Math Camp Handouts[1]

Sergio Ocampo Díaz

University of Minnesota

Contents

[1]These notes are intended to summarize the main concepts, definitions and results covered in the math refresher course for the Economics PhD of the University of Minnesota. The material is not my own, these notes only include selected sections of books or articles relevant to the first year PhD sequence. Please let me know of any errors that persist in the document. E-mail: ocamp020@umn.edu.

[2]This section takes material from solutions to some of the questions of Macroeconomic Theory preliminary exams at the University of Minnesota. The solutions were made with Dominic Smith.

Part I
Real Analysis

The first part of this notes covers the most basic results from real analysis, it does not seek to be comprehensive or original in the selection of the material, and it does not give a treatment at the graduate level. The idea is instead to cover the theorems that are most often encountered in the different graduate courses in economics, leaving further treatment of them to a graduate level course in mathematics.

In order to maintain the presentation simple and to fix ideas on how the proofs work almost all the material (until section 5) is presented for the special case of the real numbers, and no mention is done of its topological properties. I introduce the basic concepts needed to follow the derivations in the first section and then proceed to treat sequences in $\mathbb{R}$ continuity of real functions and limits of sequences of real functions. After this I present the extension of the main results to $\mathbb{R}^n$ and basic topology concepts. I skip a formal construction of the set of real numbers because of little value added to the graduate courses in economics, as well as differentiability and integrability of function because I consider the results on these two topics to be well known and because their derivation follows closely that of continuity of functions. I also avoid the treatment of series and their convergence, this is sometimes useful in economics but time doesn't allow to give a good coverage of this topic.

Overall I follow closely (almost verbatim) the presentation in Wade (2010). This book is not at all a graduate level textbook, but I have chosen it because it presents all the material that I consider relevant in a clear and approachable manner. The objective of this approach to the material is to give the students the ability to read, understand and ultimately use more advance textbooks when needed, without having to pay a high cost of entry to the concepts of real analysis.

1 Preliminaries

1.1 Distance

In almost all of the topics to be covered the concept of distance or size is present. In the real numbers this concept is given by the absolute value, because of it I begin by defining it and establishing some of its properties. Later on a more general treatment in terms of metrics and inner products will be given.

Definition 1.1. (Absolute Value) Let $a \in \mathbb{R}$ then $|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$

Proposition 1.1. *Let $a \in \mathbb{R}$ and $M \geq 0$. $|a| \leq M$ if and only if $-M \leq a \leq M$.*

Proof. To prove an if and only if one has to prove both implications:

i. If $|a| \leq M$ then $-M \leq a \leq M$.

   Since $|a| \leq M$ then by multiplying by -1 we get $-M \leq -|a|$.

   Case 1. $a \geq 0$ then $-M \leq 0 \leq a = |a| \leq M$ so $-M \leq a \leq M$.

   Case 2. $a < 0$ then $-M \leq -|a| = a \leq 0 \leq M$ so $-M \leq a \leq M$.

   We have then proven the implication in all possible cases.

ii. If $-M \leq a \leq M$ then $|a| \leq M$.

   Case 1. $a \geq 0$ then $|a| = a \leq M$.

   Case 2. $a < 0$. Note that since $-M \leq a$ we also have, $-a \leq M$. This gives $|a| = -a \leq M$.

   Again, we have then proven the implication in all possible cases.

This completes the proof since both implications have been proven.

□

Proposition 1.2. *The absolute value satisfies the following properties:*

i. ***(Positive definiteness)*** $\forall_{a \in \mathbb{R}}\, |a| \geq 0$, *moreover,* $|a| = 0$ *if and only if* $a = 0$.

ii. ***(Symmetry)*** $\forall_{a,b \in \mathbb{R}}\, |a - b| = |b - a|$

iii. ***(Triangle Inequalities)***

$$\forall_{a,b \in \mathbb{R}} \quad |a + b| \leq |a| + |b| \quad \wedge \quad |a| - |b| \leq |a - b| \quad \wedge \quad ||a| - |b|| \leq |a - b|$$

Proof. The properties are established one at time:

i. **(Positive definiteness)** By definition if $a > 0$ then $|a| = a > 0$. If $a < 0$ then $|a| = -a > 0$. Finally if $a = 0$ then $|a| = a = 0$. By the trichotomy property these are all the possible cases. it is always the case the $|a| \geq 0$ and $|a| = 0$ only if $a = 0$.

ii. **(Symmetry)** Let $a, b \in \mathbb{R}$, either $a = b$, $a > b$ or $a < b$. In the first case $a - b = b - a = 0$ and so are their absolute values. In the second case $a - b > 0$ so $|a - b| = a - b = -(b - a) = |b - a|$. The third case is proven in the same way.

iii. **(Triangle inequalities)** Let $a, b \in \mathbb{R}$.

(a) First note that for all $x \in \mathbb{R}$ it holds that $|x| \leq |x|$ so by Proposition 1.1:

$$-|a| \leq a \leq |a| \quad \wedge \quad -|b| \leq b \leq |b|$$

Summing the inequalities:

$$-(|a| + |b|) \leq a + b \leq |a| + |b|$$

Then, again, by Proposition 1.1:

$$|a + b| \leq |a| + |b|$$

(b) To prove this inequality note that:

$$|a| - |b| = |(a - b) + b| - |b|$$

Then by the inequality above:

$$|a| - |b| \leq |a - b| + |b| - |b| = |a - b|$$

which is the desired result.

(c) By Proposition 1.1 we need to verify:

$$-|a - b| \leq |a| - |b| \leq |a - b|$$

The RHS inequality is proven in the previous case. The left hand side follows in the same way, note:

$$|b| - |a| \leq |b - a| = |a - b|$$

by the second inequality and symmetry. Then by multiplying by -1 we get:

$$-|a - b| \leq |a| - |b|$$

which is the desired result.

$\square$

Now that we have defined the concept of absolute value we can show how it generalizes to the concept of metric and norm in $\mathbb{R}$.

Definition 1.2. (Metric Space) A metric space is a set E, together with a metric (that is a distance function) $\rho : E \times E \to \mathbb{R}$ such that for all $x, y, z \in E$ it holds that:

i. $\rho(x, y) \geq 0$ and $\rho(x, y) = 0$ if and only if $x = y$.

ii. $\rho(x, y) = \rho(y, x)$

iii. $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$

Remark 1.1. Note that the set $\mathbb{R}$ together with the function $\rho(x, y) = |x - y|$ is a metric space. The first property is satisfied because of positive definiteness, the second one is equivalent is symmetry and the third one follows from the triangle inequality:

$$\rho(x, y) = |x - y| = |(x - z) + (z - y)| \leq |x - z| + |z - y| = \rho(x, z) + \rho(z, y)$$

Definition 1.3. (Normed Vector Space) A normed vector space is a vector space E, together with a norm (that is a function) $\|\cdot\| : E \to \mathbb{R}$ such that for all $x, y \in E$ and $a \in \mathbb{R}$ it holds that:

i. $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$.

ii. $\|ax\| = |a| \|x\|$

iii. $\|x + y\| \leq \|x\| + \|y\|$

Remark 1.2. Note that the set $\mathbb{R}$ together with the function $\|x\| = |x|$ is a normed vector space. The first property is satisfied because of positive definiteness, the second follows from the definition of the absolute value and the third one is the triangle inequality.

Although is not going to be directly useful in what follows I include here the definition of a vector space:

Definition 1.4. (Vector Space) A vector space is a set E together with and addition and a scalar multiplication operations such that E is closed under both:

$$\forall_{x,y \in E} x + y \in E \qquad \forall_{x \in E} \forall_{a \in \mathbb{R}} a \cdot x \in E$$

The operations have to satisfy the following properties:

i. $x + y = y + x$

ii. $(x + y) + z = x + (y + z)$

iii. $a \cdot (x + y) = a \cdot x + a \cdot y$

iv. $(a + b) \cdot x = a \cdot x + b \cdot x$

v. $(ab) \cdot x = a \cdot (b \cdot x)$

vi. $\exists_{\vec{0} \in E} \forall_{x \in E} x + \vec{0} = x \quad \wedge \quad 0 \cdot x = \vec{0}$

vii. $1 \cdot x = x$

1.2 Sup and Inf

Another topic that will be visited in the following sections is that of the supremum and infimum of a set. This is of particular relevance when talking about optimization. I start by defining formally what they are.

Definition 1.5. (Supremum) Let $E \subseteq \mathbb{R}$ and $E \neq \emptyset$.

i. E is said to be **bounded above** if and only if $\exists_{M\in\mathbb{R}}\forall_{a\in E} a \leq M$.

ii. A number $M \in \mathbb{R}$ is said to be an **upper bound** for E if and only if $\forall_{a\in E} a \leq M$.

iii. A number $s \in \mathbb{R}$ is called the **supremum** of E if and only if:

   (a) s is an upper bound of E

   (b) for all $M \in \mathbb{R}$, if M is an upper bound of E, then $s \leq M$

In what follows we denote the supremum of E by $s = \sup E$. We will denote $\sup E = \infty$ if E fails to be bounded above.

Proposition 1.3. *A set $E \subseteq \mathbb{R}$ has only one supremum.*

Proof. If E has no upper bound then $\sup E = \infty$ and is unique. Assume E is bounded above and suppose for a contradiction that there exist numbers $s_1 \neq s_2$ that are both supremums of E. By definition they are both upper bounds so it must be that:

$$s_1 \leq s_2 \quad \wedge \quad s_2 \leq s_1$$

by the trichotomy property this implies $s_1 = s_2$ which is a contradiction. So it must be that $s_1 = s_2$. □

A useful property of the supremum is its approximation property.

Proposition 1.4. *Let $E \subseteq \mathbb{R}$ such that* $\sup E < \infty$.

$$\forall_{\epsilon>0}\exists_{a\in E} \sup E - \epsilon < a \leq \sup E$$

Proof. Suppose for a contradiction that the implication doesn't hold. That is:

$$\exists_{e>0}\forall_{a\in E} a \leq \sup E - e$$

So $\sup E - e$ is an upper bound for E, then it must be that the supremum is lower than or equal to it:

$$\sup E \leq \sup E - e$$

which implies $e \leq 0$. But this is a contradiction since $e > 0$ by assumption. So the implication holds. □

Remark 1.3. Its easy to see that if $E \subseteq \mathbb{N}$ and its bounded above then $\sup E \in E$. Just use the approximation property and set $\epsilon < 1$.

The definition and approximation property for the infimum can be inferred easily from the above treatment. We now proceed to establish some properties of the real numbers that use the infimum and supremum.

Theorem 1.1. (Archimedean principle) Given positive real numbers a and b, there exists an integer $n \in \mathbb{N}$ such that $b < na$.

Proof. If $b < a$ then set $n = 1$ and the result follows. If $a < b$ define the set $E = \{k \in \mathbb{N} | ka \leq b\}$. We know that $E \neq \emptyset$ since $1 \in E$. Since $a > 0$ it follows that $k \leq {}^{b}\!/\!_{a}$ for all $k \in E$, so E is bounded above. Then E has a supremum, but since $E \subseteq \mathbb{N}$ we know $\sup E \in E$. Let $n = \sup E + 1$, clearly $n \notin E$ since its larger than $\sup E$, so by definition $na > b$ as desired. $\square$

Theorem 1.2. (Density of the rationals) If $a, b \in \mathbb{R}$ and $a < b$ then there exists $q \in \mathbb{Q}$ such that $a < q < b$.

(the strategy of the proof is to find how much space is there between a and b, then we can locate a fraction that varies in units less than that space.

Proof. By assumption $b - a > 0$, then by the Archimedean principle there exists $n \in \mathbb{N}$ such that $n\,(b - a) > 1$. Note that this implies that ${}^{1}\!/\!_{n} < b - a$.

Case 1. $b > 0$. Consider the set $E = \{k \in \mathbb{N} | b \leq {}^{k}\!/\!_{n}\}$, this is set is nonempty by the Archimedean principle. Since $E \subseteq \mathbb{N}$ then it has a least element (naturals are bounded below and ordered), let this element be k_0. Set $m = k_0 - 1$ and $q = {}^{m}\!/\!_{n}$. Note that $m \notin E$ this can happen in two ways, either $m \leq 0$ or ${}^{m}\!/\!_{n} < b$, in either case $q < b$.
Finally note that:

$$a = b - (b - a) \leq \frac{k_0}{n} - (b - a) < \frac{k_0}{n} - \frac{1}{n} = \frac{m}{n} = q$$

which gives the inequality: $a < q < b$.

Case 2. $b \leq 0$. Choose by the Archimedean principle $k \in \mathbb{N}$ such that $k + b > 0$. By the first case there exists $r \in \mathbb{Q}$ such that $k + a < r < k + b$. Let $q = r - k$, clearly $q \in \mathbb{Q}$ and $a < q < b$.

$\square$

The last property talks about the monotonicity of sets in $\mathbb{R}$.

Proposition 1.5. *Let $A, B \subset \mathbb{R}$ and $A, B \neq \emptyset$ and $A \subseteq B$ then:*

i. If $\sup B < \infty$, *then* $\sup A \leq \sup B$.

ii. If $\inf B > -\infty$, *then* $\inf A \geq \inf B$.

Proof. Since $A \subseteq B$ any upper bound of B is an upper bound of A, inf particular $\sup B$ is an upper bound for A, it then follows from the definition of supremum that $\sup A \leq \sup B$. Clearly $-A \subseteq -B$, then, by the first part $\sup(-A) \leq \sup(-B)$. Note that $\inf A = -\sup(-A)$, so the result follows:

$$\inf A = -\sup(-A) \geq -\sup(-B) = \inf B$$

$\square$

2 Sequences in $\mathbb{R}$

2.1 Limits

Sequences and their limits will be used extensively to build the concept of continuity. A sequence is a function $f : \mathbb{N} \to \mathbb{R}$ so that for each $n \in \mathbb{N}$ there is a real number $x_n = f(n)$. Its common to represent the sequence by its values $\{x_n\}$ instead of the function that generates them. What interests us is to determine if as $n \to \infty$ the numbers x_n approach a fixed point a. Formally:

Definition 2.1. (Convergent Sequence) A sequence $\{x_n\}$ is said to converge to a real number a ($x_n \to a$) if and only if:

$$\forall_{\epsilon>0}\exists_N\forall_{n\geq N}\,|x_n - a| < \epsilon$$

a is called the limit of x_n.

Proposition 2.1. *A sequence $\{x_n\}$ has at most one limit.*

Proof. Suppose for a contradiction that $\{x_n\}$ has two limits a and b such that $a \neq b$. Then by definition for $\epsilon > 0$ there exists natural numbers N_1 and N_2 such that$|x_n - a| \leq \epsilon/2$ if $n \geq N_1$ and $|x_n - b| \leq \epsilon/2$ if $n \geq N_2$. Then for $n \geq N = \max\{N_1, N_2\}$ both conditions hold. Using the triangle inequality:

$$|a - b| = |(a - x_n) + (x_n - b)| \leq |a - x_n| + |x_n - b| < \epsilon$$

Since ϵ was arbitrarily chosen we have, for all $\epsilon > 0$ it holds that $|a - b| < \epsilon$ which implies $a = b$, a contradiction. □

A sequence might not converge and this can happen in different ways, one that will be useful to characterize is when the sequence diverges, formally this is:

Definition 2.2. (Divergent Sequence) A sequence $\{x_n\}$ can diverge to $+\infty$ or $-\infty$:

i. $\{x_n\}$ is said to diverge to $+\infty$ if and only if:

$$\forall_{M\in\mathbb{R}}\exists_{N\in\mathbb{N}}\forall_{n\geq N}x_n > M$$

We denote this by $x_n \to +\infty$.

ii. $\{x_n\}$ is said to diverge to $-\infty$ if and only if:

$$\forall_{M\in\mathbb{R}}\exists_{N\in\mathbb{N}}\forall_{n\geq N}x_n < M$$

We denote this by $x_n \to -\infty$.

Since a sequence has infinitely many terms one could think on sequences made from selection of terms of an original sequence, formally these selections are called subsequences. They are specially useful when there are properties that a selection of the sequence satisfies, but that are not satisfied or difficult to prove for the whole sequence.

Definition 2.3. (Subsequence) A subsequence $\{x_{n_k}\}$ of a sequence $\{x_n\}$ is a sequence indexed by $k \in \mathbb{N}$ such that for $n_k < n_{k+1}$. This means that the sequence x_{n_k} selects elements of x_n without repetition and maintaining the order of the original sequence.

Note that n_k is in itself a sequence (in the natural numbers) and that it is strictly increasing, so as $k \to \infty$ so does n_k. Thus if x_n is approaching a number when n is 'large' so will x_{n_k}.

Proposition 2.2. *Let $\{x_n\}$ be a sequence and $\{x_{n_k}\}$ **any** subsequence. IF $x_n \to a$ as $n \to \infty$ then $x_{n_k} \to a$ as $k \to \infty$.*

Proof. Let $\epsilon > 0$, since $x_n \to a$ there exists N such that $|x_n - a| < \infty$ for all $n > N$.

Now note that since $n_k < n_{k+1}$ it must hold that $k \geq n_k$ (this can be formally shown by induction), then let $K = N$, for $k \geq K$ it holds that $n_k \geq k \geq N$, so by the hypothesis: $|x_{n_k} - a| < \epsilon$ for all $k > K$.

Since this holds for arbitrary ϵ the convergence of $\{x_{n_k}\}$ is established. □

A sequence of real numbers can drift in the space, but some of them are bounded in the sense that their elements are never larger (or lower) than a certain value. This concept is made precise below:

Definition 2.4. (Bounded Sequence) Let $\{x_n\}$ be a sequence of real numbers:

i. $\{x_n\}$ is said to be bounded above if and only if $\exists_{M\in\mathbb{R}} \forall_n x_n \leq M$.

ii. $\{x_n\}$ is said to be bounded below if and only if $\exists_{m\in\mathbb{R}} \forall_n x_n \geq m$.

iii. $\{x_n\}$ is said to be bounded if and only if $\exists_{C\in\mathbb{R}} |x_n| \leq C$.

Note now that if a sequence is converging its values are approaching a given number, so after a certain index N the sequence is bounded by a $\lim x_n + e$ for some $e > 0$. The values before N form a finite list and therefore have a maximum. These two facts are used to establish the following proposition.

Proposition 2.3. *Every convergent sequence is bounded.*

Proof. Let $\{x_n\}$ be a convergent sequence such that $x_n \to a$ and consider $\epsilon = 1$. By assumption of convergence there exists N such that $|x_n - a| < \epsilon$ for $n \geq N$. By the triangle inequality:

$$|x_n| - |a| \leq |x_n - a| < 1$$

So for $n \geq N$ $|x_n| < |a| + 1$ and for $n < N$ $|x_n| \leq M = \max\{|x_1|, \ldots, |x_{N-1}|\}$. Let $C = \max\{|a| + 1, M\}$, we have proven that $\forall_n |x_n| < C$, so $\{x_n\}$ is bounded. □

The converse to this proposition is clearly false since not all bounded sequences converge (think of an alternating sequence). But there is a partial converse provided by the Bolzano Weierstrass Theorem which will be established later. For now we turn to establish some properties of limits of sequences. The same properties and proofs will be extended to functions when covering continuity of real valued functions.

Theorem 2.1. (Squeeze) Let $\{x_n\}$, $\{y_n\}$ and $\{w_n\}$ be real sequences.

i. *If $x_n \to a$ and $y_n \to a$ and there exists $N_0 \in \mathbb{N}$ such that $x_n \le w_n \le y_n$ for $n \ge N_0$ then $w_n \to a$.*

ii. *If $x_n \to 0$ and y_n is bounded then $x_n y_n \to 0$.*

Proof. We prove each statement separately:

i. Let $\epsilon > 0$, by convergence of x_n and y_n there exists two numbers N_x and N_y such that

$$|x_n - a| \le \epsilon \qquad \wedge \qquad |y_n - a| < \epsilon$$

for $n \ge N = \max\{N_x, N_y, N_0\}$. By Proposition 1.1 these inequalities imply:

$$a - \epsilon < x_n < a + \epsilon \qquad \wedge \qquad a - \epsilon < y_n < a + \epsilon \qquad x_n \le w_n \le y_n$$

Joining we get:

$$a - \epsilon < x_n \le w_n \le y_n < a + \epsilon$$

Which by Proposition 1.1 implies $|w_n - a| < \epsilon$ for all $n \ge N$, thus proving convergence.

ii. Let $\epsilon > 0$ and C a bound for y_n, then there exists $N \in \mathbb{N}$ such that $|x_n| \le \epsilon/C$ for all $n \ge N$. Then $|x_n y_n - 0| = |x_n|\,|y_n| \le |x_n|\,C < \epsilon$ thus proving convergence.

□

The last two results make clear that when talking about convergence only the behavior of the sequence for large n matters, since the beginning of the sequence can be conveniently dropped when needed.

A common practice is to construct a sequence that converges to a certain point of relevance, like boundary point of a set or the supremum of it. The following proposition shows that this can always be done for the supremum or infimum of a set (provided it is finite), it is also true that it can be done with a boundary point but the result is postponed until the concept of openness of a set is discussed.

Proposition 2.4. *$E \subset \mathbb{R}$ and $\sup E < \infty$. Then there exists a sequence $\{x_n\}$ such that $x_n \in E$ for all n and $x_n \to \infty$.*

Proof. For each $n \in \mathbb{N}$ consider $\epsilon = \frac{1}{n}$, by the property of approximation to supremum (Proposition 1.4) there exists $w \in E$ such that $\sup E - \epsilon < w \le \sup E$, let this w be called w_n. So the sequence $\{w_n\}$ satisfies $\sup E - \frac{1}{n} < w_n \le \sup E$. Note that $x_n = \sup E - \frac{1}{n}$ and $y_n = \sup E$ are sequences that satisfy $x_n \to \sup E$ and $y_n \to \sup E$, so by the squeeze theorem we get the result $w_n \to \sup E$ and $w_n \in E$ for all n. □

Limits of sequences are known for being well behaved in the sense that they respect the basic arithmetic operations.

Proposition 2.5. *Let $\{x_n\}$ and $\{y_n\}$ be convergent sequences and $\alpha \in \mathbb{R}$ a number. Then:*

i. $\lim(x_n + y_n) = \lim x_n + \lim y_n$

ii. $\lim(\alpha x_n) = \alpha \lim x_n$

iii. $\lim(x_n y_n) = (\lim x_n)(\lim y_n)$

iv. *If* $y_n \neq 0$ *and* $\lim y_n \neq 0$ *then* $\lim \frac{x_n}{y_n} = \frac{\lim x_n}{\lim y_n}$

Proof. For the proof let $x_n \to x$ and $y_n \to y$:

i. Let $\epsilon > 0$ as before there is a number $N \in \mathbb{N}$ such that $|x_n - x| < \frac{\epsilon}{2}$ and $|y_n - y| < \frac{\epsilon}{2}$, so by the triangle inequality:

$$|(x_n + y_n) - (x + y)| \leq |x_n - x| + |y_n - y| < \epsilon$$

for all $n \geq N$, thus proving convergence of $(x_n + y_n) \to (x + y)$.

ii. Note that is equivalent to show that $\alpha x_n \to \alpha x$ or that $(\alpha x_n - \alpha x) \to 0$, but the sequence $z_n = \alpha(x_n - x)$ converges to zero since $(x_n - x) \to 0$ and α is bounded.

iii. Note that since $\{x_n\}$ and $\{y_n\}$ converge they are bounded. What we want to show is that $(x_n y_n - xy) \to 0$. Note that:

$$x_n y_n - xy = x_n(y_n - y) + (x_n - x)y$$

Since x_n is bounded and $y_n - y \to 0$ then $x_n(y_n - y) \to 0$ and similarly $(x_n - x)y \to 0$. Then by part one their sum converges to zero as well proving the result.

iv. We first prove the following result: if $y_n \to y$ then $\frac{1}{y_n} \to \frac{1}{y}$.
First note that if $\{y_n\}$ is bounded and non-zero there exists $c, C \geq 0$ such that $c \leq |y_n| \leq C$ Then $\frac{1}{C} \leq \left|\frac{1}{y_n}\right| \leq \frac{1}{c}$, so $\left\{\frac{1}{y_n}\right\}$ is a bounded sequence.
Now note that since $y_n \to y$ for any $\epsilon > 0$ there exists N such that $|y_n - y| \leq c|y|\epsilon$ for $n \geq N$, and that $\left|\frac{1}{y_n} - \frac{1}{y}\right| = \frac{|y_n - y|}{|y_n||y|}$. Because of boundedness we get:

$$\left|\frac{1}{y_n} - \frac{1}{y}\right| = \frac{|y_n - y|}{|y_n||y|} \leq \frac{|y_n - y|}{c|y|} < \epsilon$$

Proving convergence. Then the result follows by applying previous result on multiplication to x_n and $\frac{1}{y_n}$.

