the other firms? (Explain the propagation mechanism)

ii. Can you explain what this network is, economically?

(k) Consider W such that $w_{1,n} = 1 - \alpha$ and $w_{i+1,i} = 1 - \alpha$ for $i < n$.

i. How does a shock from an arbitrary firm's z_i affect the other firms? (Explain the propagation mechanism)

ii. Can you explain what this network is, economically?

(l) Consider the following W.

$$W = (1 - \alpha) \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

i. How does a shock from an arbitrary firm's z_i affect the other firms? (Explain the propagation mechanism)

ii. Can you explain what this network is, economically?

(m) Recall Hulten's theorem.

i. Are there any wedges in this model?

ii. Do you think the "first-order" caveat of Hulten's theorem will cause problems?

(5) Believe it or not, this problem has barely scratched the surface of this area of the literature. The good news is that lots of the network results are fairly portable. If you want to learn more, I suggest looking at the work by Baqaee and Farhi.

References

Acemoglu, Daron et al. (2012). "The Network Origins of Aggregate Fluctuations". In: *Econometrica* 80.5, pp. 1977–2016.

Allen, Treb and Costas Arkolakis (2014). "Trade and the Topography of the Spatial Economy". In: *Quarterly Journal of Economics* 4, pp. 1085–1139.

Carvalho, Vasco M. and Alireza Tahbaz-Salehi (2019). "Production Networks: A Primer". In: *Annual Reviews of Economics* 11, pp. 635–663.

Evans, Lawrence C. (2010). *Partial Differential Equations, Second Edition.* United States of America: American Mathematical Society.

Fajgelbaum, Pablo D and Edouard Schaal (2020). "Optimal Transport Networks in Spatial Equilibrium". In: *Econometrica* 88.4, pp. 1411–1452.

Gabaix, Xavier (2011). "The granular origins of aggregate fluctuations". In: *Econometrica* 79.3, pp. 733–772.

Hubbard, John and Barbara Hubbard (2009). *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach, 4th edition.* Ithaca, NY: Matrix Editions.

Hulten, Charles R. (1978). "Growth Accounting with Intermediate Inputs". In: *Review of Economic Studies* 45.3, pp. 511–518.

Luenberger, David G. (1969). *Optimization by Vector Space Methods.* New York, NY: John Wiley and Sons, Inc.

Royden, Halsey and Patrick Fitzpatrick (2010). *Real Analysis, 4th edition.* Boston: Prentice Hall.

Rudin, Walter (1976). *Principles of Mathematical Analysis.* New York: McGraw-Hill.

Spivak, Michael (1965). *Calculus on manifolds.* New York: W.A. Benjamin.

Strogatz, Steven H. (2018). *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry and Engineering.* Boca Raton, FL: CRC Press.