Math Camp Part III

Matthew J. H. Murphy[1]
mjhm@sas.upenn.edu

Last updated: August 15, 2024

[1]University of Pennsylvania

Contents

Chapter 1

Introduction

This is my first year preparing these notes. As a result, there will be errors. Please let me know if you find any. There may be a prize for the student who finds the most errors (measured by my whims). Also, please let me know what material could use more attention, whether in terms of time, delivery, or exercises.

Many of the problems that are used in these notes come from problem sets that I myself wrote. Solutions are being updated as issues arise: please let me know if you find any errors. Other exercises are interesting problem that I have taken from previous years' exams, or from previous years' exercises.

1.1 Main Topics and References

Here is a (partial) list of all of the professors and textbooks that have influenced the construction of these notes. I am thankful for their instruction as well as their exercises.

- Edward Bierstone: some of the optimization theory.
  - ▷ Michael Spivak, *Calculus on Manifolds*
  - ▷ Gerald B. Folland, *Advanced Calculus*
- Jonathan Korman: most of the optimization theory.
  - ▷ David G. Luenberger and Yinyu Ye, *Linear and Nonlinear Programming*
- Fabio Pusateri: most of the measure theory.
  - ▷ Charles C. Pugh, *Real Mathematical Analysis*
- Gun Ho Jang: some of the probability theory section.

I am also grateful to Alberto Ramírez de Aguilar for his slides from previous years' math camps. I have used those to guide my construction of those notes, especially the probability theory section. I also used his notes on measure theory to compare to the treatment with which I am familiar. His notes informed much of the probability theory sections as well. As noted, I have also drawn from his collection of exercises.

1.2 Things left to do

examples and exercises...

- i think that most of the optimization theory section is actually done, except for the examples and exercises that i want to include.
  - ▷ this is high priority.
- probability theory, inequalities, etc. especially jensen's inequality is going to be an important thing to talk about. apart from that i think that using alberto's notes is going to be sufficient for us. (be cognizant of time)
  - ▷ i need to think about what i want to do in terms of convergence results and the inequalities
  - ▷ am already at a lot of pages, not sure how much more I need to do

1.3 Other Points

Throughout, ZF+C is used. Although, unless I have missed something, the Axiom of Choice is only used briefly in the Measure Theory section.

There are several sections that I might change in future years. Especially, I am not sure if the functional optimization section is as useful as I make it out to be. If, at the end of the year, you find that it would have been better to cover linear programming or dynamic programming, please let me know.

Finally, if you don't enjoy the math, at least enjoy the music.

Chapter 2

Optimization

As economists, it is important to ask why we care about math. Our goal is in general to push our understanding of the ways that humans behave. We apply this understanding to many varied settings: how people make choices about education and healthcare, how people choose where to live, and how policies affect these choices, etc.

Because we cannot perfectly control for the environment, and perfectly isolate the different mechanisms at play in human decision-making, economists use models as the framework for discussing and investigating the trade-offs at play in peoples' decisions.

What these models almost universally have in common is a posited objective for individuals. There is something that agents *want*. Sometimes we will be explicit in what this is (i.e., a firm seeks profit) and sometimes we are not explicit in this (i.e., when we talk about utility). These models also give agents choices. Let $\mathcal{X}$ be the set of choices that an individual has, and let $x \in \mathcal{X}$ be a specific choice that an individual can make. Often we will say that $\mathcal{X} \subseteq \mathbb{R}^n$, which is a given parameterization of the choice set (e.g., years of schooling).

Just as we parameterize the choices, we also parameterize the objective of agents by some function that looks like $f : \mathcal{X} \to \mathbb{R}$. Again, sometimes this will be natural (choices about production will lead to given profits) and sometimes this parameterization will not be natural (as in the case of individual utility). By giving this parameterization to agents, we are saying that an agent's goal is to solve the following problem:

$$\max_{x \in \mathcal{X}} f(x)$$

What does this mean? For a choice x to maximize f among all choices that are available to an agent, we need the following to be true:

$$f(x) \geq f(y) \quad \text{for all} \quad y \in \mathcal{X}$$

For the most part, we are not that interested in the value of $f(x)$ itself, even for this maximizer x. If our interest is predominantly in the choices that an individual makes, then what we are most interested is

$$\mathcal{X}^* := \text{argmax}_{x \in \mathcal{X}} f(x) = \{x \in \mathcal{X} | f(x) \geq f(y) \text{ for all } y \in \mathcal{X}\}$$

This is the set of *all* maximizers of f on the set $\mathcal{X}$.

Natural questions arise here:

- Do we have $\mathcal{X}^* \neq \emptyset$?
- Do we have $|\mathcal{X}^*| = 1$?

- If $|\mathcal{X}^*| > 1$, then what is the value of $|\mathcal{X}^*|$?

The purpose of this chapter is to develop tools to solve the problem above. We will also investigate properties that will ensure that we have maximizers, and also that we have unique maximizers.

2.1 Unconstrained Optimization

Charlotte Cardin, *99 Nights*

In this section we look at the problem in as much generality as we can. Here, we assume that $f : \mathbb{R}^n \to \mathbb{R}$, so that $\mathcal{X} = \mathbb{R}^n$. As a first observation, it is obvious that we need to place some conditions on our functions in order to have something intelligent to say in general. If our parameterization f is poorly behaved, then we will not be able to develop general tools.

Assumption 2.1.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be our objective function. We (unless noted) assume that $f \in \mathcal{C}^2$. That is, f is continuous and twice continuously differentiable.

Example 2.1.1. Consider the following functions defined on $[0, 1]$:

$$\begin{aligned} f(x) &:= \mathbb{1}\{x \in \mathbb{Q}\} \\ g(x) &:= x(1-x) + \mathbb{1}\{x = 0.75\} \\ h(x) &:= -\left|x - \frac{1}{2}\right| \end{aligned}$$

The first is just a poorly behaved function. The second is better behaved, but the point of discontinuity makes it so that any general analysis is impossible. The third similarly is actually a very nice function, but its lack of differentiability at $1/2$ makes to impossible to make use of differentiation.

This will ensure that we are able to use the technical tools of multivariate calculus. Now, we need to introduce some definitions. There are two different types of maxima which will be of interest to us:

Definition 2.1.2. A *global maximum* of f is an $x \in \mathbb{R}^n$ such that $f(x) \geq f(y)$ for all $y \in \mathbb{R}^n$.

This is the definition that we used above when we were motivating what our goal was. In general, this is the object in which we are interested. We care about what choice does the *best* for an agent, and so we care about all of their possible choices. At the same time, the tools that we are going to use are predominantly going to be the tools of calculus. These tools are inherently *local*. They tell us about what is happening near a certain function, not what is happening globally.[1]

Definition 2.1.3. A *local maximum* of f is an $x \in \mathbb{R}^n$ such that there exists a neighbourhood U of x such that $f(x) \geq f(y)$ for all $y \in \mathbb{R}^n$.

[1]This is the beauty of complex analysis, where the local behaviour of a function perfectly determines what happens elsewhere. This is an entirely unrelated point, and evidences that I wrote this section first.

Example 2.1.2. Adding to our list of running questions, it is natural to ask:

$$\text{Does} \quad x \in \{\text{local maxima}\} \Rightarrow x \in \{\text{global maxima}\} \quad ?$$

It can easily be shown that this is not generally true. See Figure 2.1. However, we will be able to derive some conditions for which this will be the case. I defer this question to later. For the next portion, we will deal only with local maxima.

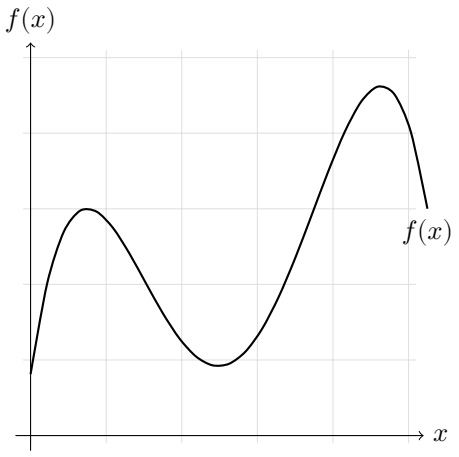


Figure 2.1: Example showing that local maxima need not be global maxima

We now present the first result of this section, which you should all be able to recognize as an immediate generalization of a first-year calculus result.

Theorem 2.1.4 (First Order Necessary Condition)**.** *Suppose that x is a local maximum of f. Then we have that $\nabla f(x) = 0$.*

Proof. This follows from a passing through to one dimension, and then an immediate application of the definition of a derivative. Suppose that we have that $\nabla f(x) \neq 0$. Then at least one entry of $\nabla f(x)$ is nonzero. Suppose without loss that it is the first entry. Also suppose without loss that the first entry is $c > 0$. Then, by the definition of a derivative, that for every small ϵ there exists a $\delta > 0$ such that $y := x + \frac{\delta}{2} e_1$ we have

$$\left| \frac{f(y) - f(x)}{||y - x||} - c \right| < \epsilon$$

By simply moving things around, we have that

$$\frac{\delta}{2}(c - \epsilon) < f(y) - f(x) < \frac{\delta}{2}(c + \epsilon)$$

So, we have that $f(y) > f(x)$ as long as we take $\epsilon < c$. As we can take δ arbitrarily small, we have that x is not the maximum of f on any of its neighbourhoods, so that it is not a local maximum.[2] If $c < 0$, then by simply considering $y \det x - \frac{\delta}{2} e_1$ the same conclusion is reached. □

[2] As a matter of form, one shouldn't use proofs by contradiction if one can make it a proof by contraposition instead. This proof would not have benefited from my assuming at the outset that x is a local maximum and then saying we have derived a contradiction at the end.

Two points:

1. We needed to have that there exists a neighbourhood of x on which f is defined. Of course, when the set under consideration is given by $\mathbb{R}^n$ (or homeomorphic to $\mathbb{R}^n$) then this is not a problem. When we move to looking at constrained optimization, then we will need to adjust this, as the statement will then only be true for any x in the interior of the constraint set.

2. This is very much only a necessary condition. It is certainly *not* sufficient. Consider everyone's favourite counterexample in x^3. It certainly has zero derivative at 0, but 0 is certainly not a local maximum.

Notice that the intuition for the proof is identical in the one-dimensional case as in the multi-dimensional case. It is natural to then ask if the second- order condition in one dimension applies can be generalized to higher dimensions. It can (and the interpretation is exactly the same), but in order to state it we need to concern ourselves with some more tedious notation.

Definition 2.1.5. Let $H \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then we say that H is *negative semi-definite* if we have, for all $x \in \mathbb{R}^n$ that $\langle x, x\rangle_H := x^T H x \leq 0$. We say that it is *negative definite* if the inequality is strict for all x. In the first case we write $H \leq 0$ and in the second we write $H < 0$.

An equivalent characterization of the above I think provides more intuition.

Lemma 2.1.6. *A symmetric matrix H is negative semi-definite if and only if all of its eigenvalues are weakly negative. A symmetric matrix H is negative definite if and only if all of its eigenvalues are strictly negative.*

Proof. As H is symmetric, it is orthogonally diagonalizable. From here the result is immediate (decompose a vector as the linear combination of orthogonal eigenvectors). □

Example 2.1.3. In order to see why this intuition is useful, we consider again a one-dimensional example. In one dimension, even if $f'(x) = 0$, if $f''(x) > 0$, then we know that we cannot have a local maximum. For example, consider the difference between x^2 and $-x^2$. So, the second derivative provides another necessary condition for a function to have a local maximum at x. However, again as x^3 tells us, there are issues if we have that $f''(x) = 0$ (making a weak inequality a necessary, but not sufficient, condition). We can summarize these two points in the following results. In higher dimensions, the only difference now is that we need to consider what happens every possible one-dimensional ray.

Theorem 2.1.7 (Second Order Necessary Condition)**.** *Suppose that x is a local maximum of f. Then it must be that $\nabla^2 f(x) \leq 0$.*

This summarizes the intuition for why zero is not a maximum for x^2. The next result shows that in some cases we can make claims about sufficiency (dealing with the example of x^3).

Theorem 2.1.8 (Second Order Sufficient Condition)**.** *Suppose that at x we have $\nabla f(x) = 0$ and $\nabla^2 f(x) < 0$. Then we have that x is a strict local maximum, where strict here means that for a neighbourhood U of x we have that $f(x) > f(y)$ for all $y \in U \setminus \{x\}$.*

Proof. Apply a Taylor expansion around x. Then we have that, in a direction v we have that, for all $s \in \mathbb{R}$

$$f(x+sv) = f(x) + s\langle \nabla f(x), v\rangle + \frac{1}{2}s^2 \cdot \langle v, v\rangle_{\nabla^2 f(x)} + o(s^2)$$

where you should know what this "little-o" notation means. Because $\nabla f(x) = 0$, the second term disappears. And, because we have that $\nabla^2 f(x) < 0$, the third term is negative. Hence, by dividing by s^2 and taking $s \to 0$, the claim is proved. □

Summary: We have defined the different types of maxima for a function f. We also looked at some necessary conditions for a point to be a local maximum, and have additionally derived one sufficient condition for a point to be a local maximum of a function.

Actually Optimizing: With these tools, how could we actually go about solving the problem that we want to solve? Well, we know two conditions that must be satisfied for a point to be a local maximum. By taking the total derivative of f and setting it equal to zero, we will reduce our problem to looking only at a relatively small number of points (typically a measure zero set, and often a finite number of points). For these points, we can then look at the hessian of f. If the hessians for any of them are negative definite then we have found a local maximum. If they are negative semi-definite, we will need to be a little bit smarter (and this is where the artistry comes in). If they have positive eigenvalues we can rule them out. This will leave us typically with a small number of points for consideration. From these, how do we get a global maximum? Simple: compare all of the local maxima, and the one with the greatest value is the global maximum.

Example 2.1.4. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ that is given by

$$f(x, y) = x^2 + 4x + 3xy - y^2$$

1. Find candidates for local maxima/minima through first-order conditions.
2. Which of these candidates satisfy the second-order necessary (sufficient) conditions?
3. Are there global maxima and minima?

Solution.

1. We have that

   $$\nabla f(x, y) = \begin{pmatrix} 2x + 4 + 3y \\ 3x - 2y \end{pmatrix}$$

   Only solutions to this system of equation are possible candidates. We must have that $y = \frac{3}{2}x$, so that $x = -\frac{8}{13}$ and $y = -\frac{12}{13}$ is the only candidate.

2. We check the value of the Hessian here.

   $$\nabla^2 f(x, y) = \begin{pmatrix} 2 & 3 \\ 3 & -2 \end{pmatrix}$$

   It can be verified that the eigenvalues of this matrix are $\sqrt{13}$ and $-\sqrt{13}$, so that this point is neither a maximum nor a minimum.

3. No, clearly not. Consider $(n, 0)$ and $(0, n)$ as $n \to \infty$.

□

2.2 Equality Constraints

Half Moon Run, *Salt*

The above section gives us the machinery that we are going to need even when we look at more complicated settings. In general, all that we are going to change is to restrict the directional vector v to be "feasible." We are again considering a function f defined on $\mathbb{R}^n$, but now restrict ourselves to a subset of $\mathbb{R}^n$ and look for local maxima on this set. What is the interpretation? There is some (differentiable) manifold of interest, and we only care about what happens to f on this subset. Suppose that f converts different activities into some kind of output (e.g., money). Then a common example is going to be restricting ourselves to looking at a fixed amount of time spent on each of the activities ($x_1 + ...x_n = h$ for some fixed h). Most of what we care about in economics is this "constrained" optimization: reduce computing power restricted to some amount of accuracy, maximize welfare subject to budget constraints, etc.

I am guessing that at some point during all of your previous education you have seen a Lagrangian before. It is a powerful tool of optimization, and will be extremely important in basically every optimization problem that you do throughout first year.

That being said, I think that very seldom is the underlying theory discussed in detail. So, I want to start with the theory, and then progress into some examples of how the theory can be applied.

Theorem 2.2.1. *Let $U \subseteq \mathbb{R}^n$ be open, and $g : U \to \mathbb{R}^p$ be a $\mathcal{C}^1$ function: $g = (g_1, ..., g_p)$. Similarly, let $f : U \to \mathbb{R}$ be differentiable. Suppose that f has a local extreme value on $g^{-1}(0)$ at a point a, where $Dg(a)$ has rank p (i.e., the inverse is smooth at a). Then, there exist $\lambda_1, ..., \lambda_p \in \mathbb{R}$ such that $\nabla f(a) = \sum_{i=1}^{p} \lambda_i \nabla g_i(a)$.*

Notice first that this gives us a problem in $n + p$ unknowns (the a and the λ_i's) and $n+p$ equations (the ∇_i's and the g's). It turns out that from here we just need to apply the implicit function theorem (which you should have seen), and then some basic linear algebra (which you should have seen).

Proof. Now, because we have that the rank of Dg is given by p, there exist $i_1, ..., i_p$ such that

$$\det \frac{\partial g}{\partial (x_{i_1}, ..., x_{i_p})} \neq 0$$

without loss, assume that it is the last p. We now write $x = (u, v) = (u_1, ..., u_{n-p}, v_1, ..., v_p)$.

Hence, by the implicit function theorem, we can solve $g(u, v)$ for $v = h(u)$ in a neighbourhood of a, where h is a $\mathcal{C}^1$ function. The function $\partial(u) = f(u, h(u))$ has an extreme value at $(a_1, ..., a_{n-p})$.

We have that $D\varphi(u) = Df(u, h(u))D(u, h(u))$, where this last matrix is of dimension $n \times n - p$. We get that

$$\nabla \varphi(u) = \nabla f(u, h(u)) \cdot \begin{pmatrix} I \\ Dh(u) \end{pmatrix}$$

As the left-hand side is equal to zero at $(a_1, ..., a_{n-p})$, then we know that we have that

$$\nabla f(a, h(a)) \cdot \begin{pmatrix} I \\ Dh(a) \end{pmatrix} = 0 = \nabla g(a, h(a)) \cdot \begin{pmatrix} I \\ Dh(a) \end{pmatrix}$$

This is because of how h is defined: because it is to be such that g is always 0, the derivative of g with respect to the parameterization h must be zero. Finally, we apply some basic linear algebra.

First, because of the assumption that we have Dg having rank p, we know that all of the ∇g_i are linearly independent. Now, notice that the matrix

$$\begin{pmatrix} I_{n-p} \\ Dh(a) \end{pmatrix}$$

Will have full rank, because of the identity matrix in the first $n-p$ rows. Hence, the left-kernel of this map has dimension p (as it is surjective from $\mathbb{R}^n$ to $\mathbb{R}^{n-p}$). Because all of the ∇g_i lie in this kernel, and are linearly independent, they form a basis for this kernel. We conclude. □

We call these λ the Lagrange multipliers. They have their interpretation as "Shadow Prices" on the constraints, because λ_i is the differential increase in f if the constraint g_i were relaxed differentially. Hence, the constraint g_i imposes a "cost" of λ_i on f.

Intuition: Why is this construction important? I think that it highlights what is happening when we are doing Lagrangian optimization. For each of the binding constraints, we are restricting our search to points that lie on this constraint. That reduces the problem to looking at a subspace.[3] As we move along this constraint, we will hit a zero of ∇f (because it is a local minimum where the constraint binds). However, when we back this out we have an additional term that comes from needing to bind on the constraint itself.

Example 2.2.1. Why do we assume that Dg has rank p? Consider a simple case where $g : \mathbb{R}^2 \to \mathbb{R}$, with $g(x, y) = x^2$. Then the constraint set $g^{-1}(0)$ is just the y-axis. Now consider $f(x) = x - y^2$. This has a nontrivial maximum on the constraint set at the origin, but here $\nabla f(0, 0) \neq (0, 0)^T$. The rank condition ensures that we are getting enough information from g.

Geometric Interpretation: A different (perhaps better) way to think about what is happening is interpreting the final result itself. What we are saying is that the gradient of f, which corresponds to the direction of steepest ascent,[4] moves in exactly the same direction as the constraint. Because we are holding the constraint fixed, we are interested in the tangent space, so that we are perpendicular. For f to achieve a (constrained) local maximum, then we need there to be no way to ascend without violating the constraint (draw a picture!).

[3] This subspace corresponds to the manifold generated by that constraint.

[4] see Section ??

Why is this useful? To understand this, it is probably best to go through some examples. Notice that as with many of the results of the previous section, the above provides a necessary condition, but not a sufficient condition. More than this, the above theorem doesn't tell you how to deal with multiple constraints. In general, there will be some binding constraints and some non-binding constraints. So, it requires some level of thought in order to reduce the problem into one that is manageable and we can apply the theorem's results. Notice even then that solving the system of equations isn't in general easy.

2.2.1 Examples

Example 2.2.2. Show that the maximum value of the function $f(x, y, z) = x^2y^2z^2$ on the sphere $x^2 + y^2 + z^2 = r^2$ is $r^6/27$.

Solution. This is an abstract problem, but I think it highlights a basic example of for what we need to check. The first step is to translate the problem into the language that we have used. That is, f is our objective that we are maximizing, and

$$g(x, y, z) = x^2 + y^2 + z^2 - r^2$$

our constraint function.

Because we are looking for solutions on the sphere, by the above theorem we can look for (x, y, z, λ) that are solutions to the following system of equations.

$$\begin{aligned} 2xy^2z^2 + 2\lambda x &= 0 \\ 2x^2yz^2 + 2\lambda y &= 0 \\ 2x^2y^2z + 2\lambda z &= 0 \\ x^2 + y^2 + z^2 &= r^2 \end{aligned}$$

Now we need to solve this system of equations. We first note again that the above theorem is one of necessity and not sufficiency, so we need to rule out some problems. Notice that $\lambda, x, y = 0$, and $z = r$ is going to satisfy the system, but will *not* be a maximum. Indeed, we cannot have any of the $x, y, z = 0$, for there f will be equal to zero. We can apply this to the above, and get a reduced system:

$$\begin{aligned} y^2z^2 + \lambda &= 0 \\ x^2z^2 + \lambda &= 0 \\ x^2y^2 + \lambda &= 0 \end{aligned}$$

The other possibility is that $|x| = |y| = |z|$, so that $3x^2 = r^2$, which then yields that the maximum is going to be achieved at $x^2y^2z^2 = (r^2/3)^3 = r^6/27$.

□

Example 2.2.3. Consider the function

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

and the set

$$\{(x, y) \in \mathbb{R}^2 | y \leq -2x - 1\}$$

Find the local maxima of f on the set, and also find which local maxima are global maxima.

I choose this example for several reasons. First, it highlights the difference between local and global minima. Second, it highlights that inequality constraints aren't really all that different from equality constraints, at least some of the time. Finally, it highlights a little how being smart (observant) can save you some energy, but doing things rigourously can be painful.

Solution. First, translating this into our language, we look at $g(x, y) = -2x - 1 - y$ as our constraint function, and we are interesting in the set $g(x, y) \geq 0$. I look at $-f$ and hence search for minima instead of maxima. Notice that of course the problem is identical!

Notice that this function can be extended to a C^∞ function on $\mathbb{R}^2$, so it makes sense to compute its gradient generally.

We have that

$$\frac{\partial f}{\partial x} = -4x(y - x^2)$$
$$\frac{\partial f}{\partial y} = 2(y - x^2)$$

Now, these are simultaneously zero only when $y = x^2$. On our set, we must have that $x^2 + 2x + 1 = (x + 1)^2 \geq 0$ so that $x = -1, y = 1$. What this means is that we have no interior points as being local minimums, since at these interior points we would need to have $\nabla f = 0$.

I now check to confirm that the only possible local minimum on the boundary is $(-1, 1)$. Take any point (x, y) other than this on the boundary. In particular, for this point we have $x \neq 1$. I show that in some feasible direction, f is decreasing. To confirm this, I show that

$$\left\langle \nabla f(x, y), \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\rangle < 0 \text{ or } \left\langle \nabla f(x, y), \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right\rangle < 0,$$

so that the point (x, y) violates the necessary first order condition for a minimum. These directions are exactly the vectors along the boundary.