$\square$

A final property of limits is that they respect order in the following sense:

Proposition 2.6. *Let* $\{x_n\}$ *and* $\{y_n\}$ *be convergent sequences such that there exists* $N_0 \in \mathbb{N}$ *for which if* $n \geq N_0$ *then* $x_n \leq y_n$. *Then* $\lim x_n \leq \lim y_n$.

Proof. Suppose for a contradiction that $x_n \leq y_n$ for $n \geq N_0$ but $\lim x_n > \lim y_n$. Since both sequences converge there exists an $N > N_0$ such that $|x_n - \lim x_n| < \epsilon$ and $|y_n - \lim y_n| < \epsilon$ for $\epsilon = \frac{\lim x_n - \lim y_n}{2} > 0$. Using Proposition 1.1:

$$x_n > \lim x_n - \epsilon \qquad \wedge \qquad \lim y_n + \epsilon > y_n$$

Note that:

$$\lim x_n - \epsilon = \frac{\lim x_n + \lim y_n}{2} = \lim y_n + \epsilon$$

So joining:

$$x_n > y_n$$

which contradicts the assumption. So the result holds. □

2.2 Theorems

In this subsection I cover three important results in real analysis, namely the monotone convergence theorem, the Bolzano-Weierstrass theorem and the completeness of $\mathbb{R}$. These theorems talk about conditions to establish convergence of a series in $\mathbb{R}$.

The monotone convergence theorem establishes that a monotone and bounded sequence is necessarily convergent. To get this result we first define formally monotonicity of a sequence.

Definition 2.5. (Monotone Sequence) Let $\{x_n\}$ be a sequence of real numbers.

i. $\{x_n\}$ is said to be increasing if and only if $x_1 \leq x_2 \leq \dots$.

ii. $\{x_n\}$ is said to be decreasing if and only if $x_1 \geq x_2 \geq \dots$.

iii. $\{x_n\}$ is said to be monotone if and only if its either increasing or decreasing.

Clearly strictly increasing and strictly decreasing sequences are defined with the corresponding strict inequalities. More importantly its not necessary for the sequence to be always monotone, as before it suffices for a sequence to be monotone for 'large' n.

Theorem 2.2. (Monotone convergence) If $\{x_n\}$ is monotone and bounded then $x_n \to x$ and $|x| < \infty$.

Remark 2.1. Note that if a sequence is increasing it suffices to ask from it to be bounded above, since it is immediately bounded below by x_1. Similarly, if a sequence is bounded below it suffices to ask from it to be bounded below since its bounded above by x_1.

Proof. Suppose first that that $\{x_n\}$ is increasing and bounded above. Let $E = \{x_n | n \in \mathbb{N}\}$ be the set of all values of the sequence and $x = \sup E$. Since $\{x_n\}$ is bounded above there exists $C < \infty$ such that $|x_n| < C$ for all n, so C is an upper bound for E and thus $x \leq C < \infty$ by definition of supremum. Now let $\epsilon > 0$, by the property of approximation to suprema (Proposition 1.4) there exists $N \in \mathbb{N}$ such that

$$x - \epsilon < x_N \leq x$$

Since $\{x_n\}$ is increasing $x_N \leq x_n$ for all $n \geq N$, so $x - \epsilon < x_n$, moreover, since x is the supremum of E, $x_n \leq x$ for all n. So for all $n \geq N$:

$$x - \epsilon < x_n < x$$

Since $\epsilon > 0$:

$$-\epsilon < x_n - x < \epsilon \longrightarrow |x_n - x| < \epsilon$$

proving convergence of x_n to x.

The proof for a decreasing sequence bounded below follows the same step but using the infimum of set E. It also follows directly from the fact that if $\{x_n\}$ is increasing then $\{-x_n\}$ is decreasing and that $\inf \{x_n | n \in \mathbb{N}\} = \sup \{-x_n | n \in \mathbb{N}\}$, then by the first part of the theorem we get the result. $\square$

Before establishing the stronger result of the Bolzano-Weierstrass theorem we first need to define the notion of nested sets and the nested interval property.

Definition 2.6. (Nested Intervals) A sequence of sets $\{I_n\}$ is said to be nested if and only if $I_1 \supseteq I_2 \supseteq \dots$

Proposition 2.7. *If $\{I_n\}$ is a nested sequence of nonempty, closed and bounded intervals then*

$$E = \bigcap_{n \in \mathbb{N}} I_n = \{x | \forall_n x \in I_n\}$$

is nonempty.
Moreover, if $|I_n| \to 0$ then E contains exactly one number.

Proof. Let $I_n = [a_n, b_n]$, since I_n is nested it must be that the sequence $\{a_n\}$ is increasing and bounded above by b_1, similarly $\{b_n\}$ is a decreasing sequence bounded below by a_1. By the monotone convergence theorem $a_n \to a$ and $b_n \to b$, moreover since limits respect order and $a_n \le b_n$ for all n it follows that $a \le b$, moreover $a_n \le a \le b \le b_n$ by construction. So $x \in E$ if and only if $x \in [a, b]$. (one direction of the claim is immediate, the second one can be proven using the property of approximation to sup and inf).
IF $|I_n| \to 0$ then $b_n - a_n \to 0$, so:

$$\lim b_n - \lim a_n = 0 \longrightarrow b = a$$

which gives the result since $E = [a, a] = \{a\}$. □

Now we can present the Bolzano-Weierstrass theorem, a partial converse to the boundedness of convergent sequences.

Theorem 2.3. (Bolzano-Weierstrass) Every bounded sequence has a convergent subsequence.

Proof. Let $\{x_n\}$ be a bounded sequence. Then there exists numbers $a, b \in \mathbb{R}$ such that $x_n \in [a, b]$ for all n. Let $I_1 = [a, b]$, note that all I_1 contains all the elements of the sequence, let $n_1 = 1$ so that $x_{n_1} = x_1 \in I_1$.

Now consider the intervals $I^{'} = \left[a, \frac{a+b}{2}\right]$ and $I^{''} = \left[\frac{a+b}{2}, b\right]$, clearly $I_1 = I^{'} \cup I^{''}$ and either $I^{'}$ or $I^{''}$ (or both) contain infinitely many elements of the sequence. Then let the interval that still contains infinitely many elements of the sequence be I_2 and pick $n_2 > n_1$ such that $x_{n_2} \in I_2$. Note that the measure o the interval is now $|I_2| = \frac{a+b}{2}$.

Repeat this process to construct sequences of sets and natural numbers $\{I_k\}$, $\{n_k\}$ such that $I_k \supset I_{k+1}$, $|I_k| = \frac{a+b}{2^k}$, $n_k < n_{k+1}$ and $x_{n_k} \in I_k$. All this properties can be verified formally by induction.

Note that the sequence $\{I_k\}$ is formed then by nested intervals and that $|I_k| \to 0$, so there exists a single element $x \in \bigcap I_k$. Now we prove that the subsequence $\{x_{n_k}\}$ converges to x, note that for all k $x \in I_k$ so:

$$0 \le |x_{n_k} - x| \le I_k = \frac{b-a}{2^k}$$

By the squeeze theorem $|x_{n_k} - x| \to 0$ thus proving convergence. □

The final result presented here is that of completeness of the real numbers. The completeness of a space relates convergent sequences with Cauchy sequences which will be defined below. Not all spaces are complete, but completeness is particularly important to establish the convergence of a sequence when its limit is not know, and later on to prove the contraction mapping theorem.

Definition 2.7. (Cauchy Sequence) A sequence $\{x_n\}$ is said to be Cauchy if and only if:

$$\forall_{\epsilon>0}\exists_{N\in\mathbb{N}}\forall_{n,m\geq N}\,|x_n - x_m| < \epsilon$$

So a Cauchy sequence is characterized by having its elements closer together. The following proposition establishes that a convergent sequence is necessarily Cauchy. Intuitively if the sequence is going towards its limit then its elements have to be close to it, and thus close to each other.

Proposition 2.8. *If $\{x_n\}$ is convergent then it is Cauchy.*

Proof. Let x be the limit of $\{x_n\}$ and take $\epsilon > 0$, then there exists $N \in \mathbb{N}$ such that for $n, m \geq N$:

$$|x_n - x| < \frac{\epsilon}{2} \quad \wedge \quad |x_m - x| < \frac{\epsilon}{2}$$

Then by the triangle inequality:

$$|x_n - x_m| = |(x_n - x) + (x - x_m)| \leq |x_n - x| + |x_m - x| < \epsilon$$

Which proves that $\{x_n\}$ is Cauchy. □

Now we prove the completeness of the real numbers.

Theorem 2.4. (Cauchy) Let $\{x_n\}$ be a sequence of real numbers. $\{x_n\}$ is convergent if and only if $\{x_n\}$ is Cauchy.

Proof. Note that by the previous proposition we only need to prove that Cauchy sequences converge.

We first prove that a Cauchy sequence is bounded. Let $\{x_n\}$ be Cauchy, then for $\epsilon = 1$ there exists N such that $|x_N - x_m| < 1$ for all $m \geq N$. Then by the triangle inequality:

$$|x_m| - |x_N| < |x_N - x_m| < 1 \quad \longrightarrow \quad |x_m| < 1 + |x_N|$$

Then $\{x_n\}$ is bounded by $M = \max\{1 + |x_N|, |x_1|, |x_2|, \ldots, |x_{N-1}|\}$.

Since $\{x_n\}$ is bounded it has a convergent subsequence by the Bolzano-Weierstrass theorem. Let $\{x_{n_k}\}$ be such subsequence and x its limit. Now let $\epsilon > 0$ and note that, since $\{x_n\}$ is Cauchy, there is a number N_1 such that:

$$\forall_{n,m\geq N_1}\,|x_n - x_m| < \frac{\epsilon}{2}$$

And since $\{x_{n_k}\}$ is convergent there exists a number K_2 such that:

$$\forall_{k\geq K_1}\,|x_{n_k} - x| < \frac{\epsilon}{2}$$

So for k such that $k \geq K$ and $n_k \geq N_1$ we have:

$$|x_n - x| = |(x_n - x_{n_k}) + (x_{n_k} - x)| \leq |x_n - x_{n_k}| + |x_{n_k} - x| < \epsilon$$

Note that this holds for all $n \geq N$, thus proving convergence. □

2.3 LimSup and LimInf

Some properties of sequences can be further characterized in terms of the limit supremum and limit infimum. These objects give information about the behavior of the sequence for large n, that is, they are the limits of the supremum and infimum of the tail of the sequence. More formally:

Definition 2.8. (Limsup Liminf) Let $\{x_n\}$ be a real sequence. The limit supremum and limit infimum are extended real numbers such that:

$$\limsup x_n = \lim_{n\to\infty}\left(\sup_{k\geq n} x_k\right) \qquad \wedge \qquad \liminf x_n = \lim_{n\to\infty}\left(\inf_{k\geq n} x_k\right)$$

These objects can also be constructed as follows: let $\{s_n\}$ and $\{t_n\}$ be sequences of extended real numbers such that:

$$s_n = \sup_{k\geq n} x_k = \sup\{x_k|k\geq n\} \qquad \wedge \qquad t_n = \inf_{k\geq n} x_k = \inf\{x_k|k\geq n\}$$

Since the sets $E_n = \{x_k|k \geq n\}$ satisfy $E_n \supseteq E_{n+1}$ by the monotone property of sup and inf ew get $s_n \geq s_{n+1}$ and $t_n \leq t_{n+1}$, so the sequences are monotone but potentially not bounded. If they are bounded by the monotone convergence theorem there are numbers s and t such that $s_n \to s$ and $t_n \to t$, if the sequences are unbounded then s or t are $+\infty$ of $-\infty$. By definition $s = \limsup x_n$ and $t = \liminf x_n$.

The following results are stated without a proof.

Proposition 2.9. *Let $\{x_n\}$ be a sequence and $s = \limsup x_n$ and $t = \liminf x_n$. There are subsequences $\{x_{n_k}\}$ and $\{x_{n_l}\}$ such that $x_{n_k} \to s$ as $k \to \infty$ and $x_{n_l} \to t$ as $l \to \infty$.*

The proof of this proposition is carried out by cases since the s (or t) can be either equal to $+\infty$, $-\infty$ or to a real number $-\infty < s < +\infty$. The first two cases can be solved using the squeeze theorem and the third case is requires more care and is approached using the approximation property of supremum.

Proposition 2.10. *Let $\{x_n\}$ be a sequence and x an extended real number. $x_n \to x$ if and only if $x = \limsup x_n = \liminf x_n$.*

Proof. If $x_n \to x$ then by proposition 2.2 all its subsequences converge to x. By the previous proposition there are subsequences that converge to $\limsup x_n$ and $\liminf x_n$ since sequences can have at most one limit it must be that $x = \limsup x_n = \liminf x_n$.

Now assume that $x = \limsup x_n = \liminf x_n$ holds. There are two cases:

Case 1. $x = \pm\infty$. wlog we consider $x = \infty$. Then it must be that for all $M > 0$ there is a number N such that $\inf_{k\geq n} x_k > M$ for all $n \geq N$, since $\inf_{k\geq n} x_k$ is a lower bound for the set $\{x_k|k \geq n\}$ it follows that $x_n > M$ for all $n \geq N$, so $x_n \to \infty$, proving the result.

Case 2. $-\infty < x < \infty$. Now we show that $\{x_n\}$ is Cauchy and hence, by completeness, convergent. Let $\epsilon > 0$. Since $\lim(\sup x_n) = x$ and $\lim(\inf x_n) = x$ there is an N such that:

$$\sup_{k\geq N} x_k - x < \frac{\epsilon}{2} \qquad \wedge \qquad x - \inf_{k\geq N} x_k < \frac{\epsilon}{2}$$

The absolute value is dispensed with since the sequence $\{\sup x_n\}$ is decreasing and the sequence $\{\inf x_n\}$ increasing.
Now let $m, n \geq N$ and wlog let $x_n \geq x_m$, we have:

$$|x_n - x_m| = x_n - x_m = (x_n - x) + (x - x_m) \leq \left(\sup_{k\geq N} x_k - x\right) + \left(x - \inf_{k\geq N} x_k\right) < \epsilon$$

Thus proving that $\{x_n\}$ is Cauchy and convergent. But we know by the previous proposition that there are subsequences that converge to $\limsup x_n = x$, so by proposition 2.2 all subsequences and the sequence has to converge to x.

This finishes the proof □

Its clear that the limsup is greater than or equal that elements of a sequence for large n, in the same way the liminf is lower than or equal. This gives the following property.

Proposition 2.11. *Let $\{x_n\}$ be a real sequence and $\{x_{n_k}\}$ a subsequence. If $x_{n_k} \to x$ then*

$$\liminf x_n \leq x \leq \limsup x_n$$

The next property follows from the same fact.

Proposition 2.12. *A sequence $\{x_n\}$ is bounded above if and only if $\limsup x_n < \infty$ and bounded below if and only if $\liminf x_n > -\infty$.*

Finally the limsup and liminf respect inequalities:

Proposition 2.13. *Let $\{x_n\}$ and $\{y_n\}$ be sequences such that $x_n \leq y_n$ for large n, then:*

$$\liminf x_n \leq \liminf y_n \qquad \wedge \qquad \limsup x_n \leq \limsup y_n$$

3 Continuity on $\mathbb{R}$

3.1 Limits

Here we study the behavior of a real functions, functions whose domain and range are subsets of the real numbers. It is first necessary to define the limit of a function at a point, this is, to what value does a function tends to when its evaluated near a point in the domain.

Definition 3.1. (Limit of a Function) Let $a \in \mathbb{R}$ and I an open interval that contains a. Let f be a real function defined on I(except possible at a). $f(x)$ is said to converge to y as $x \to a$ if and only if:

$$\forall_{\epsilon>0}\exists_{\delta>0} 0 < |x-a| < \delta \longrightarrow |f(x) - y| < \epsilon$$

We write $y = \lim\limits_{x \to a} f(x)$.

This definition of the limit of a function resembles that of the limit of a sequence, in fact one would think that if there is a sequence $\{x_n\}$ such that $x_n \to a$ then it should follow that $f(x_n) \to a$ as $n \to \infty$. This leads to the sequential characterization of limits.

Theorem 3.1. (Sequential characterization of limits) *Let $a \in \mathbb{R}$ and I an open interval that contains a. Let f be a real function defined on I(except possible at a). f has a limit at a $\left(y = \lim\limits_{x \to a} f(x)\right)$ if and only if $f(x_n) \to y$ for all sequence $\{x_n\}$ such that $x_n \in I \setminus \{a\}$ and $x_n \to a$.*

Proof. Suppose $y = \lim\limits_{x \to a} f(x)$ and let $\epsilon > 0$, then there exists $\delta > 0$ such that

$$0 < |x-a| < \delta \longrightarrow |f(x) - y| < \epsilon$$

Let $\{x_n\}$ be such that $x_n \in I \setminus \{a\}$ and $x_n \to a$, then there exists N such that $|x_n - a| < \delta$ for $n \geq N$, then it holds that:

$$|f(x_n) - y| < \epsilon$$

for all $n \geq N$. This proves that $f(x_n) \to y$, for all $\{x_n\}$ such that $x_n \to a$.

To prove the converse suppose that $f(x_n) \to y$, for all $\{x_n\}$ such that $x_n \to a$. Suppose for a contradiction that $y \neq \lim\limits_{x \to a} f(x)$, then there exists $\epsilon_0 > 0$ such that for all δ the implication above doesn't hold. In particular for each $n \in \mathbb{N}$ there is a value $x_n \in I \setminus \{a\}$ such that for $\delta_n = \frac{1}{n}$:

$$0 < |x_n - a| < \delta_n \quad \wedge \quad |f(x_n) - y| \geq \epsilon_0$$

As $n \to \infty$ we have $\delta_n \to 0$, so by the squeeze theorem the sequence $\{x_n\}$ converges to a. Then by assumption $f(x_n) \to y$, which implies that there exists N such that $|f(x_n) - y| < \epsilon_0$ for all $n \geq N$. But by construction $|f(x_n) - y| \geq \epsilon_0$ for all n. This is a contradiction. $\square$

This last theorem is of great importance since it allows to extend all the results obtained for sequences of real numbers to limits of functions. For instance:

Proposition 3.1. *Let $a \in \mathbb{R}$ and I an open interval that contains a. Let f and g be real functions defined on I (except possible at a). If $f(x)$ and $g(x)$ converge as $x \to a$, then so do the functions $(f+g)(x)$, $(fg)(x)$, $(\alpha f)(x)$ and $(f/g)(x)$ (when the limit of $g(x)$ is nonzero). In fact:*

i. $\lim\limits_{x\to a}(f+g)(x) = \lim\limits_{x\to a} f(x) + \lim\limits_{x\to a} g(x)$.

ii. $\lim\limits_{x\to a}(fg)(x) = \left(\lim\limits_{x\to a} f(x)\right)\left(\lim\limits_{x\to a} g(x)\right)$.

iii. $\lim\limits_{x\to a}(\alpha f)(x) = \alpha\left(\lim\limits_{x\to a} f(x)\right)$.

iv. $\lim\limits_{x\to a}(f/g)(x) = \lim\limits_{x\to a} f(x) / \lim\limits_{x\to a} g(x)$.

Other properties of sequences, like the squeeze theorem and the limit of an inequality can also be extended to functions. The proof just requires to call the equivalent theorem for sequences.

Proposition 3.2. *Let $a \in \mathbb{R}$ and I an open interval that contains a. Let f, g and h be real functions defined on I (except possible at a). Then:*

i. If $g(x) \le h(x) \le g(x)$ for all $x \in I \setminus \{a\}$ and $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = y$ then $\lim\limits_{x\to a} h(x) = y$.

ii. If $|g(x)| < M$ for all $x \in I \setminus \{a\}$ and $f(x) \to 0$ as $x \to a$ then $\lim\limits_{x\to a} f(x)\, g(x) = 0$.

And,

Proposition 3.3. *Let $a \in \mathbb{R}$ and I an open interval that contains a. Let f and g be real functions defined on I (except possible at a). If f and g have limits as $x \to a$ and $g(x) \le f(x)$ for all $x \in I \setminus \{a\}$ then $\lim\limits_{x\to a} g(x) \le \lim\limits_{x\to a} f(x)$.*

There are functions that cannot be defined in an open interval, to handle them its useful to define one-sided limits.

Definition 3.2. (One Sided Limits) Let $a \in \mathbb{R}$.

i. A function f is said to converge to y when x approaches a from the right if and only if f is defined on some open interval I with left endpoint a and

$$\forall_{\epsilon>0}\exists_{\delta>0} a < x < a+\delta \longrightarrow |f(x) - y| < \epsilon$$

Then we write $y = \lim\limits_{x\to a^+} f(x)$.

ii. A function f is said to converge to y when x approaches a from the left if and only if f is defined on some open interval I with left endpoint a and

$$\forall_{\epsilon>0}\exists_{\delta>0} a - d < x < a \longrightarrow |f(x) - y| < \epsilon$$

Then we write $y = \lim\limits_{x\to a^-} f(x)$.

Proposition 3.4. $\lim\limits_{x\to a} f(x) = y$ *if and only if* $\lim\limits_{x\to a^+} f(x) = \lim\limits_{x\to a^-} f(x) = y$.

Proof. If $\lim\limits_{x\to a} f(x) = y$ then for any $\epsilon > 0$ there exists δ such that $0 < |x-a| < \delta \longrightarrow |f(x) - y| < \epsilon$. But in particular, if x is such that $a < x < a + \delta$ then $0 < |x-a| < \delta$, so by the implication $|f(x) - y| < \epsilon$, proving $\lim\limits_{x\to a^+} f(x) = y$. In the same way:

$$a - \delta < x < a \longrightarrow 0 < |x-a| < \delta \longrightarrow |f(x) - y| < \epsilon$$

so we have $\lim\limits_{x\to a^-} f(x) = y$.

To prove the converse let $\lim\limits_{x\to a^+} f(x) = \lim\limits_{x\to a^-} f(x) = y$. Then for $\epsilon > 0$ there are δ_1 and δ_2 such that

$$a < x < a + \delta_1 \longrightarrow |f(x) - y| < \epsilon$$

and

$$a - \delta_2 < x < a \longrightarrow |f(x) - y| < \epsilon$$

So for $\delta = \min\{\delta_1, \delta_2\}$ it holds that if x is such that $0 < |x-a| < \delta$ then either the first statement or the second one hold, thus $|f(x) - y| < \epsilon$, proving that $\lim\limits_{x\to a} f(x) = y$ $\square$

3.2 Continuity

The definition of continuity is basically that a function coincides with its limit at every point. Since some functions are only defined when approaching a point from one direction the definition must take that into account.

Definition 3.3. (Continuous Function) Let $E \subseteq \mathbb{R}$ and $E \neq \emptyset$ and $f : E \to \mathbb{R}$. f is said to be continuous at a point $a \in E$ if and only if:

$$\forall_{\epsilon>0}\exists_{\delta>0}\, |x-a| < \delta \quad \wedge \quad x \in E \longrightarrow |f(x) - f(a)| < \epsilon$$

f is said to be continuous on E if and only if it is continuous at every $x \in E$.

Note that the definition includes $x \in E$ as part of the hypothesis for the implication, this takes care of the one sided limits. The definition also makes clear the following result.

Proposition 3.5. *Let I be an open interval that contains a point a and $f : I \to \mathbb{R}$ a function. f is continuous at $a \in I$ if and only if*

$$f(a) = \lim_{x \to a} f(x)$$

Proof. For δ small $|x-a| < \delta$ implies $x \in I$, then the condition for continuity is the same as that for the limit of a function. $\square$

What is important about the previous result is that it allows for a sequential characterization of continuity in much the same way as that for limits.

Proposition 3.6. *Let $E \subseteq \mathbb{R}$ and $E \neq \emptyset$ and $f : E \to \mathbb{R}$. f is continuous at $a \in E$ if and only if for all sequence $\{x_n\}$, if $x_n \in E$ and $x_n \to a$ then $f(x_n) \to f(a)$.*

Among other things the last two results will make possible to extend results obtained for sequences and limits to continuous functions. For instance:

Proposition 3.7. *Let $E \subseteq \mathbb{R}$ and $E \neq \emptyset$, and $f, g : E \to \mathbb{R}$ functions. If f and g are continuous then so are $f + g$, αf and f/g if $g(x) \neq 0$ for $x \in E$.*

Note that the proof of this results just calls upon that of limits of functions.

Finally a result on composition of functions is presented.

Definition 3.4. (Composition of Functions) Suppose $A, B \subseteq \mathbb{R}$ and $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$. If $f(A) \subseteq B$ then the composition of g with f is the function $g \circ f : A \to \mathbb{R}$ defined by:

$$(g \circ f)(x) = g(f(x))$$

The main result about composition of functions is that the limit and the composition can be interchanged if the functions are continuous. Formally:

Proposition 3.8. *Suppose $A, B \subseteq \mathbb{R}$ and that $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ with $f(A) \subseteq B$.*

i. *If* $A = I\backslash\{a\}$*, where* I *is a non-degenerate interval that either contains* a *or has it as an endpoint, and* $L = \lim\limits_{x\to a} f(x)$ *exists and belong to* B*, and* g *is continuous at* $L \in B$*, then:*

$$\lim_{x\to a} (g \circ f)(x) = g\left(\lim_{x\to a} f(x)\right)$$

ii. *If* f *is continuous at* $a \in A$ *and* g *is continuous at* $f(a) \in B$ *then* $g \circ f$ *is continuous at* a*.*

Proof. Suppose $x_n \in I\backslash\{a\}$ and $x_n \to a$. Since $f(A) \subseteq B$ then $f(x_n) \in B$. By sequential characterization of limits $f(x_n) \to L$. Finally since g is continuous at L it follows that $g(f(x_n)) = (g \circ f)(x_n) \to g(L) = g\left(\lim\limits_{x\to a} f(x)\right)$. Then appealing to the sequential characterization of limits again the proof is completed. To prove the second part replace L by $f(a)$ and use continuity. $\square$

3.3 Theorems

Now we proceed to establish two properties of continuous functions, the extreme value theorem and the intermediate value theorem. The first one is of particular importance for the application to optimization below.

First the concept of boundedness of a function is formalized.

Definition 3.5. (Bounded Function) Let $E \subseteq \mathbb{R}$. A function $f : E \to \mathbb{R}$ is said to be bounded on E if and only if

$$\exists_{M\in\mathbb{R}}\forall_{x\in E}\,|f(x)| \leq M$$

The extreme value theorem establishes that a continuous function on a closed and bounded interval is always bounded. This result will be interpreted later meaning that a continuous function on a compact set attains its maximum and its minimum.

Theorem 3.2. (Extreme value) If I is a closed and bounded interval and $f : I \to \mathbb{R}$ a continuous function on I, then f is bounded on I, moreover if

$$M = \sup_{x\in I} f(x) \quad \wedge \quad m = \inf_{x\in I} f(x)$$

then there exist points $x_m, x_M \in I$ such that:

$$f(x_M) = M \quad \wedge \quad f(x_m) = m$$

Proof. Suppose for a contradiction that f is not bounded, then there exists $x_n \in I$ such that $|f(x_n)| > n$ for all $n \in \mathbb{N}$. Since I is bounded we know by the Bolzano-Weierstrass theorem that there exists a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to a$ as $k \to \infty$, since a is closed we know that $a \in I$ (since limits respect inequalities and x_{n_k} is in the interval). Then $f(a) \in \mathbb{R}$, in particular $f(a) < \infty$, but taking limits on $|f(x_{n_k})| > n_k$ we get $|f(a)| = \infty$ which is a contradiction. Then f has to be bounded.

Since f is bounded then M and m are finite. To prove that there exists $x_M \in I$ such that $f(x_M) = M$ suppose for a contradiction that $f(x) < M$ for all $x \in I$. Then the function

$$g(x) = \frac{1}{M - f(x)}$$

is continuous (by composition of functions, sum and scalar multiplication). By the first part it must also be bounded on I. Then there exists $C > 0$ such that:

$$|g(x)| = g(x) \leq C$$

which implies:

$$f(x) \leq M - \frac{1}{C}$$

for all $x \in I$. But this contradicts M being the supremum since $M - \frac{1}{C}$ is an upper bound and is lower than M.

The existence of x_m is established in a similar fashion. $\square$

The final result to be presented treats a 'smoothness' property of continuous functions, intuitively one can see that a condition to be continuous is not to have jumps. This intuition gives rise to the following two results, the first one establishes that a continuous function does not change its behavior abruptly and the second one that it does not skip any values when going to one point to another.

Theorem 3.3. (Sign preserving property) Let $f : I \to \mathbb{R}$ where I is an open and non-degenerate interval. If f is continuous at $x_0 \in I$ and $f(x_0) > 0$ then:

$$\exists_{\epsilon,\delta>0} |x - x_0| < \delta \longrightarrow f(x) > \epsilon$$

Proof. Let $\epsilon = {}^{f(x_0)}/_2 > 0$, by continuity there is $\delta > 0$ such that $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \epsilon$, then by proposition 1.1:

$$-\frac{f(x_0)}{2} < f(x) - f(x_0) < \frac{f(x_0)}{2}$$

which implies from the left inequality:

$$0 < \epsilon = \frac{f(x_0)}{2} < f(x)$$

which is the desired result. □

Theorem 3.4. (Intermediate value theorem) Let I be a non-degenerate interval and $f : I \to \mathbb{R}$ continuous on I. If $a, b \in I$ with $a < b$, and y_0 lies in between $f(a)$ and $f(b)$ then there is an $x_0 \in (a, b)$ such that $f(x_0) = y_0$.

Proof. wlog suppose that $f(a) < y_0 < f(b)$ and consider the set $E = \{x \in [a, b] \,|\, f(x) < y_0\}$, since $a \in E$ and $E \subset [a, b]$ then E is nonempty and bounded, hence it admits a supremum, call it x_0, clearly $x_0 \in [a, b]$, moreover $x_0 \neq a$ and $x_0 \neq b$.

Case 1. $x_0 \neq a$. To prove this suppose for a contradiction that $x_0 = a$, then it must be that $E = \{a\}$, so $f(x) \geq y_0$ for all $x \in (a, b]$. Then let $x_n = a + {}^1/_{K+n}$, for K large enough so that $a + {}^1/_{K+1} \leq b$. It follows that $x_n \to a$, and by continuity of f that $f(x_n) \to f(a)$. Yet for all n it holds that $f(x_n) \geq y_0 > f(a)$, then since limits respect inequalities: $\lim f(x_n) > f(a)$, a contradiction. So $x_0 \neq a$.

Case 2. $x_0 \neq b$. To prove this suppose for a contradiction that $x_0 = b$, then it by the property of approximation to supremum there is a sequence $x_n \in E$ such that $x_n \to b$, so by continuity $f(x_n) \to f(b)$, but since $x_n \in E$ it also holds that $f(x_n) < y_0 < f(b)$, since limits respect inequalities $\lim f(x_n) = f(b) \leq y_0 < f(b)$, which is a contradiction. So $x_0 \neq b$.

Now it is left to establish that $f(x_0) = y_0$. By the property of approximation to supremum there is a sequence $x_n \in E$ such that $x_n \to x_0$. Since $f(x_n) < y_0$ and f is continuous it holds that $f(x_0) \leq y_0$. Equality is shown by contradiction.

Suppose $f(x_0) < y_0$. Then $y_0 - f(x)$ is continuous and its value at $x = x_0$ is positive. By the sign preserving property we can choose ϵ and $\delta > 0$ such that $|x - x_0| < \delta$ implies $y_0 - f(x) > \epsilon$. But this contradicts x_0 being the supremum since $x' \in (x_0, x_0 + \delta)$ satisfies the hypothesis and thus $f(x) < y_0$, so $x \in E$ but $x > x_0$. Then it must be that $f(x_0) = y_0$. □

The intermediate value theorem has special importance in economics since it allows, among other things, to guarantee the existence of an equilibrium. This is done through an immediate consequence of the theorem:

Corollary 3.1. Let I be a non-degenerate interval and $f : I \to \mathbb{R}$ continuous on I. If $a, b \in I$ with $a < b$, and $f(a) f(b) < 0$ then there exists a $x_0 \in (a, b)$ such that $f(x_0) = 0$.

This corollary is not only useful to guarantee existence of equilibrium but also to guarantee solutions to root finding problems in broader settings.

3.4 Uniform continuity and equicontinuity

Sometimes it is necessary to establish stronger versions of continuity, in the definition of continuity above for each ϵ one needs to find a particular δ for which the continuity condition holds, this gives room for behavior in the function that is sometimes undesirable.

One (stronger) version of continuity is that of uniform continuity. In contrast to continuity, uniform continuity is not defined at a point but directly on a set, formally:

Definition 3.6. (Uniform Continuity) Let $E \subseteq \mathbb{R}$ be nonempty and $f : E \to \mathbb{R}$. Then f is said to be uniformly continuous on E if and only if:

$$\forall_{\epsilon>0}\exists_{\delta>0}\, |x_1 - x_2| < \delta \longrightarrow |f(x_1) - f(x_2)| < \epsilon$$

Note that the previous version of continuity fixed a point in the domain and asked that if another point was close to it, then the image was close to the image of the former point. Uniform continuity asks that if any two points are no more than δ away from one another then their images cannot be more than ϵ away.

Clearly any uniformly continuous functions is also continuous, just by fixing $x_2 = a$, but not all continuous functions are uniformly continuous, the key to find a counterexample is to note that uniform continuity demands that the images of points that are close together are also close. This translates in $\mathbb{R}$ to a statement about the slope of the function, if it is too steep then the images will be far apart. So a function that is unbounded lends itself well for the counterexample.

Example 3.1. Consider $f(x) = \ln x$ on $(0, 1]$.

f is clearly continuous. Let $\epsilon > 0$ then we want $|\ln x - \ln a| < \epsilon$ which is satisfied if $\max\{{}^{x}/_{a}, {}^{a}/_{x}\} < e^{\epsilon}$, for $x > a$ then $x < ae^{\epsilon}$ so $0 < x - a < ae^{\epsilon} - a$, for $x < a$ then $ae^{-\epsilon} < x$ so $ae^{-\epsilon} - a < x - a < 0$. Let $\delta = \min\left(a\left(1 - e^{-\epsilon}\right), a\left(e^{\epsilon} - 1\right)\right)$, then by proposition 1.1 if $|x - a| < \delta$ we have $a\left(e^{-\epsilon} - 1\right) \leq -\delta < x - a < \delta \leq a\left(e^{\epsilon} - 1\right)$ which implies from above $|\ln x - \ln a| < \epsilon$.

f is not uniformly continuous. Suppose it is, let $\epsilon > 0$ then there is δ for which the uniform continuous continuous condition holds. Then let $x \in (0, 1)$ and $x' = x + \min\{1 - x, \delta\}$. clearly $\left|x - x'\right| \leq \delta$ so $\left|\ln x - \ln x'\right| < \epsilon$ but $\left|\ln x - \ln x'\right| = \ln \frac{x'}{x} = \ln\left(1 + \frac{\min\{1-x,\delta\}}{x}\right)$. Then there is an $x \in (0, 1)$ such that $\left|\ln x - \ln x'\right| > \epsilon$:

$$\begin{aligned}
\ln\left(1 + \frac{\min\{1 - x, \delta\}}{x}\right) &= \epsilon + 1 \\
\frac{\min\{1 - x, \delta\}}{x} &= e^{\epsilon+1} - 1
\end{aligned}$$

Then

$$x = \frac{\delta}{e^{\epsilon+1} - 1} \qquad \vee \qquad x = \frac{1}{e^{\epsilon+1}}$$

Both this values are in $(0, 1)$.

Another (better) example is:

Example 3.2. Consider $f(x) = x^2$ on $\mathbb{R}$. Suppose f is uniformly continuous and let δ be such that:

$$|x_1 - x_2| < \delta \longrightarrow |f(x_1) - f(x_2)| < 1$$

By the Archimedean principle there is n such that $\delta n > 1$, let $x_2 = n$ and $x_1 = n + \frac{\delta}{2}$, so they satisfy the hypothesis. Then:

$$1 > \left| f\left(n + \frac{\delta}{2}\right) - f(n) \right| = n^2 + n\delta + \frac{\delta^2}{4} - n^2 = n\delta + \frac{\delta^2}{4} > n\delta > 1$$

which is a contradiction.

These examples relied on the fact that the functions were unbounded on the specified domain. If a function is continuous and is defined on a closed and bounded interval then it is also bounded. It would seem then that such functions will also be uniformly continuous on those intervals. The following result formalizes this.

Proposition 3.9. *Suppose that I is a closed and bounded interval. If $f : I \to \mathbb{R}$ is continuous on I, then f is uniformly continuous on I.*

Proof. Suppose for a contradiction that f is not uniformly continuous, so:

$$\exists_{\epsilon>0}\forall_{\delta>0}\exists_{x_1,x_2} |x_1 - x_2| < \delta \quad \wedge \quad |f(x_1) - f(x_2)| > \epsilon$$

In particular for $\delta = {}^1\!/_n$ there are points x_n and y_n such that $|x_n - y_n| < {}^1\!/_n$ and $|f(x_n) - f(y_n)| > \epsilon$. Then we have sequences $\{x_n\}$ and $\{y_n\}$ on I, by the Bolzano-Weierstrass theorem $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$ and $\{y_{n_k}\}$ has a convergent subsequence $\left\{y_{n_{k_j}}\right\}$. Since $x_{n_k} \to x$ we also have $x_{n_{k_j}} \to x$, so the two subsequences converge, then by the squeeze theorem we get $|x_n - y_n| \to 0$ and since the absolute value is a continuous function we get $x_n - y_n \to 0$ which implies that $x = y$ and $f(x) = f(y)$, but $|f(x_n) - f(y_n)| > \epsilon$ for all n, so we have a contradiction. $\square$

Finally we note that there is a strong link between uniform continuity and Cauchy sequences:

Proposition 3.10. *Let $E \subseteq \mathbb{R}$ and $f : E \to \mathbb{R}$ be uniformly continuous on E. If $x_n \in E$ is Cauchy then so is $f(x_n)$.*

Proof. Let $\epsilon > 0$ then by uniform continuity there is $\delta > 0$ such that

$$\left| x - x^{'} \right| < \delta \longrightarrow |f(x) - f(x)| < \epsilon$$

Since $\{x_n\}$ is Cauchy there is N such that for $n, m \geq N$ we have $|x_n - x_m| < \delta$, then we get: $|f(x_n) - f(x_m)| < \epsilon$ for all $n, m \geq N$. This proves that $\{f(x_n)\}$ is Cauchy. $\square$

Equicontinuity is a different concept since it does not treat a function but instead a family of functions. The idea is that for a given ϵ there is a single δ for which all members of the family satisfy the uniformly continuous condition.

Definition 3.7. A family of functions $\mathcal{F}$ defined on a set E is equicontinuous if and only if:

$$\forall_{\epsilon>0}\exists_{\delta>0}\forall_{f\in\mathcal{F}} |x_1 - x_2| < \delta \longrightarrow |f(x_1) - f(x_2)| < \epsilon$$

4 Sequences of Functions

Consider now a space of function. That is, a space in which its elements are not numbers but functions. One can take then a sequence of these functions and ask if the sequence is converging to another function. To do this it is first necessary to establish what it means for two functions to be close. Here we treat two ways of establishing this convergence and (briefly) discuss their implications for how the properties of the functions in the sequence are passed (or not) to the limiting function. The first is pointwise convergence and the second one is uniform convergence, both are important in economics (and in statistics).

Definition 4.1. (Pointwise Convergence) Let $E \subseteq \mathbb{R}$ be nonempty. A sequence of functions $f_n : E \to \mathbb{R}$ is said to converge pointwise on E if and only if the limit $\lim f_n(x) = f(x)$ exists for all $x \in E$.

Note that in the definition $f(x)$ is just denoting the limit of $f_n(x)$ for a fixed x, yet this definition allows to say when a sequence of functions $\{f_n\}$ converges to a function f:

Definition 4.2. (Pointwise Convergence) Let $E \subseteq \mathbb{R}$ be nonempty. A sequence of functions $f_n : E \to \mathbb{R}$ is said to converge pointwise to a function f if and only if:

$$\forall_{x \in E} \forall_{\epsilon > 0} \exists_N \forall_{n \geq N} \left| f_n(x) - f(x) \right| < \epsilon$$

In pointwise convergence the number N for which a sequence is close (ϵ away) to the limit is allowed to vary with the point at which the function is evaluated (x), in uniform convergence this condition is strengthened so that there is a number N for which the functions in the sequence are close to the limit for all point in the domain. Formally:

Definition 4.3. (Uniform Convergence) Let $E \subseteq \mathbb{R}$ be nonempty. A sequence of functions $f_n : E \to \mathbb{R}$ is said to converge uniformly to f if and only if:

$$\forall_{\epsilon > 0} \exists_N \forall_{n \geq N} \forall_{x \in E} \left| f_n(x) - f(x) \right| < \epsilon$$

A further characterization of uniform convergence is given by its relation to Cauchy sequences. Intuitively since the functions f_n are going close to f for all x they must be also close to each other. This is made formal below:

Theorem 4.1. (Uniform Cauchy criterion) Let $E \subseteq \mathbb{R}$ be nonempty and $f_n : E \to \mathbb{R}$ a sequence of functions. f_n converges uniformly on E if and only if:

$$\forall_{\epsilon} \exists_N \forall_{n,m \geq N} \forall_x \left| f_n(x) - f_m(x) \right| < \epsilon$$

Proof. Suppose first that $f_n \to f$ uniformly and let $\epsilon > 0$, then there exits N such that for $n, m \geq N$:

$$\left| f_n(x) - f(x) \right| < \frac{\epsilon}{2} \quad \wedge \quad \left| f_m(x) - f(x) \right| < \frac{\epsilon}{2}$$

for all $x \in E$. Then by the triangle inequality:

$$\left| f_n(x) - f_m(x) \right| \leq \left| f_n(x) - f(x) \right| + \left| f(x) - f_m(x) \right| < \epsilon$$

establishing that $\{f_n\}$ is Cauchy on E.

Now suppose that the condition above holds, then for each $x \in E$ the sequence of numbers $\{f_n(x)\}$ is Cauchy and by completeness of the real numbers it converges to a number $f(x)$. Now consider:

$$|f_n(x) - f_m(x)| < \epsilon$$

For a given x this holds, also, since the absolute value and the sum are continuous functions and limits respect inequalities we have:

$$\begin{aligned} \lim_{m\to\infty} |f_n(x) - f_m(x)| &< \epsilon \\ \left| f_n(x) - \lim_{m\to\infty} f_m(x) \right| &< \epsilon \\ |f_n(x) - f(x)| &< \epsilon \end{aligned}$$

Since this holds for all $n \geq N$ and all $x \in E$ we have established uniform convergence of f_n to f. □

Clearly if a sequence of functions converges uniformly it also converges pointwise, and the converse is not true, the reason for strengthening the definition of convergence is that most properties of the functions of the sequence are not transferred to the limit under pointwise convergence but they are under uniform convergence. For example the pointwise limit of continuous or differentiable functions is not necessarily continuous or differentiable.

Example 4.1. Let $f_n(x) = x^n$ defined on $E = [0,1]$, then for all $x \in [0,1)$ we have $f_n(x) \to 0$ and $f_n(1) \to 1$, so the limits exist hence f_n is convergent the limiting function is then:

$$f(x) = \begin{cases} 0 & \text{if } x \in [0,1) \\ 1 & \text{if } x = 1 \end{cases}$$

While $f_n(x) = x^n$ is clearly continuously differentiable on E, $f(x)$ is discontinuous at 1 and hence also fails to be differentiable there.

Moreover, even when the limiting distribution is differentiable or integrable the limit of the derivate or the integral need not coincide with the derivate or integral of the limiting function.

Example 4.2. Let $f_n(x) = {}^{x^n}/_n$ defined on $[0,1]$, then $f_n(x) \to 0$ pointwise on E so $f(x) = 0$. The derivative of f_n is $f_n'(x) = x^{n-1}$ and thus $f_n'(1) \to 1$ but $f'(x) = 0$ for all $x \in E$, so:

$$\lim_{n\to\infty} f_n'(x) \neq \left(\lim_{n\to\infty} f_n(x)\right)$$

Uniform convergence allows to get the results we wanted. Continuity is preserved as proven in the following proposition.

Proposition 4.1. *Let $E \subseteq \mathbb{R}$ be nonempty and $f_n \to f$ uniformly on E. If f_n is continuous at some $x_0 \in E$ then f is also continuous at x_0.*

Proof. Let $\epsilon > 0$ and choose N such that if $n \geq N$ then:

$$|f_n(x) - f(x)| < \frac{\epsilon}{3}$$

for all $x \in E$. Since f_N is continuous choose $\delta > 0$ such that:

$$|x - x_0| < \delta \longrightarrow |f_N(x) - f_N(x_0)| < \frac{\epsilon}{3}$$

Now by the triangle inequality, if $|x - x_0| < \delta$ we have:

$$\begin{aligned} |f(x) - f(x_0)| &= |(f(x) - f_N(x)) + (f_N(x) - f_N(x_0)) + (f_N(x_0) - f(x_0))| \\ &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)| \\ &< \epsilon \end{aligned}$$

where the first and last elements are bounded by uniform convergence and the middle one by continuity. ∎

Corollary 4.1. *Let $E \subseteq \mathbb{R}$ be nonempty and $f_n \to f$ uniformly on E. If f_n is continuous on E then f is also continuous on E.*

Uniform continuity also preserves boundedness of functions. In particular under uniform convergence the boundedness of the individual functions of the sequence can be extended to uniform boundedness.

Definition 4.4. (Uniformly Bounded Functions) A sequence of functions $\{f_n\}$ is said to be uniformly bounded if and only if there exists $M \in \mathbb{R}$ such that $|f_n(x)| < M$ for all n and all x.

Proposition 4.2. *Let $\{f_n\}$ be a sequence of bounded functions, if $f_n \to f$ uniformly then $\{f_n\}$ is also uniformly bounded and f is bounded.*

Proof. First prove that f is bounded. Since $f_n \to f$ uniformly it also converges pointwise, so there exits N such that:

$$|f_N(x) - f(x)| < 1$$

Using the triangle inequality:

$$|f(x)| < 1 + |f_N(x)| \leq 1 + M_N$$

where the second inequality follows from f_n being bounded. Then $|f(x)|$ is bounded by $1 + M_N$.

To prove that $\{f_n\}$ is uniformly bounded note that by uniform convergence there is an N such that:

$$|f_n(x) - f(x)| < 1$$

for all $n \geq N$, and by the triangle inequality:

$$|f_n(x)| - |f(x)| < 1$$

which implies

$$|f_n(x)| < 1 + M_f$$

where M_f is the bound of f, which is exists by the first part of the proof. Because of uniform convergence this bound works for all x, and all $n \geq N$. Letting M_n be the bound of function f_n we have that

$$M = \max\{M_1, M_2, \ldots, M_{N-1}, 1 + M_f\}$$

is a bound for all f_n. This establishes the result. $\square$

Uniform convergence also allows to interchange integrals and derivatives:

Proposition 4.3. *Suppose that $f_n \to f$ uniformly on a closed interval $[a, b]$. If f_n is integrable on $[a, b]$ then f is also integrable and:*

$$\lim \int_a^b f_n(x)\, dx = \int_a^b \lim f_n(x)\, dx$$

Proposition 4.4. *Let (a, b) be a bounded interval and suppose that f_n converges at some x_0 (as in pointwise convergence for a given point). If f_n is differentiable for all n and $f_n^{'}$ converges uniformly on (a, b) then there exists f such that $f_n \to f$ uniformly and:*

$$\lim f_n^{'}(x) = (\lim f_n(x))^{'}$$

Proof. Let $c \in (a, b)$ and define:

$$g_n(x) = \begin{cases} \frac{f_n(x) - f_n(c)}{x - c} & \text{if } x \neq c \\ f_n^{'}(c) & \text{if } x = c \end{cases}$$

clearly

$$f_n(x) = f_n(c) + (x - c)\, g_n(x)$$

- We first prove that g_n converges uniformly on (a, b). Let $\epsilon > 0$ and $n, m \in \mathbb{N}$ and $x \neq c$. By the mean value theorem there is number ξ between x and c such that:

  $$g_n(x) - g_m(x) = \frac{(f_n(x) - f_m(x)) - (f_n(c) - f_m(c))}{x - c} = f_n^{'}(\xi) - f_m^{'}(\xi)$$

  Since $\left\{f_n^{'}\right\}$ converges uniformly, by the uniform Cauchy criterion there is an N such that for $n, m \geq N$:

  $$|g_n(x) - g_m(x)| = \left|f_n^{'}(\xi) - f_m^{'}(\xi)\right| < \epsilon$$

  this holds for all $x \neq c$, but also for $x = c$ since then $g_n(c) = f^{'}(c)$. Note that this is the definition of $\{g_n\}$ being uniformly convergent.