On the boundary, we have $y = -2x - 1$. Hence $\nabla f(x, y) = \begin{pmatrix} 4x(x+1)^2 \\ -2(x+1)^2 \end{pmatrix}$ We have then that

$$\left\langle \nabla f(x, y), \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\rangle = 4(x + 1)^3 \text{ and that } \left\langle \nabla f(x, y), \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right\rangle = -4(x + 1)^3$$

As $x \neq -1$ by hypothesis, at least one of these is negative, violating the necessary first-order condition for a minimum.

I now show that $(-1, 1)$ is in fact a local (and global) minimum for this set. It is in fact obvious to see because it is the only point on the set for which $y = x^2$, and hence is the only point on the set for which $f(x, y) \leq 0$. However, I show this as well by showing that it satisfies the necessary second-order condition. We have that

$$\nabla^2 f(-1, 1) = \begin{pmatrix} 8 & 4 \\ 4 & 2 \end{pmatrix}$$

This can easily be seen to have one eigenvalue of zero (with eigenvector $(1\ -2)$) and also one eigenvalue of 10 (with eigenvector $(2\ 1)$). Hence, the Hessian is positive semi-definite so that it satisfies the necessary second-order condition.

Now, suppose that we are interested in finding the global minima on the set. We could do it in two ways, for instance.

First, the smart way: f is positive everywhere on the set except for at this point.

Second, suppose that we didn't make this observation. Then, we would need to show this in a different way. For example, we could show this by showing that along any ray, the function is strictly increasing. Why do I highlight this? If you were operating on autopilot, then it would be easy to fall into the trap of plowing ahead without thinking. Taking thirty seconds is better with these types of problems ins important. But I digress.

Consider any feasible ray emanating from $(-1, 1)$, parameterized by x or y (whichever is appropriate on the line). We can consider f as a function $\mathbb{R} \to \mathbb{R}$ on this line. I show that the derivative of f on this line in the direction of $(-1, 1)$ is negative so that $f(x, y) > f(-1, 1)$. First, on the vertical line $x = -1$, we have that $g(y) = f(-1, y)$ has $g'(y) = 2(y - 1)$. For $y < 1$, this derivative is always negative. Hence, any point (x, y) in the set lying on this vertical line has $f(x, y) > f(-1, 1)$.

Now, consider any line given by $y = cx + c + 1$. Then, for $g_c(x) = f(x, cx + c + 1)$, we have $g_c'(x) = 2(c - 2x)(c(x + 1) + 1 - x^2)$. We must show that if $c \geq -2$ and if $x < -1$ then $g_c'(x) < 0$ and if $c \leq -2$ and if $x > -1$ then $g_c'(x) > 0$. (This is not hard to visualize, as you just look at the line and what direction we need to move in)

Notice that generally we have that the derivative is zero when $x = -1$, when $x = c/2$, or when $x = 1 + c$.

Consider the first case. Then we have that $c/2, 1 + c \geq -1$, so that the derivative is only zero when $x \geq -1$. This means in particular that because the derivative is a cubic with a positive leading coefficient (2), the derivative is zero when x is smaller than this first zero, so that in particular for $x < -1$ the derivative is negative.

In the second case we can apply similar logic. As $c/2, 1 + c \leq -1$, the derivative is only zero when $x \leq -1$. This means that the derivative is positive when $x > -1$, proving the claim.

□

Example 2.2.4. Maximize $f(x, y) = x + \sqrt{y}$ subject to the constraint that $ax + by = c$. What do you notice?

Summary: Equality constraints represent the first digression from the standard optimization. Here we are restricting to a smaller subset (of dimension $n - p$) and then applying our same methods. If, in order to increase the value of the function we need to leave the subspace, then we have a maximum. This is what the Lagrangian method is capturing.

2.3 Inequality Constraints

Men I Trust, *Untourable Album*

Here we introduce in more generality the inequality constraints of the above example. The theorem above required us to be in a situation where we were looking at equality constraints. The Kuhn-Tucker theory of this section extends that to inequality constraints. I want to emphasize that there is literally zero difference between this theory and that theory, except that this gives the notation in a little more generality, and we need to be a little bit more careful about directions here.

Before I introduce the theorems themselves, I want to explain a little what I mean by the similarities between his and that theory. When we are dealing with inequality constraints, at any individual point x we are dealing with some constraints that are binding ($g_i = 0$), and some constraints that are not binding ($g_i < 0$).

Let us now think locally about the constraint set at x. Each binding constraint at x restricts us to a subspace of one lower dimension (this is just the implicit function, following exactly from the previous section). On the other hand, if a constraint is not binding, then we know that locally around x it is also not going to be binding. This means that we can in some sense ignore them when looking at the point x.

A good first guess is that we just need to iterate over the different combinations of constraints, see which are binding and which are not, then just double-check to make sure that we are on the correct side of the non-binding constraints. This naive approach is actually pretty close to what we need to do. However, there is an additional subtlety here, in that we can always move *into* one of the binding constraints being non-binding. This comes back to the direction to which our gradient is pointing. It is not now enough that it is orthogonal to our constraint set, it must also be pointing *out* of our set (draw a picture!).

These two constraints form the additional basis that we need for the general Kuhn-Tucker theory. They are the "Complementary Slackness" condition, and the "Sign" condition on the multipliers. We summarize these ideas in the below theorem. Notice that the proof follows from (1) the proof in the equality condition case and (2) the discussion above.

Theorem 2.3.1. *Suppose that we aim to maximize the function $f(x)$ on the set defined by $g(x) \leq 0$ (with $g : \mathbb{R}^n \to \mathbb{R}^p$). Then if x is a local maximum of the problem and Dg is full rank at x, there exists a $\mu \in \mathbb{R}^p$ such that*

1. $\nabla f(x) = \mu \cdot Dg$
2. *μ is non-negative.*
3. $\mu_i g_i(x) = 0$.

The first condition is exactly the same as the above theorem. The second condition is just the observation that we need ∇f to be pointing *out* of the set. The final condition is that we need to ignore constraints that aren't binding.

A note on signs. It can be difficult to remember, for me at least, what conditions on the sign are required. For example, the condition is often written differently, as instead

$$\nabla f(x) + \mu \cdot Dg = 0$$

This would then flip the condition and require that μ is non-positive. In general, it comes down to personal preference; as long as you are internally consistent and know what you are doing then however you write it is fine.

I like my notation because I think that it emphasizes the geometric interpretation. Specifically, it frames the condition as ∇f being in the cone generated by all of the ∇g_i (because the condition is explicitly writing it as a linear combination of these vectors). Because we are looking at a case when $g(x) \leq 0$, we have that ∇g_i is pointing *out* of the constraint set. Hence, we need ∇f to be in the same direction as these vectors. If it were in the opposite direction, we could increase the objective by moving into the constraint set. This is exactly why we need $\mu \geq 0$. Rewriting things in different ways (i.e., $g \geq 0$, or having them both be on the same side) comes down to whether we need f pointing in the same direction as the constraint gradients or opposite.

A note on the importance of being full-rank. That Dg is full rank is even more important now than before. If g isn't full rank, then it can be hard to determine whether what is or is not a feasible direction. Consider $g(x) = x^2$ compared to $g(x) = -x^2$.

2.3.1 Feasible Directions and Sufficient Conditions

It is useful to think about the above conditions as being slight generalizations of the conditions listed in Section 2.1. In that section, our conditions for local maxima require that we if we move along any direction, the value of the objective function decreases. The conditions in Sections 2.2 are just the same, except there we need to restrict our movement to be only along the zero-set of the constraint. In this section, the dimension of the manifold in general changes, so we need to be a little more careful about which directions to which we can move.

It is perhaps useful to think about this in terms of feasible directions, which were introduced in Example 2.2.3.

Consider a point x for which some of the constraints are binding, and some of the constraints are not binding. Then, the directions that are going to be feasible to move from x (picture) are going to be those points which exactly do not move in the directions of the gradients of the binding constraints. By "feasible" I do not mean necessarily that they are actually in the set (because if exactly tangent, then can still be outside), but that they form the boundary of all points that are in the set (picture). Explicitly, we can define

$$\mathcal{F}(x, g) \coloneqq \{z \in \mathbb{R}^n | \, \langle z, \nabla g_i \rangle \leq 0, \ \forall i \text{ s.t. } g_i(x) = 0\}$$

By the conditions above, we know that if any of these inequalities are strict then the function will be strictly decreasing in direction z. This will be a first-order effect. If we are interested in the second-order effects, then we need to consider what happens in directions where these inequalities are all equalities. I.e., we focus on the set

$$\Omega(x, g) \coloneqq \{z \in \mathbb{R}^n | \, \langle z, \nabla g_i \rangle = 0, \ \forall i \text{ s.t. } g_i(x) = 0\}$$

We have effectively isolated the directions along which the gradient of f is neither increasing nor decreasing. I.e., if we have that x satisfies the necessary condition to be a local maximum, then we will have that

$$\Omega(x, g) \subseteq \{z \in \mathbb{R}^n | \, \langle z, \nabla f \rangle = 0\}$$

It turns out that these are exactly the directions for which we need to check that there is no positive second-order effect.

First, we isolate the second-order effect. To do that, we write

$$\mathcal{L}(x, \mu) = f(x) - \langle \mu, g(x) \rangle$$

Theorem 2.3.2. *If x is a local maximum of the problem with corresponding multipliers μ', then we must have that*

$$\langle z, z \rangle_{\nabla^2_x \mathcal{L}(x, \mu')} \leq 0, \ \forall z \in \Omega(x, g)$$

Theorem 2.3.3. *If x is such that there exist a valid μ (i.e., satisfying the complementary slackness and sign conditions) such that*

1. $\nabla f(x) - \mu \cdot Dg(x) = 0$

2. $\langle z, z\rangle_{\nabla_x^2 \mathcal{L}(x,\mu')} < 0,\ \forall z \in \Omega(x, g)$

then x is a strict local maximum.

Are these conditions useful? In practice, not really, because they are so tedious to check. We need to first go through the motions of calculating the multipliers, then calculating the feasible directions (in higher dimensions this is basically infeasible), then calculating the value of the adjusted inner product. So, we would like conditions that help us to get around this in general. This remains the goal of the next section.

Summary: This is the same as before, except now we need to do a little more accounting in terms of signs (reflecting directions of relaxation). Sufficient conditions become tedious to check, but are possible to determine.

2.3.2 Examples

Example 2.3.1. Maximize $f(x, y) := x + \sqrt{y}$ subject to

$$g(x, y) := \begin{pmatrix} ax + by - c \\ -x \\ -y \end{pmatrix}$$

How does this differ from Example 2.2.4? Do we need to consider some cases here?

Example 2.3.2. Consider the problem

$$\max f(x, y) = xy$$

subject to the constraints

$$\begin{aligned} g(x, y) =& x + y \leq 1 \\ h(x, y) =& x^2 + y^2 - 1 = 0 \end{aligned}$$

(a) Which feasible points have full-rank derivative for the constraints?

(b) Find the candidate(s) for maximizer.

(c) Are the second order condition satisfied at the points you found in part (b)?

(d) Find the maximizer for this problem. Is it a global maximum (on the constraint set)?

Solution.

(a) Notice that $\nabla g = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\nabla h = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$. As $\nabla h \neq 0$, if g is inactive, every feasible point is regular. So we must have, for a point to be non-regular, that $x = y$ and g being active. The only points on the unit circle that have $x = y$ are with $x = \pm\frac{\sqrt{2}}{2}$, and g is not active at $(\pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{2}}{2})$ (for the positive point, it isn't feasible, and it isn't active at the negative point). Hence all feasible points are regular.

(b) We have, using the Kuhn-Tucker method, that we have at a local minimum

$$\begin{pmatrix} y \\ x \end{pmatrix} + 2\lambda \begin{pmatrix} x \\ y \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0$$

where $\mu(x+y-1) = 0$ and $\mu \leq 0$ (we are maximizing, and they are on the same side). When g is active, on the unit circle we have $x = 1 - y$ and $x^2 + y^2 = 1$, so that either $x = 1, y = 0$ or $x = 0, y = 1$. In either case we have $f = 0$, so these are clearly not local maximums.

If g is inactive, our situation reduces to $y = 2\lambda x$ and $x = 2\lambda y$. Hence $y = 4\lambda^2 y$, so that $\lambda = \pm 1/2$. If it is positive, then $x = y$, so that $x = y = \frac{-\sqrt{2}}{2}$. If it is negative, we have $x = -y$, so that we are at $(\pm\frac{\sqrt{2}}{2}, \mp\frac{\sqrt{2}}{2})$. These are our candidates for a maximum.

(c) In the first case $(\frac{-\sqrt{2}}{2}, \frac{-\sqrt{2}}{2})$, we want to have $\nabla^2 f + \frac{1}{2}\nabla^2 h = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is negative semi-definite (as we are maximizing) on

$$T_{(\frac{-\sqrt{2}}{2}, \frac{-\sqrt{2}}{2})} S^1 = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

As we have

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0$$

we have that the second order necessary condition is satisfied as well for this point. In the other cases, we have $\lambda = -1/2$, and the tangent space is given by

$$\operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

We have then by the same computation that

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0$$

So that these points also satisfy the second order necessary condition of being negative semi-definite.

(d) The maximizer is $\left(\frac{-\sqrt{2}}{2}, \frac{-\sqrt{2}}{2}\right)$ as the other two candidates give negative values so we can reject them, and this gives a positive value. It is the maximize on the constrained set (it is compact and this point is the only candidate, so it must be the max). It is clearly not a global maximum on $\mathbb{R}^2$ (consider the line $y = x$ as $x \to \infty$, clearly f is unbounded on this line). It is, however, the maximum on all of the circle, by symmetry (its negative is also a maximum, but not in the constraint set).

□

This examples highlights a few things. First, it gives us another explicit computation of the second order conditions. It also gives us an example of combining equality and inequality constraints.

How does the problem change if we make the equality constraint on h an inequality constraint? Do we get another local maximum? Is it a global maximum?

Example 2.3.3. Let Q be a 2×2 symmetric positive definite matrix, and let $a, x \in \mathbb{R}^2$. Assume that the first entry of a, defined by a_1, is less than zero. Consider the optimization problem

$$\min \langle (x-a), (x-a) \rangle_Q$$

subject to the constraints

$$\begin{aligned} x_1 &\leq x_2 \\ -x_1 &\geq x_2 \end{aligned}$$

Find the candidate(s) for minimizers using the First order conditions, and check whether they satisfy the second order conditions.

Solution. Now, the Kuhn-Tucker conditions are

$$2Q(x-a) + \mu_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

with $\mu_1(x_1 - x_2) = \mu_2(x_1 + x_2) = 0$ and $\mu_1, \mu_2 \geq 0$. Notice that the Hessians of both our inequality conditions are zero, as they are both linear. Hence, the second order condition for any candidate will be satisfied, as the matrix in question will be $2Q$, which is positive definite on all of $\mathbb{R}^2$, so any candidate that satisfies the first order conditions will be a local minimum.

If neither of the two inequality constraints are active, we clearly have $x = a$ as the point satisfying our first order condition. Now, we must make sure that the inequality constraints are satisfied. Hence, if we have $a_1 \leq a_2$ and $-a_1 \geq a_2$, i.e., $a_2 \in [a_1, -a_1]$ (which is possible, as $a_1 < 0$), then $x = a$ is our minimum.

Now, suppose that we have the first condition as active (that is, $x_1 = x_2$), but not the second condition. If this is the case, let $Q = \begin{pmatrix} d & b \\ b & c \end{pmatrix}$, so that our condition becomes

$$\begin{aligned} d(x_1 - a_1) + b(x_2 - a_2) + \mu_1 &= 0 \\ b(x_1 - a_1) + c(x_2 - a_2) - \mu_1 &= 0 \end{aligned}$$

This yields (after adding), as $x_1 = x_2$, we have

$$\begin{aligned} d(x_1 - a_1) + b(x_1 + x_2 - a_1 - a_2) + c(x_2 - a_2) &= 0 \\ (d + 2b + c)x_1 &= da_1 + b(a_1 + a_2) + ca_2 \\ x_1 = x_2 &= \frac{da_1 + b(a_1 + a_2) + ca_2}{d + 2b + c} \end{aligned}$$

Note that after substituting into the equation, we find that $\mu_1 = \frac{(cd - b^2)(a_1 - a_2)}{d + 2b + c}$. I claim that this is positive only when $a_1 > a_2$. This is because Q positive definite, so that $cd - b^2 > 0$. And, taking $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, we have $x^T Q x = d + 2b + c > 0$. Hence, $\mu_1 > 0$ if and only if $a_1 > a_2$. Now we need this point to be negative (so that the other condition holds). That is, we need $(d + b)a_1 + (c + b)a_2 < 0$.

If only the second condition is active, we have $x_1 = -x_2$. This yields

$$\begin{aligned} d(x_1 - a_1) + b(x_2 - a_2) + \mu_2 &= 0 \\ b(x_1 - a_1) + c(x_2 - a_2) + \mu_2 &= 0 \end{aligned}$$

This yields (after subtracting), as $x_1 = -x_2$, we have

$$\begin{aligned} d(x_1 - a_1) + b(-x_1 + x_2 + a_1 - a_2) - c(x_2 - a_2) &= 0 \\ (d - 2b + c)x_1 &= da_1 - b(a_1 - a_2) + ca_2 \\ x_1 = -x_2 &= \frac{da_1 - b(a_1 - a_2) - ca_2}{d - 2b + c} \end{aligned}$$

This yields $\mu_2 = \frac{(cd-b^2)(a_1+a_2)}{d-2b+c}$. This is positive exactly when $a_1 > -a_2$. This is as in the first case, taking $x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Hence, in each of the three possibilities ($a_2 < a_1, a_2 \in [a_1, -a_1], a_2 > -a_1$), we have exactly one candidate for a local minimum, which by the second order condition is the local minimum. We need this point to be negative (so that the other condition holds). That is, we need $(d - b)a_1 + (b - c)a_2 < 0$.

If both conditions are active, we have that the only possible point is $(0, 0)$. Here, we must have that

$$-2Qa + \mu_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

This becomes

$$\begin{pmatrix} -2(da_1 + ba_2) + \mu_1 + \mu_2 \\ -2(ba_1 + ca_2) - \mu_1 + \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This is just a linear system of equations, and yields $\mu_1 = \frac{1}{2}\left((d - b)a_1 + (b - c)a_2\right)$ and $\mu_2 = \frac{1}{2}\left((d + b)a_1 + (b + c)a_2\right)$. Now, these μ_i are both positive, only when the candidates for the local min with one active condition are not feasible (see the conditions above!). Hence, we get the following candidates for local minima, with different cases. The second order condition for each of these points is satisfied.

$$\begin{aligned} x_1 = x_2 = \frac{da_1 + b(a_1 + a_2) + ca_2}{d + 2b + c} &\text{ if } a_1 > a_2 \text{ and } (d + b)a_1 + (c + b)a_2 < 0 \\ x_1 = -x_2 = \frac{da_1 - b(a_1 - a_2) - ca_2}{d - 2b + c} &\text{ if } a_2 > -a_1 \text{ and } (d - b)a_1 + (c - b)a_2 < 0 \\ (0, 0) &\text{ if } (d + b)a_1 + (c + b)a_2 > 0 \text{ and } (d - b)a_1 + (c - b)a_2 > 0 \\ &\quad \text{and } a_2 \notin [a_1, -a_1] \\ x = a &\text{ if } a_2 \in [a_1, -a_1] \end{aligned}$$

□

This is a tedious example perhaps, so not worth it to go through all of the algebra. However, I think that it is still useful to discuss the different cases. Draw a picture, and skip over most (all?) of the algebra.

2.4 Concavity and Quasi-Concavity

Gabrielle Shonk, *Across the Room*

As it stands, our approach has to be find all local maxima and compare them. This can be time-consuming, and also require numerically comparing values that may be nontrivial to compute. This, as well as some of the discussion above, motivates that we should look for conditions on our functions such that finding local maxima will be equivalent to finding global maxima.

It turns out that what we want is concavity. Recall the definition of concavity:

Definition 2.4.1. We say that a function $f : U \to \mathbb{R}$ is concave, for U convex, if for all $x, y \in U$ and $\lambda \in (0, 1)$ we have that

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$$

We say that f is strictly concave if the above inequality is strict whenever $x \neq y$.

There is an equivalent definition that requires that the subgraph of f be convex in $U \times \mathbb{R}$, but it is obvious that the definitions are equivalent.

Recall that a function being concave implies that the tangent space to a function at any point lies above the graph of the function. Indeed, that is the defining trait of differentiable concave functions.[5] There is also a characterization in terms of its Hessian.

Lemma 2.4.2. *Let $f : U : \mathbb{R}$, with U convex, with f twice differentiable. Then*

1. *f is concave if and only if $\nabla^2 f$ is negative semi-definite for all $x \in U$*
2. *If $\nabla^2 f$ is negative definite for all $x \in U$, then f is strictly concave.*

Notice that the second is not an if and only if, as the Hessian can have some zero eigenvalues and have the function be strictly concave. For example, consider $f(x) = -x^2$. You should already have learned that. If you didn't then that is bad.

Regardless, this characterization in terms of the Hessian is extremely useful. Notice that the necessary second-order conditions always required that the (sometimes constrained) Hessian was negative semi-definite. If we assume that f is concave, then we will always have this condition be satisfied. Similarly, if we have that f is strictly concave, then the second-order portion of the sufficient condition is always going to be satisfied. Concavity is nice!

Notice that this second-order nature holds in the Kuhn-Tucker problem if we make each g_i convex.[6] There, we would have that the Hessian of $\mathcal{L}$ remains negative semi-definite because we would be subtracting positive semi-definite matrices.

The other (more important) benefit of having g be convex is that then $g^{-1}((-\infty, 0])$ will be a convex set. This means that we can really utilize the concavity of f on the whole domain in question (draw a picture?).[7]

So now, we are going to restrict our attention to looking at optimization problems of the type (f, g), where we have that f is concave and g_i is convex.

[5] There is an analogous notion for non-differentiable concave functions. See below, if we have time.

[6] Herein lies the difficulty with the convention that I take $g \leq 0$ to be the constraint set, as opposed to $g \geq 0$. If I did the latter, then we would want the g_i to be concave. Like many of these things, it comes down to personal preference.

[7] Maybe also discuss what is star convexity, and how that is not enough for our purposes.

Aside: Is this assumption useful? Of course, computationally these assumptions move us forward—that is why we are making them. But, are they too restrictive? In general, concavity (or its weaker counter-part quasi-concavity) is assumed very frequently across fields in economics, because the computational traction that it gives is useful. Also see below Section **??**. For example, OLS is a concave optimization problem.

We get some quick results from this discussion

Theorem 2.4.3. *Let x be a local maximum of the concave optimization problem (f, g). Then x is a global maximum.*

Note that we very much need that g is also convex. If it were not, then we could do wacky things (draw a picture).

We get some other trivial results:

Corollary 2.4.4. *In the convex optimization problem (f, g), the set of all maxima is convex.*

Corollary 2.4.5. *Let f be a concave function. Then $x \in \int(U)$ is a maximum if and only if $\nabla f(x) = 0$.*

This is just an observation of what we said above regarding the second-order conditions.

Before we state the full generalization of the results above, we need one regularity condition in order to guarantee that the function g is not too restrictive.

Definition 2.4.6. Let g be a convex function. If there exists an $x \in \mathbb{R}^n$ such that $g(x) \ll 0$, then g is said to satisfy *Slater's Condition*.

With this minor regularity condition, we get the following:

Theorem 2.4.7. *Let (f, g) be a concave optimization problem. If g satisfies Slater's Condition, then x is a maximum if and only if there exists a $\mu \leq 0$ such that*

1. $\nabla f(x) + \mu \cdot Dg(x) = 0$
2. $\mu_i g_i(x) = 0$

Again, this is just a description of what we had noticed above, that concavity gives us only a need to check first-order conditions.

2.4.1 Examples

APM462 I think will be useful?. Alberto's notes are probably going to be fairly useful for this. MWG also?

2.4.2 Expected Utility Theory

How do we think about concave and convex functions? In economics they come up all of the time. Specifically, we like to think of them in terms of decreasing returns to scale and what is called "risk" aversion.

Throughout this section, I have been abstractly saying that the agent has some objective that they are trying to maximize. Let us call that objective $u(x)$. Until now, there was no uncertainty in the agent's outcome. If they choose x, then they get $u(x)$. However, the economic situations in which we are most interested have uncertainty. If I am deciding between two crops to plant, my optimal choice will likely depend on the amount of rain

that we get throughout the summer, or when is the first frost. But when I am making the choice, I do not know the answers to those questions.