- Now we prove that f_n converges uniformly. Recall that

  $$f_n(x) = f_n(c) + (x - c)\, g_n(x)$$

Let $c = x_0$ which gives:

$$\begin{aligned} |f_n(x) - f_m(x)| &= |(f_n(x_0) - f_m(x_0)) + (x - x_0)(g_n(x) - g_m(x))| \\ &\leq |f_n(x_0) - f_m(x_0)| + |x - x_0| |g_n(x) - g_m(x)| \\ &\leq |f_n(x_0) - f_m(x_0)| + |b - a| |g_n(x) - g_m(x)| \end{aligned}$$

Since $f_n(x_0)$ converges by hypothesis, we know by completeness of the real numbers that for $\epsilon > 0$ there is a number N_1 such that for $n, m \geq N_1$:

$$|f_n(x_0) - f_m(x_0)| < \frac{\epsilon}{2}$$

and by uniform convergence of g_n there is a number N_2 such that for $n, m \geq N_2$:

$$|g_n(x) - g_m(x)| < \frac{\epsilon}{2(b-a)}$$

Joining, for $n, m \geq \max\{N_1, N_2\}$ we have:

$$|f_n(x) - f_m(x)| < \epsilon$$

establishing uniform convergence of f_n by the uniform Cauchy criterion.

- Now fix $c \in (a, b)$ and let f and g be the limits of $\{f_n\}$ and $\{g_n\}$. We want to show that $f'(c) = \lim f_n'(c)$, since c is arbitrary this gives the desired result.

  Note that by definition g_n is continuous at c then g is continuous at c by proposition 4.1. Since $g_n(c) = f_n'(c)$ we can write:

  $$\lim_{n\to\infty} f_n'(c) = \lim_{n\to\infty} g_n(c) = g(c) = \lim_{x\to c} g(x)$$

  But we also have:

  $$\frac{f(x) - f(c)}{x - c} = \lim_{n\to\infty} \frac{f_n(x) - f_n(c)}{x - c} = \lim_{n\to\infty} g_n(x) = g(x)$$

  Then:

  $$(\lim f_n(c))' = f'(c) = \lim_{x\to c} \frac{f(x) - f(c)}{x - c} = \lim_{x\to c} g(x)$$

  This completes the proof since we can join equalities to get:

  $$\lim_{n\to\infty} f_n'(c) = \lim_{x\to c} g(x) = (\lim f_n(c))'$$

$\square$

5 Topology of $\mathbb{R}^n$

5.1 Definitions

The final topic in real analysis is to study the basic concepts that cover the topology of $\mathbb{R}^n$, with few exceptions the results obtained before will extend trivially and wont be presented. The focus is on the the definition and properties of open/closed sets and their properties and relation with continuity. To start with the formal treatment a notion of a norm in $\mathbb{R}^n$ will be given.

Definition 5.1. (Norm) A norm is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ such that:

i. For all $x \in \mathbb{R}^n$ it holds that $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$.

ii. For all $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ it holds that $\|\alpha x\| = |a| \|x\|$.

iii. The triangle inequality holds in its two forms:

$$\|x + y\| \leq \|x\| + \|y\| \qquad \wedge \qquad \|x\| - \|y\| \leq \|x - y\|$$

Note that the absolute value is a proper norm for $\mathbb{R}$ and that there are many possible norms in $\mathbb{R}^n$ that are extensions of the absolute value:

- Euclidean norm: $\|x\| = (x \cdot x)^{1/2} = \sqrt{x_1^2 + \cdots + x_n^2}$
- L^1 norm: $\|x\|_1 = \sum |x_k|$
- Sup norm: $\|x\|_\infty = \max \{|x_1|, \ldots, |x_n|\}$

In what follows I will always use the norm as the Euclidean norm but most results are valid for any arbitrary norm that satisfies the three conditions above. One useful property is the Cauchy-Schwartz inequality:

Proposition 5.1. ***(Cauchy-Schwartz inequality)*** *Let $x, y \in \mathbb{R}^n$ then $|x \cdot y| \leq \|x\| \|y\|$ under the Euclidean norm.*

Proof. First note that for any $t \in \mathbb{R}$ we have:

$$0 \leq \|x - ty\|^2 = (x - ty) \cdot (x - ty) = (x \cdot x) - 2t (x \cdot y) + t^2 (y \cdot y) = \|x\|^2 - 2t (x \cdot y) + t^2 \|y\|^2$$

Then set $t = \frac{x \cdot y}{\|y\|^2}$, note that this can only be done if $y \neq 0$. Then we have:

$$\begin{aligned} 0 &\leq \|x\|^2 - \frac{(x \cdot y)^2}{\|y\|^2} \\ (x \cdot y)^2 &\leq \|x\|^2 \|y\|^2 \\ (x \cdot y) &\leq \|x\| \|y\| \end{aligned}$$

When $y = 0$ the inequality holds as a equality trivially. $\square$

Now to start with the topological properties of $\mathbb{R}^n$ its necessary to introduce the concept of the ball, or neighborhood, which will allow us to talk about what is happening in the surroundings of a point in the space.

Definition 5.2. (Open Ball) Let $a \in \mathbb{R}^n$ and $\epsilon > 0$. The open ball centered at a with radius ϵ is the set:

$$B_\epsilon(a) = \{x \in \mathbb{R}^n | \, \|x - a\| \leq \epsilon\}$$

Note that this definition resembles a lot the one used for convergence of sequences and continuity. One can define these concepts for $\mathbb{R}^n$ as follows:

Definition 5.3. (Convergence) A sequence $\{x_k\}$ in $\mathbb{R}^n$ is said to converge to a number $a \in \mathbb{R}^n$ $(x_k \to a)$ if and only if:

$$\forall_{\epsilon>0} \exists_N \forall_{k \geq N} x_k \in B_\epsilon(a)$$

It is said to be bounded if and only if:

$$\exists_{M \in \mathbb{R}} \forall_k \|x_k\| < M$$

And it is said to be Cauchy if and only if:

$$\forall_\epsilon \exists_N \forall_{k,m \geq N} \|x_k - x_m\| < \epsilon$$

Definition 5.4. (Limit of Function) Let $a \in \mathbb{R}^n$ and I an open set that contains a. Let f be a real function defined on I (except possible at a). $f(x)$ is said to converge to y as $x \to a$ if and only if:

$$\forall_{\epsilon>0} \exists_{\delta>0} x \in B_\delta(a) \longrightarrow f(x) \in B_\epsilon(y)$$

We write $y = \lim_{x \to a} f(x)$. This can also be expressed as follows:

$$\forall_\epsilon \exists_\delta f(B_\delta(a)) \subseteq B_\epsilon(y)$$

Or equivalently:

$$\forall_\epsilon \exists_\delta B_\delta(a) \subseteq f^{-1}(B_\epsilon(y))$$

This trivial extension of the definitions is behind the extension of all the previous theorems and propositions to the multidimensional case. The proof of these equivalence is left to be proven below.

Now we introduce the definition of an open and a closed set:

Definition 5.5. (Open Set) A set $V \subseteq \mathbb{R}^n$ is said to be open if and only if:

$$\forall_{x \in V} \exists_{\epsilon>0} B_\epsilon(x) \subseteq V$$

Definition 5.6. (Closed Set) A set $E \subseteq \mathbb{R}^n$ is said to be closed if and only if $E^c = \mathbb{R}^n \backslash E$ is open.

Openness and closedness are preserved by standard set operations. The following properties are stated without proof:

Proposition 5.2. *Let $n \in \mathbb{N}$.*

i. If $\{V_\alpha\}_{\alpha\in A}$ is any collection of open subsets of $\mathbb{R}^n$ then $\bigcup_{\alpha\in A} V_\alpha$ is also open.

ii. If $\{V_1, \dots, V_p\}$ is a finite collection of open subsets of $\mathbb{R}^n$ then $\bigcap_{k=1}^{p} V_k$ is also open.

iii. If $\{E_\alpha\}_{\alpha\in A}$ is any collection of closed subsets of $\mathbb{R}^n$ then $\bigcap_{\alpha\in A} E_\alpha$ is also closed.

iv. If $\{E_1, \dots, E_p\}$ is a finite collection of closed subsets of $\mathbb{R}^n$ then $\bigcup_{k=1}^{p} E_k$ is also closed.

v. If V is open and E is closed then $V\backslash E$ is open and $E\backslash V$ is closed.

Sets can also be called open or closed relative to a given subset of $\mathbb{R}^n$, this is of particular importance when dealing with functions that are not defined over the whole space. Then we are only interested in the topological properties of its effective domain. The relevant definitions are:

Definition 5.7. (Relatively Open/Closed Set) Let $E \subseteq \mathbb{R}^n$. A set U is said to be open/closed relative to E if and only if there is an open/closed set A such that $U = E \cap A$.

The definition of relative openness allows us to treat a more interesting concept, that of connectivity. This concepts characterizes when one can go between any two points of a set without leaving it.

Definition 5.8. (Connected Set) Let $E \subseteq \mathbb{R}^n$.

i. A pair of sets U, V is said to separate E if and only if $U, V \neq \emptyset$, they are relatively open in E, $E = U \cup V$ and finally $U \cap V = \emptyset$.

ii. A set E is said to be connected if and only if E cannot be separated by any pair of relatively open sets U and V.

An easier condition to check to prove a set is not connected is if there are open sets A and B such that $E \cap A \neq \emptyset$, $E \cap B \neq \emptyset$, $E \subseteq A \cup B$ and $A \cap B = \emptyset$. Then set $U = E \cap A$ and $V = E \cap B$.

Now we can turn to the definitions of interior and closure of a set which are, respectively the largest open set contained in a set and the smallest closed set that contains it:

Definition 5.9. (Interior of a Set) Let $E \subseteq \mathbb{R}^n$. The interior of E is:

$$E^{\circ} = \bigcup \{V | V \subseteq E \quad \wedge V \text{ is open in } \mathbb{R}^n\}$$

The closure is defined similarly as:

$$\overline{E} = \bigcap \{V | E \subseteq V \quad \wedge V \text{ is closed in } \mathbb{R}^n\}$$

Clearly the following properties follow:

Remark 5.1. Let $E \subseteq \mathbb{R}^n$ then:

i. $E^{\circ} \subseteq E \subseteq \overline{E}$.

ii. $E^{\circ} = E$ if and only if E is open.

iii. $\overline{E} = E$ if and only if E is closed.

A related concept is that of the boundary, as it name indicates its the part of the space that is contact with both the set and its complement.

Definition 5.10. (Boundary of a Set) Let $E \subseteq \mathbb{R}^n$. The boundary of E is defined as:

$$\partial E = \{x \in \mathbb{R}^n | \forall_{\epsilon} \quad B_{\epsilon}(x) \cap E \neq \emptyset \quad \wedge \quad B_{\epsilon}(x) \cap E^c \neq \emptyset\}$$

The following result follows:

Proposition 5.3. *Let $E \subseteq \mathbb{R}^n$ then $\partial E = \overline{E} \backslash E^{\circ}$.*

Proof. The proof has two parts:

i. $x \in \overline{E}$ if and only if $B_{\epsilon}(x) \cap E \neq \emptyset$ for all $\epsilon > 0$

   (a) Suppose for a contradiction that $x \in \overline{E}$ and that there is an $\overline{\epsilon}$ such that $B_{\overline{\epsilon}}(x) \cap E = \emptyset$. Then $(B_{\overline{\epsilon}}(x))^c$ is closed and moreover $E \subseteq (B_{\overline{\epsilon}}(x))^c$. Then by definition of closure $\overline{E} \subseteq (B_{\overline{\epsilon}}(x))^c$, so $\overline{E} \cap B_{\overline{\epsilon}}(x) = \emptyset$ which implies $x \notin \overline{E}$, this is a contradiction.

   (b) We want to show $\forall_{\epsilon>0} B_{\epsilon}(x) \cap E \neq \emptyset \longrightarrow x \in \overline{E}$ which is equivalent to show that $x \notin \overline{E} \longrightarrow \exists_{\overline{\epsilon}>0} B_{\epsilon}(x) \cap E = \emptyset$.
   Suppose that $x \notin \overline{E}$, then $x \in \overline{E}^c$ which is open, so there exists a $\overline{\epsilon} > 0$ such that $B_{\overline{\epsilon}}(x) \subseteq \overline{E}^c$. Note that $B_{\overline{\epsilon}}(x) \cap E \subseteq B_{\overline{\epsilon}}(x) \cap \overline{E}$ since $E \subseteq \overline{E}$, and finally that $B_{\overline{\epsilon}}(x) \cap \overline{E} = \emptyset$ since $B_{\overline{\epsilon}}(x) \subseteq \overline{E}^c$.
   Then we have found a $\epsilon > 0$ such that $B_{\epsilon}(x) \cap E = \emptyset$ for any arbitrary point $x \notin \overline{E}$.

ii. $x \notin E^{\circ}$ if and only if $\forall_{\epsilon>0} B_{\epsilon}(x) \cap E^c \neq \emptyset$.

   (a) We want to show $x \notin E^{\circ} \longrightarrow \forall_{\epsilon>0} B_{\epsilon}(x) \cap E^c \neq \emptyset$ which is equivalent to $\exists_{\epsilon>0} B_{\epsilon}(x) \cap E^c = \emptyset \longrightarrow x \in E^{\circ}$.
   Let $\overline{\epsilon} > 0$ be such that $B_{\overline{\epsilon}}(x) \cap E^c = \emptyset$, then it must be that $B_{\overline{\epsilon}}(x) \subseteq E$, but $B_{\overline{\epsilon}}(x)$ is open, hence $B_{\overline{\epsilon}}(x) \subseteq E^{\circ}$, since $x \in B_{\overline{\epsilon}}(x)$ we get the result.

   (b) We want to show $\forall_{\epsilon>0} B_{\epsilon}(x) \cap E^c \neq \emptyset \longrightarrow x \notin E^{\circ}$
   Suppose for a contradiction that $\forall_{\epsilon>0} B_{\epsilon}(x) \cap E^c \neq \emptyset$ but $x \in E^{\circ}$. Since $x \in E^{\circ}$ there exists an open set $V \subseteq E$ such that $x \in V$. Since V is open there exists $\overline{\epsilon} > 0$ such that $B_{\overline{\epsilon}}(x) \subseteq V \subseteq E$, so $B_{\overline{\epsilon}}(x) \cap E^c = \emptyset$, which is a contradiction.

□

5.2 Sequences, closedness and compactness

There are many topological properties that are related to convergence of sequences, in particular closedness and compactness of a set will be determined by the behavior of sequences that lie in the set. In order to establish these relations its best to characterize the link between convergence in $\mathbb{R}^n$ and convergence in $\mathbb{R}$, the following proposition establishes the main tool:

Proposition 5.4. *Let $a \in \mathbb{R}^n$ and $\{x_k\}$ a sequence with $x_k \in \mathbb{R}^n$ for all k. The j^{th} element of an element of the vectors a and x_k are denoted $a(j)$ and $x_k(j)$ respectively.*

$$(x_k \to a) \iff \forall_j \left(x_k(j) \to a(j)\right)$$

Proof. Let $j \in \{1, \ldots, n\}$, then:

$$0 \leq |x_k(j) - a(j)| \leq \|x_k - a\| \leq \sqrt{n} \max_{l \in \{1,\ldots,n\}} \{|x_k(l) - a(l)|\}$$

By the squeeze theorem $x_k(j) \to a(j)$ for all j if and only if the real sequence $\|x_k - a\| \to 0$. But this happens in and only if $x_k \to a$. This completes the proof. □

Using this proposition its possible to extend almost all the results established on $\mathbb{R}$ to higher dimensional Euclidean spaces. I will then skip these theorems and results and concentrate in the ones relating convergence of sequences and topological properties of sets in $\mathbb{R}^n$. For this lets first define what a limit point is:

Definition 5.11. (Limit Point) Let $E \subseteq \mathbb{R}^n$, $x \in \mathbb{R}^n$ is a limit point of E if and only if for all $\epsilon > 0$ there exists a point $y \in E$ such that $y \neq x$ and $\|y - x\| < \epsilon$ (equivalently $y \in B_\epsilon(x) \setminus \{x\}$).

Remark 5.2. From the definition of a limit point is clear that one can also characterize them as points $x \in \mathbb{R}^n$ such that there exists $x_k \in E \setminus \{x\}$ and $x_k \to x$.

Proposition 5.5. *Let $E \subseteq \mathbb{R}^n$. E is closed if and only if E contains all its limits points.*

Proof. If E is empty the result is vacuously true. Wlog we assume $E \neq \emptyset$.

i. Suppose for a contradiction that E is closed, x is a limit point, but $x \notin E$. Then, since E is closed E^c is open and $x \in E^c$, so there exists $\epsilon > 0$ such that $B_\epsilon(x) \subseteq E^c$. Since x is a limit point there exists $y \in E$ such that $y \in B_\epsilon(x)$ and $y \neq x$. But this is a contradiction since $y \in E$ and $y \in B_\epsilon(x) \subseteq E^c$. Then it has to be that if E is closed and x is a limit point, then $x \in E$.

ii. Suppose for a contradiction that E contains all of its limits points and that it is not closed. Then E^c is not open and there exists a point $x \in E^c$ such that for all ϵ the ϵ-Ball centered in x is not contained in E^c, i.e. $\forall_\epsilon B_\epsilon(x) \cap E \neq \emptyset$, so for all ϵ there is a $y \in E$ such that $y \in B_\epsilon(x) \cap E$ and since $x \notin E$, $y \in B_\epsilon(x) \setminus x$. Thus x is a limit point of E, which implies by assumption that $x \in E$, a contradiction. Then it has to be that if E contains all of its limits points the it is also closed.

(a) Alternatively one can construct a convergent sequence in E with limit x: for $k = 1, 2, \ldots$ we can define $x_k \in B_{1/k}(x) \cap E$, note that $x_k \neq x$ and $x_k \in E$ for all k. By construction $x_k \to x$, so x is a limit point.

$\square$

This result is of great usefulness when dealing with closed sets, since they contain the limit of all of their convergent sequences. A stronger result is obtained below for compact sets, all of their sequences have a convergent subsequence whose limit is in the set. To get to this result we first define formally what a compact set is:

Definition 5.12. (Compact Set) Let $E \subseteq \mathbb{R}^n$.

i. An open covering of E is a collection $\{V_\alpha\}_{\alpha \in A}$ such that each V_α is open and $E \subseteq \bigcup\limits_{\alpha \in A} V_\alpha$.

ii. The set E is said to be compact if and only if every open covering of E has a finite sub-cover.

What is behind the result for sequences is the Heine-Borel theorem, that allows for a sharp characterization of compact sets in Euclidean spaces I state without proof:

Theorem 5.1. (Heine-Borel) Let $E \subseteq \mathbb{R}^n$. E is compact if and only if E is closed and bounded.

Corollary 5.1. *If $E \subseteq \mathbb{R}^n$ is compact and $x_k \in E$ then there exists a subsequence $\{x_{k_l}\}$ and $x \in E$ such that $x_{k_l} \to x$ as $l \to \infty$.*

Proof. By the Heine-Borel theorem since E is compact it is also closed and bounded, then $\{x_k\}$ is a bounded sequence. By the Bolzano-Weierstrass theorem x_k has a convergent subsequence, and by proposition 5.5 the limit belongs to E. $\square$

Some other facts of compact sets are listed below.

Proposition 5.6. *If $E \subseteq \mathbb{R}^n$ is a compact set then it is closed.*

Proposition 5.7. *If E is a compact set and $H \subseteq E$ is closed then H is compact.*

Proof. When $E \subseteq \mathbb{R}^n$ the result follows directly from the Heine-Borel theorem, since E is compact its bounded, so H is bounded an by hypothesis closed, hence compact.

For a general compact set E note the following: Let $\mathcal{V} = \{V_\alpha\}_{\alpha \in A}$ be an open cover of H. Note that H^c is open (since H is closed), then $\mathcal{V} \cup H^c$ is an open cover of E (i.e. $E \subseteq H^c \cup \left(\bigcup\limits_{\alpha \in A} V_\alpha\right)$). Since E is compact then there exists a finite set A_0 such that

$$E \subseteq H^c \cup \left(\bigcup_{\alpha \in A_0} V_\alpha\right)$$

But since H cannot be covered by H^c and $H \subseteq E$ it follows that $H \subseteq \bigcup\limits_{\alpha \in A_0} V_\alpha$. So $\{V_\alpha\}_{\alpha \in A_0}$ is a finite subcover for H. H is compact. $\square$

5.3 Continuity and topology

As mentioned above open sets allow for new characterizations of continuous functions.

Proposition 5.8. *Let $n, m \in \mathbb{N}$ and $f : \mathbb{R}^n \to \mathbb{R}^m$. The following three conditions are equivalent.*

i. f is continuous on $\mathbb{R}^n$.

ii. If $V \subseteq \mathbb{R}^m$ is open then $f^{-1}(V)$ is open in $\mathbb{R}^n$.

iii. If $E \subseteq \mathbb{R}^m$ is closed then $f^{-1}(E)$ is closed in $\mathbb{R}^n$.

Proof. The proof is carried out by relating pairs of conditions.

i. f is continuous on $\mathbb{R}^n$ then preimages of open sets are open.

Let f be continuous and $V \subseteq \mathbb{R}^m$ open. If $f^{-1}(V) = \emptyset$ then the implication is verified. Suppose $f^{-1}(V) \neq \emptyset$ then there exists $a \in f^{-1}(V)$. Since V is open there exists $\epsilon > 0$ such that $B_\epsilon(f(a)) \subseteq V$. Since f is continuous there exists $\delta > 0$ such that $f(B_\delta(a)) \subseteq B_\epsilon(f(a))$, then $f(B_\delta(a)) \subseteq V$ and $B_\delta(a) \subseteq f^{-1}(V)$, proving openness.

ii. If $V \subseteq \mathbb{R}^m$ is open then $f^{-1}(V)$ is open in $\mathbb{R}^n$ implies that If $E \subseteq \mathbb{R}^m$ is closed then $f^{-1}(E)$ is closed in $\mathbb{R}^n$.

Let E be closed, then E^c is open, then $f^{-1}(E^c)$ is open, then note that $(f^{-1}(E))^c = f^{-1}(E^c)$, so $f^{-1}(E)$ is closed.

iii. If $E \subseteq \mathbb{R}^m$ is closed then $f^{-1}(E)$ is closed in $\mathbb{R}^n$ then f is continuous on $\mathbb{R}^n$.

Let $a \in \mathbb{R}^n$ and $\epsilon > 0$. Let $E = (B_\epsilon(f(a)))^c$, E is a closed set, then $f^{-1}(E)$ is closed by hypothesis. As before: $f^{-1}((B_\epsilon(f(a)))^c) = (f^{-1}(B_\epsilon(f(a))))^c$, so $f^{-1}(B_\epsilon(f(a)))$ is open, since $a \in f^{-1}(B_\epsilon(f(a)))$ by construction there is a $\delta > 0$ such that $B_\delta(a) \subseteq f^{-1}(B_\epsilon(f(a)))$ which is equivalent to $f(B_\delta(a)) \subseteq B_\epsilon(f(a))$ proving continuity.

□

From the proposition above we have that the inverse image of continuous functions preserves openness and closedness, but the following examples show that it does not preserve boundedness and connectedness.

Example 5.1. Let $f(x) = \frac{1}{x^2+1}$ and $E = (0, 1]$. Clearly f is continuous and E is bounded, but $f^{-1}(E) = (-\infty, \infty)$ is not.

Example 5.2. Let $f(x) = x^2$ and $V = (1, 4)$. Clearly f is continuous and E is connected, but $f^{-1}(E) = (-2, -1) \cup (1, 2)$ is not.

Also it is important to note that the image of continuous functions is not as well behaved as the preimage. The following examples make this clear:

Example 5.3. Let $f(x) = x^2$ and $V = (-1, 1)$. Clearly f is continuous and V is open, but $f(V) = [0, 1)$ is neither open or closed.

Example 5.4. Let $f(x) = 1/x$ and $E = [1, \infty)$. Clearly f is continuous and E is closed, but $f(E) = (0, 1]$ is neither open or closed.

Despite these problems the image of a continuous functions does preserve compactness as proven below.

Proposition 5.9. *Let $n, m \in \mathbb{N}$. If $H \subseteq \mathbb{R}^n$ is compact and $f : H \to \mathbb{R}^m$ continuous, then $f(H)$ is compact in $\mathbb{R}^m$.*

Proof. By the Heine-Borel theorem it suffices to prove that $f(H)$ is closed and bounded.

Let y be a limit point of $f(H)$ then there exist $\{y_k\}$ such that $y_k \in f(H)$ and $y_k \to y$. By definition of image, for each k, there exists $x_k \in H$ such that $y_k = f(x_k)$. Since H is compact there exists a convergent subsequence $\{x_{k_l}\}$ such that $x_{k_l} \to x \in H$, sine y_k converges it must be that $f(x_{k_l}) \to y$. By continuity

$$y = \lim y_{k_l} = \lim f(x_{k_l}) = f(\lim x_{k_l}) = f(x)$$

but since $x \in H$ then $y \in f(H)$, thus proving that $f(H)$ contains all of its limit points. Hence it is closed.

To prove boundedness suppose for a contradiction that $f(H)$ is not bounded. Then choose $x_k \in H$ such that $\|f(x_k)\| \geq k$ for all k. As before there is a convergent subsequence $x_{k_l} \to x \in H$, since the norm and f are continuous we have:

$$\|f(x)\| = \|f(\lim x_{k_l})\| = \infty$$

But since $x \in H$, $f(x) \in \mathbb{R}^m$ and thus $\|f(x)\| < \infty$, this is a contradiction. $\square$

The image of a continuous function also preserves connectedness:

Proposition 5.10. *Let $n, m \in \mathbb{N}$. If $H \subseteq \mathbb{R}^n$ is connected and $f : H \to \mathbb{R}^m$ continuous, then $f(H)$ is connected in $\mathbb{R}^m$.*

To finish consider this alternative proof of the extreme value theorem.

Theorem 5.2. ***(Extreme value theorem - Weierstrass)*** *If $H \subseteq \mathbb{R}^n$ is compact and $f : H \to \mathbb{R}$ continuous, then:*

$$M = \sup\{f(x) \,|x \in H\} \qquad \wedge \qquad m = \inf\{f(x) \,|x \in H\}$$

are finite real numbers. Moreover there exists points x_M and x_m such that $M = f(x_M)$ and $m = f(x_m)$.

Proof. Consider the supremum. Since H is compact and f is continuous then $f(H)$ is compact, hence bounded, so $M < \infty$. By the approximation property to the supremum one can construct a converging sequence to M, since $f(H)$ is closed then $M \in f(H)$ then there exists x_M such that $f(x_M) = M$. $\square$

Part II
Convex Analysis

This part of the course aims to introduce basic concepts of convex analysis. As the rest of the course it does not provide a deep account of the subject, instead the most relevant definitions and some of the most used results are presented.

The definitions and basic properties of convex sets and hyperplanes are presented first, this treatment follows mostly what is presented in the first part of Rockafellar (1997). The treatment ends with a statement of the separating and the supporting hyperplane theorems, most proofs here are omitted since they won't add to the course and they are in general too lengthly to be presented in detail.

Then convex and concave functions are defined and special attention is given to their continuity properties, establishing then the existence of directional derivatives and characterizing convex and concave functions with their first and second derivatives. Finally quasi-convex and quasi-concave functions are introduced and their most important properties are discussed.

This material is mostly based on chapters 7 and 8 of Sundaram (1996), extracting the portions relevant for the results shown in Sections 13 and 15 below. The material is presented before those topics are covered (changing the order with respect to Sundaram) because of its similarity with the topics already covered in the Real Analysis section of the course. I also introduce some other results and definitions from Rockafellar (1997), sections 1 to 3, trying to maintain a unified notation and to keep the course as concise as possible.

6 Convex sets

A convex set is a special kind of connected set whose elements can be joined with a straight line. This property makes convex sets extremely well behaved and allows to strengthen results previously outlined in the notes.

Definition 6.1. (Convex set) A set C in a vector space is convex if and only if for all $x, y \in C$ and $\lambda \in (0,1)$ it follows that $\lambda x + (1-\lambda)\, y \in C$.

It will be useful to define what a convex combination of elements is:

Definition 6.2. (Convex Combination) Let $x_i \in \mathbb{R}^n$ for $i = 1, \ldots, m$ and $\lambda_i \geq 0$ such that $\sum\limits_{i=1}^{m} \lambda_i = 1$. The vector sum $\lambda_1 x_1 + \ldots + \lambda_m x_m$ is called a convex combination of the elements $\{x_i\}$. A special case obtains when $m = 2$.

The question arises if a convex set contains only the convex combinations of its elements up to $m = 2$ or if it contains any convex combination. The answer is given by the following characterization.

Proposition 6.1. *A set $C \subseteq \mathbb{R}^n$ is convex if and only if it contains all the convex combinations of its elements.*

Proof. By definition of convexity the proposition holds for $m = 2$. For arbitrary m the result is established by induction.

Let $m > 2$. The induction hypothesis states that any convex combination of $m - 1$ elements of C belongs to C. Now consider $\lambda_1, \ldots, \lambda_m$ with $\lambda_i \geq 0$ and $\sum \lambda_i = 1$, it must holds that at there is a $\lambda_i \neq 1$, let it be λ_1. Now take $x_1, \ldots, x_m \in C$ and consider its convex combination $x = \lambda_1 x_1 + \ldots + \lambda_m x_m$. We want to show that $x \in C$.

Define $\lambda_j' = \frac{\lambda_j}{1-\lambda_1}$ for $j \geq 2$ and $y = \lambda_2' x_2 + \ldots + \lambda_m' x_m$. Clearly $\lambda_j' \geq 0$ and $\sum \lambda_j' = 1$, then by the induction hypothesis $y \in C$. Note that $x = \lambda_1 x_1 + (1 - \lambda_1)\, y$.

Now C is convex if and only if $x = \lambda_1 x_1 + (1 - \lambda_1)\, y \in C$, since $\lambda_1 \in (0,1)$ is arbitrary. $\square$

There are some convex sets of special importance, among them hyperplanes are used frequently:

Definition 6.3. (Hyperplane) A set $H \subseteq \mathbb{R}^n$ is a hyperplane if and only if there exists $\beta \in \mathbb{R}$ and $b \in \mathbb{R}^n$ such that $H(b, \beta) = \{x \in \mathbb{R}^n | x \cdot b = \beta\}$. Note that H has dimension $n - 1$ and that β and b are unique up to a common nonzero multiple.

Hyperplanes look like points in $\mathbb{R}$, lines in $\mathbb{R}^2$ and planes in $\mathbb{R}^3$ and its clear that they divide the whole space into two sides, these sides are called half-spaces.

Definition 6.4. (Half-Space) A hyperplane $H \subseteq \mathbb{R}^n$ characterized by $\beta \in \mathbb{R}$ and $b \in \mathbb{R}^n$ generates the following two half spaces:

$$\{x \in \mathbb{R}^n | x \cdot b \leq \beta\} \qquad \wedge \qquad \{x \in \mathbb{R}^n | x \cdot b \leq \beta\}$$

Another type of convex set that is frequently encountered is the convex hull of a set. This is the smallest convex set that contains a set, formally:

Definition 6.5. (Convex Hull) Let $E \subseteq \mathbb{R}^n$ be an arbitrary set. The convex hull of E, $\operatorname{co}(E)$, is defined as the intersection of all convex sets C_i such that $E \subseteq C_i$. By proposition 6.3 $\operatorname{co}(E)$ is convex.

Proposition 6.2. *Let $E \subseteq \mathbb{R}^n$, then $co(E) = \{x \in \mathbb{R}^n | \exists_{x_i \in E} \exists_{\lambda_i \geq 0} x = \sum \lambda_i x_i \quad \wedge \quad \sum \lambda_i = 1\}$.*

Proof. By definition the elements of E belong to $\operatorname{co}(E)$, $E \subseteq \operatorname{co}(E)$. Then by proposition 6.1 all convex combinations of the elements of E belong to $\operatorname{co}(E)$. This establishes

$$\left\{x \in \mathbb{R}^n | \exists_{x_i \in E} \exists_{\lambda_i \geq 0} x = \sum \lambda_i x_i \quad \wedge \quad \sum \lambda_i = 1\right\} \subseteq \operatorname{co}(E)$$

To establish the other direction it suffices to show that the set of convex combinations of the elements of E is convex. It follows that it is a convex set that contains E, it must be smallest such set since removing any element will also remove convexity.

Let $x = \lambda_1 x_1 + \ldots + \lambda_m x_m$ and $y = \mu_1 y_1 + \ldots + \mu_r y_r$ be two convex combinations with $x_i, y_i \in E$. Consider $\gamma \in (0,1)$ and $\gamma x + (1-\gamma) y$, and define $\gamma_i = \gamma \lambda_i$ for $i = 1, \ldots m$ and $\gamma_i = (1-\gamma) \mu_{i-m}$ for $i = m+1, \ldots, m+r$. Clearly $\gamma_i \geq 0$ and $\sum \gamma_i = 1$, then the sum

$$\gamma x + (1-\gamma) y = \gamma_i x_i + \ldots + \gamma_m x_m + \gamma_{m+1} y_1 + \ldots + \gamma_{m+r} y_r$$

is a convex combination of elements of E, this proves convexity and finishes the proof. $\square$

This proposition allows to characterize a special type of convex hull:

Definition 6.6. (Simplex) A set $C \subseteq \mathbb{R}^n$ is an m-dimensional simplex if and only if it can be expressed as $\operatorname{co}(\{b_1, \ldots, b_m\})$ for b_i affinely independent. The vectors b_i are called the vertices of the simplex, and an element of the simplex is of the form $\lambda_1 b_1 + \ldots + \lambda_m b_m$ for $\lambda_i \geq 0$ and $\sum \lambda_i = 1$.

Two important properties of convexity is that it is preserved under arbitrary intersection, under sum of sets, scalar multiplication and translation :

Proposition 6.3. *Let $\{C_i\}_{i \in I}$ where I is an arbitrary set of indexes, $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.*

i. Consider $C = \bigcap C_i$. If C_i is convex for all i then C is convex.

ii. Consider $C = C_i + C_j$. If C_i and C_j are convex then C is convex.

iii. Consider $C = C_i + a$. If C_i is convex then C is convex.

iv. Consider $C = \alpha C_i$. If C_i is convex then C is convex.

Proof. The proof follows from the definition of convexity.

i. Let $x, y \in C = \bigcap C_i$, then for all i it holds that $x, y \in C_i$, since C_i is convex it holds that $\forall_{\lambda \in (0,1)} \lambda x + (1-\lambda) y \in C_i$, this is true for all i, then $\forall_{\lambda \in (0,1)} \lambda x + (1-\lambda) y \in C$ proving convexity.

ii. Let $x, y \in C$ then there exist $x_i, y_i \in C_i$ and $x_j, y_j \in C_j$ such that $x = x_i + x_j$ and $y = y_i + y_j$. Let $\lambda \in (0,1)$, then $\lambda x + (1-\lambda) y = (\lambda x_i + (1-\lambda) y_i) + (\lambda x_j + (1-\lambda) y_j)$. Since C_i and C_j are convex $\lambda x_i + (1-\lambda) y_i \in C_i$ and $\lambda x_j + (1-\lambda) y_j \in C_j$, then their sum belongs to C proving convexity.

iii. Let $x, y \in C$, then $x - a, y - a \in C_i$, then for any $\lambda \in (0,1)$ we have $\lambda x + (1-\lambda) y = \lambda (x-a) + (1-\lambda)(y-a) + a$, by convexity of C_i $\lambda (x-a) + (1-\lambda)(y-a) \in C_i$, then $\lambda x + (1-\lambda) y \in C_i + a$.

iv. Let $x, y \in C$, if $\alpha = 0$ then $x = y = 0$ and C is convex. If $\alpha \neq 0$ then $x/\alpha, y/\alpha \in C_i$. Let $\lambda \in (0,1)$ and consider $\lambda x + (1-\lambda) y = \alpha (\lambda x/\alpha + (1-\lambda) y/\alpha)$, by convexity $\lambda x/\alpha + (1-\lambda) y/\alpha \in C_i$, thus proving convexity since $\lambda x + (1-\lambda) y \in \alpha C_i$.

□

Corollary 6.1. *Let $b_i \in \mathbb{R}^n$ and $\beta_i \in \mathbb{R}$ for $i \in I$, where I is an arbitrary index set. Then the set:*

$$C = \{x \in \mathbb{R}^n | \forall_i x \cdot b_i \leq \beta_i\}$$

is convex.

Proof. It follows form C being the arbitrary intersection of half spaces. □

This corollary is of special interest in economics since it implies that any system of linear inequalities and equations has a convex solution set. This corollary can be generalized when considering a set of the form

$$C = \{x \in \mathbb{R}^n | \forall_i f_i(x) \leq \beta_i\}$$

for f_i a (proper) quasi-convex function.

Finally we deal with the use of hyperplanes to separate convex sets. It turns out that one can always separate two convex sets whose (relative) interiors are disjoint. We first define formally what it means to separate the sets and then we state an existence result that guarantees separation. This result is of special use in economics to answer existence questions on equilibrium prices (for general equilibrium theory) or optimal contracts.

Definition 6.7. (Separating Hyperplane) Let $C_1 \in \mathbb{R}^n$ and $C_2 \in \mathbb{R}^n$ be two sets. $H = \{x \in \mathbb{R}^n | x \cdot b = \beta\}$ is a separating hyperplane of C_1 and C_2 if and only if:

$$\forall_{x \in C_1} x \cdot b \leq \beta \qquad \wedge \qquad \forall_{y \in C_2} y \cdot b \geq \beta$$

that is, if each set is contained in one of the half spaces induced by the hyperplane. The separation is said to be strong if at least one of the inequalities holds strictly.

It turns out that a separating hyperplane exists if and only if $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset$, see Rockafellar (1997, Thm 11.3, pp 97). The following two theorems make this precise, they are taken from Sundaram (1996) but more detailed (and general) versions can be found in Rockafellar (1997, Ch 11).

Theorem 6.1. ***(Separating Hyperplane) Let*** *$C_1, C_2 \subseteq \mathbb{R}^n$ be convex sets and $x \in \mathbb{R}^n$.*

i. If $x \notin C_1$, there exists b and β with $\|b\| = 1$ such that $H(b, \beta)$ separates C_1 and x strongly.

ii. If $C_1 \cap C_2 = \emptyset$, there exists b and β with $\|b\| = 1$ such that $H(b, \beta)$ separates C_1 and C_2 strongly.

(a) The condition can be relaxed to $ri(C_1) \cap ri(C_2) = \emptyset$ but then separation is not strong.

Proof. Sundaram (1996, sec 1.6.1, pp 56) □

An important consequence of this theorem is the following dual representation of a closed convex set.

Proposition 6.4. *A closed convex set C is the intersection of the closed half-spaces that contain it.*

Proof. Wlog assume $\emptyset \neq C \neq \mathbb{R}^n$. Let $a \notin C$, then by the separating hyperplane theorem there exists a hyperplane such that C is contained in one of its half-spaces and a is not. Then the intersection of half spaces containing C contains no points other than those in C. □

One can see that this characterization involves choosing some hyperplanes that are 'tangent' to the set C, those that only touch C on its boundary. The formal definition of tangency for convex sets is that of supporting hyperplane.

Definition 6.8. (Supporting Hyperplane) A hyperplane $H(b, \beta)$ with $b \neq 0$ is a supporting hyperplane of set C if and only if $x \cdot b \leq \beta$ for all $x \in C$, and there exists $x \in C$ such that $x \cdot b = \beta$. A supporting half-space is then a half-space generated by a supporting half space.

Its clear that a supporting hyperplane has a close relation with a linear function $(x \cdot b)$ taking a maximum on a set. Thus the supporting hyperplanes of a set can be characterized with the support function of a set. The supporting hyperplane of a set C in direction b is:

$$\overline{H}(b) = \{x \in \mathbb{R}^n | x \cdot b = \delta(b|C)\}$$

Proposition 6.5. *The closure of a convex set C is the intersection of its supporting half-spaces.*

7 Convex and concave functions

7.1 Definitions

Convex sets give rise to convex and concave functions which are defined depending of the properties of their epigraphs and subgraphs, these are:

Definition 7.1. (Epigraph and Subgraph) Let $f : \Gamma \to \mathbb{R}$ be a real valued function with domain $\Gamma \subseteq \mathbb{R}^n$. The epigraph and subgraph of f are sets in $\mathbb{R}^n \times \mathbb{R}$ of the form:

$$\text{epi}f = \{(x, y) \in \Gamma \times \mathbb{R} | y \geq f(x)\} \qquad \text{sub}f = \{(x, y) \in \Gamma \times \mathbb{R} | y \leq f(x)\}$$

Graphically the epigraph of f is the set above the values of f in the space $\mathbb{R}^n \times \mathbb{R}$ and the subgraph the set below them.

Now its possible to define what a convex and a concave function are:

Definition 7.2. (Convex and Concave Functions) Let $f : \Gamma \to \mathbb{R}$ be a real valued function with domain $\Gamma \subseteq \mathbb{R}^n$.

i. f is said to be **convex** if and only if epi$f$ is a convex set in $\mathbb{R}^n \times \mathbb{R}$.

ii. f is said to be **concave** if and only if sup$f$ is a convex set in $\mathbb{R}^n \times \mathbb{R}$.

iii. f is said to be **affine** if and only if f is convex and concave.

The following proposition gives another characterization of convex and concave functions that allows to avoid the use of sets.

Proposition 7.1. *Let $\Gamma \subseteq \mathbb{R}^n$ be a convex set and $f : \Gamma \to \mathbb{R}$ a function.*

*i. f is **convex** if and only if for all $x, y \in \Gamma$ and $\lambda \in (0, 1)$*

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

*ii. f is **concave** if and only if for all $x, y \in \Gamma$ and $\lambda \in (0, 1)$*

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

Proof. I only prove the first statement of the proposition.

i. (If f is convex then $f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)$)
Let f be convex, then epi$f$ is a convex set, since $(x, f(x)) \in \text{epi}f$ and $(y, f(y)) \in \text{epi}f$ it follows that for any $\lambda \in (0, 1)$ it holds that $(\lambda x + (1 - \lambda) y, \lambda f(x) + (1 - \lambda) f(y)) \in$ epi$f$. By the definition of epi$f$ we get the result:

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

ii. (If $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$ then f is convex)
To show that f is convex we must show that epi$f$ is convex. Let $(x,z) \in$ epi$f$ and $(y,w) \in$ epi$f$ and $\lambda \in (0,1)$. By hypothesis $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$. By definition of epi$f$ we have $z \geq f(x)$ and $w \geq f(y)$, then by multiplying the first inequality by λ and the second one by $(1-\lambda)$ and summing we get: $\lambda f(x) + (1-\lambda) f(y) \leq \lambda z + (1-\lambda) w$. Joining inequalities we get: $f(\lambda x + (1-\lambda) y) \leq \lambda z + (1-\lambda) w$, then, by definition of epi$f$ we get $(\lambda x + (1-\lambda) y, \lambda z + (1-\lambda) w) \in$ epi$f$.

□

With this proposition it becomes possible to define strictly convex and strictly concave functions as those for which the inequalities above hold strictly.

Another useful result relates convex to concave functions in a trivial way:

Proposition 7.2. *Let $\Gamma \subseteq \mathbb{R}^n$ be a convex set and $f : \Gamma \to \mathbb{R}$ a function. f is convex if and only if $-f$ is concave.*

Proof. The proof is immediate from the previous proposition.

Let f be a convex function, this happens if and only if

$$\begin{aligned} f(\lambda x + (1-\lambda) y) &\leq \lambda f(x) + (1-\lambda) f(y) \\ -f(\lambda x + (1-\lambda) y) &\geq \lambda(-f(x)) + (1-\lambda)(-f(y)) \end{aligned}$$

Which happens if and only if $-f$ is concave. □

Another characterization of convex and concave function is Jensen's inequality:

Proposition 7.3. (Jensen's Inequality) *Let $\Gamma \subseteq \mathbb{R}^n$ be a convex set and $f : \Gamma \to \mathbb{R}$ a function. f is convex if and only if $f(\lambda_1 x_1 + \ldots + \lambda_k x_k) \leq \lambda_1 f(x_1) + \ldots + \lambda_k f(x_k)$ for $x_1, \ldots, x_k \in \Gamma$ and $\lambda_i \geq 0$ and $\sum \lambda_1 = 1$.*

Proof. By applying proposition 7.1 repetitively to pairs of elements. □

Note that the previous proposition is an if and only if statement.

Finally we introduce a special type of convex function that plays a role in optimization and duality theory:

Definition 7.3. (Support Function) The support function $\delta(\cdot|C)$ of a convex set $C \subseteq \mathbb{R}^n$ is defined as:

$$\delta(x|C) = \sup_{y \in C} \quad \{x \cdot y\}$$

The argument x is the direction in which C is supported. δ is clearly convex by properties of the supremum.

7.2 Properties of convex and concave functions

Convex and concave functions have a lot nice properties, in what follows some of the properties of concave functions are presented. The first property has to do with the continuity of concave functions.

Proposition 7.4. *Let $f : \Gamma \to \mathbb{R}$ be a concave function on a convex set $\Gamma \subseteq \mathbb{R}^n$. Let $x \in \Gamma$ such that there exists an open neighborhood of x, $V \subseteq \Gamma$, and a real number M such that $|f(y)| \leq M$ for all $y \in V$. Then f is continuous at x.*

Proof. Consider $\{x_k\}$ in V with $x_k \to x$. Since V is open there is an $r > 0$ such that $B_r(x) \subset V$. Let $\alpha \in (0, r)$ and consider the set $A = \{z | \, \|x - z\| = \alpha\}$, that is the boundary of the ball $B_\alpha(x)$. Since $x_k \to x$ pick K large enough so that $\|x - x_k\| < \alpha$ for $k > K$. Note that $A \subset V$.

For $k > K$ there exists a $z_k \in A$ such that $x_k = \lambda_k x + (1 - \lambda_k) z_k$, for some $\lambda_k \in (0, 1)$. Since $x_k \to x$ and $\|x - z_k\| = \alpha > 0$ it must be that $\lambda_k \to 1$ as $k \to \infty$.

By concavity of f it follows that:

$$f(x_k) = f(\lambda_k x + (1 - \lambda_k) z_k) \geq \lambda_k f(x) + (1 - \lambda_k) f(z_k)$$

Taking limits we have:

$$\begin{aligned} \liminf f(x_k) &\geq \liminf (\lambda_k f(x) + (1 - \lambda_k) f(z_k)) \\ &= f(x) + \liminf_{k \to \infty} ((1 - \lambda_k) f(z_k)) \\ &= f(x) \end{aligned}$$

Since $z_k \in A \subset V$ we have $|f(z_l)| \leq M$, then $\liminf ((1 - \lambda_k) f(z_k)) = \lim ((1 - \lambda_k) f(z_k)) = 0$ recalling that $\lambda_k \to 1$.

In a similar way there exists numbers $\theta_k \in (0, 1)$ such that $x = \theta_k x_k + (1 - \theta_k) z_k$, by concavity:

$$f(x) = f(\theta_k x_k + (1 - \theta_k) z_k) \geq \theta_k f(x_k) + (1 - \theta_k) f(z_k)$$

As before $\theta_k \to 1$, then by taking limits:

$$\begin{aligned} f(x) &\geq \limsup (\theta_k f(x_k) + (1 - \theta_k) f(z_k)) \\ &\geq \limsup (\theta_k f(x_k)) \\ &\geq \limsup (f(x_k)) \end{aligned}$$

where, as before $\lim ((1 - \theta_k) f(z_k)) = \limsup ((1 - \theta_k) f(z_k)) = 0$ and, since $\theta_k \to 1$ we get the last step.

We have now that $\liminf f(x_k) \geq f(x) \geq \limsup (f(x_k))$, which implies $\lim f(x_k) = f(x)$ and thus continuity of f at x. □

Note that it was necessary to use lim inf and lim sup in the proof since we needed to establish the existence (and value) of $\lim f(x_k)$.

Corollary 7.1. *Let $f : \Gamma \to \mathbb{R}$ be a concave function on a convex set $\Gamma \subseteq \mathbb{R}^n$. Suppose that there exists an open set, $V \subseteq \Gamma$, and a real number M such that $|f(y)| \leq M$ for all $y \in V$. Then f is continuous on V.*

Stronger results can be more easily obtained for functions of one variable. For instance its easy to show that a concave function on an open interval is continuous on all of the interval[3]. There are also implications for the differentiability of the function.