As economists, how do we model the agent's optimal choice in this setting? It is already a stretch to assume that we know the agent's optimal choice in the absence of uncertainty. How the agent aggregates across uncertainty is even harder to reasonably know.

Instead, we often make a very unrealistic assumption that the agent aggregates linearly across uncertainty. If there are two possible "states" (Rainy, Sunny) that are equally likely, then we assume that the agent's utility from planting Canola is given by

$$\frac{1}{2}u(C,R) + \frac{1}{2}u(C,S)$$

Why is concavity useful here? Suppose that our objective is actually of the following form:

$$u(y(p,w))$$

where y is the yield of a crop, p is the crop that we plant, and w is the weather. If u is concave, then we are "risk averse:" if we are deciding between wheat and canola, and on average they give the same yield, then we will prefer to plant the crop that is less dependent on the weather (in this case, wheat). Why is this? (draw a picture). Concavity gives us a powerful tool for modelling preferences for stability.

This also has implications for smoothing consumption across periods in intertemporal models, decreasing returns to scale, and other settings.

2.4.3 Non-differential Convex Optimization

We have this general theorem for all concave functions defined on an open set:

Theorem 2.4.8. *If $f : U \to \mathbb{R}$ is concave, then f is differentiable almost everywhere and ∇f is continuous almost everywhere.*

First, questions about "almost everywhere" will have to wait until we cover measure theory. For now, just think about this being a really really small set.

So, in any compact interval, there are going to be few points at which our function of interest is not going to be differentiable. In a somewhat surprising turn, our theory can be somewhat extended to cover these cases.

What we need to do is look at what the defining characteristic of the derivative is, and then try to extend it to our situation. For any one-dimensional function, the derivative defines the slope of the tangent line at a point. In turn, the tangent line is the line that lies either entirely below or entirely above the function in a neighbourhood of the point.

Because convex functions are continuous on any open set, we can define the "sub-differential" to be the set of all slopes whose corresponding lines lie entirely under the graph of the function. This is going to be the definition that we want to use for our theory. A *sub-differential* is a generalization of the gradient to non-differentiable convex functions.

Definition 2.4.9. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. The vector $g \in \mathbb{R}^n$ is a *sub-differential* (or *sub-gradient*) of f at x_0 if, for all $x \in \mathbb{R}^n$ we have that

$$\langle g, x - x_0 \rangle \leq f(x) - f(x_0)$$

Proposition 2.4.10. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then a sub-differential exists at every point in $\mathbb{R}^n$.*

Proof. Left as an exercise (Exercise 5.2.13). □

In that exercise you will also use the sub-gradient to prove some nice results. The following proposition shows how the sub-differential relates to the gradient of f when f is differentiable.

Proposition 2.4.11. *Let f be convex. Then, f is differentiable at x_0 if and only if there is a unique sub-differential at x_0. Moreover, this sub-differential is equal to $\nabla f(x_0)$.*

Proof. See Exercise 5.2.13. □

In terms of optimizing, the conditions are exactly the same, we just have to be careful about the directions that we can move.

In the unconstrained case, we can use the example of $|x|$ as a motivator:

Lemma 2.4.12. *Suppose that f is a convex function defined on an open set. Then if f has a minimum at x, we have that*

$$0 \in \partial f$$

Notice that we just need 0 to be in the sub-differential, we can have that other vectors lie in the set.

When we move to the constrained case we get a similar condition:

Theorem 2.4.13 (First order Necessary Condition for Minimum)**.** *Suppose that f is convex, and g_i are convex. Consider the maximization problem*

$$\min f(x)$$

Subject to

$$g_i(x) \leq 0, \quad \textit{for all} \quad i$$

Then if x is a local minimum of f on the constraint set, there exists a $\mu \leq 0$ and a vector $v \in \partial f(x)$ such that

$$v \in \sum \mu_i \partial g_i(x)$$

When we are looking for maximizers, we need to be more careful. Because of the convexity of the lower level sets of f, we have additional movement for which we need to account. The condition becomes as follows.

Theorem 2.4.14 (First order Necessary Condition for Maximum)**.** *Suppose that f is convex, and g_i are convex. Consider the maximization problem*

$$\max f(x)$$

Subject to

$$g_i(x) \leq 0, \quad \textit{for all} \quad i$$

Then if x is a local maximum of f on the constraint set, for each $v \in \partial f(x)$, we have that there exist μ_i satisfying complementary slackness and sign conditions such that

$$v \in \sum \mu_i \partial g_i(x)$$

This isn't a super important section, but it is interesting and I like it. The following examples highlight in practice how these tools can be used

Example 2.4.1. Consider the following optimization problem:

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

subject to

$$g(x,y) = |x-2| + |y+3| - 1 \leq 0$$

(a) Find the subdifferentials of f and g.

(b) Solve the minimization problem using subdifferentials.

(c) Draw a diagram in $\mathbb{R}^2$ of the feasible set $g \leq 0$ and the level sets of f.

(d) Is there a way to frame the problem in terms of differentiable functions?

Solution. (a) As f is differential, we know exactly that $\partial f = \{\nabla f\} = \left\{ \begin{pmatrix} 2x \\ -1 \end{pmatrix} \right\}$. Now, as g is the sum of a function $q(x) = |x-2|$ and a function $p(y) = |y+3| - 1$, we have that $\partial g = \partial q \times \partial p$. Hence, we have that

$$\partial g(x,y) = \begin{cases} \{(-1,-1)\} & \text{if} \quad x < 2, y < -3 \\ \{-1\} \times [-1,1] & \text{if} \quad x < 2, y = -3 \\ \{(-1,1)\} & \text{if} \quad x < 2, y > -3 \\ [-1,1] \times \{-1\} & \text{if} \quad x = 2, y < -3 \\ [-1,1]^2 & \text{if} \quad x = 2, y = -3 \\ [-1,1] \times \{1\} & \text{if} \quad x = 2, y > -3 \\ \{(1,-1)\} & \text{if} \quad x > 2, y < -3 \\ \{1\} \times [-1,1] & \text{if} \quad x > 2, y = -3 \\ \{(1,1)\} & \text{if} \quad x > 2, y > -3 \end{cases}$$

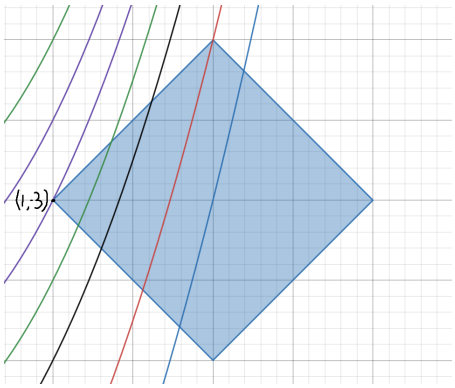
(b) We first check to see if there is a minimum on the interior, that is, where $\partial f \ni 0$. Now, this cannot happen, as $-1 \neq 0$.

Hence, we know that the minimum occurs somewhere on the boundary, where we have a minimum when $-\nabla f \in \lambda \partial g$, where $g = 0$ and $\lambda > 0$. As $-\nabla f = \begin{pmatrix} -2x \\ 1 \end{pmatrix}$, we must have that $y \geq -3$ (so that the second term could be positive). As $x \geq 1$ (so that the first term in the gradient of f is negative), we must also have that $x \leq 2$ (so that the first term in a subgradient of g could be negative).

Now, if $x = 2$, we must have that, to be on the boundary, $y = -2$, so that $\begin{pmatrix} -4 \\ 1 \end{pmatrix} \notin \lambda[-1,1] \times \{1\}$.

If $x < 2, y = -3$, the only boundary point is $(1, -3)$. In order to have $\begin{pmatrix} -2 \\ 1 \end{pmatrix} \in \lambda\{-1\} \times [-1, 1]\}$, we could have $\lambda = 2$. But notice then that we have $-\nabla f \in \lambda \partial g$, so that this point is a minimizer. This completes the question; $(1, -3)$ is the minimizer.

(c) This does agree with our conclusion, as the left corner is at $(1, -3)$.



(d) Yes!

□

Example 2.4.2. Solve the following convex minimization problem:

$$\min f(x, y) = \max\{|x|, y + 4\}$$

subject to

$$\begin{aligned} g_1(x, y) =& (x-1)^2 + (y-1)^2 - 1 \leq 0 \\ g_2(x, y) =& |x| + y - b \leq 0 \end{aligned}$$

Notice that the answer should depend on $b > 0$!

Solution. We first compute the subdifferential of f. Notice that $f(x, y) = \max\{f_1(x, y), f_2(x, y)\}$, where $f_1(x, y) = |x|$ and $f_2(x, y) = y + 4$. We have then that

$$\partial f_1(x, y) = \begin{cases} \{(-1, 0)\} & \text{if} \quad x < 0 \\ [-1, 1] \times \{0\} & \text{if} \quad x = 0 \\ \{(1, 0)\} & \text{if} \quad x > 0 \end{cases} \qquad \partial f_2(x, y) = \{(0, 1)\}$$

However, consider that on the ball $(x-1)^2 + (y-1)^2 - 1 \leq 0$, we have that $0 \leq x \leq 2$ and $0 \leq y \leq 2$, so that $y + 4 > |x|$. This means that on the area of interest, we will have that $\partial f(x, y) = \{(0, 1)\}$ always.

Now, consider that on the area of interest we have

$$\partial g_1(x, y) = \begin{pmatrix} 2(x-1) \\ 2(y-1) \end{pmatrix}$$

$$\partial g_2(x, y) = \begin{cases} [-1, 1] \times \{1\} & \text{if} \quad x = 0 \\ \{(1, 1)\} & \text{if} \quad x > 0 \end{cases}$$

Notice that we can drop the case where $x < 0$ here because we know that given the first constraint we have $x \geq 0$.

Now, consider that as the subdifferential contains only one vector, and this vector is always non-zero, we know that we will have a solution only on the boundary. In what follows, λ_1, λ_2 are arbitrary non-negative scalars

We check when $\begin{pmatrix} 0 \\ -1 \end{pmatrix} \in \lambda_1 \partial g_1(x, y)$. Then, as $\lambda_1 \neq 0$ clearly, we must have $g_1 = 0$ with the first entry of the vector in the subdifferential being 0, so that $x = 1$. If $y = 2$, then the second entry will always be positive, which is not allowable. However, if $y = 0$, then $\begin{pmatrix} 0 \\ -1 \end{pmatrix} \in 2\partial g_1(1, 0)$. So, if we have $b \geq 1$, (i.e., $(1, 0)$ is feasible) then $(1, 0)$ is a minimizer for f on the set of interest.

Notice that we never have $\begin{pmatrix} 0 \\ -1 \end{pmatrix} \in \lambda_2 \partial g_2(x, y)$, as each vector in $\partial g_2(x, y)$ has a positive second entry. Hence, all that is left to check is when $\begin{pmatrix} 0 \\ -1 \end{pmatrix} \in \lambda_1 \partial g_1(x, y) + \lambda_2 \partial g_2(x, y)$. First, if $x = 0$ we have that $y = 1$. Here, elements of this sum take the form

$$\begin{pmatrix} -2\lambda_1 + \lambda_2 w \\ \lambda_2 \end{pmatrix}, w \in [-1, 1]$$

Notice that the second term can never be negative, so that we have no minimum here.

Now, if $x > 0$, as $g_2(x, y) = 0$ in this scenario, that elements of this sum take the form

$$\begin{pmatrix} 2\lambda_1(x-1) + \lambda_2 \\ 2\lambda_1(y-1) + \lambda_2 \end{pmatrix}$$

Now, as we need the second term to be negative while the first term needs to be zero, we need $1 \geq x > y$.

Notice that when $x > 0$, both of our constraints are symmetric in terms of x, y. Solutions are explicitly given by $x = b - y$, so that (from g_1) we have $x = \frac{2b \pm \sqrt{-4b^2+16b-8}}{4}$. Note this means that our values of y are $y = \frac{2b \mp \sqrt{-4b^2+16b-8}}{4}$. Now, we will have x be the larger of the two, so that if $-4b^2 + 16b - 8 > 0$ (corresponds to $b > 2\left(1 - \frac{\sqrt{2}}{2}\right)$, and $x = \frac{2b \pm \sqrt{-4b^2+16b-8}}{4} \leq 1$ (corresponds to $b \leq 1$, by easily expanding), our minimizer is given exactly by

$$x = \frac{2b + \sqrt{-4b^2 + 16b - 8}}{4}$$
$$y = \frac{2b - \sqrt{-4b^2 + 16b - 8}}{4}$$

If $b \geq 1$, then our minimizer is given as in the part above, $(1, 0)$.

Now, all of these calculations have been assuming that Slater's condition is satisfied. That is exactly the case when b is larger than the smaller root of the quadratic $-4b^2+16b-8$ (because when this quadratic is first zero, it intersects the circle only once). This corresponds to $b = 2\left(1 - \frac{\sqrt{2}}{2}\right)$. Here we cannot use subdifferentials, but because there is only one point in our constraint set: $(1 - \frac{\sqrt{2}}{2}, 1 - \frac{\sqrt{2}}{2})$; this point is clearly the minimizer. $\square$

Duality
In the future, it is probably good to have a section on this. For this though, we probably do need an introduction to linear programming.

2.5 Parametric Optimization

Field Guide, *Make Peace with That*

Concavity gives us a lot of structure. It turns out that we can extend the structure beyond a single optimization problem to a family of optimization problems. Suppose now that we have a vector $\theta \in \Theta$ that parameterizes our optimization problem. That is, our primitives become two maps $f : \mathbb{R}^n \times \Theta \to \mathbb{R}$ and $g : \mathbb{R}^n \times \Theta \to \mathbb{R}^p$ that are each twice differentiable. The problem becomes

$$\max_x f(x, \theta) \text{ subject to}$$
$$g(x, \theta) \leq 0$$

Here, notice that as we are optimizing over x, θ is a fixed parameter of the problem. We have that $\Theta \subseteq \mathbb{R}^m$ is our parameter space. With this we will get a constraint set for each θ (denoted $G(\theta)$), and also a solution set (denoted $\mathcal{S}(\theta)$) for each θ.

Our goal is going to be to characterize $\mathcal{S}(\theta)$ for each θ, and then see how it varies as we vary θ. Notice that for fixed θ we can simply apply the tools that we have developed above. It is the latter question that is more interesting for us to solve.

Now, recall some of the definitions corresponding to correspondences.

Definition 2.5.1. Let $G : \Theta \rightrightarrows \mathbb{R}^m$ be a correspondence. We say that G is *Compact-Valued* if $G(\theta)$ is a compact set for all θ. We say that G is *Convex-Valued* if $G(\theta)$ is a convex set for all θ.

Notice that these definitions are not about the graph of G, but just about its values at a specific point. Also recall the definition of upper hemi-continuity.

Definition 2.5.2. Let G be a correspondence. We say that G is *upper hemi-continuous* (uhc) at θ if for all sequences $\{\theta_n\}$ such that $\theta_n \to \theta$, and all sequences x_n such that $x_n \in G(\theta_n)$ there exists a subsequence θ_{n_k} and $x \in G(\theta)$ such that $x_{n_k} \to x$. We say that G is uhc if it is uhc at each θ

What this means is that limit points cannot simply disappear as we vary θ. Another way of thinking about this is, if G is Compact-Valued and Θ is closed, then the graph of G must be also be closed. We now state the corresponding definition of lower hemi-continuity.

Definition 2.5.3. Let G be a correspondence. We say that G is *lower hemi-continuous* (lhc) at θ if for every sequence $\{\theta_n\}$ such that $\theta_n \to \theta$ and $x \in G(\theta)$ there exists a subsequence θ_{n_k} and $x_{n_k} \in G(\theta_{n_k})$ such that $x_{n_k} \to x$. We say that G is lhc if it is lhc at each θ

What this means is that limit points cannot simply *appear* (draw a picture, compare with uhc). We also finally have the corresponding notion of continuity.

Definition 2.5.4. Let G be a correspondence. We say that G is continuous at θ if it is both uhc and lhc at θ. We say that G is continuous if it is continuous at each θ.

With this (hopefully review) out of the way, we can start to state some results. First, we do not impose concavity on the optimization problem for each θ, we just impose continuity restrictions.

Theorem 2.5.5 (Maximum Theorem)**.** *Let $f : \mathbb{R}^n \times \Theta \to \mathbb{R}$ be continuous and $G : \Theta \rightrightarrows \mathbb{R}^n$ be a compact-valued and continuous correspondence. Define*

$$v(\theta) := \max_x f(x, \theta) \text{ s.t. } x \in G(\theta)$$
$$\mathcal{S}(\theta) := \operatorname*{argmax}_x f(x, \theta) \text{ s.t. } x \in G(\theta)$$

Then we have that $v(\theta)$ is continuous on Θ and $\mathcal{S}$ is a compact-valued uhc correspondence on Θ.

This shouldn't be too surprising actually, because the continuity of all of the objects (and the compact-valued nature of G) ensures that nothing surprising happens when we change θ. When we additionally add concavity, then things start to become nicer.

Theorem 2.5.6 (Maximum Theorem with Concavity)**.** *Let $f : \mathbb{R}^n \times \Theta \to \mathbb{R}$ be continuous, twice differentiable, and concave and let and $g : \mathbb{R}^n \times \Theta \to \mathbb{R}^p$ be convex, twice differentiable, and such that it implies that $G : \Theta \rightrightarrows \mathbb{R}^n$ is a compact-valued and continuous correspondence. Define*

$$v(\theta) := \max_x f(x, \theta) \text{ s.t. } x \in G(\theta)$$
$$\mathcal{S}(\theta) := \operatorname*{argmax}_x f(x, \theta) \text{ s.t. } x \in G(\theta)$$

Then we have that $v(\theta)$ is a concave function Θ and $\mathcal{S}$ is a compact-valued, convex-valued, and uhc correspondence on Θ. If, in addition, f is strictly concave, then v is strictly concave and $\mathcal{S}$ is a continuous function on Θ.

These theorems are nice because they give us some structure on how v and $\mathcal{S}$ change when we vary parameters. At the same time, it is not extremely useful in terms of determining how the objective changes as we change these parameters. The next result tells us that the result is actually quite intuitive. First-order effects are all that matter; we don't need to worry about how optimal choices change when we change the parameters.

Theorem 2.5.7 (Envelope)**.** *Let $f : \mathbb{R}^n \times \Theta \to \mathbb{R}$ be continuous, twice differentiable, and concave and let and $g : \mathbb{R}^n \times \Theta \to \mathbb{R}^p$ be convex, twice differentiable, and such that it implies that $G : \Theta \rightrightarrows \mathbb{R}^n$ is a compact-valued and continuous correspondence that satisfies Slater's condition. Define again*

$$v(\theta) := \max_x f(x, \theta) \text{ s.t. } x \in G(\theta)$$

Then we have that

$$\nabla_\theta v(\theta) = \left(\nabla_\theta \mathcal{L}(x, \mu, \theta)\right)\big|_{x=x(\theta), \mu=\mu(\theta)}$$

2.6 Functional Optimization

YOASOBI, *THE BOOK / THE BOOK 2*[8]

[8] Am I cheating with this? Maybe, but they are both very short albums, and I make the rules.

This is the least standard section for this part of the course. I do think that there are times when it comes up in practice, and so I think it is worth spending some time on it.

Another name for this section is Calculus of Variations. We take a step away from what we have done mostly through this chapter, and instead of maximizing in a finite number of dimensions we now look at maximizing entire functions. However, I want to emphasize that here, as above, the methods are going to be similar.

We (clearly) do not have time to go through an entire course on functional optimization. Instead, I want to highlight some of the simple techniques and emphasize that there is little difference between what we have done previously and what we are doing here. Things become slightly more difficult with regards to terminology, and we will need to upgrade our technical machinery, but the intuition for the techniques remains exactly the same. Functional optimization is something that it is important to know about in my thinking (it has applications in all of our fields of study).

Let us return to the material. We are interested in the following problem:

$$\max_{u \in \mathcal{A}} \mathcal{F}[u]$$

where $\mathcal{F}$ is a "functional," which is a function on the space of functions.

Example 2.6.1. Let $\mathcal{A} := \{u : [0,1] \to \mathbb{R} | u \in \mathcal{C}^1, u(0) = u(1) = 1\}$. That is, $\mathcal{A}$ contains all continuously differentiable functions on $[0,1]$ whose endpoints are exactly 1.

We would then write $\mathcal{F} : \mathcal{C}^1[0,1] \to \mathbb{R}$. That is, $\mathcal{F}$ takes continuously differentiable functions defined on the interval and returns a value. What is the most common type of functional that we see?

- Evaluation at a point, or some function of evaluations
- Evaluation of the derivative.
- Definite integration.

As one specific example, we could set

$$\mathcal{F}[u] = -\frac{1}{2} \int_0^1 [u^2(x) + (u')^2(x)] \, dx$$

This problem can look extremely intimidating, because it is. So, what should our plan be? As a proposal:

1. Derive some necessary 1st order conditions at a maximum
2. Find a u^* that satisfies the necessary conditions
3. Verify that the candidate u^* is indeed a maximum

Because the problem in its current form looks so difficult, it is going to be a good idea to reduce the problem from one infinite-dimensional problem to (many) one-dimensional problems.

Example 2.6.1 (continued)**.** Consider the setup above. Fix some $v \in \mathcal{C}^1$ such that $v(0) = v(1) = 0$. Notice that $v \notin \mathcal{A}$, but it has the property that if $u \in \mathcal{A}$ then for all $s \in \mathbb{R}$ we have that $u + sv \in \mathcal{A}$.

Now, consider a u_* that is a maximizer for $\mathcal{F}$. We must have that

$$\mathcal{F}[u_* + sv] \leq \mathcal{F}[u_*]$$

So, if we define $f(s) \coloneqq \mathcal{F}[u_* + sv]$, then we will have that 0 is a maximizer for f. (See how we have started to convert the problem?) In particular, because f is clearly differentiable (it is in fact a polynomial), we have

$$f'(0) = 0 \Rightarrow -\frac{1}{2}\int_0^1 2v(u_* + sv) + 2v'(u_*' + sv')dx\Big|_{s=0} = 0$$

In particular, we can rewrite this as the following condition

$$\int_0^1 vu_* + v'u_*'dx = 0 \quad \forall \text{ test functions } v$$

If we further assume that $u \in \mathcal{C}^2$, then we can use integration by parts to rewrite

$$\int_0^1 v'u_*'dx = vu_*'\Big|_0^1 - \int_0^1 vu_*''dx$$

Plugging this back in, we have that a necessary condition is that

$$\int_0^1 v(u_* - u_*'')dx = 0$$

recall that this is *for all* test functions v. It is intuitive (we will formalize this in a minute) that this implies that we get

$$u_* \equiv u_*''$$

This of course gives us functions of the form $ae^x + be^{-x}$. Fitting our ending values gives us a candidate of

$$u_*(x) = \frac{1 - e^{-1}}{e - e^{-1}}e^x + \frac{e - 1}{e - e^{-1}}e^{-x}$$

At this stage this is just a guess! But it is easy to show that in this case we get sufficiency. If u_* satisfies the conditions above, we have that

$$\mathcal{F}[u_* + sv] = \mathcal{F}[u_*] + d\int_0^1 [u_*v + u_*'v']\,dx + \frac{s^2}{2}\int_0^1 [v^2 + (v')^2]\,dx \geq F[u_*]$$

We want to formalize the above a little bit more.

Lemma 2.6.1. *Let g be continuous on $[a, b]$. If $\int_a^b g(x)v(x)dx = 0$ for all v test functions, then we have that $g \equiv 0$ on $[a, b]$.*

Proof. Exercise. Have you ever heard of a bump function? □

We now turn to a more general class of problems. The set-up will be as follows:

$$\mathcal{A} = \{u : [a,b] \to \mathbb{R} | u \in \mathcal{C}^1, u(a) = A, u(b) = B\}$$

And we have that

$$\mathcal{F}[u] = \int_a^b L(x, u(x), u'(x))dx$$

We call the function

$$L(x,z,p) : [a,b] \times \mathbb{R}^2 \to \mathbb{R}$$

the "Lagrangian." For notation, we will write

$$L_z(x,z,p) := \frac{\partial}{\partial z} L(x,z,p)$$
$$L_p(x,z,p) := \frac{\partial}{\partial p} L(x,z,p)$$

Unfortunately, because this is an entirely new problem, we need to introduce some new notation. It will have analogues in the finite-dimensional case, I promise.