Proposition 7.5. *Let $g : \Gamma \to \mathbb{R}$ be a concave function with $\Gamma \subseteq \mathbb{R}$ an open and convex set. Let $x_1, x_2, x_3 \in \Gamma$ with $x_1 < x_2 < x_3$. Then:*

$$\frac{g(x_2) - g(x_1)}{x_2 - x_1} \geq \frac{g(x_3) - g(x_1)}{x_3 - x_1} \geq \frac{g(x_3) - g(x_2)}{x_3 - x_2}$$

with strict inequality if g is strictly concave.

Proof. Let $\alpha = \frac{x_2 - x_1}{x_3 - x_1}$, clearly $\alpha \in (0,1)$, $1 - \alpha = \frac{x_3 - x_2}{x_3 - x_1}$ and $\alpha x_3 + (1-\alpha) x_1 = x_2$. Since g is concave we have:

$$g(x_2) = g(\alpha x_3 + (1-\alpha) x_1) \geq \alpha g(x_3) + (1-\alpha) g(x_1)$$

With strict inequality if g is strictly concave. Replacing for α and rearranging:

$$\begin{aligned} g(x_2) &\geq \frac{x_2 - x_1}{x_3 - x_1} g(x_3) + \left(1 - \frac{x_2 - x_1}{x_3 - x_1}\right) g(x_1) \\ g(x_2) - g(x_1) &\geq \frac{x_2 - x_1}{x_3 - x_1} \left(g(x_3) - g(x_1)\right) \\ \frac{g(x_2) - g(x_1)}{x_2 - x_1} &\geq \frac{g(x_3) - g(x_1)}{x_3 - x_1} \end{aligned}$$

This establishes one of the inequalities, expressing x_1 and x_3 as linear combinations of the other points the other inequalities are obtained. □

Corollary 7.2. *Let $g : \mathbb{R} \to \mathbb{R}$ be concave and $x \in \mathbb{R}$. Then the quotient $\frac{g(x+b) - g(x)}{b}$ is non-increasing for $b > 0$. Moreover, the limit of the quotient as $b \to 0^+$ exists in the extended real line.*

Proof. Let $b_1 < b_2$, and consider $x_1 = x$, $x_2 = x + b_1$ and $x_3 = x + b_2$, then by the first inequality above:

$$\frac{g(x + b_1) - g(x)}{b_1} \geq \frac{g(x + b_2) - g(x)}{b_2}$$

which proves that the quotient is non-increasing. The existence of the limit follows from the quotient being a monotone function. If its bounded then the limit exists by the monotone convergence theorem, if its not then the limit is infinity. □

Corollary 7.3. *Let $g : \mathbb{R} \to \mathbb{R}$ be concave and $x \in \mathbb{R}$. Then the quotient $\frac{g(x) - g(x+b)}{-b}$ is non-increasing for $b < 0$. Moreover, the limit of the quotient as $b \to 0^-$ exists in the extended real line.*

[3] The proof above has no change except for the fact that z does not depend on k.

Proof. Let $b_1 < b_2$, and consider $x_1 = x - b_1$, $x_2 = x + b_2$ and $x_3 = x$, then by the second inequality above:

$$\frac{g(x) - g(x + b_1)}{-b_1} \geq \frac{g(x) - g(x + b_2)}{-b_2}$$

which proves that the quotient is non-increasing. The existence of the limit follows from the quotient being a monotone function. If its bounded then the limit exists by the monotone convergence theorem, if its not then the limit is infinity. □

With this result it is possible to establish the existence of directional derivatives for concave functions of one variable.

Proposition 7.6. *If $g : \mathbb{R} \to \mathbb{R}$ is concave then all (one-sided) directional derivatives exist at all points. They might take infinite value.*

Proof. Let $x, y \in \mathbb{R}$. The directional derivative of g at x in the direction of y is:

$$Df(x, y) = \lim_{t \to 0^+} \frac{g(x + ty) - g(x)}{t}$$

There are three possible cases for the direction y:

Case 1. Suppose $y > 0$.

$$\frac{g(x + ty) - g(x)}{t} = \frac{g(x + ty) - g(x)}{ty} y$$

Clearly $ty \to 0^+$ if and only if $t \to 0^+$. Let $b = ty$, then

$$\lim_{t \to 0^+} \frac{g(x + ty) - g(x)}{t} = y \lim_{b \to 0^+} \frac{g(x + b) - g(x)}{b}$$

by corollary 7.2 we know that the limit in the RHS exists, then the directional derivative exists in any direction $y > 0$.

Case 2. Suppose $y < 0$.

$$\frac{g(x + ty) - g(x)}{t} = \frac{g(x) - g(x + ty)}{-ty} y$$

Clearly $ty \to 0^-$ if and only if $t \to 0^-$. Let $b = ty$, then

$$\lim_{t \to 0^+} \frac{g(x + ty) - g(x)}{t} = y \lim_{b \to 0^-} \frac{g(x) - g(x + b)}{-b}$$

Again, by corollary 7.3 we know that the limit in the RHS exists, then the directional derivative exists in any direction $y < 0$.

Case 3. Finally if $y = 0$ the derivative in direction y is trivially equal to 0 since:

$$\frac{g(x + ty) - g(x)}{t} = \frac{g(x) - g(x)}{t} = 0$$

□

This result can be extended to higher dimensions:

Proposition 7.7. *Let $\Gamma \subseteq \mathbb{R}^n$ be open and convex. If $f : \Gamma \to \mathbb{R}$ is concave, then f possesses all directional derivatives at all points in Γ.*

Proof. Let $x \in \Gamma$ and $h \in \mathbb{R}^n$. Define $g(t) = f(x + th)$ for $t \geq 0$, since Γ is open g is well defined in a neighborhood of 0. Also note that:

$$\frac{f(x+th) - f(x)}{t} = \frac{g(t) - g(0)}{t}$$

For any $\alpha \in (0,1)$:

$$\begin{aligned}
g\left(\alpha t + (1-\alpha) t'\right) &= f\left(x + \left(\alpha t + (1-\alpha) t'\right) h\right) \\
&= f\left(\alpha (x + th) + (1-\alpha)\left(x + t'h\right)\right) \\
&\geq \alpha f(x+th) + (1-\alpha) f\left(x + t'h\right) \\
&= \alpha g(t) + (1-\alpha) g\left(t'\right)
\end{aligned}$$

then g is concave on $\mathbb{R}_+$. Then by corollary 7.2 the limit:

$$Df(x,h) = \lim_{t \to 0^+} \frac{f(x+th) - f(x)}{t} = \lim_{t \to 0^+} \frac{g(t) - g(0)}{t}$$

exists, this proves the existence of the directional derivative of f at x in an arbitrary direction h. □

The question remains if a concave function is differentiable on its domain. The answer is given without a proof in the following proposition:

Proposition 7.8. *Let $\Gamma \subseteq \mathbb{R}^n$ be open and convex. If $f : \Gamma \to \mathbb{R}$ is concave, then f is differentiable almost everywhere (on all Γ except for a set of measure zero). Moreover the derivative of f is continuous wherever it exists.*

Having established that a concave function is almost everywhere continuously differentiable it is now possible to study some properties of the derivatives of a convex functions. These properties allow for a characterization of concave (and convex) functions based on first and second derivatives of the function.

Proposition 7.9. *Let $\Gamma \subseteq \mathbb{R}^n$ be open and convex. If $f : \Gamma \to \mathbb{R}$ be differentiable on Γ. f is concave on Γ if and only if*

$$\forall_{x,y \in \Gamma} \quad Df(x)(y-x) \geq f(y) - f(x)$$

Similarly, f is convex on Γ if and only if

$$\forall_{x,y \in \Gamma} \quad Df(x)(y-x) \leq f(y) - f(x)$$

Proof. Sundaram (1996, sec. 7.5, pp. 190). □

Proposition 7.10. *Let $\Gamma \subseteq \mathbb{R}^n$ be open and convex. If $f : \Gamma \to \mathbb{R}$ be a C^2 function on Γ.*

i. f is concave on Γ if and only if $D^2 f(x)$ is a negative semidefinite matrix for all $x \in \Gamma$.

ii. f is convex on Γ if and only if $D^2 f(x)$ is a positive semidefinite matrix for all $x \in \Gamma$.

iii. If $D^2 f(x)$ is negative definitive for all $x \in \Gamma$ then f is strictly concave.

iv. If $D^2 f(x)$ is positive definitive for all $x \in \Gamma$ then f is strictly convex.

Proof. Sundaram (1996, sec. 7.6, pp. 191). □

These two propositions are extremely useful, the first one specially for cases in which one knows the function is concave (convex) and needs to know the behavior of its derivative, the second one for cases in which one knows the function and needs conditions for it to be concave (convex). This is shown in the following example.

Example 7.1. Let $f : \mathbb{R}^2_{++} \to \mathbb{R}$ be $f(x, y) = x^a y^b$ with $a, b > 0$.

Using the proposition 7.10 f is concave if the Hessian matrix is negative semidefinite. The jacobian is:

$$Df(x, y) = \left[\begin{array}{cc} a x^{a-1} y^b & b x^a y^{b-1} \end{array}\right]$$

The Hessian is:

$$D^2 f(x) = \left[\begin{array}{cc} a(a-1) x^{a-2} y^b & ab x^{a-1} y^{b-1} \\ ab x^{a-1} y^{b-1} & b(b-1) x^a y^{b-2} \end{array}\right]$$

The determinant of the matrix is:

$$\begin{aligned} \left|D^2 f(x)\right| &= a(a-1) b(b-1) x^{2(a-1)} y^{2(b-1)} - a^2 b^2 x^{2(a-1)} y^{2(b-1)} \\ &= ((a-1)(b-1) - ab) ab x^{2(a-1)} y^{2(b-1)} \\ &= (1 - a - b) ab x^{2(a-1)} y^{2(b-1)} \end{aligned}$$

The determinant is then negative if $a + b < 1$, positive if $a + b > 1$ and zero if $a + b = 1$. The diagonal terms are negative if $a < 1$ and $b < 1$. Then f is strictly concave if $a + b < 1$ and $a, b < 1$, it is concave if $a + b = 1$ and $a, b < 1$ and is neither concave nor convex when $a + b > 1$.

8 Quasi-Convex and quasi-concave functions

8.1 Definitions

Another relevant (and more general) family of functions is that of quasi-convex and quasi-concave functions. These are defined with the upper and lower contour sets of the function.

Definition 8.1. (Contour sets) Let $f : \Gamma \to \mathbb{R}$ be a real valued function an $\Gamma \subseteq \mathbb{R}^n$. The upper and lower contour sets of f at $a \in \mathbb{R}$ are:

$$U_f(a) = \{x \in \Gamma | f(x) \geq a\} \qquad L_f(a) = \{x \in \Gamma | f(x) \leq a\}$$

Quasi-concave and quasi-convex function are then defined as follows:

Definition 8.2. (Quasi-Convex and QuasiConcave Functions) Let $f : \Gamma \to \mathbb{R}$ be a real valued function with domain $\Gamma \subseteq \mathbb{R}^n$.

i. f is said to be **quasi-convex** if and only if $L_f(a)$ is a convex set in $\mathbb{R}^n$ for all $a \in \mathbb{R}$.

ii. f is said to be **quasi-concave** if and only if $U_f(a)$ is a convex set in $\mathbb{R}^n$ for all $a \in \mathbb{R}$.

As with convex and concave functions there are alternative characterizations of quasi-convex and quasi-concave functions.

Proposition 8.1. *Let $\Gamma \subseteq \mathbb{R}^n$ be a convex set and $f : \Gamma \to \mathbb{R}$ a function.*

*i. f is **quasi-convex** if and only if for all $x, y \in \Gamma$ and $\lambda \in (0, 1)$*

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

*ii. f is **quasi-concave** if and only if for all $x, y \in \Gamma$ and $\lambda \in (0, 1)$*

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

Proof. I only prove the first statement of the proposition.

i. Let f be quasi-convex, then $L_f(a)$ is convex for all a. Consider $x, y \in \Gamma$ and $\lambda \in (0, 1)$. Wlog let $f(x) \geq f(y)$ and $a = f(x)$. For the chosen a $L_f(a)$ is convex and $x, y \in L_f(a)$ by construction. By convexity $\lambda x + (1 - \alpha) y \in L_f(a)$, then by definition

$$f(\lambda x + (1 - \lambda) y) \leq a = f(x) = \max\{f(x), f(y)\}$$

which is the desired result.

ii. Suppose that $f(\lambda x + (1 - \lambda) y) \leq \max\{f(x), f(y)\}$ for all $x, y \in \Gamma$ and $\lambda \in (0, 1)$. Choose $a \in \mathbb{R}$ and consider $L_f(a)$, if the set is empty or a singleton the proof follows vacuously, then wlog choose $x, y \in L_f(a)$. These points satisfy $f(x) \leq a$ and $f(y) \leq a$, then $\max\{f(x), f(y)\} \leq a$. By hypothesis we then have

$$f(\lambda x + (1 - \lambda) y) \leq a$$

which implies that $\lambda x + (1 - \lambda) y \in L_f(a)$, proving that $L_f(a)$ is convex for all a.

□

A function will be called strictly quasi-convex (concave) if the inequalities of proposition 8.1 hold strictly. A result that parallels proposition 7.2 follows trivially.

Proposition 8.2. *Let $\Gamma \subseteq \mathbb{R}^n$ be a convex set and $f : \Gamma \to \mathbb{R}$ a function. f is quasi-convex if and only if $-f$ is quasi-concave.*

Proof. The proof is immediate from the previous proposition.

Let f be a quasi-convex function, this happens if and only if

$$\begin{aligned} f\left(\lambda x+(1-\lambda)y\right) &\leq \max\left\{f\left(x\right),f\left(y\right)\right\} \\ -f\left(\lambda x+(1-\lambda)y\right) &\geq -\max\left\{f\left(x\right),f\left(y\right)\right\} \\ -f\left(\lambda x+(1-\lambda)y\right) &\geq \min\left\{-f\left(x\right),-f\left(y\right)\right\} \end{aligned}$$

Which happens if and only if $-f$ is quasi-concave. □

As noted above the family of quasi-convex (concave) functions is a generalization of that of convex (concave) functions. The following proposition makes this precise.

Proposition 8.3. *Let $f : \Gamma \to \mathbb{R}$ be a function where $\Gamma \subseteq \mathbb{R}^n$. If f is concave on Γ it is also quasi-concave. If f is convex on Γ it is also quasi-convex.*

Proof. Suppose f is convex, then for all $x, y \in \Gamma$ and $\lambda \in (0,1)$ we have:

$$\begin{aligned} f\left(\lambda x+(1-\lambda)y\right) &\leq \lambda f\left(x\right)+(1-\lambda)f\left(y\right) \\ &\leq \lambda\max\left\{f\left(x\right),f\left(y\right)\right\}+(1-\lambda)\max\left\{f\left(x\right),f\left(y\right)\right\} \\ &= \max\left\{f\left(x\right),f\left(y\right)\right\} \end{aligned}$$

which finishes the proof. □

Quasi-concave (convex) functions are also useful in that their class is preserved by any monotone non-decreasing (increasing) transformation, while concaveness (convexity) is weakened to quasi-concaveness (convexity) unless the transformation is also concave (convex) and the original function monotone non-decreasing.

Proposition 8.4. *Let $f : \Gamma \to \mathbb{R}$ be quasi-concave and $\phi : \mathbb{R} \to \mathbb{R}$ monotone non-decreasing. Then $\phi \circ f$ is quasi-concave.*

Proof. Let $x, y \in \Gamma$ and $\lambda \in (0,1)$. Note that since ϕ is monotone non-decreasing if $f(x) \geq f(y)$ then $\phi(f(x)) \geq \phi(f(y))$, then $\min\{\phi(f(x)), \phi(f(y))\} = \phi(\min\{f(x), f(y)\})$. The conclusion follows since by assumption:

$$f\left(\lambda x+(1-\lambda)y\right)\geq\min\left\{f\left(x\right),f\left(y\right)\right\}$$

Applying ϕ:

$$\begin{aligned} \phi\left(f\left(\lambda x+(1-\lambda)y\right)\right) &\geq \phi\left(\min\left\{f\left(x\right),f\left(y\right)\right\}\right) \\ \phi\left(f\left(\lambda x+(1-\lambda)y\right)\right) &\geq \min\left\{\phi\left(f\left(x\right)\right),\phi\left(f\left(y\right)\right)\right\} \end{aligned}$$

This completes the proof. □

8.2 Properties of quasi-convex and quasi-concave functions

A lot of the nice properties of convex (concave) functions cannot be generalized to quasi-convex (concave) functions. For instance functions of the latter family are not necessarily continuous in the interior of its domain and can have local optima that are not global. There are still some properties regarding their first and second derivatives that are presented below.

Proposition 8.5. *Let $f : \Gamma \to \mathbb{R}$ be a C^1 function and $\Gamma \subseteq \mathbb{R}^n$ be open and convex.*

i. Then f is quasi-concave on Γ if and only if for all $x, y \in \Gamma$:

$$f(y) \geq f(x) \longrightarrow Df(x)(y-x) \geq 0$$

ii. Then f is quasi-convex on Γ if and only if for all $x, y \in \Gamma$:

$$f(y) \geq f(x) \longrightarrow Df(x)(y-x) \leq 0$$

Proof. I prove only the necessity part of the first claim:

Let f be quasi-concave on Γ and wlog let $x, y \in \Gamma$ such that $f(y) \geq f(x)$. Let $t \in (0,1)$ and by quasi-concaveness:

$$f(x + t(y-x)) = f((1-t)x + ty) \geq \min\{f(x), f(y)\} = f(x)$$

Joining:

$$\begin{aligned} f(x+t(y-x)) &\geq f(x) \\ \frac{f(x+t(y-x)) - f(x)}{t} &\geq 0 \end{aligned}$$

Taking limits as $t \to 0^+$ the LHS converges to the directional derivative of f in the direction $y - x$, then:

$$\begin{aligned} Df(x, y-x) &\geq 0 \\ Df(x)(y-x) &\geq 0 \end{aligned}$$

which is the desired result. The last step follows from Theorem 1.56 in Sundaram (1996). □

A characterization of quasi-concave (convex) functions is also available using the bordered hessian of the function. In what follows let:

$$\tilde{H}_f(x) = \begin{bmatrix} 0 & Df(x) \\ Df(x)' & D^2(x) \end{bmatrix}$$

and $C_k(x)$ the k^{th} order principal minor of $\tilde{H}_f(x)$,

The following proposition is stated without a proof.

Proposition 8.6. *Let $f : \Gamma \to \mathbb{R}$ be a C^2 function and $\Gamma \subseteq \mathbb{R}^n$ be open and convex.*

i. If f is quasi-concave on Γ then $(-1)^k |C_k(x)| \geq 0$ for $k = 1, \ldots, n$.

ii. If $(-1)^k |C_k(x)| > 0$ for $k = 1, \ldots, n$, then f is quasi-concave on Γ.

iii. If f is quasi-convex on Γ then $|C_k(x)| \leq 0$ for $k = 1, \ldots, n$.

iv. If $(-1)^k |C_k(x)| < 0$ for $k = 1, \ldots, n$, then f is quasi-convex on Γ.

Part III

Optimization

The material in this section follows closely Sundaram (1996) and except minor changes in notation and a selection of material and examples to be presented there are no actual differences between the book and the material presented here.

Topics in optimization are covered in five sections. The first one just introduces the problem of optimization, the second one uses results from Real Analysis to establish conditions for existence of a solution (the extreme value theorem), then necessary conditions for an optimum in unrestricted and restricted problems are covered in the next three sections. Finally the implications of convexity and quasi-convexity over optimization problems are presented. These last sections are the most important ones in the topic of optimization problems, since the implications of convexity are often overlook and the treatment of constraint maximization problems is well understood by most students.

9 Introduction

In general we will refer as an optimization problem a maximization or minimization problem parametrized by parameter θ to a problem of the form:

$$v(\theta) = \max_{x \in \Gamma(\theta)} f(x, \theta) \qquad \text{or} \qquad v(\theta) = \min_{x \in \Gamma(\theta)} f(x, \theta)$$

where $\theta \in \Theta$, $\Gamma(\theta) \subseteq \mathbb{R}^n$ is the feasible set and $f : \mathbb{R}^n \times \Theta \to \mathbb{R}$ is a function. The term $v(\theta)$ is denoted as the value of the problem and is a function of the parameter θ and the set

$$G(\theta) = \{x \in \Gamma(\theta) \,|\, f(x, \theta) = v(\theta)\} = \underset{x \in \Gamma(\theta)}{\text{argmax}} f(x, \theta)$$

is the solution of the problem, the set of argmax (argmin).

Until parametric problems are studied in detail the explicit mention of the parameters does not add much to the discussion, hence they are omitted in almost all of what follows.

Because of the following result all the material will be devoted to maximization problems:

Proposition 9.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and $-f$ another function such that its value at x is $-f(x)$. x is a maximum of f on Γ if and only if x is a minimum of $-f$ on Γ and z is a minimum of f on Γ if and only if z is a maximum of $-f$ on Γ.*

Proof. Let x be a maximum of f on Γ then $f(x) \geq f(y)$ for all $y \in \Gamma$. The inequality implies that $-f(x) \leq -f(y)$ for all $y \in \Gamma$, this is the definition of x being a minimum of $-f$. The second part of the proposition by noting that $-(-f) = f$, so that a relabeling of the functions is enough. $\square$

Using this result its easy to transform any maximization problem into a minimization problem which gives validity to all the results obtained under maximization to a general optimization problem. Another useful result is that the solution to the problem is invariant to strictly increasing transformations, while the value of the problem is the only thing transformed:

Proposition 9.2. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function. Then x is a maximum of f on Γ if and only x is a maximum of $\varphi \circ f$ on Γ.*

Proof. Let x be a maximum of f on Γ, then for all $y \in \Gamma$ we have $f(x) \geq f(y)$, since φ is strictly increasing $\varphi(f(x)) \geq \varphi(f(y))$, then x is a maximum for $\varphi \circ f$.

Now let x be a maximum of $\varphi \circ f$, we have $\varphi(f(x)) \geq \varphi(f(y))$ for all $y \in \Gamma$. Suppose for a contradiction that x does not maximize f on Γ, then there exists $z \in \Gamma$ such that $f(z) > f(x)$, since φ is strictly increasing $\varphi(f(z)) > \varphi(f(x))$ which contradicts x being a maximizer of $\varphi \circ f$. $\square$

Now we proceed to characterize the solution of the problems in three steps, first establishing if a solution exists, second finding tools to identify the solution and third obtaining properties of the solution from the fundaments of the problem, that is, which properties does the set of optimizers and the value of the problem exhibit, given properties of f and Γ.

10 Existence

A set of (broad) conditions that guarantee the existence of a solution to a problem (the existence of an optimizer) is given above by Weierstrass theorem (Theorem 5.2). When the function f is continuous on x (for a given θ) and Γ is a compact subset of $\mathbb{R}^n$ the existence of a maximum and a minimum of the function is guaranteed, as well as elements in Γ that attain such values.

The result of Weierstrass' theorem is extremely useful because of its relatively mild conditions on the problem, when those conditions are not satisfied it is necessary to evaluate each problem individually, since the conditions on the theorem are only sufficient, that is, failure to meet does not rule out the existence of a solution.

To highlight the usefulness of the result I reproduce here two example from Sundaram (1996) one in which the conditions are satisfied directly, and another to which Weierstrass' theorem does not apply, but that is modified to establish the existence of a solution.

Example 10.1. (Utility Maximization)

$$v = \max_{x \in B(p,I)} u(x) \qquad B(p,I) = \left\{x \in \mathbb{R}^n_+ | p \cdot x \leq I\right\}$$

where $p \geq 0$ and $I \geq 0$ and u is a continuous function. For Weierstrass' theorem to apply and guarantee a solution we need to establish the compactness of the budget set $B(p,I)$, this can be shown for $p \gg 0$.

Note that if the agent were to spend the entire income in good i the maximum she can spend is $x_i = {}^{I}\!/_{p_i}$. Then the feasible consumption bundles are bounded above by $(\xi, \ldots, \xi)$ where

$$\xi = \max\left\{\frac{I}{p_1}, \ldots, \frac{I}{p_n}\right\}$$

(that is $x \in B(p,I)$ implies $0 \leq x \leq (\xi, \ldots, \xi)$), hence B is bounded.

To see that it is also closed note that B is the intersection of closed sets which is closed, or more directly recall that a set is closed if and only if it contains all of its limits points. Let $\left\{x^k\right\}$ be a sequence in B such that $x^k \to x$. Since $x^k \in B$ we have $x^k \geq 0$, by the squeeze theorem $x \geq 0$. Also $p \cdot x^k \leq I$, and since $x^k \to x$ we have $x_i^k \to x_i$, so $p_i x_i^k \to p_i x_i$. Then:

$$p \cdot x^k = \sum p_i x_i^k \to \sum p_i x_i = p \cdot x$$

which gives the result, by taking limits on $p \cdot x^k \leq I$.