Definition 2.6.2. Let $u \in \mathcal{A}$. Suppose that there exists a function g_u on $[a,b]$ such that

$$\frac{\partial}{\partial s}\Big|_{s=0} \mathcal{F}[u + sv] = \int_a^b g_u(x)v(x)dx$$

for all test functions v. Then, we call g_u the *variational derivative* of $\mathcal{F}$ at u, and we denote g_u by

$$\frac{\partial \mathcal{F}}{\partial u}(u)(\cdot) \text{ or } \frac{\partial \mathcal{F}}{\partial u}(u)$$

How do we think of this? Well, it is really just the analogue of ∇f in the finite dimensional case. What do I mean by this? Well, in finite dimensions, the gradient of a function is exactly defined by the following property:

$$\frac{\partial}{\partial s}\Big|_{s=0} f(u + sv) = \nabla f(u) \cdot v = \langle \nabla f, v \rangle$$

The only difference here is that our inner product has changed. Instead of the simple dot product in finite dimensions, we now contend with the integral.

Example (2.6.1) we had that $\frac{\partial \mathcal{F}}{\partial u} = u - u''$.

We can now state a generalization of the necessary condition that we derived in the above example.

Lemma 2.6.3. *Let $\mathcal{A} = \{u : [a,b] \to \mathbb{R} | u \in \mathcal{C}^1, u(a) = A, u(b) = B\}$. Suppose that for $u_* \in \mathcal{A}$, and all test functions v, we have that $u_* + v \in \mathcal{A}$. Then, if u_* maximizes $\mathcal{F}$ on $\mathcal{A}$, $\frac{\partial \mathcal{F}}{\partial u}(u_*)$ exists and is continuous on $[a,b]$ and we have that $\frac{\partial \mathcal{F}}{\partial u}(u_*) \equiv 0$ on $[a,b]$.*

Proof. As $u_* + sv \in \mathcal{A}$, as u_* maximizes, we have that $\mathcal{F}[u_* + sv] \leq \mathcal{F}[u_*]$. Then, define $f(s) \coloneqq \mathcal{F}[u_* + sv]$. We have that $f'(0) = 0$. Because by definition $f'(0) = \int_a^b \frac{\partial \mathcal{F}}{\partial u}(u_*)v$, we are done. Proving existence formally requires a little more work, so I omit this. □

If we put some structure on $\mathcal{F}$ then we can get significantly stronger results.

Theorem 2.6.4. *Let $\mathcal{A} = \{u : [a,b] \to \mathbb{R} | u \in \mathcal{C}^1, u(a) = A, u(b) = B\}$. Suppose that we can write*

$$\mathcal{F}[u] = \int_a^b L(x, u(x), u'(x))dx$$

for some Lagrangian $L \in \mathcal{C}^2$.[9] *Then we have that $\frac{\partial \mathcal{F}}{\partial u}(u)$ exists, is continuous, and given by*

$$\frac{\partial \mathcal{F}}{\partial u}(u)(x) = -\frac{\partial}{\partial x}L_p(x, u, u') + L_z(x, u, u')$$

This is called the Euler-Lagrange equation.

Proof. If we assume that $u \in \mathcal{C}^2$ then the proof becomes a straightforward computation (so I leave it as an exercise). Relaxing to $u \in \mathcal{C}^1$ is much more difficult. □

Example (2.6.1) Here we have that $L = \frac{1}{2}(z^2 + p^2)$. Then we have that $L_z = z$ and $L_p = p$. So, we get that $\frac{\partial \mathcal{F}}{\partial u}(u) = u - u''$.

Corollary 2.6.5. *Let u_* be a maximizer of $\mathcal{F}$ on $\mathcal{A}$, where $\mathcal{F}$ is defined with a Lagrangian L that is suitably smooth. Then we have that*

$$-\frac{\partial}{\partial x}L_p(x, u_*(x), u_*'(x)) + L_z(x, u_*(x), u_*'(x)) = 0$$

Example 2.6.2. One of my favourite examples. Let us show that the minimum (reasonably smooth) distance between two points is a straight line.

Specifically, we have that

$$\mathcal{A} = \{u : [a,b] \to \mathbb{R} | u \in \mathcal{C}^1\}$$
$$\mathcal{F}[u] = \int_a^b \sqrt{1 + (u')^2(x)}dx$$

We then have that

$$L_z \equiv 0$$
$$L_p \equiv \frac{p}{\sqrt{1+p^2}}$$

As we have that L_p must be a constant (its derivative is zero), we have that the derivative of any minimizer must be constant. Hence we get that $u_* = \alpha x + \beta$.

I will give a few exercises as problems, but the idea throughout is just to apply the same techniques that we have developed here.

[9] I think this may be able to be relaxed to $L \in \mathcal{C}^1$, but I am not certain.

2.6.1 Equality Constraints

We focus on two types of equality constraints: "Isoperimetric Problems" and "Holonomic Problems." The methods that we are going to use are basically a direct extension of the finite-dimensional Lagrangian methods.

Isoperimetric Problems

This is a straightforward extension of the equality constraints methods we looked at previously. To translate that setting into that context, we have two functionals:

$$\mathcal{F}[u] = \int_a^b L^F(x, u, u')dx$$
$$\mathcal{G}[u] = \int_a^b L^G(x, u, u')dx$$

The problem that we are trying to solve is

$$\begin{aligned} &\max_u \ \mathcal{F}[u] \\ &\text{s.t. } G[u] = c \\ &\quad u \in \mathcal{A} \coloneqq \{u \in \mathcal{C}^1, u(a) = A, u(b) = B\} \end{aligned}$$

Notice that here $\mathcal{G}$ is just from $\mathcal{C}^1[a, b] \to \mathbb{R}$, but we could easily make the range $\mathbb{R}^n$. This is equivalent to moving from one constraint to n, just as before.

We get the corresponding result, as we had in the finite dimensional case.

Theorem 2.6.6. *If $u_* \in \mathcal{A}$ has $\frac{\partial \mathcal{G}}{\partial u}(u_*) \not\equiv 0$, then for a maximizer u_* we must have that*

$$\frac{\partial \mathcal{F}}{\partial u}(u_*) + \lambda \frac{\partial \mathcal{G}}{\partial u}(u_*) \equiv 0 \quad \text{on } [a, b]$$

for some $\lambda \in \mathbb{R}$.

I state the theorem without proof, but again this is just the generalization of the finite dimensional case.

Example 2.6.3. Suppose that the problem is to maximize

$$\int_a^b u(x)dx \quad \text{subject to} \quad \int_a^b \sqrt{1 + (u')^2(x)}dx = d$$

With the condition that

$$u \in \mathcal{A} \coloneqq \{u : [-a, a] \to \mathbb{R} | u \in \mathcal{C}^1, u(-a) = u(a) = 0\}$$

We get first order conditions being that we have

$$-1 + \lambda \frac{\partial}{\partial x}\left[-\frac{u'}{\sqrt{1 + (u')^2}}\right] = 0$$

We can rewrite this as

$$\lambda^2 \frac{(u')^2}{1+(u')^2} = (c-x)^2$$

Now, take some u that satisfies the above. I claim that this means that u satisfies

$$(x-c_1)^2 + (u(x)-c_2)^2 = \lambda^2$$

for some $c_1, c_2 \in \mathbb{R}$. This would mean exactly that u lies on a circle. This equation holds if and only if

$$u' = \frac{c_1 - x}{u - c_2} \iff (u')^2(u-c_2)^2 = (x-c_1)^2$$

It can easily be shown that this is equivalent to (plugging into the equation for the circle)

$$(1+(u')^2) = \frac{\lambda^2}{(u-c_2)^2} = \frac{\lambda^2 w (u')^2}{(x-c_1)^2}$$

Hence, we have that solutions lie on a circle.

Example 2.6.4. Suppose that we would like to connect the points $(0,b)$ and $(a,0)$ and enclose an area of $s := \frac{ab}{2}$ with a solid that has the smallest area of revolution. That is,

$$\begin{aligned} \min \mathcal{F}[u] =& 2\pi \int_0^a u(x)\sqrt{1+(u')^2(x)}dx \\ \text{s.t. } \mathcal{G}[x] =& \int_0^a u(x)dx = s \end{aligned}$$

Our set of consideration is going to be

$$u \in \mathcal{A} := \{u : [0,a] \to \mathbb{R} | u \in \mathcal{C}^1, u(0) = b, u(a) = 0\}$$

Let us write out L^F, L^G:

$$L^F = z\sqrt{1+p^2} \quad L^G = z$$

This gives us the Euler-Lagrange equation

$$-\frac{\partial}{\partial x}\left[\frac{uu'}{\sqrt{1+(u')^2}}\right] + \sqrt{1+(u')^2} + \lambda = 0$$

It can be verified that solutions are functions of the form $u(x) = \alpha x + \beta$, where $\lambda = \frac{-1}{\sqrt{1+\alpha^2}}$.

In general, solutions to the Euler-Lagrange equations are hard to compute, because they will be a second-order ODE. That doesn't mean that they are impossible to solve though. And often, if the Euler-Lagrange equation is well-behaved then the local uniqueness for ODEs in general will give us a unique candidate.

Holonomic Problems

Suppose now that we are in a multi-dimensional space. That is, we are interested in vector-value functions.

We aim to maximize

$$F[x,y,z] = \int_a^b L(t,x,y,z,\dot{x},\dot{y},\dot{z})dt$$

Subject to the constraint that the curve (x,y,z) lies on the surface defined by $H(x,y,z)=0$.

We get the following conditions (as long as H is regular: $\nabla H \neq 0$ on the surface $H=0$):

$$\begin{pmatrix} \frac{\partial F}{\partial x}(x,y,z)(t) \\ \frac{\partial F}{\partial y}(x,y,z)(t) \\ \frac{\partial F}{\partial z}(x,y,z)(t) \end{pmatrix} + \lambda(t) \begin{pmatrix} H_x(x,y,z) \\ H_y(x,y,z) \\ H_z(x,y,z) \end{pmatrix} = 0$$

Notice that we have additional degrees of freedom in the λ because our constraints are tighter.

Chapter 3

Measure Theory

Measure theory forms the foundation for how we think about uncountably infinite settings. These settings allow us to think about settings that are significantly more complicated than other settings. Perhaps surprisingly, these settings often afford us more tractability than discrete environments.

The most obvious example is thinking about situations where agents are "small." In any discrete economy, agents choices' will have some impact on aggregate outcomes. This impact may be small, but it will not be zero. At the same time, for realism it does not seem that this is a reasonable model of the world. Very seldom do we think of individuals as considering the aggregate impact of their own choices. Of course, there are some times when individuals over-weigh their aggregate impact (e.g., voting or the Loblaw's boycott in Canada this summer), but there also the discrete model does not capture the behaviour.

In order to model the situation makes sense to think of individuals as having "measure zero" impact on the aggregate economy. That is, individuals have no impact on anything, and so will act only to maximize their own objective (we can think of many situations where this doesn't apply, but that is outside the scope of this introduction).

Another example where this is important is in thinking of probability theory, which is, of course, the second part of this chapter. Discrete probabilities are notoriously hard to work with, and we know from the Central Limit Theorem that as things grow large, everything looks normal anyway. So, we would like a robust way of thinking about continuous probability spaces.

This chapter goes through the motivation of measures, σ-algebras, how we think about extensions of continuity and integration in this setting, and then goes on to discuss probability spaces. These tools are more technical than the previous chapter (whose tools seem directly applicable), but are just as important. In many ways it is their nuance that makes them important.

3.1 Measures

The Wombats, *Beautiful People Will Ruin Your Life*[1]

One of the motivation for this section is we want to have some idea of the "size" of sets. If we want to talk about models where there is a continuum of agents, then we need to have a serious discussion about how we think about sets. When we are dealing with non-trivial sets, how do we know what proportion of the population they represent?

[1]Fun fact: the lead singer of the Wombats' name is Matt Murphy.

What we will do is start with our intuition for what we think of as size (discuss) and then generalize this to a wider class of functions called "measures." In order to get there we will need some fairly serious math, but it is important to not lose the intuition at any point.

This is where I will take a somewhat non-standard approach to introducing measure theory. I am going to start with a lengthy foray into "length" in $\mathbb{R}$ before we get into the general definitions that are normally associated with measure theory. Suppose that we have an interval in $\mathbb{R}$, $I = (a, b)$. The most natural notion of size is its length: $|I| := b - a$. Now suppose that we have a more complicated set (draw – not in $\mathbb{R}^2$!). A natural way to think about length is the total length of intervals necessary to cover this set.

Definition 3.1.1. We define the *Lebesgue outer measure* of $A \subseteq \mathbb{R}$ as

$$m^*(A) := \inf\Big\{ \sum |I_k| \Big| \{I_k\} \text{ an open covering of } A \text{ by intervals} \Big\}$$

If for all countable open coverings of A we have $\sum |I_k| = \infty$, we define $m^*(A) := A$.

Why do we call this the "outer" measure? Because exactly we are covering A from the outside. It is natural to think about a converse, where we measure from the inside. We will see later what happens when we look at the inner measure, but for now we focus on the outer measure. We get first some basic results.

Proposition 3.1.2. *We have the following:*

(a) $m^*(\emptyset) = 0$ *and* $m^*(\mathbb{R}) = \infty$

(b) m^* *is monotonic: if* $A \subseteq B$ *then we have* $m^*(A) \leq m^*(B)$

(c) m^* *is countably sub-additive: for* $A_1, \ldots$ *a countable set we have* $m^*\left(\bigcup A_i\right) \leq \sum m^*(A_i)$

Proof. Do it on the board. □

For do we extend this to larger spaces? We simply use the higher-dimensional analogue of intervals: boxes.

Definition 3.1.3. In $\mathbb{R}^n$, we define for a box $B = \prod I_j$ that $|B| := \prod |I_j|$. We then get

$$m^*(A) := \inf\Big\{ \sum |B_k| \Big| \{B_k\} \text{ an open covering of } A \text{ by boxes} \Big\}$$

We have some more basic results about "zero" sets. That is, sets that have outer measure zero.

Proposition 3.1.4. *We have the following:*

(a) The subset of a zero set is a zero set

(b) The countable union of a zero set is a zero set

(c) The plane $\{x | x_i = a\}$ *is a zero set*

Notice that at this stage it is not even obvious that general hyper-planes are zero sets. This highlights the trickiness of dealing with this loose definition. We would need to construct a specific open covering using boxes that converges in value to zero. Later we will see that linear transformations act like they should, but at this stage we know very little about how the outer measure operates.

It is not even trivial that we have the following result.

Lemma 3.1.5. *$m^*(\overline{B}) = \prod |I|$, where $\overline{B} = \prod \overline{I}$.*

Proof. Use the Lebesgue number property of open coverings of compact sets to get the lower bound. That is, for any open covering by boxes, there exists a Lebesgue number $\lambda > 0$. For this λ, divide the set $\overline{B}$ into (open) rectangles of this size. From here we can just count. The upper bound is immediate from a shrinking argument. □

We would need to show that there is no way to save space in the box by a smart covering. Of course, this is the result that we suspect. The takeaway is that in general measures are very difficult to work with, but tend to operate the way that we expect them to once we get past those difficulties.

Corollary 3.1.6. *Closed/open intervals do not matter for how we measure the outer measure.*

The above was really an involved example using $\mathbb{R}$ as the baseline. We want to generalize these ideas to other spaces, and so we will need a more general definition. We will eventually get to a definition of a measure, but for now we start with a definition of "outer measure."

Definition 3.1.7. Suppose that we have a set X. An outer measure on X is a function $\omega : 2^X : [0, \infty)$ such that

(a) $\omega(\emptyset) = 0$

(b) ω is monotonic: if $A \subseteq B$ then we have $\omega(A) \leq \omega(B)$

(c) ω is countably sub-additive: for $A_1, ...$ a countable set we have $\omega\left(\bigcup A_i\right) \leq \sum \omega(A_i)$

The outer measure is defined on *every* subset of a set X. We can think of the outer measure as a coarse notion of length. What separates this definition from the definition of a "measure" is that we have no way of insuring that weird things don't happen. For example, are there subsets of $[0, 1]$ that are "dense" enough to have the same outer measure as $[0, 1]$, but that are non-intersecting? Eventually, we want to replace (c) with a different statement that makes our definition additive, not sub-additive.

With this example in mind, it turns out that the only nice property that we need for the outer measure is that it outer measure behaves well when we intersect a set with it and its compliment.

Definition 3.1.8. A set E is measurable with respect to outer measure ω if for all $A \subseteq X$ we have that

$$\omega(A) = \omega(A \cap E) + \omega(A \cap E^C)$$

If $\omega = m^*$ then we call E *Lebesgue measurable*. We denote the Lebesgue measure on the set of Lebesgue measurable sets by $mE := m^*E$.

From here, it is important to do a little detour into σ-algebras, because it turns out that this will be the structure of interest for us when we are thinking about the collection of measurable sets. It is important to keep in mind that although we are especially interested in Lebesgue measurable sets, the definitions and structure all apply to general spaces and measures, as we will see when we progress.

Summary: We have tried to generalize the notion of "length" for arbitrary spaces. This came out as the definition of an outer measure. We started a discussion of what it means for a set to be measurable with respect to this measure.

3.2 σ-Algebras

Annalisa, *E POI SIAMO FINITI NEL VORTICE*[2]

As with notions of topology, what we want is a notion of a "good" set. Our idea is that "good" sets are going to exactly be the measurable sets. How our definition of "good" differs from the topological setting is that each good set is going to have a corresponding number, its "size." In order to accommodate thinking about size, we need to alter our definition of what is a good set.

Agents on their own should not matter, but if a large group of agents together act then they should have some impact on aggregate outcomes. In order to formalize that notion, we need some way of talking about a proportion of agents. Consider first use indexing agents on the unit interval: $i \in [0, 1]$.

Why is this notion of good different than the topological notion of good? If A is a "good" set, i.e., if we can measure the proportion of agents that lie in A, then it is obvious that $[0, 1] \setminus A$ will also be a "good" set. This is because the proportion in this set is easy to measure: it is just 1 minus the proportion in A. This lies in stark contrast to topology, where openness lies in direct contrast to closedness.

But, if closed sets are also good, what is stopping us from generating every set? In $\mathbb{R}$ with the standard topology the reason is subtle. In general, we do not allow arbitrary unions, but instead we only allow *countable* unions of good sets.

Why is this restriction subtle in $\mathbb{R}$? Because countable unions (of basis sets) are equivalent to arbitrary unions when we think about the generation of the standard topology: the standard topology on $\mathbb{R}$ it is second-countable. So, our considering complements has significant bite here, it forces us to restrict our ideas.

Another good motivating example of what a "good" set is will become clearer when we discuss probability theory. We can think of the set $[0, 1]$ representing states of the world. "Good" sets then, are exactly sets that we can know for certain or not know (events). E.g., it is raining, it is snowing. If we know when it is raining, then we know when it is not raining. If we know when it is raining and when it is snowing, we know when it is raining or snowing.

With these ideas, it is time to define the type of object that is going to exactly capture what properties our "good" sets are going to have:

Definition 3.2.1. A collection $A \subseteq 2^X$ is a *σ-algebra* if the following conditions hold:

1. $X \in A$
2. If $E, F \in A$ then we have that $E \setminus F \in A$
3. If $E_1, ..., E_n, ... \in A$, then $\bigcup_{i=1}^{\infty} E_i \in A$

Notice that the second and third conditions together imply that countable intersections are also allowed. For any set X we get two trivial σ-algebras, just as we did in topology:

[2]She should have won Sanremo 2024.

Example 3.2.1. Given a set X, both $\{\emptyset, X\}$ and 2^X are σ-algebras.

σ-algebras are an arbitrary mathematical structure, so it is hard in general to think about non-trivial examples. To this end, it will be helpful to talk about a useful class of σ-algebras. Just as in topology, where we could talk about picking a basis and generating a topology around that basis, we can generate a σ-algebra around a collection set.

Definition 3.2.2. Given a $B \subseteq 2^X$, we define the σ-algebra generated by B (denoted $A(B)$) to be the smallest σ-algebra containing B. That is, it is the collection of subsets of X satisfying the following conditions:

1. $B \subseteq A(B)$
2. $A(B)$ is a σ-algebra
3. If A' is a σ-algebra containing B, then $A(B) \subseteq A'$

We have the following result:

Proposition 3.2.3. *$A(B)$ defined above exists and is unique. Specifically, letting*

$$\mathcal{A}(B) := \{A' \text{ } \sigma\text{-algebra} | B \subseteq A'\}$$

we have that $A(B)$ can be explicitly derived as

$$A(B) = \bigcap_{A' \in \mathcal{A}(B)} A'$$

I leave the above as an exercise. Using this definition, it is easy to think of examples of σ-algebras (relate this to the "knowing" example).

Example 3.2.2. Some discrete stuff—draw a picture.

Example 3.2.3. Consider $\mathbb{R}$ endowed with the standard topology. Consider the σ-algebra that is generated by all open sets.

What do we get with this construction? Of course, we will have all open intervals (a, b) because they are open. However, we will also get all closed intervals $[a, b]$ and half-open intervals $(a, b]$ and $[a, b)$. Because we get closed sets, we also get all singletons. So, we get any countable set.

It turns out that this last example will be extremely important.

Definition 3.2.4. We call the σ-algebra as described in Example 3.2.3 the *Borel σ-algebra*. We will denote it by $\mathcal{B}(\mathbb{R})$ or simply by $\mathcal{B}$. If $E \in \mathcal{B}$ then we will call E *Borel-measurable*[3] or simply *Borel*.

There is a natural question to ask. We have added many sets to our standard topology. Have we added all of them?

Theorem 3.2.5. $\mathcal{B} \neq 2^{\mathbb{R}}$.

[3]The fact that we call these sets measurable is not random of course.

Proof. See Section 3.3.3. This is an interesting question, and not trivial. Very poorly behaved sets are measurable. □

Example 3.2.4. Consider the same with $\mathbb{R}^n$ with $n > 1$. The generating set can now equivalently be cubes or spheres. We will denote this σ-algebra as $\mathcal{B}_n$.

One final peace of terminology that we will use. It is sometimes nice to denote a set together with a σ-algebra on that set.

Definition 3.2.6. A tuple (X, A) such that $X \neq \emptyset$ and A is a σ-algebra on X is called a *measurable space*. Given a measurable space (X, A), we will call any set $E \in A$ measurable.

We have now defined measures and measurable spaces. It is natural to ask what is the connection between the two. The next section discusses this.

3.3 Measurability

Peach Pit, *Being so Normal*

We now get some meaningful results on the relationship between the two previous sections.

Theorem 3.3.1. *Let X be a set with an outer measure ω. Then the collection of measurable sets M with respect to ω is a σ-algebra. Moreover, $\omega|_M$ is countably additive (on disjoint subsets things are additive) and zero sets are measurable.*

Proof. We first show that zero sets are measurable. Notice that we will always have that

$$\omega(E) \leq \omega(E \cap Z) + \omega(E \cap Z^c)$$

this just follows from the sub-additivity of outer measures. But then if Z has measure 0, we will also have

$$\omega(E \cap Z) + \omega(E \cap Z^c) \leq \omega(E)$$

because the subset of a measure zero set must be measure zero.

To show that it is closed under finite union, we first show that it is closed under subtraction. This requires just De Morgan's laws and writing things out clearly. Start with two measurable E_1, E_2. We have

$$\begin{aligned}\omega(X) =& \omega(X \cap E_1) + \omega(X \cap E_1^c) \\ =& \omega(X \cap E_1 \cap E_2) + \omega(X \cap E_1 \cap E_2^c) + \omega(X \cap E_1^c \cap E_2) + \omega(X \cap E_1^c \cap E_2^c) \\ =& \omega(X \cap (E_1 \cap E_2^c)) + \omega(X \cap (E_1 \cap E_2^c)^c)\end{aligned}$$

We then have that it is closed under finite union because $A \cap B = (A^c \setminus B)^c$.

To show that we are closed under infinite union, it is enough to show that it holds for infinite disjoint union, by the above.

Suppose that $E = \bigcup^\infty E_i$, where all of the E_i are disjoint. Define, for each n, $F_n := \bigcup^n E_i$. We know that each F_n is measurable by above. Then, for each n we have that

$$\omega(X) = \omega(X \cap F_n) + \omega(X \cap F_n^c) \geq \sum^n \omega(X \cap E_i) + \omega(X \cap E^c)$$

But notice that we also have that $\omega(X) \leq \sum \omega(X \cap E_i) + \omega(X \cap E^c)$. Since the above holds for all n, we must have that there is equality (either the sum converges or it diverges; either way it is obvious). Countable additivity follows from a similar argument. □

This result gives us another way to think about measures. The outer measure approach is one that starts from basics, using our basic intuition for what our notion of size should be. There is another approach that starts with a σ-algebra, and then defines a "measure" on this set.

Definition 3.3.2. Let (X, A) be a measurable space. A *measure* on X is a function $\mu : S \to \mathbb{R} \cup \{\infty\}$ such that:

1. $\mu(\emptyset) = 0$
2. $\mu(E) \geq 0$ for all $E \in S$
3. μ is countably additive on disjoint subsets

Definition 3.3.3. Let (X, A) be a measurable space, and μ a measure on A. We call the tuple (X, A, μ) a *measure space*.

There is a natural question of whether the two notions are identical. Of course, for any of "our" measures, one gets a measure by this definition. The major difference is the σ-algebra that one gets. The discussion below on the Lebesgue measure starts to get at this.

3.3.1 Lebesgue Measurability

Now, of course we are headed to showing that the Lebesgue measurable sets are exactly the sets that are Borel (plus or minus zero sets—this is important!). We prove this in a few steps.

First, we show that half-spaces are measurable.

Lemma 3.3.4. *Half-spaces are measurable.*

Proof. Exercise. It is easy □

Corollary 3.3.5. *Boxes are measurable.*

Corollary 3.3.6. *Open and closed sets are measurable.*

Proof. Take a countable basis of boxes. □

Definition 3.3.7. Let G_δ be the set of the countable intersection of open sets.
Let F_σ be the set of the countable union of closed sets.

Theorem 3.3.8. *A set E is measurable if and only if there exist $G \in G_\delta$ and $F \in F_\sigma$ such that $F \subset E \subset G$ and $m(G \setminus F) = 0$.*

Proof. Assume that E is bounded and measurable, so that it is contained in some box R. Abusing notation, set $E^c = R \setminus E$. Let $\mathcal{R}_n$ be an open cover of E such that $\sum_{R \in \mathcal{R}_n} |R| \leq m(E) + \frac{1}{n}$.

Define recursively $U_m = \bigcup_{\mathcal{R}_m} R \cap U_{m-1}$. Do the same process for E^c, and define those sets V_m. Because E is measurable, we have that $m(U_m \setminus E) = m(U_m) - m(E) \to 0$. Define

$G \coloneqq \bigcap U_m$. By the continuity of measure, we have that $m(G) = m(E)$. Similarly, define $K_m \coloneqq R\setminus$ is closed and increasing, with $K_m \subset E$. The result follows with $F \coloneqq \bigcup K_m$.

If E is not bounded, we need to do a little more work.

I leave the other direction as an exercise, as it was left to me. □

Aside: So, except for a zero set, all measurable sets are Borel. Can we show that all zero sets are Borel? The answer, which may be surprising, is no. See a little bit later (Section 3.3.2). This may seem like a semantic point (and it is), but the point is that often times people start directly with the definition of the Lebesgue measure on Borel sets. While it is certainly defined on Borel sets, the Borel σ-algebra is not the entire set of measurable sets. It is "almost" all Lebesgue measurable sets (which is why often people don't care), but it is but a point in all measurable sets.

We also can say a little bit more about what happens in higher dimensions

Theorem 3.3.9. *Let $E \subseteq \mathbb{R}^n, F \subseteq \mathbb{R}^m$. Then if A, B are measurable, we have that $A \times B$ is measurable in $\mathbb{R}^{n+m}$ with $m(A \times B) = m(A) \cdot m(B)$. This is with the convention that $\infty \cdot 0 = 0 \cdot \infty = 0$.*

Proof. Suppose first that $E = Z$ is a zero set. Notice that $Z \times \mathbb{R}^m$ will be a zero set because we can cover $Z \times [-k, k]^m$ by sets with measure at $\epsilon/2^k$. Taking the union over $k \in \mathbb{N}$ gives the result.

If E, F are both cubes, then this is trivial. Now, if they are both open sets, we can get disjoint cubes $\{R_i^E\}$ such that $m(E \setminus \cup R_i^E) = 0$, and similarly for F. This case then follows from the two previous cases.

Now, for general E, F, if they are bounded we apply Theorem 3.3.8 and the result follows. For unbounded we can simply take the union over increasing bounded subsets. □

3.3.2 The Cantor Set

To show that not all measurable sets are Borel, we must do two things. First, we must observe that the cardinality of the Borel σ-algebra is the cardinality of the reals. This observation follows from $\mathbb{R}$ being second-countable, and all Borel sets being a countable number of operations away from an open set.

Second, we construct a measure zero set that is uncountable. Why will this prove the claim? Well, we know that (1) $|2^{\mathbb{R}}| > |\mathbb{R}|$ and (2) the subset of any measure zero set has measure zero. That is all.

We now begin the construction of the Cantor set. Make sure to draw a picture.

There are two equivalent constructions. I will first go through the graphical interpretation, and then explain why this is equivalent to the ternary expansion definition.

1. Start with the unit interval $C_0 = [0, 1]$
2. Take out the middle third. I.e., $C_1 = ([0, 1/3] \cup [2/3, 1])$
3. Take out the middle third of each smaller interval:

$$C_2 = ([0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1])$$

4. Repeat the process indefinitely:

$$C = \bigcap_{i=0}^{\infty} C_i$$

What is another way to envision this? Just as we can write out the binary expansion of any real number, we can also write out the ternary expansion. For any $x \in [0,1]$ we have that

$$x = \sum_{i=1}^{\infty} x_i 3^{-i}$$

for some x_i. The construction of the Cantor set is exactly removing at the i-th stage any number x that in its ternary expansion has $x_i = 1$. So, we are left exactly with those numbers that have only 0's and 2's in their ternary expansion.[4]

Lemma 3.3.10. *The Cantor set is uncountable. That is, we have $|C| = |\mathbb{R}|$.*

Proof. Use the ternary expansion definition of the set. Can we make a bijection to the unit interval when we represent those numbers in their binary expansion? □

Theorem 3.3.11. *Not all measurable sets are Borel. The set of measurable sets has cardinality given by $|2^{\mathbb{R}}|$.*

3.3.3 Vitali Sets: Non-measurability

We have just shown in the above section that the set of measurable sets has cardinality given by $|2^{\mathbb{R}}|$. The set of all subsets of $\mathbb{R}$ also has cardinality given by $|2^{\mathbb{R}}|$. A natural question is whether every set is measurable. I do not know whether this is surprising or not, but the answer is no. Not every set is measurable.

Theorem 3.3.12. *There exists a non-measurable set.*

Proof. Note that this relies heavily on the Axiom of Choice. In some ways it is obvious that it is going to be needed. Non-measurable sets are extremely poorly behaved, and so analytically writing down a non-measurable set is going to require some funny business.

First, start with the unit interval $[0,1]$. Define a relation $\sim$ on $[0,1]$ that is given by

$$x \sim y \iff x - y \in \mathbb{Q}$$

It can easily be seen that $\sim$ is an equivalence relation.

Now, construct a set V (a "Vitali" set) by choosing exactly one element from each equivalence class.[5]

We show that V is not measurable. We do this by showing that if V were measurable, then it would simultaneously need to have measure zero and positive measure.

Consider now, for each $r \in \mathbb{Q} \cap [-1,1]$, the set $V_r := \{v + r | v \in V\}$. We will have that $V_r \subseteq [-1,2]$. Notice as well that $m^*(V_r) = m^*(V)$ and that $V_r \cap V_{r'} = \emptyset$ for all $r \neq r'$.

Consider as well that there will be countably many of these V_r, and that

$$[0,1] \subseteq \bigcup_r V_r \subseteq [-1,2]$$

[4] What about $\frac{1}{3}$? Well, we can write it as 0.1 or as 0.0222..., can we not?

[5] AoC-ed.

The first inclusion follows from the definition of V, and the second inclusion follows from algebra. Finally, notice that V is measurable if and only if V_r is measurable (for each r). Then, if V were measurable (towards a contradiction), we would have that

$$m([0,1]) \leq m\left(\bigcup_r V_r\right) \leq m([-1,2])$$

But then, we would have that

$$1 \leq \sum_r m(V_r) = \sum_r m(V) \leq 3$$

We cannot simultaneously have these inequalities satisfied. The first implies that $m(V) > 0$, so that $\sum_r m(V) = \infty$ (because it is the infinite sum of a positive number). The second implies that $m(V) = 0$, so that $\sum_r m(V) = 0$. We conclude that V is *not* measurable. □

This construction is not unique to the unit interval. In fact, it can be shown that any positive measure measurable set contains a non-measurable set. More than this, the set of non-measurable sets also has cardinality $|2^{\mathbb{R}}|$.

3.4 Measurable Functions

Gold Spectacles, *Gold Spectacles*

We now have a notion of "good" sets. Just as in topology, we can now define what it means for a function to be "good" in this context. The definition here is going to be identical in notion, just with now corresponding notions.

Definition 3.4.1. Suppose that (X, A_X) and (Y, A_Y) are two measurable spaces. We say that a map $f : X \to Y$ is *measurable* with respect to A_X, A_Y if the pre-image of every measurable set is measurable:

$$f^{-1}(E_Y) \in A_X \text{ for all } E_Y \in A_Y$$

What is our intuition for topological continuity? We know that it is equivalent to sequential continuity, so that things are "close" to one another remain "close" to one another after we apply the function (of course, this is just the $\epsilon - \delta$ definition of continuity from first-year calculus).

There is a similar intuition here, but we can't apply exactly the same logic. There is no notion of "closeness" in a measurable space. Instead, what I think is good to think about is a notion of differentiation.

Suppose that $x, x' \in X$ are such that for all $E \in A$ we have that $x, x' \in E$ or $x, x' \notin E$. That is, from the point of view of A we cannot differentiate x, x'. Then, if f is measurable we must have that A_Y cannot differentiate $f(x)$ and $f(x')$. If we had some $E_Y \in A_Y$ such that $f(x) \in E_Y$ and $f(x') \notin E_Y$, then we could use f to somehow gain additional information about x, x'. What measurability is forcing upon our structure is that information cannot be gained when we apply a function.

Aside: There is another (equivalent) definition of measurability of functions to $\mathbb{R}$ that I think is more intuitive, and also will be useful for thinking about the expanded notion of integration in this context. We will define it just for positive functions (which will be our interest in probability spaces anyway), but the definition can be easily extended to general functions to $\mathbb{R}$.

Definition 3.4.2. Let (X, A_X) be a measurable space. Define the undergraph of a function $f : M \to [0, \infty)$ to be the set

$$\mathcal{U}f := \{(x, y) \in X \times [0, \infty) | 0 \leq y \leq f(x)\}$$

We say that f is measurable $\mathcal{U}f$ is (Borel-)measurable with respect to the product measure.[6]

We are also going to be interested in a specific class of measurable functions, where we restrict the measurable spaces to just being $\mathbb{R}$.

Definition 3.4.3. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is measurable with respect to $\mathcal{B}, \mathcal{B}$. Then we say that f is *Borel-measurable*.

We get some basic results, which have their direct analogues from topology.

Proposition 3.4.4. *Let (X, A_X) and (Y, A_Y) be measurable spaces with $f : X \to Y$. If there exists a collection $B \subseteq 2^Y$ such that*

1. $\mathcal{A}(B) = A_Y$
2. $f^{-1}(E) \in A_X$ *for all* $E \in B$

then f is measurable.

Corollary 3.4.5. *If $f : \mathbb{R} \to \mathbb{R}$ is continuous,[7] then f is Borel-measurable.*

Proof. Exercise. □

We have another theorem, where we look just at slices.

Theorem 3.4.6. *Let (X, A_X) be a measurable space. Let $f : X \to \mathbb{R}$. Then the following are equivalent:*

1. *f is measurable*
2. $\{x \in X | f(x) > \alpha\} \in A_X$ *for all* $\alpha \in \mathbb{R}$
3. $\{x \in X | f(x) \leq \alpha\} \in A_X$ *for all* $\alpha \in \mathbb{R}$

Proof. This follows from basic set theory, so I leave it as an exercise. □

With the above results, we can start to talk about specific convergence results that are the primary goal of this portion.

I close this section with a specific definition that will clean up our language somewhat is the following. We can think of a measurable function p from a measurable space to $\{0, 1\}$ being a "proposition." If $p(x) = 1$, then p is "true" at x, and otherwise it is "false" at x.

Definition 3.4.7. Let (X, A, μ) be a measure space equipped with a , with $p : X \to \{0, 1\}$ be a measurable function. We say that p happens "almost everywhere" if

$$\mu(p^{-1}(0)) = 0$$

[6]Similar to the product topology, the outer measure is defined on the correct notion of "boxes." From there we take the infimum as before. This isn't always unique, but all of the measures with which we will deal (in particular all finite measures, and the Lebesgue measure) it is.

[7]With respect to the standard topology. However, we could take some other topologies here...

3.5 Lebesgue Integration and Convergence Theorems

San Cisco, *Under the Light*

Just as we had two definitions for a function to be measurable, we will have two (equivalent) definitions of the integral.

The first definition will correspond to the second definition of measurability (Definition 3.4.2) I introduce it first because it requires less notation, and I think it corresponds to the graphical interpretation that we have for the integral as the "area under the graph."

Definition 3.5.1. Let (X, A, μ) be a measure space, and let $f : X \to [0, \infty)$ be a measurable function. Let m be the corresponding product measure. Then, we say that f is *integrable* if $m(\mathcal{U}f) < \infty$, and define

$$\int_X f := m(\mathcal{U}(f))$$

The set of all integrable functions from $\mathbb{R}$ to $\mathbb{R}$ is denoted $\mathcal{L}^1$. We define

$$\mathcal{L}^p := \{f | f^p \text{ integrable}\}$$

The other definition corresponds loosely to the Reimann sum intuition that we get from first-year calculus. Unfortunately, like that definition, this definition requires a little more machinery.

Definition 3.5.2. Let (X, A) be a measurable space. A function $s : X \to \mathbb{R}$ is called an *A-simple* function if:

1. s is measurable
2. There exists a finite set $C \subseteq \mathbb{R}$ such that $s(X) \in C$

I.e., they are functions that only take finitely many values.

Theorem 3.5.3. *A function $s : X \to \mathbb{R}$ is A-simple if and only if there is a finite partition $E_1, ..., E_m$ of X and $c_1, ..., c_m \in \mathbb{R}$ such that*

$$s(x) \equiv \sum_{i=1}^{m} c_i \mathbb{1}(x \in E_i)$$

We call such representation a canonical representation of s.

With this definition and result, we can explicitly state the other definition of the Lebesgue integral.

Definition 3.5.4. Let (X, A, μ) be a measure space, and let $s : X \to \mathbb{R}$ be a simple and measurable function, with $s(x) = \sum_{i=1}^{m} c_i \mathbb{1}(x \in E_i)$ a canonical representation. Then, we define the Lebesgue integral of s to be

$$\int_X s d\mu := \sum c_i \mu(E_i)$$

Now, let $f : X \to \mathbb{R}_+$ be any measurable function. We define

$$\int_X f d\mu := \sup_{s \leq f,\, s\ A-\text{simple}} \left(\int_X s d\mu \right)$$

For simple functions, this is an obvious definition. For arbitrary functions, we take all simple functions that are less than f, and take the supremum over all of their integrals. This definition is really the generalization of Riemann sums from first-year calculus. There we used rectangles. Here we need to use more general sets because the functions can be more complicated. For example, consider $f(x) = \mathbb{1}(x \in \mathbb{Q})$.

Proposition 3.5.5. *The two definitions are equivalent.*

Proof. See Theorem 3.3.9. This is really just a slight generalization of that theorem to the situation where one of the elements in a product space is not necessarily $\mathbb{R}$. □

When I was learning Lebesgue Integration, I had the differences between Riemann and Lebesgue integration described to me as follows. Suppose that it is you job to count the number of tokens in a room. All of the tokens are stacked in piles. In Riemann integration, what you are doing is going to each stack of tokens, counting them individually, and then adding that sum to your running total. In Lebesgue integration, what you are doing is counting how many stacks have one token, then how many have two tokens (multiplying by 2 and then adding to your running sum) and so on. I leave it to you to determine which method you would prefer.

We can similarly get the definition of an integral over a region.

Definition 3.5.6. The *restricted integral* of f to a set E is defined by

$$\int_E f := \int_X f \cdot \mathbb{1}_E$$

We can get some results regarding the integral that should appear obvious.[8]

Proposition 3.5.7. *Throughout, fix (X, A, μ) and two measurable functions $f, g : X \to \mathbb{R}_+$. We have that*

1. *If $f(x) \leq g(x)$ for all $x \in X$, then $\int f d\mu \leq \int g d\mu$*
2. *If $E, F \in S$ with $E \subseteq F$, we have that $\int_E f d\mu \leq \int_F f d\mu$*
3. *If $\mu(E) = 0$, then $\int_E f d\mu = 0$*
4. *If f is such that $\int f d\mu = 0$, then $f = 0$ almost everywhere*
5. *$\int_E (f + g)\, d\mu = \int_E f d\mu + \int_E g d\mu$*

3.5.1 Convergence Theorems

One of the beauties of measurable functions is that the point-wise limit of measurable functions is measurable.[9] Many functions that aren't Riemann-integrable are the point-wise limit of Riemann-integrable functions. Because we have this closedness of the set of measurable functions, we can talk about the convergence of integrals. We all want to be able to exchange limits and integrals, and there are good guidelines for when we can when we are dealing with measurable functions.

[8] Proving them is not always trivial. The undergraph definition makes some of these very difficult to compute. This highlights why the two definitions should be seen as complementary, because they make proving different things often easier.

[9] This hasn't actually been proven, but it is a corollary of the theorems below. Ask if you are unsure.

Theorem 3.5.8 (Monotone Convergence Theorem)**.** *Suppose that $(f_n) \nearrow f$ is a non-decreasing sequence of measurable functions that converges point-wise to f almost everywhere. Then we have that*

$$\lim_{n\to\infty} \int_E f_n \, d\mu = \int_E f \, d\mu$$

Proof. If measures are upwardly continuous, then this follows immediately from the undergraph definition of the integral. □

This theorem is extremely powerful in that the conditions are easy to check, and it has wide applicability.

For example, it makes computing integrals fairly straightforward. We need check only for *one* sequence of simple functions that converges to a function f in order to compute the integral of f.

One of the most useful applications is that it gives us an easy way to compute the integrals of many sums.

Corollary 3.5.9. *If (f_n) are all measurable, then we have that*

$$\int_E \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu$$

Why do we need the monotonicity condition?

Example 3.5.1. Consider the sequence of functions $f_n : [0,1] \to \mathbb{R}$ given by

$$f_n(x) = \begin{cases} n & \text{if} \quad x \leq \frac{1}{n} \\ 0 & \text{else} \end{cases}$$

We have that $f_n \to f(x) \equiv 0$ point-wise clearly. However, we have that

$$\int_{[0,1]} f_n = 1 \quad \text{for all } n$$

but

$$\int_{[0,1]} 0 = 0$$

Does the direction of the monotonicity really matter? It turns out that it does, because of issues with integrability.[10]

Example 3.5.2. Consider a sequence of functions $f_n : [0,1] \to \mathbb{R}$ given by

$$f_n(x) = \begin{cases} 0 & \text{if} \quad x = 0 \\ \frac{1}{nx} & \text{else} \end{cases}$$

Then we will have that $f_n \searrow 0$ point-wise clearly. However, we have that

$$\int_{[0,1]} f_n = \infty \quad \text{for all } n$$

[10]Recall that a functions is integrable if it is measurable and its undergraph has finite measure.

This second example motivates the following lemma

Lemma 3.5.10. *Suppose that $f_n \searrow f$ with each f_n integrable. Then we have that*

$$\lim_{n\to\infty} \int_E f_n = \int_E f$$

Proposition 3.5.11. *Suppose that f_n are all measurable. Then we have that*

$$\underline{f}_n(x) \coloneqq \inf_{m\geq n} f_m(x)$$

is measurable.

Proof. Notice that

$$\mathcal{U}(\underline{f}_n) = \bigcap_{m\geq n} \mathcal{U}(f_m)$$

These are all measurable, proving the claim. □

Both of these examples motivates the following theorem.

Theorem 3.5.12 (Lebesgue Dominated Convergence)**.** *Suppose that f_n are all measurable, that $f_n \to f$ almost everywhere, and there exists a g integrable such that $f_n \leq g$ almost everywhere for all n. Then we have that f is integrable, and that*

$$\lim_{n\to\infty} \int_E f_n d\mu = \int_E f$$

Proof. Define

$$\underline{f}_n(x) \coloneqq \inf_{k\geq n} f_k(x)$$
$$\overline{f}_n(x) \coloneqq \sup_{k\geq n} f_k(x)$$

These functions are both clearly measurable (use the undergraph definition) Then, we have that $g \geq \overline{f_n}(x) \geq \underline{f_n}(x)$ for all n, so that all are integrable.

We conclude by first noting that $\overline{f}_n \searrow f$ and $\underline{f}_n \nearrow f$ and by applying Lemma 3.5.10 and Theorem 3.5.8. □

Corollary 3.5.13. *The point-wise limit of measurable functions is measurable.*

Proof. Notice that if $f_n \to f$, then we have that $\underline{f}_n \nearrow f$. The undergraph of f is clearly then measurable. □

There is one last convergence "lemma" that I would like to show you. It is a little more situational, but it is still relevant to at least have seen.

Lemma 3.5.14 (Fatou's Lemma)**.** *Suppose that (f_n) is a sequence of measurable functions. Then we have that*

$$\int_E \liminf_n f_n \leq \liminf \int_E f_n$$

Proof. Now, we have that $\liminf_n f_n = \lim_{n\to\infty} \underline{f}_n =: f$. As we know that $\underline{f}_n \leq f_n$, we are done once we "stare at the definition of liminfs." □

3.5.2 $\mathcal{L}_p$ Spaces

I want to return briefly to Definition 3.5.1. Specifically, I want to return to the latter part of the definition where I defined $\mathcal{L}_p$ spaces. These spaces are actually very important in a wide-range of applications. Specifically, I want to define a (natural) norm on these spaces, and then talk about two notions of convergence.

Definition 3.5.15. Let (X, A, μ) be a measure space and $p \in [1, \infty)$. Then we define

$$||\cdot||_p : \mathcal{L}_p \to \mathbb{R}$$

as a norm[11] by

$$||f||_p = \left(\int |f|^p d\mu \right)^{1/p}$$

Definition 3.5.16 (Convergence in $\mathcal{L}_p$)**.** Let (X, A, μ) be a measure space and $p \in [1, \infty)$. Let (f_n) be a sequence in $\mathcal{L}_p$ and $f \in \mathcal{L}_p$. Then we say that (f_n) converges to f in $\mathcal{L}_p$ if

$$||f_n - f||_p \to 0$$

We will write $f_n \to_p f$.

We have a final notion of convergence.

Definition 3.5.17 (Convergence in Measure)**.** Let (X, A, μ) be a measure space, and let (f_n) be a sequence of measurable functions. Then we say that (f_n) converges to f in measure μ (written $f_n \to_\mu f$) if for all $\epsilon > 0$ we have

$$\lim_{n\to\infty} \mu\left(\left\{x \in X \middle| \, |f_n(x) - f(x)| \geq \epsilon\right\}\right) = 0$$

We have one result that I will state here that will become relevant when we move to talking about probability theory (soon). It just discusses how the different types of convergence are nested within one another.

Theorem 3.5.18. *If $f_n \to_p f$ for some $p \geq 1$, then we have that $f_n \to_\mu f$.*

Proof. Left as an exercise. □

[11] Technically this is a "semi-norm" because it cannot distinguish because of zero-sets, but I rather consider this a norm on the equivalence classes of functions that differ only on a measure zero set.