Example 10.2. (Cost Minimization)

$$v(w,y) = \min_{x \in \Gamma(y)} w \cdot x \qquad \Gamma(y) = \left\{x \in \mathbb{R}^n_+ | g(x) \geq y\right\}$$

where $w \geq 0$ is a vector of input prices and $y \geq 0$ is a minimum level of production (or utility) and g is a continuous function. Note that the feasible set is unbounded for monotone functions g, but it is otherwise closed and the objective function is clearly continuous.

The objective now is to show that this problem is equivalent to one in which the feasible set is compact. Let $\overline{x} \in \Gamma$ and define $\overline{c} = w \cdot \overline{x}$ a given cost level. It is clear that since this cost level is attainable the optimal cost level cannot be higher that $\overline{c}$. Now define $\xi_i = \frac{2\overline{c}}{w_i}$, the firm cannot optimally use more than ξ_i units of input i since then the total cost would exceed $\overline{c}$ for sure (since the other inputs are nonnegative. Then without loss of generality the feasible set can be reduce to:

$$\bar{\Gamma} = \left\{x \in \mathbb{R}^n_+ | g(x) \geq y \quad \wedge \quad x_i \leq \xi_i\right\}$$

which is bounded, and since its the intersection of Γ with closed sets $\{x|x_i \leq \xi_i\}$ it is also closed. Then $\bar{\Gamma}$ is compact.

To prove closedness more carefully let $\left\{x^k\right\}$ be a sequence in $\overline{\Gamma}$ such that $x^k \to x$ then by continuity of g:

$$g(x) = g\left(\lim x^k\right) = \lim g\left(x^k\right) \geq y$$

and

$$x_i = \lim x_i^k \geq \xi_i$$

which proves that $x \in \overline{\Gamma}$.

11 Unconstrained optima

When the solution of a problem satisfies $G \subset \text{int}\Gamma$ the optima is called unconstrained. This is when the solution belongs to the interior of the feasible set. When this happens the characteristics of the feasible set become 'irrelevant' since locally it behaves as $\mathbb{R}^n$ (there is an open ball around the solution that is completely contained in the feasible set). This local behavior allows to characterize necessary conditions of the solution. Conditions that all elements of G must satisfy.

Proposition 11.1. (First order conditions) Suppose $x^\star \in \text{int}\Gamma \subseteq \mathbb{R}^n$ is a local maximum of f on Γ and suppose that f is differentiable at $x^\star$, then $Df(x^\star) = 0$.

Proof. The proof is elementary and goes in two steps, first establishing the result for $\mathbb{R}$ and then for an arbitrary $\mathbb{R}^n$.

Case 1. Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable at $x^\star$. For any sequence $y^k \to x^\star$ we have:

$$\lim_{k\to\infty} \frac{f(y^k) - f(x^\star)}{y^k - x^\star} = f^{'}(x^\star)$$

Consider now two sequences, $y^k \to x^\star$ and $z^k \to x^\star$ such that $y^k < x^\star$ and $z^k > x^\star$. For sufficiently large k it must be that $y^k, z^k \in B(x^\star, \epsilon)$ for ϵ such that the ball is in the interior of Γ. Since $x^\star$ is a maximizer $f(y^k) - f(x^\star) \leq 0$ and $f(z^k) - f(x^\star) \leq 0$ for large enough k. This gives:

$$\frac{f(y^k) - f(x^\star)}{y^k - x^\star} \geq 0 \geq \frac{f(z^k) - f(x^\star)}{z^k - x^\star}$$

Taking limits we get the result $f^{'}(x^\star) = 0$.

Case 2. Let $f : \mathbb{R}^n \to \mathbb{R}$ attain a local maximum at $x^\star$ and be differentiable at that point. The proof will show that $\frac{\partial f}{\partial x_i}(x^\star) = 0$, which gives $Df(x^\star) = \left[\frac{\partial f}{\partial x_1}(x^\star), \ldots, \frac{\partial f}{\partial x_n}(x^\star)\right]^{'} = 0$.
Let e_i be the unit vector in direction i and define $g(t) = f(x^\star + te_i)$. For any sequence $t_k \to 0$ we have:

$$\frac{g(t_k) - g(0)}{t_k} = \frac{f(x^\star + te_i) - f(x^\star)}{t_k}$$

and the RHS of the expression converges to $\frac{\partial f}{\partial x_i}(x^\star)$, then g is differentiable at 0 and $g^{'}(0) = \frac{\partial f}{\partial x_i}(x^\star)$.
Now note that the distance between $x^\star + te_i$ and $x^\star$ is $\|x^\star + te_i - x^\star\| = |t|$, so for sufficiently low t we have $x^\star + te_i \in \text{int}\Gamma$. Since $x^\star$ is a maximum in Γ this gives $g(t) \leq g(0)$, showing that g attains a maximum at $t = 0$.
Finally applying case 1 it must be that $g^{'}(0) = 0$, completing the proof.

□

This proposition allows to find potential optima by checking points that satisfy these first order conditions.

12 Constrained Optimization

12.1 Lagrange

When the feasible set of a problem is characterized by equality constraints the theory of Lagrange multipliers allows us to find necessary (first order) conditions for a feasible point to be an optimizer. In what follows the feasible set will have the form:

$$\Gamma = U \cap \{x \in \mathbb{R}^n | g(x) = 0\}$$

where $U \subseteq \mathbb{R}^n$ is open and $g : \mathbb{R}^n \to \mathbb{R}^k$ represents the k equality constraints of the problem.

To state the result of this section let $\rho(A)$ be the rank of a matrix A of dimension $n \times k$ and recall that the jacobian of a vector valued valued function $f : \mathbb{R}^n \to \mathbb{R}^k$ is Df, a matrix of dimensions $k \times n$, where each row is the jacobian of one of the outputs of the function.

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix} \qquad Df(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & & & \\ \frac{\partial f_k(x)}{\partial x_1} & \frac{\partial f_k(x)}{\partial x_2} & \cdots & \frac{\partial f_k(x)}{\partial x_n} \end{bmatrix}$$

Theorem 12.1. (Lagrange) Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^k$ be C^1 functions. Suppose $x^\star$ is a local optimizer of f on $\Gamma = U \cap \{x \in \mathbb{R}^n | g(x) = 0\}$. Suppose also that $Dg(x^\star)$ has rank k (that is $\rho(Dg(x^\star)) = k$). Then there exists a vector $\lambda^\star \in \mathbb{R}^k$ such that:

$$Df(x^\star) + \sum_{i=1}^{k} \lambda_i^\star Dg_i(x^\star) = 0$$

Proof. Sundaram (1996, sec. 5.6, pp. 135). □

This theorem gives necessary conditions for a (local) optimizer, but they are more complicated than those of proposition 11.1, in that now it is also necessary to find the vector of Lagrange multipliers λ. Moreover the full column rank condition on $Dg(x^\star)$ is not a trivial one, while all optimizers in an unconstrained problem would satisfy the first order condition as in proposition 11.1, there can be optimizers for which Lagrange multipliers fail to exist in a constrained problem. For this consider the following example:

Example 12.1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be given by:

$$f(x, y) = -y \qquad g(x, y) = y^3 - x^2$$

and the optimization problem to be:

$$\max f(x, y) \qquad \text{s.t. } g(x, y) = 0$$

The constraint gives $y^3 = x^2$, since $x^2 \geq 0$ it follows that $y \geq 0$, so the function attains a unique global maximum on Γ at $y = x = 0$. Yet, $Dg(x, y) = \begin{bmatrix} -2x & 3y^2 \end{bmatrix}$ and $Dg(0, 0) = \begin{bmatrix} 0 & 0 \end{bmatrix}$ so $\rho(Dg(0, 0)) = 0 < 1$. Moreover $Df(x, y) = \begin{bmatrix} 0 & -1 \end{bmatrix}$, so:

$$Df(0, 0) + \sum_{i=1}^{k} \lambda_i^\star Dg_i(0, 0) = \begin{bmatrix} 0 & -1 \end{bmatrix} \neq 0$$

for all possible λ, so the global optimum does not satisfy the Lagrange conditions.

In practice it is necessary to find the candidate solutions to the problem, for doing this it is useful to represent the problem with the maximization of a Lagrangian, a function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ of the form:

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{k} \lambda_i g_i(x)$$

The optimizers of this functions satisfy the FOC as in proposition 11.1. These conditions are:

$$D\mathcal{L}(x^\star, \lambda^\star) \quad = \quad 0$$

Note that the conditions with respect to x are the same as those in Lagrange's theorem:

$$Df(x^\star) + \sum_{i=1}^{k} \lambda_i^\star Dg_i(x^\star) = 0$$

while the conditions with respect to λ are:

$$g(x^\star) = 0$$

thus guaranteeing that $x^\star \in \Gamma$. Thus an optimizer of the Lagrangian problem is feasible and satisfies the necessary FOC of Lagrange's theorem. Moreover, the value of the maximization of the Lagrangian is the same as the value of the original problem, since at any optimizer $g(x^\star) = 0$ and thus $\mathcal{L}(x^\star, \lambda^\star) = f(x^\star)$.

As before the use of the Lagrange depends on the optimizer of f satisfying the full column rank constraint. For example, as shown in Example 12.1, there are no Lagrange multipliers that satisfy the Lagrange FOC.

Example 12.2. Consider again Example 12.1. The Lagrangian is:

$$\mathcal{L} = -y + \lambda\left(y^3 - x^2\right)$$

and the FOC are:

$$\begin{aligned} -2\lambda x &= 0 \\ -1 + 3\lambda y^2 &= 0 \\ y^3 - x^2 &= 0 \end{aligned}$$

By the second equation $\lambda \neq 0$, so for the first equation to be satisfied it must be that $x = 0$, then by the third equation $y = 0$, but this violates the second equation. The system has no solution.

Another case is when multiple solutions exist to the Lagrange FOC, but none of them are the solution to the original problem.

Example 12.3. Let $f(x, y) = 2x^3 - 3x^2$ and $g(x, y) = (3 - x)^3 - y^2$. Consider the problem of maximizing f.

The constraint implies that only $x \leq 3$ is feasible, since $y^2 \geq 0$ and by the constraint $(3 - x)^3 = y^2$. Also it can be shown that f is nonpositive for $x \in \left(-\infty, \frac{3}{2}\right]$ and strictly positive and increasing for $x > \frac{3}{2}$. Then the global maximum of f is attained at $(x, y) = (3, 0)$. The problem is that the full column rank constraint fails at this point since:

$$Dg(x, y) = \left[\begin{array}{cc} -3(3 - x)^2 & -2y \end{array}\right] \qquad Dg(3, 0) = \left[\begin{array}{cc} 0 & 0 \end{array}\right]$$

The one were to follow through with the Lagrangian the objective function would be:

$$\mathcal{L}(x, y, \lambda) = 2x^3 - 3x^2 + \lambda\left((3 - x)^3 - y^2\right)$$

with FOC:

$$\begin{aligned} 6x^2 - 6x - 3\lambda(3 - x)^2 &= 0 \\ -2\lambda y &= 0 \\ (3 - x)^3 - y^2 &= 0 \end{aligned}$$

For the second condition to be met we must have either $\lambda = 0$ or $y = 0$, buy if $y = 0$ the third equation implies $x = 3$ which violates the first equation, so it must be that $\lambda = 0$. From the first equation ether $x = 0$ or $x = 1$, the third equation gives the value of y, either $\sqrt{27}$ or $\sqrt{8}$, respectively. But neither $\left(0, \sqrt{27}\right)$ or $\left(1, \sqrt{8}\right)$ are the global optimum of the problem.

12.2 Kuhn-Tucker

When the feasible set of a problem is characterized by inequality constraints the theory of Lagrange multipliers can be extended to find necessary (first order) conditions for a feasible point to be an optimizer. In what follows the feasible set will have the form:

$$\Gamma = U \cap \{x \in \mathbb{R}^n | h(x) \geq 0\}$$

where $U \subseteq \mathbb{R}^n$ is open and $h : \mathbb{R}^n \to \mathbb{R}^k$ represents the k inequality constraints of the problem.

When dealing with inequality constraints it is necessary to distinguish between those that are slack at a point and those that bind. As is foreseeable the ones that are slack do not matter for characterizing the optimum (and thus the necessary conditions), while those that bind behave like equality constraints, and are handled in the same way as the constraints in the Lagrange problem. In what follows let $\overline{k} \leq k$ be the number of binding constraints at a point $x \in \Gamma$, that is $\overline{k}$ constraints have $h_i(x) = 0$ and $k - \overline{k}$ have $h_i(x) > 0$.

Theorem 12.2. (Kuhn-Tucker) Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ be C^1 functions. Suppose $x^\star$ is a local optimizer of f on $\Gamma = U \cap \{x \in \mathbb{R}^n | h(x) \geq 0\}$. Let $\overline{k}$ be the number of effective or binding constraints at $x^\star$ and suppose also that $Dh(x^\star)$ has rank $\overline{k}$ (that is $\rho(Dh(x^\star)) = \overline{k}$). Then there exists a vector $\lambda^\star \in \mathbb{R}^k$ such that:

$$\begin{aligned} Df(x^\star) + \sum_{i=1}^{k} \lambda_i^\star Dh_i(x^\star) &= 0 \\ \lambda_i^\star h_i(x^\star) &= 0 \\ \lambda_i^\star &\geq 0 \end{aligned}$$

Proof. Sundaram (1996, sec. 5.4, pp. 165). □

Note that the first condition is the same as the FOC of the Lagrangian before, this condition was derived for problems with equality constraints. The second set of conditions, that hold for $i = 1, \ldots, k$, are called complementary slackness conditions, they make sure that slack constraint are ignored in the FOC by assigning a multiplier of zero to them. It is not a surprise that the rank condition on the constraint's jacobian only requires the rank to be equal to the number of binding constraints.

As before the FOC can be derived from an associated Lagrangian, the difference lies in the need of the extra 'complementary slackness' conditions that replace the FOC of the Lagrangian with respect to the multipliers.

When handling problems with equality and inequality constraints it is then possible to pose the Lagrangian as before and add the complementary slackness conditions only for the inequality constraints. The rank condition has to be verified as before.

13 Convexity and Optimization

13.1 Optimization under convexity

Previous results on optimization are significantly strengthened by convexity. As is shown below, it ensures that all local optima are global optima, and under strict convexity that the global optima is unique. Convexity also makes the FOC of proposition 11.1 sufficient for an optimum.

Proposition 13.1. *Let $\Gamma \subseteq \mathbb{R}^n$ be convex and $f : \Gamma \to \mathbb{R}$ be concave on Γ.*

i. Any local maximum of f is a global maximum.

ii. The set of argmax is either empty of convex.

iii. If f is strictly concave then the set of argmax is either empty or a singleton.

Proof. The proposition is proven by parts

i. Suppose for a contradiction that $x^\star$ is a local maximum that is not a global maximum.

   There exists $r > 0$ such that for all $y \in B_r(x) \cap \Gamma$ it holds that $f(x) \geq f(y)$. There also exists $z \in \Gamma$ such that $f(z) > f(x)$. Since Γ is convex the point $\lambda x + (1-\lambda) z \in \Gamma$ for all $\lambda \in (0,1)$.
   For λ close to one it must be that $\lambda x + (1-\lambda) z \in B_r(x)$, then $f(x) \geq f(\lambda x + (1-\lambda) z)$. Finally, by concavity of f it also holds that:

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

   Joining and rearranging:

   $$f(x) \geq f(z)$$

   which is a contradiction.

ii. Let x_1 and x_2 be maximizers of f, then $f(x_1) = f(x_2) = \overline{f}$. Consider now $\lambda x_1 + (1-\lambda) x_2$ for $\lambda \in (0,1)$. By concavity:

   $$f(\lambda x_1 + (1-\lambda) x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2) = \overline{f}$$

   Since $\overline{f}$ is the global maximum of f on Γ it must be that $f(\lambda x_1 + (1-\lambda) x_2) = \overline{f}$. Then $\lambda x_1 + (1-\lambda) x_2 \in \operatorname{argmax} f$, this proves convexity of the set.

iii. Suppose f is strictly concave and that $x_1, x_2 \in \operatorname{argmax} f$ and $x_1 \neq x_2$. As before, consider $\lambda x_1 + (1-\lambda) x_2$ for $\lambda \in (0,1)$. By strict concavity:

   $$f(\lambda x_1 + (1-\lambda) x_2) > \lambda f(x_1) + (1-\lambda) f(x_2) = \overline{f}$$

   Which contradicts $\overline{f}$ being the maximum of f over Γ. Then it must be that $x_1 = x_2$.

□

Proposition 13.2. *Let $\Gamma \subseteq \mathbb{R}^n$ be convex and $f : \Gamma \to \mathbb{R}$ be concave and differentiable on Γ. x is an unconstrained maximum of f on Γ if and only if $Df(x) = 0$.*

Proof. Necessity of $Df(x) = 0$ has been shown in proposition 11.1 for any local maximum, then it also holds for a global maximum.

To establish sufficiency recall by proposition 7.9 that concavity of f implies:

$$\forall_{x,y\in\Gamma} \quad Df(x)(y-x) \geq f(y) - f(x)$$

If x is such that $Df(x) = 0$ then this condition is:

$$\forall_{y\in\Gamma} \quad f(x) \geq f(y)$$

which makes x a global maximum. □

Finally convexity has implications over the Kuhn-Tucker theorem, mainly that under some additional conditions the FOC of the theorem are also sufficient for a maximum.

Theorem 13.1. (Kuhn-Tucker under convexity) Let $f : U \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ be C^1 and concave functions where U is open and convex.

Then $x^\star$ maximizes f over $\Gamma = U \cap \{x \in \mathbb{R}^n | h(x) \geq 0\}$ if

$$\begin{aligned} Df(x^\star) + \sum_{i=1}^{k} \lambda_i^\star Dh_i(x^\star) &= 0 \\ \lambda_i^\star h_i(x^\star) &= 0 \\ \lambda_i^\star &\geq 0 \end{aligned}$$

If moreover there exists $\tilde{x} \in U$ such that $\forall_i h_i(\tilde{x}) > 0$ the above conditions are also necessary, if $x^\star$ maximizes f over Γ then the conditions are met.

Proof. Sundaram (1996, sec. 7.7, pp. 194). □

The condition that there exists a point in U such that all restrictions are slack at that point is called Slater's condition. This condition is only needed to establish necessity of the Kuhn-Tucker conditions, not sufficiency, but necessity was already proven under the column rank condition and without concaveness in Theorem 12.2. The condition on h being concave ensures that the set Γ is convex and that the associated Lagrangian is a concave function.

13.2 Optimization under quasi-convexity

Two results are presented on optimization under quasi-convexity, first that strict quasi-concave (convex) functions preserve the uniqueness of the optimizers in unrestricted problems, second additional conditions are introduced to the Kuhn-Tucker theorem for sufficiency of the FOC.

Proposition 13.3. *Suppose $f : \Gamma \to \mathbb{R}$ is strictly quasi-concave and Γ is convex. Any local maximum of f is also global and the set of argmax of f is either empty or a singleton.*

Proof. Suppose x is a local maximum of f on Γ but not a global maximum. Then there exists $r > 0$ such that $f(x) \geq f(y)$ for all $y \in B_r(x) \cap \Gamma$, there also exists $z \in \Gamma$ such that $f(x) < f(z)$. Consider now $\lambda \in (0,1)$ such that $\lambda x + (1-\lambda) z \in B_r(x)$. By hypothesis $f(x) \geq f(\lambda x + (1-\lambda) z)$, but by strict quasi-concaveness:

$$f(\lambda x + (1-\lambda) z) > \min\{f(x), f(z)\} = f(x)$$

which is a contradiction. Then if x is a local maximum it must be a global maximum too.

If the set $\operatorname{argmax} f$ is empty or a singleton the second part follows trivially, if not choose $x, y \in \operatorname{argmax} f$, by definition $f(x) = f(y) = \overline{f}$. Let $\lambda \in (0,1)$, then by strict quasi-concaveness:

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

which contradicts $\overline{f}$ being the maximum of f over Γ. This completes the proof. □

Theorem 13.2. (Kuhn-Tucker under quasi-convexity) Let $f : U \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ be C^1 and quasi-concave functions where U is open and convex.

Then $x^\star$ maximizes f over $\Gamma = U \cap \{x \in \mathbb{R}^n | h(x) \geq 0\}$ if

$$\begin{aligned} Df(x^\star) + \sum_{i=1}^{k} \lambda_i^\star Dh_i(x^\star) &= 0 \\ \lambda_i^\star h_i(x^\star) &= 0 \\ \lambda_i^\star &\geq 0 \end{aligned}$$

And at least one of the following two conditions holds:

i. $Df(x^\star) \neq 0$.

ii. f is concave.

Proof. Sundaram (1996, sec. 8.8, pp. 220). □

13.3 An example

Consider a finite horizon, discrete time, consumption-savings problem where the agent can either consume or save (invest) in capital that will be productive in the following period. The agent derives utility from consumption according to utility function u and discounts the future at a constant rate $\beta < 1$. Production only uses capital and the technology is described by a function f.

The problem of an agent endowed with k_0 units of capital is:

$$v(k_0) = \max_{\{c_t, k_{t+1}\}_{t=0}^{T}} \sum_{t=0}^{T} \beta^t u(c_t) \qquad \text{s.t. } c_t + k_{t+1} \leq f(k_t) \quad c_t, k_t \geq 0 \quad k_0 \text{ given}$$

A solution to this problem is a sequence $\{c_t, k_{t+1}\}_{t=0}^{T}$ for consumption and savings.

It is assumed that $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a C^1, strictly increasing, homogenous of degree one and concave function that satisfies:

$$f(0) = 0 \quad f^{'}(k) > 0 \quad \lim_{k \to 0} f^{'}(k) = \infty \quad \lim_{k \to \infty} f^{'}(k) = 0$$

It is also assumed that $u : \mathbb{R}_+ \to \mathbb{R}$ is a C^1, strictly increasing, strictly concave and bounded function that satisfies:

$$\lim_{c \to 0} u^{'}(c) = \infty$$

It is apparent from these conditions that no optimum of this problem satisfies $c_t + k_{t+1} < f(k_t)$. Suppose for a contradiction that there is an optimal plan $\{c_t, k_{t+1}\}_{t=0}^{T}$ such that $c_t + k_{t+1} < f(k_t)$ for some t. Then the alternative plan that increases consumption at time t so that $c_t + k_{t+1} = f(k_t)$ generates strictly higher utility at that time. Then the value of this alternative plan is strictly higher, contracting the optimality of the original plan. It follows that we can consider the alternative problem:

$$v(k_0) = \max_{\{k_{t+1}\}_{t=0}^{T}} \sum_{t=0}^{T} \beta^t u(f(k_t) - k_{t+1}) \qquad \text{s.t. } 0 \leq k_{t+1} \leq f(k_t) \quad k_0 \text{ given}$$

Note that since f is continuous and concave the set of feasible allocations

$$\Gamma(k_0) = \left\{k \in \mathbb{R}^T | 0 \leq k_{t+1} \leq f(k_t)\right\}$$

is closed, bounded and convex.

i. Boundedness is immediate from the finiteness of k_0. The maximum k_t can be is if the whole production is saved each period, then $k_t \in [0, f(k_0)]$ for all t, which implies $\Gamma(k_0) \subset [0, f(k_0)]^T$.

ii. Closedness follows from the continuity of f. Let $\{k^n\} \subset \Gamma(k_0)$ be such that $k^n \to k$, then for any t it must hold that $0 \leq k_{t+1}^n \leq f(k_t^n)$. Taking limits we get $0 \leq k_{t+1} \leq f(k_t)$ which implies $k \in \Gamma(k_0)$.

iii. Convexity follows from concaveness of f. Let $k, k^{'} \in \Gamma(k_0)$ and $\lambda \in (0,1)$, recall that k and $k^{'}$ are T-tuples of capital. It follows that:

$$0 \leq k_{t+1} \leq f(k_t) \quad \wedge \quad 0 \leq k^{'}_{t+1} \leq f\left(k^{'}_t\right)$$

Multiplying by λ and $1-\lambda$ and summing gives:

$$0 \leq \lambda k_{t+1} + (1-\lambda) k^{'}_{t+1} \leq \lambda f(k_t) + (1-\lambda) f\left(k^{'}_{t+1}\right)$$

Since f is concave $\lambda f(k_t) + (1-\lambda) f\left(k^{'}_{t+1}\right) \leq f\left(\lambda k_{t+1} + (1-\lambda) k^{'}_{t+1}\right)$ and the result follows.