Chapter 4

Probability Theory

As I've discussed (hopefully), there are two main applications for measure theory. The first is dealing with models where there is a continuum of agents. This has applications in basically every field, but you see it most commonly in macro.[1] The other main application is in probability theory. This is of course applicable to every field, and is extremely important foundational to why economics is important.

So, the first thing we need to do is convert the above theory into language that we can use to describe probability theoretic elements.

So, we start a set of outcomes Ω. We can think of this as describing the underlying "states of the world."

Example 4.0.1. A good running example is going to be an environment where we roll two (independent and fair) six-sided dice. Here, we are going to have that

$$\Omega = \{(x, y)|x, y \in \{1, 2, 3, 4, 5, 6\}\}$$

We then are going to bestow a σ-algebra S on this space and call this the "Sample Space." I like to think about this as sets that we can "see" in the states of the world. In the above example, we could have that S contains all of the outcomes, so that S is the discrete σ-algebra. We could also have that S distinguishes only between different sums. That is,

$$S = \sigma\left(r_n := \{(x, y)|x + y = n\} \middle| n \in [2, 12] \cap \mathcal{Z}\right)$$

We know here that there is a more complicated state space underneath S, but we can't distinguish between whether the roll was $(2, 5)$ and $(4, 3)$. One can call this the "Settlers" (of Catan) σ-algebra.[2]

A third possible σ-algebra, which we will call the "Twilight-t" σ-algebra,[3] will be defined as follows:

$$S := \left\{\{(x, y)\middle|\text{one of } x, y \geq t\}, \{(x, y)\middle|\text{both } x, y \geq t\}\right\}$$

We will call a set $E \in S$ an event.

[1] Though, in terms of the technical uses, my understanding is that it is more common in micro theory than in macro.

[2] Why not the "Monopoly"? Because there doubles matter. Also, why "Settlers" and not "Catan"? For two reasons. First, I played Settlers before it was cool, and second, if you are playing with the Cities & Knights expansion (as you should if you choose to play this sad game) then distinguishing the dice becomes important.

[3] Of course, so-named because of "Twilight Imperium."

Finally, we need our notion of measure. We will call this the probability measure and require that $p(\Omega) = 1$ (because probabilities are at most 1).

If S is the discrete σ-algebra, then we would have that $p(\{(x, y)\}) = \frac{1}{36}$ is the foundation of the probability measure. If S is the Settlers σ-algebra, then we have that

$$p(r_n) = \frac{1}{6} - \frac{1}{36}|7 - n|$$

is the foundation of this probability measure.

Definition 4.0.1. Let Ω be a set, S a σ-algebra on Ω, and $p : S \to [0, 1]$ a probability measure on S. Then we call the tuple (Ω, S, p) a *probability space*.

With these basic definitions, we can get a lot of characteristics of probability that we are used to.

4.1 Independence and Bayes' Rule

Maisie Peters, *You Signed Up For This*

As important as events are, what is much more important is to how they are related. We are generally not interested in events in isolation (because that is just a matter of measuring), but rather how different events influence the probability of other events. For example, what is the probability that the roll is a six *given* (or, *conditional*) on the roll being even?

We have a few definitions, and then some trivial results.

Definition 4.1.1. Let (Ω, S, p) be a probability space. Given $B \in S$ such that $p(B) > 0$ we can define the *conditional probability* of an event A given B as

$$p(A|B) := \frac{p(A \cap B)}{p(B)}$$

Proposition 4.1.2. *$p(\cdot|B)$ is a probability measure on S.*

Proof. Exercise in recalling the definition of a measure, and what is the number 1. □

Equipped with the definition of conditional probability, we can define now another classical term, the independence of two events. Two events are called independent if they have no relation to one another. If I roll two dice, the probability of the first being a six is independent of the second being a three.

Definition 4.1.3. Let (Ω, S, p) is a probability space. We say that $A, B \in S$ are *independent* if

$$p(A \cap B) = p(A)p(B)$$

Now suppose that $p(C) > 0$. We say that A, B are *conditionally independent given C* if

$$p(A \cap B|C) = p(A|C)p(B|C)$$

Notice that this in general is also defined for zero-probability events (which are independent of everything, including themselves!).

Lemma 4.1.4. *Events A, B are independent if and only if A, B^c are independent.*

Proof. Write out the conditions. □

We now get our first result with a name for this chapter.

Theorem 4.1.5 (Bayes' Rule)**.** *Suppose that $p(A), p(B) \in (0,1)$. Then we have that*

$$p(B|A) = \frac{p(A|B)p(B)}{p(A|B)p(B) + p(A|B^c)p(B^c)}$$

Proof. Exercise. □

4.2 Random Variables

Valley, *Lost in Translation*

In most cases, we are interested not only in the underlying state-space, but we are interested in what is called a "Random Variable," which is an association between the state-space and $\mathbb{R}$. Now, it can be easy (for me, at least) to confuse what exactly is happening here. I think that this comes up a lot because we often think about the state-space as being itself a subset of $\mathbb{R}$, and then we consider the random variable as simply inclusion (think of the dice example). My goal is to show which is in general a bad way to think about things, and to keep them separate. But I am getting ahead of myself.

Definition 4.2.1. We call a function $X : \Omega \to \mathbb{R}$ a random variable (with respect to S) if X is measurable (with respect to S and the $\mathcal{B}$).

Example 4.2.1. Suppose that we let Ω be as above. Let $X : \Omega \to \mathbb{R}$ be defined by $X(x, y) = x$. Then we have that X is a random variable with respect to the discrete σ-algebra, but not with respect to the Settlers σ-algebra.

The former is obvious because any function will be measurable. The second is because, for instance,

$$X^{-1}(\{2\}) = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)\}$$

which is not measurable with respect to the Settler's σ-algebra.

Other examples include

$$X(x,y) = \begin{cases} 0 & \text{if} \quad x, y < 3 \\ 1 & \text{if} \quad x \geq 3, y < 3 \\ 1 & \text{if} \quad x < 3, y \geq 3 \\ 2 & \text{if} \quad x, y \geq 3 \end{cases}$$

or

$$X(x,y) = (x+y)^2$$

Once we have a random variable, we want to start thinking about the distribution of the random variable X.

Definition 4.2.2. Let (Ω, S, p) be a probability space and X a random variable. Then the *distribution* of X (also called the *induced probability*) of X is the (measurable) function

$$\begin{aligned} p_X : &\mathcal{B} \to \mathbb{R} \\ &B \mapsto p(X^{-1}(B)) \end{aligned}$$

We say that X, Y have the same distribution is $p_X \equiv p_Y$.

Definition 4.2.3. A random variable X is discrete if $P(X \in \chi_0) = 1$, where $\chi_0 = \{x \in \mathbb{R} | P(X = x) > 0\}$. If X is discrete, we define the *probability mass function* of X to be $\text{pmf}_X(x) \coloneqq P(X = x)$.

Example 4.2.2. Bernoulli, Uniform for Finite, Binomial, etc. (can use the above state space).

Definition 4.2.4. A random variable is said to be (absolutely) continuous if there is a function $f : \mathbb{R} \to \mathbb{R}$ such that

$$p(X \in (a, b]) = \int_a^b f(x)dx$$

We call f the *probability density function* of X.

It is important here that the x is taking a value in the co-domain, not the domain of X: $x \in \mathbb{R}$, not Ω. Note that if X is continuous, then $\text{pdf}_X(x) = \frac{\partial}{\partial x} P(X \leq x)$.

Example 4.2.3. uniform, exponential, normal, etc.

Definition 4.2.5. For a random variable X we define the *cumulative distribution function* of X to be

$$\text{cdf}_X(x) = F_X(x) \coloneqq p(X \leq x)$$

Proposition 4.2.6. *Cumulative distribution functions have the following characteristics:*

1. *F is non-decreasing*
2. $\lim_{x \to \infty} F(x) = 1 = 1 - \lim_{x \to -\infty} F(x)$
3. *F is right-continuous*
4. *Defining*
$$F(x^-) \coloneqq \lim_{y \nearrow x} F(y)$$
gives us that $F(x^-) = p(X < x)$
5. $p(X = x) = F(x) - F(x^-)$

An equivalent definition of a continuous random variable is that F_X is a continuous function. Similarly F_X being a step function is equivalent to X being a discrete random variable.

It is often convenient for us to talk only in terms of the distributions of functions themselves. The problem with doing this and ignoring the state-space is that it can cause us to

lose sight of the underlying structure. A powerful reason for this is when we start thinking in terms of more than one random variable.

Suppose that we have two random variables X, Y on (Ω, S, p). We might be interested in how related these two distributions are. In order to think about that, we need to think about their *joint distribution*, which is going to be

$$\begin{aligned} p_{X,Y} : &\mathcal{B}^2 \to \mathbb{R} \\ &B \mapsto p((X,Y) \in B) \end{aligned}$$

So now, we care about the probability over a two-dimensional space, and are interested in how X and Y co-move. Draw some pictures using the dice example.

We can similarly get a cumulative distribution function $F_{X,Y}(x,y) = p(X \leq x, Y \leq y)$.

Definition 4.2.7. We say that X, Y are independent if $F_{X,Y} = F_X \cdot F_Y$.

This is actually a result, because the independence would be proved by looking at events in $\mathcal{B}^2$. But this is obviously true by the generating elements of that set.

Thinking about joint random variables only makes sense if there is some underlying state-space. Otherwise there is no way to talk about two random variables living together.

4.3 Moments

Virginia to Vegas, *life gets interesting...*

In the above, random variables have been defined in an abstract manner for the most part. Now, we want to think about situations when the values that random variables take have some important meaning. Thinking about the number of successes in a dice roll, or about the random variable being income gives the real number value some relevant meaning. So, we want to think about what the "average" realization is going to be.

Definition 4.3.1. Let (Ω, S, p) be a probability space. Given a random variable, we define the *expected value* of X as:

$$\mathbb{E}[X] := \int_S X dp$$

where of course this integral is the Lebesgue integral with respect to the measure p.

If X is continuous with pdf f, we can write this as

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

Here also is where the discussion of $\mathcal{L}_p$ spaces becomes important. If $\mathbb{E}[X]$ is well-defined, we say that $X \in \mathcal{L}_1(p)$.[4]

Definition 4.3.2. We define the *n-th moment* of X to be

$$\mathbb{E}[X^n] = \int_S X^n dp$$

[4]It is unfortunate that we are using p here, but it is what the notation is; alas.

and writing $\mu = \mathbb{E}[X]$, we define the n-th central moment of X to be

$$\mathbb{E}[(X-\mu)^n] = \int_S (X-\mu)^n dp$$

We will call the second central moment of X to be its *variance*.

There are also names for the third and fourth central moments (skewness and kurtosis).

Proposition 4.3.3. *If we have that $\mathbb{E}(|X|^t) < \infty$ for some $t > 0$ then we have $\mathbb{E}(|X|^s) < \infty$ for all $0 \leq s \leq t$*

Proof. Notice that $|X|^s \leq 1 + |X|^t$. □

Now, suppose that X has finite expectation. Then we have that

$$\mathbb{E}[X] = \int_0^\infty P(X > z)dz - \int_{-\infty}^0 P(X < z)dz = \int_{-\infty}^\infty x dF(x)$$

In the case that X is continuous, then this can easily be shown by integration by parts (but it holds generally). This representation is useful because it gives us the following result.

Theorem 4.3.4 (Law of the Unconscious Statistician)**.** *Let X be a random variable, and $g : \mathbb{R} \to \mathbb{R}$ measurable. Then $Y = g(X)$ is a random variable, with expectations equal to*

$$\mathbb{E}[Y] = \int g(x)dF(x) \quad \textit{or} \quad \sum g(x)\mathrm{pmf}_x(x)$$

Proof. Assume that X is non-negative and $g : \mathbb{R}_+ \to \mathbb{R}_+$. Then we have that

$$\begin{aligned}
\mathbb{E}[Y] &= \int_0^\infty P(g(X) > z)dz \\
&= \int_0^\infty \int_\Omega \mathbb{1}(g(x) > z)dF(x)dz = \int_\Omega \int_0^\infty \mathbb{1}(g(x) > z)dzdF(x) \\
&= \int_\Omega g(x)dF(x)
\end{aligned}$$

□

Note also that the expectation is a linear functional. This means that we have that

$$\begin{aligned}
\mathbb{E}[aX + b] &= a\mathbb{E}[X] + b \\
\mathbb{E}[X + Y] &= \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

Notice that the latter again is relying on the definition of a common state space.

A final result that I again appears obvious, but it is important to know.

Theorem 4.3.5. *Let X, Y be two independent random variables. Let $g, h : \mathbb{R} \to \mathbb{R}$ be measurable, such that gX and hY have finite expectation. Then we have that*

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)]$$

Notice that the converse doesn't hold in general! This is one of those things that may not seem obvious but is obvious: take $g \equiv 0 \equiv h$. A more interesting, but similarly intuitive result is that if this equality holds for all g, h, then X, Y are independent. This has the flavour of calculus of variations to me: we can use local properties to derive global conditions..

4.4 Inequalities

Billy Raffoul, *A Few More Hours at YYZ*

One of the things that we often use expectations for is to bound probabilities. I will state some and hope that this can possible be a dictionary in which you find use at some point.

Theorem 4.4.1 (Chebychev's Inequality)**.** *Suppose that X has mean μ and variance σ^2. Then we have, for all $\alpha \geq 0$, that*

$$P(|X - \mu| \geq \alpha\sigma) \leq 1/\alpha^2$$

Proof. Now, we have that

$$P(|X - \mu| \geq \alpha\sigma) = \mathbb{E}\left[\mathbb{1}\left(|X - \mu| \geq \alpha\sigma\right)\right] \leq \mathbb{E}\left[\frac{|X-\mu|^2}{\alpha^2\sigma^2}\mathbb{1}\left(|X - \mu| \geq \alpha\sigma\right)\right] = \frac{1}{\alpha^2}$$

□

Theorem 4.4.2 (Markov's Inequality)**.** *Suppose that $X \geq 0$ (that is, X only takes non-negative values) and has $\mathbb{E}[X] = \mu < \infty$. Then for all $\alpha \geq 0$ that*

$$P(X \geq \alpha) \leq \mu/\alpha$$

Proof. We have that

$$P(X \geq \alpha) = \int_0^\infty \mathbb{1}(X \geq \alpha) dF(x) \leq \int_0^\infty \frac{X}{\alpha}\mathbb{1}(X \geq \alpha) dF(x) \leq \mathbb{E}[X/\alpha] = \frac{\mu}{\alpha}$$

□

Now, the following you should have seen at some point (I hope), although almost certainly in a different context.

Theorem 4.4.3 (Cauchy-Schwarz)**.** *Let X, Y be two random variables having finite second moment. Then*

$$\left(\mathbb{E}[XY]\right)^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

The following is a mathematical result and then statistical application with which I have an unfortunate history. Once you have seen the proof though, it is quite cute and is a useful result.

Lemma 4.4.4 (Young's Inequality)**.** *Let $p, q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then we have that for all $x, ygeq0$ that*

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}$$

Proof. First, notice that the result holds for $y = 0$. Now, fix $y > 0$. Then, define the function

$$\varphi(x) \coloneqq \frac{x^p}{p} - xy + \frac{y^p}{p}$$

We wish to show that this is weakly larger than zero. Notice first that $\varphi(x)$ is differentiable at all $x \geq 0$ (as $p > 1$). Then, we can see explicitly that

$$\varphi'(0) < 0$$
$$\varphi'(x_0) = 0 \iff x_0 = y^{\frac{1}{p-1}}$$

Notice as well that this must be a minimum, as $\varphi'(0) = 0$, and the limit is at ∞. Notice that we have that $\frac{1}{p-1} = q - 1$. This means that we have

$$\varphi(x) \geq \varphi(x_0) = \frac{y^{p(q-1)}}{p} + \frac{y^q}{q} - y^q = y^q \left(\frac{1}{p} + \frac{1}{q} - 1 \right) = 0$$

This completes the proof. □

In what follows, quickly define $||X||_r \coloneqq (\mathbb{E}[|X|^r])^{1/r}$.

Theorem 4.4.5 (Hölder's Inequality)**.** *Let $p, q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then we have that*

$$\mathbb{E}[|XY|] \leq ||X||_p \, ||Y||_q$$

Proof. Let $U \coloneqq \frac{1}{||X||_p} X$ be a random variable, and similarly let $V \coloneqq \frac{1}{||Y||_q} Y$. Then we have that

$$\frac{\mathbb{E}[|XY|]}{||X||_p \, ||Y||_q} = \mathbb{E}[|UV|] \leq 1$$

This last inequality follows from Lemma 4.4.4. This proves the claim. □

The final inequality is the one that probably you will use the most? I feel at least that it comes up fairly frequently.

Theorem 4.4.6 (Jensen's Inequality)**.** *Suppose that φ is convex. Then we have that*

$$\varphi\left(\mathbb{E}[X]\right) \leq \mathbb{E}\left[\varphi(X)\right]$$

when these both exist and are finite.

Proof. Draw a picture. This is taking averages over more extreme objects, so it works. Alternatively, see Exercise 5.2.13 for a fun proof. □

Notice that the reverse inequality will hold for concave functions φ.

Example 4.4.1. If $0 < q \leq p$ and $\mathbb{E}[|X|^p] < \infty$, then we have that

$$||X||_p \geq ||X||_q$$

The proof of this comes immediately from Theorem 4.4.6.

Example 4.4.2. Suppose that $X, Y \in \mathcal{L}_p$ for $p > 1$. Then we have that

$$||X + Y||_p \leq ||X||_p + ||Y||_p$$

Proof is left as an exercise. *Hint:* can you use Lemma 4.4.4?

4.5 Convergence Results

HONNE, *Love Me / Love Me Not*

This is the last real section of this portion of the course. The purpose is to show two things. First, averages become averages. Second, everything looks normal if you have enough observations. We will require a little more notation, but it is likely that you will have seen all of these results before so I hope that it isn't overwhelming.

First we state some definitions of the different types of convergence. These are just the probability theory analogues of the types of convergence in Section 3.5:

Definition 4.5.1. A sequence of random variables X_n converges to X

(i) in *distribution* ($X_n \to_d X$) if $P(X_n \leq x) \to P(X \leq x)$ as $n \to \infty$, for all x.

(ii) in *probability* ($X_n \to_p X$) if for all $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \to 0$$

as $n \to \infty$.

(iii) *almost surely* ($X_n \to_{a.s} X$) if

$$P\Big(\limsup_{n\to\infty} |X_n - X| = 0\Big) = 1$$

(iv) in $\mathcal{L}_p$ ($X_n \to_{\mathcal{L}_p} X$) for $p > 0$ if

$$\mathbb{E}\left(|X_n - X|^p\right) \to 0$$

as $n \to \infty$.

How do we think about the difference between all of these definitions? First, we have the following relationships.

Proposition 4.5.2. *We have that*

$$X_n \to_{a.s} X \Rightarrow X_n \to_p X \Rightarrow X_n \to_d X$$
$$X_n \to_{\mathcal{L}_p} X \Rightarrow X_n \to_p X \Rightarrow X_n \to_d X$$

Example 4.5.1. To see why convergence in distribution is so weak, it is very important to remember the difference between a random variable and its distribution. Consider the setting as above, but now define two random variables:

$$X := \mathbb{1}\left(\omega \in \{1, 3, 5\}\right)$$
$$Y := \mathbb{1}\left(\omega \in \{2, 4, 6\}\right)$$

These two random variables have identical distributions, and yet are clearly unrelated in any meaningful sense.

Example 4.5.2. To see the difference between almost sure convergence and convergence in probability, I like the following example. Suppose that X_n is binary, equal to 1 if the n-th product in a line has a defect (it is a 0 if it has a defect). Suppose that X is the random variable that is always 1 (there is no defect). If $X_n \to_p X$, then this means that eventually there are no defects. However, there could still be infinitely many defects down the line, just the probability of there being a defect goes to zero. If $X_n \to_{a.s.} X$, then this means that there are only finitely many total defects.

Example 4.5.3. To see the difference between convergence in probability and convergence in $\mathcal{L}_p$, consider the example of

$$\begin{aligned} X_n :=& n \cdot \mathbb{1}\left(x \leq \frac{1}{n}\right) \\ X :=& 0 \end{aligned}$$

Here, we have that the $X_n \to X$ in probability, but for all $p \geq 1$ they do not converge in $\mathcal{L}_p$.

One nice thing to note is that these definitions tend to our notions of convergence:

Theorem 4.5.3. *Let $g \in \mathcal{C}^0$. Then we have*

(a) $X_n \to_{a.s.} X \Rightarrow g(X_n) \to_{a.s.} g(X)$

(b) $X_n \to_p X \Rightarrow g(X_n) \to_p g(X)$

(c) $X_n \to_d X \Rightarrow g(X_n) \to_d g(X)$

4.5.1 Law of Large Numbers

The law of large numbers is the first of the main results in this section. What it tells us is if you observe something enough, it begins to look like its average. There are two common versions of the theorem, the "weak" and "strong" versions. They are so named because of the type of convergence that they imply, but it is not that one is strictly better than the other. There are times when the weak convergence applies and the strong convergence does not. That being said, these cases actually apply only to extensions of the weak law (compared to what is written here). I state both because the weak law is easy to prove. Neither are written in their full generality.

First, the weak version.

Theorem 4.5.4 (Weak Law of Large numbers)**.** *Let X_n be independent random variables with mean μ. Moreover, let them be such that* $\text{Var}(X_n) \leq \sigma^2$ *for all n. Then we have that, defining*

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$$

that

$$\overline{X}_n \to_p \mu$$

Proof. By Chebyshev's inequality. □

Now, the strong version.

Theorem 4.5.5 (Strong Law of Large Numbers)**.** *Let X_n be independent random variables with mean μ and finite variances $\mathrm{Var}(X_n) \leq \sigma^2$, then we have that*

$$\overline{X}_n \to_{a.s.} \mu$$

Proof. Much more complicated, and beyond our scope. □

4.5.2 Central Limit Theorem

This second statement is that, if you squint hard enough, deviations look normal. As with the law of large numbers, there are many varying statements of this theorem. I present just one, because it is feasible:

Theorem 4.5.6 (Central Limit Theorem)**.** *Let X_n be i.i.d. random variables with mean μ and variance σ^2. Then, defining*

$$Z_n := \sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sigma}$$

we have that

$$Z_n \to_d \mathcal{N}(0, 1)$$

The Law of Large numbers states that averages go to averages. The Central Limit Theorem tells us how quickly things go to their averages.

Theorem 4.5.7 (δ-Method)**.** *Suppose that X_i is a sequence of random variables, and a_n a sequence of real numbers such that $a_n \to \infty$. Then if*

$$a_n(X_n - \mu) \to_d Z$$

for some random variables Z, and some constant μ, then for all $g \in \mathcal{C}^1$, we have

$$a_n \left(g(X_n) - g(\mu)\right) \to_d g'(\mu)Z$$

Why is this important? The application typically is to the central limit theorem (which is why I include it in this section). It allows us to easily apply transformations to random variables, and then compute what their asymptotic variance will be.

Aside: When we are discussing these X_n as sequences, we are thinking about them as being defined on the *same* state space. This can be hard to think about in standard examples, where we like to think about the X_n as different "realizations" of the same state space. For instance, in the dice example, we think about re-rolling the dice. How can we square these ideas?

The simplest way is to think about this is a very rich state-space, with some notion of independence between each of the "realizations." Take a (simpler) example of rolling a single die. We would have that

$$\Omega_i = \{1, 2, 3, 4, 5, 6\}$$

Endowed with the discrete σ-algebra, and a uniform probability measure: $p_i(\{\omega\}) = \frac{1}{6}$.

We can then think of the *entire* state space as being

$$\Omega = \Omega_1 \times \Omega_2 \times ... = \prod_{i=1}^{\infty} \Omega_i$$

With the corresponding product measure,[5] where we define

$$p\left(\prod_{i=1}^{\infty} E_i\right) = \prod_{i=1}^{\infty} p_i(E_i)$$

We can then think of the X_i as being equal to

$$X_i(\omega) = X_i\left((\omega_1, \omega_2, ...,)\right) = X_i(\omega_i)$$

This would mean that each X_i only depends on ω_i.

In general, it is often better in these settings to forego talking about the state space, since it becomes so complicated (see my footnote above). We will instead simply say that the X_n are "independent." In this case, we can't talk about complicated notions of convergence, because they depend very much on the underlying state space.

4.6 Examples of Random Variables

Bombay Bicycle Club, *Everything Else Has Gone Wrong*

We have one definition.

Definition 4.6.1. For X a random variable, we define the *moment generating function* of X: $\text{mgf}_x(t) \coloneqq \mathbb{E}[e^{tX}]$.

And also, one result.

Lemma 4.6.2. *If $\mathbb{E}[X^k]$ exists, then it is given by*

$$\mathbb{E}[X^k] = \frac{\partial^k}{\partial x^k} \text{mgf}_X(t)\Big|_{t=0}$$

Discrete Random Variables

- Bernoulli: parameterized by $p \in [0, 1]$

$$\begin{aligned} f(1) =& p \\ f(0) =& 1 - p \end{aligned}$$

Properties:

$$\begin{aligned} \mu =& p \\ \sigma^2 =& p(1-p) \\ M(t) =& pe^t + (1-p) \end{aligned}$$

[5] I don't actually know how to formalize this considering we are taking an infinite product. My guess is that you get extreme issues like in the topological setting, and need some way to deal with these issues. This aside is meant as a thought-exercise, so do not take it as a rigorous exposition.

- Binomial: parameterized by p probability of success and n number of trials

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Properties:

$$\begin{aligned} \mu =& np \\ \sigma^2 =& np(1-p) \\ M(t) =& \left[pe^t + (1-p)\right]^n \end{aligned}$$

- Geometric: parameterized by probability of success p

$$f(k) = p(1-p)^{k-1}$$

Properties:

$$\begin{aligned} \mu =& \frac{1}{p} \\ \sigma^2 =& \frac{1-p}{p^2} \\ M(t) =& \frac{pe^t}{1-(1-q)e^t} \end{aligned}$$

- Poisson: parameterized by λ

$$f(k) = \frac{1}{k!}\lambda^k e^{-\lambda}$$

Properties:

$$\begin{aligned} \mu =& \lambda \\ \sigma^2 =& \lambda \\ M(t) =& \exp\left(\lambda(e^t - 1)\right) \end{aligned}$$

- Negative Binomial: parameterized by p and n

$$f(k) = \binom{k-1}{n-1} p^n (1-p)^{k-n} \quad k \geq n$$

Properties:

$$\begin{aligned} \mu =& \frac{n}{p} \\ \sigma^2 =& \frac{n(1-p)}{p^2} \\ M(t) =& \left[\frac{pe^t}{1-(1-q)e^t}\right]^n \end{aligned}$$

Continuous Random Variables

- Uniform: parameterized by a, b

$$f(x) = \begin{cases} 0 & \text{if} \quad x \notin [a,b] \\ \frac{1}{b-a} & \text{if} \quad x \in [a,b] \end{cases}$$

Properties:

$$\begin{aligned} \mu =& \frac{a+b}{2} \\ \sigma^2 =& \frac{(b-a)^2}{12} \\ M(t) =& \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & \text{if} \quad t \neq 0 \\ 1 & \text{if} \quad t = 0 \end{cases} \end{aligned}$$

- Exponential: parameterized by λ

$$f(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \lambda e^{-\lambda x} & \text{if} \quad x \geq 0 \end{cases}$$

Properties:

$$\begin{aligned} \mu =& \frac{1}{\lambda} \\ \sigma^2 =& \frac{1}{\lambda^2} \\ M(t) =& \frac{\lambda}{\lambda - t} \end{aligned}$$

- Normal: parameterized by μ, σ^2

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Properties:

$$\begin{aligned} \mu =& \mu \\ \sigma^2 =& \sigma^2 \\ M(t) =& \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \end{aligned}$$

- Gamma: parameterized by λ and r

$$f(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x} & \text{if} \quad x \geq 0 \end{cases}$$

Properties:

$$\begin{aligned} \mu =& \frac{r}{\lambda} \\ \sigma^2 =& \frac{r}{\lambda^2} \\ M(t) =& \left(1 - \frac{t}{\beta}\right)^{-\alpha} \end{aligned}$$

◦ Beta: parameterized by α, β

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Properties:

$$\begin{aligned}
\mu =& \frac{\alpha}{\alpha+\beta} \\
\sigma^2 =& \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
M(t) =& 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{t^k}{k!}
\end{aligned}$$

Chapter 5

Exercises

5.2 Optimization Exercises

Exercise 5.2.1. True/False:

- Let $f : X \to \mathbb{R}$ be strictly concave and X convex, open. Let $A \subseteq X$ be compact and nonempty. Then f has a unique maximizer in A.
- All concave functions are continuous.

Exercise 5.2.2. Let $f \in \mathcal{C}^2(\mathbb{R})$. Suppose that there exists an $\epsilon > 0$ such that $f''(x) > \epsilon$ for all $x \in \mathbb{R}$. Prove that f has a unique global minimum. *Hint:* Show that $f'(x) = 0$ for some $x \in \mathbb{R}$.

Exercise 5.2.3. Fix an $\alpha \in \mathbb{R}$. Let $f_\alpha : \mathbb{R}^2 \to \mathbb{R}$ be defined as $f_\alpha(x, y) := (1 + \alpha)x^2 + \alpha y^2 + 2xy + 7y$.

(a) For each α, find the points, if any, satisfying the first-order necessary condition for a local minimum.

(b) For which α do those in (a) satisfy the second-order necessary condition for a local minimum?

(c) For which α do those in (b) satisfy the second-order sufficient condition for a local minimum?

(d) For which α are those in (c) a global minimum? Prove the claim.

Exercise 5.2.4. Consider the following minimization problem:

$$f(x, y, z) := \left(x - \frac{y}{3}\right)^2 + \frac{1}{4}(z + x)^2 + z^2 - 2024$$
$$h(x, y, z) := (x + y - z)^2(x + z)^2(x - 2y)$$

We aim to minimize $f(x, y, z)$ subject to $h(x, y, z) = 0$.

(a) Determine which feasible points have $\nabla h(x, y, z) \neq 0$. Why is this important to check? *Hint:* What does the full rank condition mean here?

(b) Determine the minimizer of the problem. Does problem (a) become restrictive here?

Exercise 5.2.5. On the set $\{(x, y) : x^2 + y^2 \leq 1\}$, find the extreme values of

1. $f(x,y) = 2x^2 + y^2 + 2x$
2. $f(x,y) = 3x^2 - 2y^2 + 2y$

Exercise 5.2.6. Show that the following functions have a global maximum and minimum on $\mathbb{R}^2$, and find them:

1. $f(x,y) = (x^2 - 2y^2)e^{-(x^2+y^2)}$
2. $f(x,y) = (x^2 + 2y^2)e^{-(x^2+y^2)}$

Hint: Don't do the problem twice.

Exercise 5.2.7. Find the points of the ellipse $5x^2 - 6xy + 5y^2 = 4$ for which the tangent is at the greatest distance from the origin. *Hint:* Notice that this is a double optimization problem. First, for a given point on the ellipse, we need to find the point on its tangent that is closest to the origin. Or, do we? Draw a picture and use some intuition.

Exercise 5.2.8. Let $x_1, ..., x_n$ denote nonnegative numbers. For $c > 0$, maximize the product $x_1 \cdot ... \cdot x_n$ subject to the constraint $x_1 + ...x_n = c$. From here, derive the inequality of geometric and arithmetic means,

$$(x_1 \cdot ... \cdot x_n)^{1/n} \leq \frac{x_1 + ... + x_n}{n}$$

Exercise 5.2.9. Suppose that we have a limited amount of fencing and wish to maximize the total area generated by a circular and rectangular pen.

(a) Write down the problem that we are trying to solve. What are our choice variables? Is there any way to reduce the number of choice variables?

(b) Solve this problem using inequality constraints

(c) Solve this problem using a substitution and first-year calculus. What does this teach us?

Exercise 5.2.10. Consider the following problem:

$$\max_{c,d,x,y} u(c) + u(d)$$

such that

$$\begin{aligned} c + x &\leq \overline{w} \\ d + y &\leq f(x) \\ c, d, x, y &\geq 0 \end{aligned}$$

for a fixed $\overline{w} > 0$, and u, f are nondecreasing $\mathcal{C}^2$ functions.

(a) What additional assumptions (if any) are necessary for this problem to have a solution?

(b) What additional assumptions (if any) are necessary so that the solution, when it exists, is unique?

(c) Write down the problem as a standard problem of inequality constraints and derive the standard Kuhn-Tucker conditions. What additional assumptions are needed so that these conditions completely characterize the solution set?

(d) Assume now that u is strictly increasing. Show that the optimal solution must have $c + x = \overline{w}$ and $d + y = f(x)$. Can we use this to reduce the number of constraints?

Exercise 5.2.11. In this exercise we imagine that we are a firm that is aiming to hit a target output while minimizing their cost (that is, this is the "cost minimization problem" (CMP) of the firm). Specifically, a firm produces a single output $y \geq 0$ using outputs $z \in \mathbb{R}^n_+$ given the production function $f : \mathbb{R}^n_+ \to \mathbb{R}$. The prices for inputs are given by $w \in \mathbb{R}^n_{++}$. We can state the problem as

$$\min_z \langle w, z \rangle$$

such that

$$\begin{aligned} f(z) &\geq y \\ z &\geq 0 \end{aligned}$$

We assume that $f : \mathbb{R}^n_+ \to \mathbb{R}$ is continuous and the set $\{z \in \mathbb{R}^n_+ | f(z) \geq y\}$ is nonempty.

(a) Prove that the firm's cost minimization problem has a solution for any $w \gg 0$. In what follows, denote the solution (correspondence) by $z(w)$ as a function of w.

(b) Suppose that $z^* \in z(w)$ is such that $z^* \neq 0$. Prove that $f(z^*) = y$.

(c) Let $c(w) := \langle w, z(w) \rangle$. First note that this is well-defined, even though it is an abuse of notation. We call this the firm's "cost function." Prove that $c(w)$ is concave.

(d) Assume now that $z(w)$ is single-valued for all w. Prove that $z_i(w)$ is nonincreasing in w_i for any i and $c(w)$ is nondecreasing in w_i.

Now assume that $f \in \mathcal{C}^2$ and $z(w)$ is single valued (what assumption can we make on f for the latter to hold?)

(e) Write down the Kuhn-Tucker conditions for a minimum. Under what conditions on f are these necessary *and* sufficient for a solution?

(f) Under the conditions you give in part (e), calculate $D_w z(w)$.

Exercise 5.2.12. Consider a consumer who lives for $T > 0$ periods. At the beginning of each period $1 \leq t \leq T$, this consumer must decide his consumption $c_t \geq 0$ for the current period and saving $s_t \geq 0$ for the next period, given his saving s_{t-1} from the previous period. Denote by $s_0 \geq 0$ the initial saving. The consumer maximizes his lifetime utility

$$U(s_0) := \max_{c_i, s_i} \sum_{t=1}^{T} u(c_t)$$

such that

$$\begin{aligned} c_t + s_t &\leq f(s_{t-1}) \quad \text{for all} \quad 1 \leq t \leq T \\ c_t, s_t &\geq 0 \quad \text{for all} \quad 1 \leq t \leq T \end{aligned}$$

where u is the period utility function and f is the production function. Assume $u, f \in \mathcal{C}^0$.

(a) Defining

$$C(s_0) \coloneqq \left\{ (c_1, ..., c_T, s_1, ..., s_T) \in \mathbb{R}_+^{2T} \middle| \, c_t + s_t \leq f(s_{t-1}), 1 \leq t \leq T \right\}$$

as the constraint set (parameterized by s_0). Prove that $C(s_0)$ is compact, so that the problem has a solution.

(b) Prove that C as a correspondence is continuous.

(c) Prove that U is continuous and (weakly) increasing.

(d) Assume that $u \in \mathcal{C}^1(0, \infty)$, with $u'(c) > 0$ and $\lim_{c \searrow 0} u'(c) = \infty$ (the "Inada" condition). Assume moreover that f is strictly increasing with $f(0) = 0$. Prove that any optimal solution must satisfy $c_t > 0$ for all $1 \leq t \leq T$ and $s_t > 0$ for all $1 \leq t < T$.

(e) From now on, suppose that $u, f \in \mathcal{C}^2(0, \infty)$ and are both strictly increasing and strictly concave. Moreover, suppose that u satisfies the Inada condition. Prove that the optimization problem has a unique solution.

(f) Given any $s_0 > 0$ prove that the optimal solution must satisfy

$$\frac{u'(c_t)}{u'(c_{t+1})} = f'(s_t) \quad \text{for all} \quad 1 \leq t \leq T - 1$$

What does this mean if $f(s) = s$?

(g) Consider now $T = 2$. Prove that both c_1 and c_2 strictly increase with s_0.

Exercise 5.2.13. Let $f : \mathbb{R}^n \to \mathbb{R}$ be concave (and hence also continuous).

(a) Prove that the super-graph of f, which is defined by

$$G(f) \coloneqq \{(x, y) | y \geq f(x)\}$$

is closed and convex.

(b) State the supporting hyperplane theorem.

(c) Using the above, prove Proposition 2.4.10.

(d) Prove Proposition 2.4.11.

(e) Prove the following theorem:

Theorem 5.2.14 (Benveniste and Scheinkman)**.** *Let $V : \mathbb{R}^n \to \mathbb{B}$ be concave, let $x_0 \in \mathbb{R}$, and let $B_\epsilon = B_\epsilon(x_0)$. Let $W : B_\epsilon \to \mathbb{R}$ be concave, differentiable with $W(x_0) = V(x_0)$ and $W(x) \leq V(x)$ for all $x \in B_\epsilon$, then V is differentiable at x_0 and $\nabla V(x_0) = \nabla W(x_0)$.*

(f) Let $X \subseteq \mathbb{R}^n$ be convex and compact, $A = (0, 1)$ and $f : X \times A \to \mathbb{R}$ with $f \in \mathcal{C}^1$ and concave. Define

$$V(a) \coloneqq \max_{x \in X} f(x, a)$$
$$X^*(a) \coloneqq \text{argmax}_{x \in X} f(x, a)$$

Using the above, show that V is well-defined, concave, and differentiable on A.

(g) (For after you have learned Probability Theory.) Use the above to provide another proof of Jensen's Inequality (Theorem 4.4.6).

5.3 Measure Theory Exercises

Exercise 5.3.1. Assume that $E_j, j = 1, ...$ is a countable collection of measurable subsets of $\mathbb{R}$ such that $\lim_{n\to\infty} m(E_j) = 0$.

1. Let A denote the points which belong to all but finitely many of these sets E_j. Show that A is measurable.
2. Show that $m(A) = 0$.
3. Let B denote the points which belong to infinitely many of these E_j. Show that B is measurable.
4. Is it true that $m(B) = 0$?

Exercise 5.3.2. Prove that every closed set in $\mathbb{R}$ is a G_δ set. Does it follow that every open set is an F_σ-set?

Exercise 5.3.3. Show that every hyperplane in $\mathbb{R}^n$ is a zero-measure set. *Hint:* Start with those that are parallel to the axes. (Proposition 3.1.4.)

Exercise 5.3.4. Let M be an arbitrary set. Define $\omega : 2^M \to [0, \infty]$ by $\omega(S) = |S|$. Prove that ω is a measure, and all sets are measurable. This is called the *counting measure*.

Exercise 5.3.5. Define the following function on $2^{\mathbb{N}}$.

$$\mu(E) := \begin{cases} 0 & \text{if} \quad E = \emptyset \\ \sum_{n\in E} \frac{1}{2^n} & \text{if} \quad E \neq \emptyset \end{cases}$$

Show that μ is a measure on $\mathbb{N}$. What is $\mu(\mathbb{N})$? What is $\mu(\{2k|k \in \mathbb{N}\})$?

Exercise 5.3.6. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space with μ a finite measure (that is, $\mu(\Omega) < \infty$).

(a) Show that for every nested decreasing sequence of sets $\{A_n\}_n$ in $\mathcal{F}$ we have that $\lim_{n\to\infty} \mu(A_n) = \mu(\cap_n A_n)$.

(b) Consider now a finite measure on $(\mathbb{R}, \mathcal{B})$, and define $F(x) := \mu((-\infty, x])$ for all $x \in \mathbb{R}$. Show that

(i) F is right-continuous

(ii) $\lim_{x\to-\infty} F(x) = 0$

Exercise 5.3.7. Let (X, A_X), (Y, A_Y), and (Z, A_Z) be measurable spaces. Let $f : X \to Y$ and $g : Y \to Z$ be measurable functions (with respect to these algebras). Show that $g \circ f : X \to Z$ is measurable.

Exercise 5.3.8. Show that $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) := \mathbb{1}(x \in [0, 1] \setminus \mathbb{Q})$$

is measurable. Compute

$$\int_{[0,1]} f d\mu$$

Exercise 5.3.9. Let g_n be a sequence of non-negative integrable functions which converge almost everywhere to an integrable function g. Let f_n be a sequence of measurable functions such that

$$|f_n(x)| \leq g_n(x)$$

for all x, and assume that f_n converges to a limit f almost everywhere as $n \to \infty$. Prove that if

$$\int g_n \to_{n\to\infty} \int g$$

then

$$\int f_n \to_{n\to\infty} \int f$$

Hint: Apply Fatou's Lemma to $g_n + f_n$ and $g_n - f_n$.

Exercise 5.3.10. Assume that $f : \mathbb{R} \to [0, \infty)$ is integrable. Show that integrability does not imply that

$$\lim_{|x|\to\infty} f(x) = 0$$

However, if one assumes that, in addition to being integrable, f is also uniformly continuous, is it true that

$$\lim_{|x|\to\infty} f(x) = 0?$$

Exercise 5.3.11. We show here that translation is continuous with respect to integration. That is, we show, for f integrable, that

$$\int |f(x+h) - f(x)| \to_{h\to 0} 0$$

1. Show that this holds if f is continuous of compact support.
2. Show how one can properly approximate simple functions by continuous functions of compact support.
3. Use 1 and 2 to conclude.

Exercise 5.3.12. Show that $\mathcal{L}_1(\mathbb{R})$ neither contains nor is contained by $\mathcal{L}_2(\mathbb{R})$. Keep this exercise in mind.

5.4 Probability Theory Exercises

Exercise 5.4.1. Let X be a discrete random variable whose probability mass function f_X is symmetric with respect to 0, and has $f_X(0) = 0$. Let

$$J := \begin{cases} 1 & \text{if} \quad X > 0 \\ 0 & \text{if} \quad X = 0 \\ -1 & \text{if} \quad X < 0 \end{cases}$$

Show that $|X|$ and J are independent.

Exercise 5.4.2. Let $N, X_1, X_2, ...$ be random variables where $N \in \mathbb{N}$ and the X_i are i.i.d. Moreover, suppose that X_i and N are independent, for all i. Show that

$$\mathbb{E}\left(\sum_{i=1}^{N} X_i\right) = \mathbb{E}(N)\mathbb{E}(X_i)$$

Exercise 5.4.3. Consider the sequence of random variables $X_1, ...$ with pdfs given by

$$f_n(x) = 1 + \cos(2\pi n x)$$

for $x \in [0, 1]$.

(a) Show that X_i is indeed a random variable.

(b) Show that $\{X_i\}_i$ converges in distribution. Find the cdf of the limit distribution.

Exercise 5.4.4. Suppose that $X_1 \sim N(\mu, \sigma^2)$ and $X_2 \sim N(3\mu, 4\sigma^2)$ are independent.

(a)

5.5 Review Session 1

Exercise 5.5.1 (Functional Optimization Practice)**.** Let $a \in \mathbb{R}$ and consider a function u. Solve for the distribution $p(x)$ that achieves the maximum entropy

$$\mathcal{H}(p) := -\int p(x) \log p(x) dx$$

subject ot the constraints that

$$\mathbb{E}_p[u(X)] := \int u(x)p(x)dx = a$$

In particular, show that it is in the exponential family with sufficient statistic $u(x)$.

Solution. We have that there are two equality constraints; because p is a distribution we must have that

$$\int p(x) = 1$$

So, we can write

$$\mathcal{G}_1[p] = \int u(x)p(x)dx$$
$$\mathcal{G}_2[p] = \int p(x)dx$$

We then have that

$$\frac{\partial}{\partial p}\mathcal{H}[p] = -\log p(x) - 1 \quad \frac{\partial}{\partial p}\mathcal{G}_1[p] = u(x) \quad \frac{\partial}{\partial p}\mathcal{G}_2[p] = 1$$

We can then conclude that we have

$$\log(p) + 1 = \lambda_1 u(x) + \lambda_2$$

for some $\lambda_1, \lambda_2 \in \mathbb{R}$. We conclude that we must have that

$$p(x) = c \exp\left(\lambda_1 u(x)\right)$$

for some $c \in \mathbb{R}$. We conclude that p must belong to the exponential family. □

Exercise 5.5.2 (Non-differentiable Practice 1)**.** Solve the following maximization problem using sub-differentials as a function of the parameter $b \geq 0$:

$$\max |x| + |y|$$

subject to

$$b|x| + |y| \leq 1$$

Solution. We first proceed by computing the sub-differentials of the objective function f and the constraint function g:

$$\partial |x| = \begin{cases} \{-1\} & \text{if} \quad x < 0 \\ [-1, 1] & \text{if} \quad x = 0 \\ \{1\} & \text{if} \quad x > 0 \end{cases}$$

and

$$\partial |y| = \begin{cases} \{-1\} & \text{if} \quad y < 0 \\ [-1, 1] & \text{if} \quad y = 0 \\ \{1\} & \text{if} \quad y > 0 \end{cases}$$

We can then compute

$$\partial f(x, y) = \partial |x| \times \partial |y|$$

similarly, we have that

$$\partial g(x, y) = (b \cdot \partial |x|) \times \partial |y|$$

We now look at the different cases. Suppose first that we are at a situation where neither x nor y was equal to 0. Then, at a maximum we would need to have

$$\begin{pmatrix} \pm 1 \\ \pm 1 \end{pmatrix} = \mu \begin{pmatrix} \pm b \\ \pm 1 \end{pmatrix}$$

First need that we need to have that $\mu \geq 0$. So, for equality in the second element we must have that $\mu = 1$. However, then we will not have equality for the first element, unless $b = 1$ (but of course in this case the optimization is trivial).

We now must consider the cases that $x = 0$ and $y = 0$. Notice that we cannot have both simultaneously at a maximum trivially. Suppose first that $x = 0$. Then we need, for a local maximum, for every $c_1 \in [-1, 1]$ that there exists $c_2 \in [-1, 1]$ such that

$$\begin{pmatrix} c_1 \\ \pm 1 \end{pmatrix} = \mu \begin{pmatrix} bc_2 \\ \pm 1 \end{pmatrix}$$

Again, we get that $\mu = 1$, so that we have that we need

$$c_1 = bc_2$$

So, we need to have that $c_1/b \in [-1, 1]$, so that this is a candidate for a local maximum only if $b \geq 1$. Now, we consider the case that $y = 0$. Then, similarly, for every $c_1 \in [-1, 1]$ that there exists $c_2 \in [-1, 1]$ such that

$$\begin{pmatrix} \pm 1 \\ c_1 \end{pmatrix} = \mu \begin{pmatrix} \pm b \\ c_2 \end{pmatrix}$$

We then need that $\mu = 1/b$. So, we would need $bc_1 \in [-1, 1]$, so that this is only a local maximum when we have that $b \leq 1$. □

Exercise 5.5.3 (Non-differentiable Practice 2)**.** Solve the following maximization problem using sub-differentials as a function of the parameter b:

$$\max \{x, y\}$$

subject to

$$|x - y| + b|x| \leq 1$$

Solution. We first begin by computing the sub-differentials as in the above. Notice that in general this is a much harder problem to solve without any work, because it is not obvious what the constraint set looks like. First, for the objective function f

$$\partial f(x, y) = \begin{cases} \{(1, 0)\} & \text{if} \quad x > y \\ \{(0, 1)\} & \text{if} \quad x < y \\ \{(1 - \ell, \ell) | \ell \in [0, 1]\} & \text{if} \quad x = y \end{cases}$$

To visualize this, draw a picture (and explain it). Now, we look at the constraint function. We get two portions. The first is trivial.

$$\partial |x| = \begin{cases} \{-1\} & \text{if} \quad x < 0 \\ [-1, 1] & \text{if} \quad x = 0 \\ \{1\} & \text{if} \quad x > 0 \end{cases}$$

The second is trickier:

$$\partial |x - y| = \begin{cases} \{(1, -1)\} & \text{if} \quad x > y \\ \{(-1, 1)\} & \text{if} \quad x < y \\ \{(-\ell, \ell) | \ell \in [-1, 1]\} & \text{if} \quad x = y \end{cases}$$

To visualize this, draw a picture. To actually optimize, we consider several cases (as before). First, it is obvious that we should never have a negative x. So, we can reduce to the case that $x \geq 0$ (explain). Then, suppose that we do not have $x = 0$ and we do not have that $x = y$. Then, if $x > y$, we would need to satisfy

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mu \begin{pmatrix} b+1 \\ -1 \end{pmatrix}$$

This is clearly impossible. If $x < y$, we would need to satisfy

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \mu \begin{pmatrix} b-1 \\ 1 \end{pmatrix}$$

This is also clearly impossible, unless $b = 1$. In this case, then any point with $x < y$ on the boundary will be a candidate.

Now, we check the corners. First, we suppose that $x = 0$. Then, we must satisfy $x < y$ that there exists some $c_1 \in [-1, 1]$ such that

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \mu \begin{pmatrix} bc_1 - 1 \\ 1 \end{pmatrix}$$

Here, we need to have that $\mu = 1$, so that we would need to have that $b \geq 1$. In this case, $(0, 1)$ would be a candidate for a maximizer.

The final case to check is that $x = y = \frac{1}{b}$. Here, we would need, for every $\ell_1 \in [0, 1]$ there to exist and $\ell_2 \in [-1, 1]$ such that

$$\begin{pmatrix} 1 - \ell_1 \\ \ell_1 \end{pmatrix} = \mu \begin{pmatrix} b - \ell_2 \\ \ell_2 \end{pmatrix}$$

Notice that here we would need to have that $\mu\ell_2 = \ell_1$ (so that $\ell_2 \geq 0$ as $\mu \geq 0$), which reduces the constraint to

$$\begin{aligned} 1 - \ell_1 =& \mu b - \ell_1 \\ \Rightarrow 1 =& \mu b \\ \Rightarrow \ell_2 =& b\ell_1 \end{aligned}$$

This requires that $b \leq 1$. In this case, we will have the point $(1/b, 1/b)$ being the local maximizer. □

Exercise 5.5.4 (Inequality Constraint Practice 1)**.** Let $L > 0$ be the length of a piece of wire. Suppose that you use the wire to make a circle of radius r and a square of side s. We aim to maximize the total area of the circle and the square:

$$\max A(r, s) = \pi r^2 + s^2$$

subject to the following:

$$2\pi r + 4s = L r, s \geq 0$$

Solution. We need to have then that the system that we must solve is

$$\begin{pmatrix} 2\pi r \\ 2s \end{pmatrix} + \lambda \begin{pmatrix} 2\pi \\ 4 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$$

Here we need to have that

$$\mu_1 r = \mu_2 s = 0$$

and that $\mu_i \geq 0$ for $i = 1, 2$. First, consider the case that we have $r, s \neq 0$. Then both values of μ are 0 by complementary slackness. We then solve

$$\begin{pmatrix} 2\pi r \\ 2s \end{pmatrix} + \lambda \begin{pmatrix} 2\pi \\ 4 \end{pmatrix} = 0$$

This clearly yields that $s = 2r$ (because we have an equality constraint). Hence we get a candidate maximizer when

$$2\pi r + 8r = L \Rightarrow (r, s) = \left(\frac{L}{8 + 2\pi}, \frac{L}{4 + \pi} \right)$$

We now check the other two cases before returning to the second order conditions. If we have that $x = 0$, then the problem reduces to $y = L/4$ and we get

$$\begin{pmatrix} 0 \\ L/2 \end{pmatrix} + \lambda \begin{pmatrix} 2\pi \\ 4 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0$$

We get then that this is a candidate for a maximizer. The other possibility is when $s = 0$, when would lead to a system

$$\begin{pmatrix} L \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2\pi \\ 4 \end{pmatrix} + \mu_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$$

This is the other candidate for a maximizer. Now, we look at the second order conditions. Notice that because the constraints are all linear, they contribute nothing to the second order conditions, so we must just check the Hessian of A. However, this matrix is clearly positive definite:

$$\nabla^2 A = \begin{pmatrix} 2\pi & 0 \\ 0 & 2 \end{pmatrix}$$

This means that if there are any feasible directions to check we are at a local minimum. So the interior point is a local minimum. Considering the other points, we can see which is the maximum directly by verifying numerically. (It is the point where $s = 0$). □

Exercise 5.5.5 (Equality Constraint Practice 1)**.** Let Q by a symmetric 2×2 matrix with eigenvalues given by $\lambda_1 \leq \lambda_2$. Moreover, the eigenvectors of Q are not $e_1, e_2 \in \mathbb{R}^2$. Consider the following problem:

$$\min f(x) := \frac{\langle x, x \rangle_Q}{\langle x, x \rangle}$$

subject to the constraint that $x_1 x_2 \neq 0$, where x_1, x_2 are the coordinates of the vector $x \in \mathbb{R}^2$.

(a) Show that the problem is equivalent to the problem

$$\min g(x) \deg \langle x, x \rangle_Q$$

subject to the constraints

$$\begin{aligned} \langle x, x \rangle &= 1 \\ x_1 x_2 &\neq 0 \end{aligned}$$

(b) Find the candidates for minimizers of the problem, and express the minimum value of this problem in terms of the eigenvalues.

(c) What do the second order conditions say about the candidates that you found in part (b)?

Solution. (a) We can see easily that $f(ax) = f(x)$. This is as all of the computations are linear:

$$f(ax) = \frac{(ax)^T Q(ax)}{(ax)^T(ax)} = \frac{a^2}{a^2}\frac{x^T Q x}{x^T x}$$

This means that f is constant on all rays emanating from the origin. Hence, it is sufficient to consider one representative from each ray. In this case, it is natural to consider the unit circle, as here we have $x^T x = 1$, so that our function becomes g. (See the attached picture)

(b) Consider that we can write

$$\begin{aligned} g(x) &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= ax_1^2 + cx_2^2 + 2bx_1x_2 \end{aligned}$$

Note that as e_1, e_2 are not the eigenvectors of Q, we have that $b \neq 0$. Now, the Lagrange multipliers for this (ignoring the constraint for now) yield that

$$\begin{aligned} \nabla g + \lambda \nabla(x^2 + y^2) &= 0 \\ -2\lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \nabla g \\ -2\lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= 2 \begin{pmatrix} ax_1 + bx_2 \\ bx_1 + cx_2 \end{pmatrix} = 2Q \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{aligned}$$

Notice that this means exactly that the candidates for the extrema of this function on the unit circle are those vectors x for which Qx is a scalar multiple of x. These are exactly the eigenvectors (plus-minus the eigenvectors in fact).

Ignoring the constraint was not a problem: as e_1, e_2 are not eigenvectors they are not candidates for a minimum even if we considered g on all of S^1. Hence, excluding them will not affect which points are candidates.

(c) Now, we have that for an eigenvector v_i with eigenvector λ_i, the corresponding Lagrange multiplier is $-\lambda_i$. Hence, we have that

$$\nabla^2 g + \frac{1}{\lambda_i}\nabla^2(x^2 + y^2) = 2Q - 2\lambda_i I$$

This matrix has eigenvalues of $2(\lambda_1 - \lambda_i)$ and $2(\lambda_2 - \lambda_i)$. Notice that as this matrix is symmetric, the two eigenvectors will be perpendicular. This means that the tangent space at one of the eigenvectors is exactly the span of the other eigenvector. If $\lambda_2 > \lambda_1$, then at v_2 we have that $v_1^T(2Q - 2\lambda_2 I)v_1 = 2(\lambda_1 - \lambda_2) < 0$, so that this is not a minimum. At v_1, however, we would have $v_2^T(2Q - 2\lambda_1 I)v_2 = 2(\lambda_2 - \lambda_1) > 0$.

This means that the matrix is positive definite on the tangent space, so that it is a minimum. If the two eigenvalues are equal, then $Q = \lambda I$, contradicting that e_i isn't an eigenvector. Hence we have $\lambda_2 > \lambda_1$, so that we have our minimum achieved at exactly $\pm v_1$.

□

Exercise 5.5.6 (Some other maximization problems)**.** (a)

$$\max f(x,y) = (x^2 - 2y^2)e^{-(x^2+y^2)} \min \qquad f(x,y) = (x^2 - 2y^2)e^{-(x^2+y^2)}$$

(b)

$$\max x^{1/2} + by^{1/2}$$

subject to

$$x + y \leq c$$

(c)

$$\max x^2 + by^2$$

subject to

$$\begin{aligned} x + y &\leq c \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Look at the correspondence.

Review Session 2

Exercise 5.5.7. Show that almost sure convergence does not imply convergence in $\mathcal{L}_p$.

Solution. I give two examples, both of which capture the same idea. Basically, almost sure convergence deals with the probability that we differ *at all* for each ω. It doesn't care about by how much we differ. $\mathcal{L}_p$ convergence, on the other hand, is a measure of how much, *on average*, we differ from the limit distribution X.

The first example highlights this with a story. Suppose you start in time period 1 with \$1. Each future time period, you flip a (fair) coin. If it lands on heads, you triple the amount of money that you have. If it lands on tails, then you lose everything, and in every future period you get nothing.

Let X_n be the amount of money in period n.

We have that $X_n \to_{a.s} X \equiv 0$. To see this, we can see that the events E_n such that $X_n > 0$ have the property that $E_n \subseteq E_m$ for $m \leq n$. So then, we have that

$$\mathbb{P}\left[\lim_{n\to\infty} \sup X_n = 0\right] = \mathbb{P}\left[\lim_{n\to\infty} E_n = 0\right] = 1$$

As $\cap E_n = (1, ..., 1, ...)$ which has probability zero.

However, notice that $\mathbb{E}[X_n] = \left(\frac{3}{2}\right)^n \to \infty$, so this does not converge in L_p for any $p \geq 1$.

The second example is more mathematical in foundation, and it emphasizes perhaps more clearly the mechanisms. Let $\Omega = [0,1]$, endowed with the standard Lebesgue measure. Then, define

$$X_n(\omega) = n^2 \mathbb{1}\left\{\omega \leq \frac{1}{n}\right\}$$

Then, we have that

$$\mathbb{P}\left[\lim_{n\to\infty} \sup X_n = 0\right] = \lambda((0,1]) = 1$$

But, we have that

$$\mathbb{E}[X_n] = n \to \infty$$

so again, we do not converge in $\mathcal{L}_p$. □

Exercise 5.5.8. Show that convergence in $\mathcal{L}_p$ does not imply converge almost surely.

Solution. This highlights the exact reverse of the above, as we will see. Let $\Omega = [0,1]$ endowed with the standard Lebesgue measure λ. Then, define a "typewriter" sequence

$$\begin{aligned} X_1(\omega) &:= \mathbb{1}\{\omega \in [0,1]\} \\ X_2(\omega) &:= \mathbb{1}\{\omega \in [0,1/2]\} \\ X_3(\omega) &:= \mathbb{1}\{\omega \in [1/2,1]\} \\ X_4(\omega) &:= \mathbb{1}\{\omega \in [0,1/3]\} \\ &\vdots \end{aligned}$$

Then, notice that for each $\omega \in [0,1]$, infinitely many X_n have $X_n(\omega) = 1$. This means that

$$\mathbb{P}\left[\lim_{n\to\infty} \sup X_n = 0\right] = \lambda(\emptyset) = 0$$

However, notice that if the length of the interval is $1/m$, then its p-norm will be $1/m^{1/p}$. This will clearly go to 0, and so will converge in $\mathcal{L}_p$ for any $p > 0$. □

Exercise 5.5.9. Go through some of the examples of transforming random variables.

Solution. Suppose that X is a random variable, and $g : \mathbb{R} \to \mathbb{R}$ is a monotonically increasing function.

Define $Y = g(X)$. Because g is monotonically increasing, we have that it has an inverse v: $X = v(Y)$. We can then write

$$F_Y(y) = F_X(v(y))$$

This comes just from writing this clearly. If X is continuous, and g is differentiable, then we also get

$$f_Y(y) = v'(y) f_X(v(y))$$

just from an immediate application of the chain rule.

If g is not monotonic, then we have to be a little bit more careful. For example, consider what happens when $g = x^2$. there are two inverses, and we need to be careful about the range.

Now, suppose that we have a joint distribution of X, Y, and a transformation of

$$\begin{aligned} U =& g_1(X, Y) \\ V =& g_2(X, Y) \end{aligned}$$

Now, suppose that we can invert uniquely these g_1, g_2. That is, there exists an h_1, h_2 such that if U, V are above, we have

$$\begin{aligned} X =& h_1(U, V) \\ Y =& h_2(U, V) \end{aligned}$$

Then, we have that, similar to the example of the above (just by a knowledge of higher-dimension change of variable):

$$f_{U,V}(u, v) = \left|\det(h(u, v))\right| f_{X,Y}(h_1(u, v), h_2(u, v))$$

For example, suppose that $U = X - Y$ and $V = X + Y$. Then, we have that

$$\begin{aligned} X &= \frac{1}{2}(U + V) \\ Y &= \frac{1}{2}(V - U) \end{aligned}$$

So, we have that

$$f_{U,V}(u, v) = \frac{1}{2} f_{X,Y}(u + v, v - u)$$

□

Exercise 5.5.10. Should also talk about the conditional density, and the marginal density, and how these relate to the Ω ideas, right?

Chapter 6

Selected Solutions

6.2 Solutions to Optimization Exercises

Solution. 5.2.3

(a) The first order condition is that $\nabla f_\alpha = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, this is because any possible minimum is an interior point. Hence we compute the partial derivatives (f_α is C^∞) and set them to zero.

$$\frac{\partial f_\alpha}{\partial x} = 2x(\alpha + 1) + 2y$$
$$\frac{\partial f_\alpha}{\partial y} = 2y\alpha + 2x + 7$$

From the second equation, at a local minimum we must have that $x^* = -\frac{7+2y\alpha}{2}$. Plugging this into the first equation, we must have that $-(\alpha+1)(2y^*\alpha+7)+2y^* = 0$. This yields $y^* = -\frac{7\alpha+7}{2\alpha^2+2\alpha-2}$. Subbing this value for x yields $x^* = \frac{7}{2\alpha^2+2\alpha-2}$.

(b) We must have, for any vector $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$, that $d^T\nabla^2 f_\alpha(x^*, y^*)d \geq 0$. That is, we need the eigenvalues of $\nabla^2 f_\alpha(x^*, y^*)$ to be non-negative (i.e., that it is positive semi-definite).

We can compute the Hessian of f_α as:

$$\nabla^2 f_\alpha = \begin{pmatrix} 2(\alpha+1) & 2 \\ 2 & 2\alpha \end{pmatrix}$$

Now, after dividing each entry of the polynomial by 2 (which will clearly not change the sign of the eigenvalues), the characteristic polynomial of this new matrix is given by $p(t) = \alpha(\alpha+1)t^2 + (2\alpha+1)t + 1$. The roots of this polynomial correspond to the eigenvalues of the matrix, and are given by

$$\alpha + \frac{1}{2} \pm \frac{\sqrt{5}}{2}$$

For these to both be non-negative we must have that $\alpha \geq \frac{1}{2}(\sqrt{5}-1)$. Note, this can also be seen by looking at the determinants of the principal minors, as discussed in class.

(c) Similarly to above, we need that the eigenvalues of the matrix are strictly positive. In order to achieve this, we must have that $\alpha > \frac{1}{2}(\sqrt{5}-1)$.

(d) We take the hint, to complete the square. Let $Q = \nabla^2 f_\alpha(x^*, y^*)$, and notice that

$$Q^{-1} = \frac{1}{2\alpha(\alpha+1)-2}\begin{pmatrix}\alpha & -1\\ -1 & \alpha+1\end{pmatrix}$$

Notice as well that $\begin{pmatrix}x^*\\ y^*\end{pmatrix} = Q^{-1}\begin{pmatrix}0\\ 7\end{pmatrix}$. From these together we can write, for $\alpha > \frac{1}{2}(\sqrt{5}-1)$ that

$$f_\alpha(x,y) = \frac{1}{2}\begin{pmatrix}x-x^*\\ y-y^*\end{pmatrix}^T Q\begin{pmatrix}x-x^*\\ y-y^*\end{pmatrix} - \frac{1}{2}\begin{pmatrix}x^*\\ y^*\end{pmatrix}^T Q\begin{pmatrix}x^*\\ y^*\end{pmatrix}$$

As Q is positive definite, we know that the first term is positive whenever $\begin{pmatrix}x-x^*\\ y-y^*\end{pmatrix} \neq \begin{pmatrix}0\\ 0\end{pmatrix}$. Hence, the local minimums that we identified will be global minimums. This is exactly as we have discussed in class.

□

Solution. 5.2.4

(a) "Regular" points are exactly those points for which $h(x,y,z) = 0$ and $\nabla h(x,y,z) \neq 0$. These are the full-rank points. Now, we can compute

$$\nabla h(x,y,z) = \begin{pmatrix}2(x+y-z)(x+z)^2(x-2y) + 2(x+y-z)^2(x+z)(x-2y) + (x+y-z)^2(x+z)^2\\ 2(x+y-z)(x+z)^2(x-2y) - 2(x+y-z)^2(x+z)^2\\ -2(x+y-z)(x+z)^2(x-2y) + 2(x+y-z)^2(x+z)(x-2y)\end{pmatrix}$$

I characterize those points where this gradient is zero, the feasible regular points will then be all of the other points. Note that as $h(x,y,z) = 0$, we have that $(x+y-z)(x+z)(x-2y) = 0$. Hence at any feasible point we have that the gradient looks like

$$\nabla h(x,y,z) = \begin{pmatrix}(x+y-z)^2(x+z)^2\\ -2(x+y-z)^2(x+z)^2\\ 0\end{pmatrix}$$

Hence, the regular points are those feasible points such that $x = 2y$ and have $(x+y-z)(x+z) \neq 0$, that is have $z \neq 3y, -2y$.

(b) As we know, the candidates for the minimizer will have $\nabla f = \lambda \nabla h$, and will have (x,y,z) be a regular point.

Now, we have that

$$\nabla f(x,y,z) = \begin{pmatrix}2(x-\frac{y}{3}) + \frac{1}{2}(z+x)\\ 2(x-\frac{y}{3})\\ \frac{1}{2}(z+x) + 2z\end{pmatrix} = \begin{pmatrix}\frac{5}{2}x - \frac{2}{3}y + \frac{1}{2}z\\ 2(x-\frac{y}{3})\\ \frac{5}{2}z + \frac{1}{2}x\end{pmatrix}$$

Now, setting this equal to $\lambda \nabla h$, we have that

$$\begin{aligned} \frac{5}{2}x - \frac{2}{3}y + \frac{1}{2}z &= \lambda(x+y-z)^2(x+z)^2 \\ -(x - \frac{y}{3}) &= \lambda(x+y-z)^2(x+z)^2 \\ x &= -5z \end{aligned}$$

Now, as $x = 2y$ (as we can apply Lagrange multipliers only at regular points), we have that $z = \frac{-2}{5}y$ and hence

$$\begin{aligned} 5y - \frac{2}{3}y - \frac{1}{5}y &= -2y + \frac{1}{3}y \\ y &= 0 \end{aligned}$$

Hence we also have that $x = 0, z = 0$. Note that this is not a regular point! However, it does mean that we have no regular point candidates for minimums. $(0,0,0)$ is in fact the global minimum. This is because $f(0,0,0) = -2022$, and we can only have $x = \frac{1}{3}y$, $x = -z$ and $z = 0$ when all three are zero. This completes the question.

□

Solution. 5.2.7 We first find the points on the ellipse that are the farthest from the origin. From there we argue that for their tangents, they are closest to the origin. This will prove the claim. □

Solution. 5.2.8 This is a problem where one needs to be slightly smart about ruling out corner cases on the constraint $x_i \geq 0$. Note that we can clearly not have a maximum when $x_j = c$ or $x_j = 0$, because then the product will be zero. Then, we must have that for each $j = 1, ..., n$ that

$$\prod_{i \neq j} x_i + \lambda = 0$$

In particular, because no $x_i = 0$ (this is why it was important to rule out that case, apart from making that multiplier zero by complementary slackness), we have that

$$x_{j_1} = \frac{-\lambda}{\prod_{i \neq j_1, j_2} x_i} = x_{j_2}$$

This means that all of the x_i are equal, so that we have that $x_1 = c/n$, and hence,

$$x_1 \cdot ... \cdot x_n \leq \left(\frac{c}{n}\right)^n \Rightarrow (x_1 \cdot ... \cdot x_n)^{1/n} \leq \frac{c}{n}$$

This finished the problem. □

Solution. 5.2.9

(a) Notice that if we are devoting a certain length of fencing to the regular pen, then we need to have that it is a square, as this will maximize the area given any amount of fencing. The problem then becomes:

$$\begin{aligned} &\max \pi x^2 + y^2 \\ &\text{s.t.} 2\pi x + 4y = L \\ &0 \leq x, y \leq L \end{aligned}$$

(b) We can see this clearly as a standard optimization problem with inequality constraints. We have, at a regular point solution, we have

$$\begin{pmatrix} 2\pi x \\ 2y \end{pmatrix} + \lambda \begin{pmatrix} 2\pi \\ 4 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$$

with $\mu_1 x = \mu_2 y = 0$. Now, if we have $\mu_1 \neq 0$, we have $x = 0, y = \frac{L}{4}$ and $A = \frac{L^2}{16}$. Similarly, if $\mu_2 \neq 0$, we have $x = \frac{L}{2\pi}$, so that $A = \frac{L^2}{4\pi}$. Now, if $\mu_1 = \mu_2 = 0$, we have

$$x = \lambda \, y = 2\lambda$$

Hence, we have that $2x = y$, so that $2\pi x + 8x = L$ and so $x = \frac{L}{8+2\pi}$ and $y = \frac{L}{4+\pi}$. Now, in order to see if this is a maximum, we would need that $\nabla^2 A$ is negative semi-definite. However, notice that $\nabla^2 A = \begin{pmatrix} 2\pi & 0 \\ 0 & 2 \end{pmatrix}$ which is positive definite, and hence this is a minimum, and not a maximum. Hence, the maximum occurs at $y = 0$ and $x = \frac{L}{2\pi}$, with $A = \frac{L^2}{4\pi}$.

(c) We can see easily that $y = \frac{1}{4}(L - 2\pi x)$. From here, we have that

$$A(x, y) = \pi x^2 + \frac{1}{16}(L - 2\pi x)^2$$
$$A'(x) = 2\pi x - \frac{\pi}{4}(L - 2\pi x)$$

Setting this derivative equal to zero, we find that

$$(2 + \frac{\pi}{2})x = \frac{1}{4}L$$
$$x = \frac{1}{8 + 2\pi}L$$

Which is as above. However, in this case we have $A'' > 0$, so that our point here is a minimum (as above). Hence, we must check the two extreme values $x = 0$ or $x = \frac{L}{2\pi}$. This computation is exactly as in part (b), completing the question.

□

6.3 Solutions to Measure Theory Exercises

Solution. 5.3.2 Now, if C is empty then the claim is obvious. Consider then that if C is non-empty we have that the function $D : \mathbb{R}^n \to \mathbb{R}$ given by $D(x) = d(x, C) = \min\{d(x, c), c \in C\}$ (the minimum is achieved as C is closed) is well-defined and continuous. Then, consider the family of sets $U_n = D^{-1}(-1, 1/n)$. All of the U_n are open by the continuity of D, and hence we have that the intersection of all of the U_n is C, as every point not in C is at least some positive distance away from C. Hence, we have that every closed set is G_δ. The other claim does follow immediately, for if U is open, and $C = U^c$, then $C = \bigcap U_n$ means that $U = \bigcup U_n^c$, where all of these are closed. Hence both claims are proved. □

6.4 Solutions to Probability Theory Exercises