Note that since u is continuously differentiable and strictly concave the function $U\left(\{k_{t+1}\}_{t=0}^{T}\right) = \sum \beta^t u(f(k_t) - k_{t+1})$ is also C^1 and strictly concave.

i. That U is C^1 follows from the composition and sum of continuous functions being continuous.

ii. Strict concaveness follows from the sum of strictly concave functions being concave and:

$$\begin{aligned} u\left(f\left(\lambda k_t + (1-\lambda) k^{'}_t\right) - \left(\lambda k_{t+1} + (1-\lambda) k^{'}_{t+1}\right)\right) &\geq u\left(\lambda (f(k_t) - k_{t+1}) + (1-\lambda)\left(f\left(k^{'}_t\right) - k^{'}_{t+1}\right)\right) \\ &> \lambda u(f(k_t) - k_{t+1}) + (1-\lambda) u\left(f\left(k^{'}_t\right) - k^{'}_{t+1}\right) \end{aligned}$$

where the first inequality follows from f being concave and u being strictly increasing, and the second one from u being strictly concave.

Then by proposition 13.1 there is only one solution to the problem, and by Theorem 13.1 the optimum is completely characterized by the Kuhn-Tucker conditions.

$$\begin{aligned} -u^{'}(f(k_t) - k_{t+1}) + \beta u^{'}(f(k_{t+1}) - k_{t+2}) f^{'}(k_{t+1}) + \lambda_t - \mu_t + \beta \mu_{t+1} f^{'}(k_{t+1}) &= 0 \\ -u^{'}(f(k_T) - k_{T+1}) + \lambda_T - \mu_T &= 0 \\ \lambda_t k_{t+1} &= 0 \\ \mu_t (f(k_t) - k_{t+1}) &= 0 \\ \lambda_t &\geq 0 \\ \mu_t &\geq 0 \end{aligned}$$

Since $f(0) = 0$ and $u^{'}(0) \to \infty$ the inequality constraints $0 \leq k_{t+1} \leq f(k_t)$ are never binding, except for k_{T+1}. The argument is similar to the one used to eliminate consumption. It follows that $\lambda_t = \mu_t = 0$ for $t < T$. Its easy to see that since U is strictly decreasing in k_{T+1} its optimal to set $k_{T+1} = 0$. This gives:

$$\begin{aligned} -u^{'}(f(k_t) - k_{t+1}) + \beta u^{'}(f(k_{t+1}) - k_{t+2}) f^{'}(k_{t+1}) &= 0 \\ -u^{'}(f(k_T) - k_{T+1}) + \lambda_T &= 0 \\ k_{T+1} &= 0 \\ \lambda_t &= 0 \\ \mu_t &= 0 \end{aligned}$$

The multipliers are irrelevant for the optimal allocations and they don't show up in the first equation. This allows to characterize the solution with the following set of equations:

$$\begin{aligned} -u^{'}\left(f\left(k_t\right)-k_{t+1}\right)+\beta u^{'}\left(f\left(k_{t+1}\right)-k_{t+2}\right)f^{'}\left(k_{t+1}\right) &= 0 \\ k_{T+1} &= 0 \end{aligned}$$

The first equation is the Euler equation of the problem and is a second order difference equation in k, its solution gives the optimal plan.

Part IV
Parametrized Optimization

The following sections deal with optimization problems that are indexed by a parameter. In general there might be parameters that index the feasible set over which a function is being maximized (minimized), or parameters that change the nature of the objective function. The problems we are interested in have the following form:

$$v(\theta) = \max_{x \in \Gamma(\theta)} f(x, \theta) \qquad\qquad G(\theta) = \{x \in \Gamma(\theta) \,|\, f(x, \theta) = v(\theta)\}$$

Where $v(\theta)$ is the optimal value of the problem given parameter θ, which can be multidimensional, and $G(\theta)$ is the set of solutions. Note that both the objective function f and the feasible set Γ are allowed to depend on θ.

The objective is to establish the properties of v and G as θ varies. In order to do so the concept of a correspondence must be introduced to deal with changes in the feasible set induced by θ. A correspondence is nothing but a set-valued function, and thus allows to interpret Γ, and later G, as functions of θ. The definition of a correspondence and its properties are dealt with in Section 14, almost all the attention is given to the generalization of continuity for correspondences, here the concepts of upper and lower hemi-continuity are introduced and studied.

The main result is then given by the Theorem of Maximum, Section 15, which is a theorem on parametric continuity of the solution to an optimization problem, stating sufficient conditions for v to be continuous as a function of θ. The results of the Theorem of Maximum are then strengthened though convexity assumptions, obtaining in the best scenario continuity of G as well. The section ends with applications of the Theorem of Maximum to Microeconomics.

All of the material follows chapter 9 of Sundaram (1996) with minor notes taken from chapter 3 of Stokey et al. (1989).

14 Correspondences

A correspondence is a generalization of a function, it is just a set valued function.

Definition 14.1. (Correspondence) Let $\Theta \subseteq \mathbb{R}^m$ and $X \subseteq \mathbb{R}^n$ be nonempty subsets of Euclidean spaces. A correspondence $\Gamma : \Theta \rightrightarrows X$ is a map that associates each element $\theta \in \Theta$ with a subset of X, that is $\Gamma(\theta) \subseteq X$. In other words the correspondence is a function $\Gamma : \Theta \to 2^X$, were 2^X is the power set of X.

As with functions a correspondence induces a graph in the set $\Theta \times X$, this set will be of importance latter and is defined now.

Definition 14.2. (Graph of a correspondences) Let $\Gamma : \Theta \rightrightarrows X$ be a correspondence. The graph of Γ is a subset of $\Theta \times X$ and is defined as:

$$\mathrm{Gr}(\Gamma) = \{(\theta, x) \mid \theta \in \Theta \wedge x \in \Gamma(\theta)\}$$

There are some properties of correspondences that is convenient to list from the beginning, these deal with topological properties of the image. The notion of continuity for correspondences is more tricky and will take most of the attention in what follows.

Definition 14.3. (Properties of correspondences) Let $\Gamma : \Theta \rightrightarrows X$ be a correspondence.

i. Γ is said to be **nonempty valued** if and only if $\Gamma(\theta) \neq \emptyset$ for all $\theta \in \Theta$.

ii. Γ is said to be **singled valued** if and only if $\Gamma(\theta)$ is a singleton for all $\theta \in \Theta$.

iii. Γ is said to be **closed valued** if and only if $\Gamma(\theta)$ is a closed set for all $\theta \in \Theta$.

iv. Γ is said to be **compact valued** if and only if $\Gamma(\theta)$ is a compact set for all $\theta \in \Theta$.

v. Γ is said to be **convex valued** if and only if $\Gamma(\theta)$ is a convex set for all $\theta \in \Theta$.

vi. Γ is said to be **closed** if an only if $\mathrm{Gr}(\Gamma)$ is a closed subset of $\Theta \times X$.

vii. Γ is said to have a **convex graph** if and only if $\mathrm{Gr}(\Gamma)$ is a convex set on $\Theta \times X$.

Remark 14.1. $\mathrm{Gr}(\Gamma)$ is closed if and only if for every $\theta \in \Theta$ and $\{\theta_n\} \subset \Theta$ such that $\theta_n \to \theta$ and every $\{x_n\} \subset X$ such that $x_n \in \Gamma(\theta_n)$ and $x_n \to x$ we have $x \in \Gamma(\theta)$.

$$\forall_\theta \forall_{\theta_n \to \theta} \forall_{x_n \in \Gamma(\theta_n)} x_n \to x \implies x \in \Gamma(\theta)$$

Remark 14.2. $\mathrm{Gr}(\Gamma)$ is convex if and only if for any $\theta, \theta' \in \Theta$ and any $x \in \Gamma(\theta)$ and $x' \in \Gamma(\theta')$ it holds that $\lambda x + (1-\lambda) x' \in \Gamma(\lambda\theta + (1-\lambda)\theta')$ for all $\lambda \in (0,1)$.

There is a relation between some of the properties listed above.

Proposition 14.1. *If $\Gamma : \Theta \rightrightarrows X$ has a closed graph then it is closed valued. If moreover X is compact then Γ is also compact valued.*

Proof. Suppose $\Gamma : \Theta \rightrightarrows X$ has a closed graph, and let $\theta \in \Theta$ be any point in the domain. Let $\{\theta_n\} \subset \Theta$ be such that $\theta_n = \theta$ and $\{x_n\}$ be such that $x_n \in \Gamma(\theta_n)$ for all n and $x_n \to x$.

Note that $\theta_n \to \theta$ by construction, then $(\theta_n, x_n) \to (\theta, x)$ and $(\theta_n, x_n) \in \mathrm{Gr}(\Gamma)$ for all n, since $\mathrm{Gr}(\Gamma)$ is closed it follows that $(\theta, x) \in \mathrm{Gr}(\Gamma)$ which is $x \in \Gamma(\theta)$, proving that $\Gamma(\theta)$ is a closed set.

If X is compact, then $\Gamma(\theta)$ is a closed subset of a compact set, hence it is also compact. $\square$

Correspondences, as functions, can be continuous but there are different ways in which the notion of continuity generalizes to correspondences. The difference arises from two different ways of generalizing the inverse image (or pre-image) of a function.

To be precise, recall that a function is continuous if and only if the pre-image of an open set is open. According to this a function $f : \Theta \to X$ is continuous at $\theta \in \Theta$ if and only if for all open set $V \subseteq X$ such that $f(\theta) \in V$ there exists an open set $U \subseteq \Theta$ such that for all $\theta' \in U$ it holds that $f\left(\theta'\right) \in V$, or $f(U) \subseteq V$.

Is the last step of the previous definition what induces the different generalizations, since, for a correspondence Γ, $\Gamma(U)$ does not have a clear interpretation, each element of U has as image a set, so the image of a set is not well defined. There are then two ways to approach this, both use different interpretations of what it means $f\left(\theta'\right) \in V$ for all $\theta' \in U$. A correspondence is u.h.c. if the image of each point is completely contained in the original open set, and is l.h.c. if the image is partially contained in the original open set. Formally:

Definition 14.4. (hemi-continuous correspondence) Let $\Gamma : \Theta \rightrightarrows X$ be a correspondence.

i. Γ is said to be upper hemi-continuous **(u.h.c.)** at a point $\theta \in \Theta$ if and only if for all open sets $V \subseteq X$ such that $\Gamma(\theta) \subseteq V$, there exists an open set $U \subseteq \Theta$ such that $\theta \in U$ and for all $\theta' \in U$ it holds that $\Gamma\left(\theta'\right) \subseteq V$.

ii. Γ is said to be lower hemi-continuous **(l.h.c.)** at a point $\theta \in \Theta$ if and only if for all open sets $V \subseteq X$ such that $\Gamma(\theta) \cap V \neq \emptyset$, there exists an open set $U \subseteq \Theta$ such that $\theta \in U$ and for all $\theta' \in U$ it holds that $\Gamma\left(\theta'\right) \cap V \neq \emptyset$.

iii. Γ is said to be continuous at a point $\theta \in \Theta$ if it is both u.h.c. an l.h.c.

The following examples are taken from Sundaram (1996). The first one shows a correspondence that is u.h.c. but not l.h.c. and the second one the opposite.

Example 14.1. Let $\Theta = X = [0, 2]$ and define $\Gamma : \Theta \rightrightarrows X$ as:

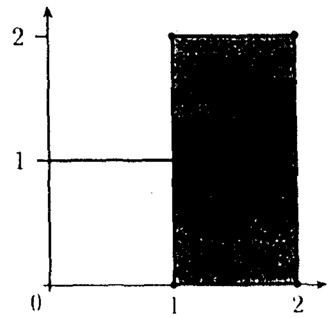
$$\Gamma(\theta) = \begin{cases} \{1\} & \theta < 1 \\ X & \theta \geq 1 \end{cases}$$

The graph of this correspondence is presented in Figure 14.1.

i. Γ is both u.h.c. and l.h.c. for $\theta < 1$.

   (a) Since Γ is single valued for $\theta < 1$ it holds that $\Gamma(\theta) \subseteq V$ if and only if $\Gamma(\theta) \cap V \neq \emptyset$ for all $\theta < 1$. Then Γ is u.h.c. if and only if Γ is l.h.c.

Figure 14.1: A u.h.c. correspondence that is not l.h.c.



(b) Γ is u.h.c. for $\theta < 1$:

Let $\theta < 1$ and $V \subseteq X$ such that $\Gamma(\theta) \subseteq V$, then $1 \in V$. To show u.h.c. we need to find $U \subseteq \Theta$ such that $\theta \in U$ and $\Gamma(\theta') \subseteq V$ for all $\theta' \in U$. Consider $\epsilon = \frac{1-\theta}{2}$ and let $U = B_\epsilon(\theta) \cap (0, \infty)$, clearly U is open, $\theta \in U$ and $\Gamma(\theta') = \{1\} \subseteq V$ for all $\theta' \in U$. This proves u.h.c.

By the remark above Γ is also l.h.c. for $\theta < 1$.

ii. Γ is both u.h.c. and l.h.c. for $\theta > 1$.

(a) Γ is u.h.c. for $\theta > 1$:

Let $\theta > 1$ and V an open set such that $\Gamma(\theta) \subseteq V$, then $X \subseteq V$ and hence $\Gamma(\theta') \subseteq V$ for all $\theta' \in \Theta$, so Γ is clearly u.h.c. since any open set U satisfies the requirement. ($\Gamma(\theta') \subseteq X \subset V$).

(b) Γ is l.h.c. for $\theta > 1$:

Let $\theta > 1$ and V an open set such that $\Gamma(\theta) \cap V \neq \emptyset$, then $V \subseteq X$. Since $\Gamma(\theta') = X$ for all $\theta > 1$ any open set U such that $U \subset (0, 2)$ satisfies the requirement. For $\theta' \in U$ it holds that $\Gamma(\theta') = X$ and $X \cap V \neq \emptyset$.

iii. Γ is u.h.c. but not l.h.c. at $\theta = 1$.

(a) Γ is u.h.c. at $\theta = 1$:

$\Gamma(1) = X$, then consider any open set $V \supset X$, clearly any open set U that contains $\theta = 1$ also satisfies $\Gamma(\theta') \subseteq X \subset V$ for $\theta' \in U$.

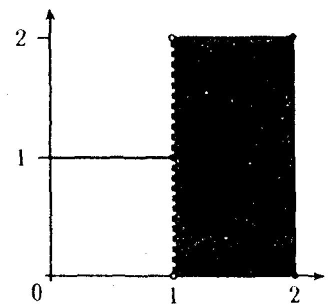
(b) Γ is not l.h.c. at $\theta = 1$:

Consider $V = (3/2, 5/2)$, clearly $\Gamma(1) = X$ and $V \cap X = V \neq \emptyset$. Yet any open set U that contains 1 also contains $\theta' < 1$, for those points $\Gamma(\theta') = 1 \notin V$, so $\Gamma(\theta') \cap V = \emptyset$, violating the condition for l.h.c.

Example 14.2. Let $\Theta = X = [0, 2]$ and define $\Gamma : \Theta \rightrightarrows X$ as:

$$\Gamma(\theta) = \begin{cases} \{1\} & \theta \leq 1 \\ X & \theta > 1 \end{cases}$$

Figure 14.2: A l.h.c. correspondence that is not u.h.c.



The graph of this correspondence is presented in Figure 14.2.

i. Γ is both u.h.c. and l.h.c. for $\theta \neq 1$. Note that for $\theta \neq 1$ this correspondence is equal to the one in the previous example.

ii. Γ is l.h.c. but not u.h.c. at $\theta = 1$.

   (a) Γ is l.h.c. at $\theta = 1$:
   $\Gamma(1) = \{1\}$, then consider any open set V such that $1 \in V$, since for all $\theta \in \Theta$ it holds that $1 \in \Gamma(\theta)$ the result follows, $\Gamma\left(\theta'\right) \cap V \neq \emptyset$ for all $\theta' \in \Theta$ and in particular for $\theta' \in U \subseteq \Theta$.

   (b) Γ is not l.h.c. at $\theta = 1$:
   Consider $V = (1/2, 3/2)$, clearly $\Gamma(1) \subseteq V$. Yet any open set U that contains 1 also contains $\theta' > 1$, for those points $\Gamma\left(\theta'\right) = X$, so $\Gamma\left(\theta'\right) \supset V$, violating the condition for u.h.c.

As with continuity in functions there is a sequential characterization of u.h.c. and l.h.c. this characterization is more useful that the definition in terms of open sets. It is in fact the way these properties are defined in Stokey et al. (1989).

Proposition 14.2. ***(Sequential characterization u.h.c.)*** *A compact valued correspondence $\Gamma : \Theta \rightrightarrows X$ is u.h.c. at $\theta \in \Theta$ if and only if for every $\{\theta_n\} \subset \Theta$ such that $\theta_n \to \theta$ and every sequence $\{x_n\} \subset X$ such that $x_n \in \Gamma(\theta_n)$ there exits a convergent subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to x \in \Gamma(\theta)$.*

$$\forall_{\theta_n \to \theta} \forall_{x_n \in \Gamma(\theta_n)} \exists_{\{x_{n_k}\}} x_{n_k} \to x \in \Gamma(\theta)$$

Proof. Sundaram (1996, sec. 9.1, pp. 231). □

Proposition 14.3. ***(Sequential characterization l.h.c.)*** *A correspondence $\Gamma : \Theta \rightrightarrows X$ is l.h.c. at $\theta \in \Theta$ if for all $x \in \Gamma(\theta)$ and all sequences $\{\theta_n\} \subset \theta$ such that $\theta_n \to \theta$ there exits a sequence $\{x_n\} \subset X$ such that $x_n \in \Gamma(\theta_n)$ and $x_n \to x$.*

$$\forall_{\theta_n \to \theta} \forall_{x \in \Gamma(\theta)} \exists_{x_n \in \Gamma(\theta_n)} x_n \to x$$

Remark 14.3. The previous proposition follows Stokey et al. (1989) and not Sundaram (1996), this is done because of two reasons. First the version given has proven to be easier to handle and more useful for economic applications and second it is the one used in Minnesota. The version given coincides with a necessary condition of l.h.c. correspondences. The sufficient condition, as stated in Sundaram (1996) asks for there to be a subsequence of $\{\theta_{n_k}\}$ for which there is a sequence $\{x_{n_k}\}$ such that $x_{n_k} \to x$.

Now we can revisit the examples given above with the sequential definition of hemi-continuity, this definition is particularly useful to disprove hemi-continuity.

Example 14.3. Let $\Theta = X = [0, 2]$ and define $\Gamma : \Theta \rightrightarrows X$ as:

$$\Gamma(\theta) = \begin{cases} \{1\} & \theta < 1 \\ X & \theta \geq 1 \end{cases}$$

The graph of this correspondence is presented in Figure 14.1.

To show that Γ is not l.h.c. at $\theta = 1$ choose $x = 2$ and let $\theta_n = 1 - {}^1\!/_n$, then $\forall_n \Gamma(\theta_n) = \{1\}$, so any sequence $x_n \in \Gamma(\theta_n)$ is of the form $x_n = 1$ for all n. Then for any sequence $x_n \to 1 \neq x$, this disproves l.h.c.

Example 14.4. Let $\Theta = X = [0, 2]$ and define $\Gamma : \Theta \rightrightarrows X$ as:

$$\Gamma(\theta) = \begin{cases} \{1\} & \theta \leq 1 \\ X & \theta > 1 \end{cases}$$

The graph of this correspondence is presented in Figure 14.2.

To show that Γ is not u.h.c. at $\theta = 1$ note that $\Gamma(\theta) = 1$ and let $\theta_n = 1 + {}^1\!/_n$, then choose $x_n = 2$ for all n, clearly $x_n \in \Gamma(\theta_n) = X$ since $\theta_n > 1$. Since $x_n \to 2$ any subsequence of $\{x_n\}$ converges to 2 as well. Then there is no subsequence that converges to $x = 1 \in \Gamma(\theta)$. This disproves u.h.c.

Finally there is a strong relation between u.h.c. and closed correspondences (those with closed graph). This relation is not an identity as the following two propositions make clear.

Proposition 14.4. ***(u.h.c and Closed graph)*** *Let $\Gamma : \Theta \rightrightarrows X$. If Γ is u.h.c, then Γ is closed (has a closed graph).*

Proof. Let Γ be u.h.c. Take $\theta \in \Theta$, $\theta_n \to \theta$ and $\{x_n\} \subset X$ such that $x_n \in \Gamma(\theta_n)$ and $x_n \to x$. Since Γ is u.h.c. there is a convergent subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \to x^{'} \in \Gamma(\theta)$. Since $x_n \to x$ it follows that $x = x^{'}$ and then $x \in \Gamma(\theta)$. Then Γ is closed. □

Proposition 14.5. ***(Closed graph and u.h.c.)*** *Let $\Gamma : \Theta \rightrightarrows X$. If X is compact and Γ is closed (has a closed graph), then Γ is u.h.c.*

Proof. Let X be compact and Γ closed. First note that this implies that Γ is compact valued (since closed graph implies closed valued). Take $\theta \in \Theta$, $\theta_n \to \theta$ and $\{x_n\} \subset X$ such that $x_n \in \Gamma(\theta_n)$. Since X is compact $\{x_n\}$ has a convergent subsequence $x_{n_k} \to x$. Since Γ is closed it follows that $x \in \Gamma(\theta)$. Then Γ is u.h.c. □

A final result is presented. It is useful for certain proofs.

Proposition 14.6. ***(Cartesian product of closed correspondences)*** *Let* $\Gamma_1 : \Theta_1 \rightrightarrows X_1$ *and* $\Gamma_2 : \Theta_2 \rightrightarrows X_2$ *be closed and define* $\Gamma : \Theta_1 \times \Theta_2 \rightrightarrows X_1 \times X_1$ *as* $\Gamma(\theta_1, \theta_2) = \Gamma_1(\theta_1) \times \Gamma(\theta_2)$. *Then* Γ *is closed.*

Proof. Let $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$, $(\theta_{1n}, \theta_{2n}) \to (\theta_1, \theta_2)$ and $\{(x_{1n}, x_{2n})\} \subset X_1 \times X_2$ such that $(x_{1n}, x_{2n}) \in \Gamma(\theta_{1n}, \theta_{2n})$ and $(x_{1n}, x_{2n}) \to (x_1, x_2)$. Then $x_{1n} \in \Gamma_1(\theta_{1n})$, $x_{2n} \in \Gamma_2(\theta_{2n})$ and $x_{1n} \to x_1$, $x_{2n} \to x_2$. Since Γ_1 and Γ_2 are closed we know that $x_1 \in \Gamma_1(\theta_1)$ and $x_2 \in \Gamma_2(\theta_2)$. This implies $(x_1, x_2) \in \Gamma(\theta_1, \theta_2)$, hence Γ is closed. $\square$

15 Theorem of the Maximum

Recall from Section 9 that a maximization problem parametrized by θ is posed as:

$$v(\theta) = \max_{x \in \Gamma(\theta)} f(x, \theta)$$

where $\theta \in \Theta$, $\Gamma : \Theta \rightrightarrows X$ is a correspondence that gives the feasible choice set given the parameter and $f : \mathbb{R}^n \times \Theta \to \mathbb{R}$ is a function. The term $v(\theta)$ is denoted as the value of the problem and is a function of the parameter θ and the solution to the problem is a correspondence $G : \Theta \rightrightarrows \mathbb{R}$ that gives the set of argmax of $f(\cdot, \theta)$ on $\Gamma(\theta)$.

$$G(\theta) = \{x \in \Gamma(\theta) \,|\, f(x, \theta) = v(\theta)\} = \underset{x \in \Gamma(\theta)}{\operatorname{argmax}} f(x, \theta)$$

The objective of this section is to characterize conditions on the primitives of the problem (f and Γ) that guarantee certain properties on the solution of the problem (v and G). The main focus is on continuity of the value and the set of optimizers, this result is given by the Theorem of the Maximum (or Maximum Theorem), the proof presented below is that of Stokey et al. (1989).

Theorem 15.1. ***(Maximum)*** *Let $\Theta \subseteq \mathbb{R}^m$ and $X \subseteq \mathbb{R}^n$, let $f : \Theta \times X \to \mathbb{R}$ be a continuous function and $\Gamma : \Theta \rightrightarrows X$ a nonempty, compact valued, continuous correspondence. Define:*

$$v(\theta) = \max_{x \in \Gamma(\theta)} f(x, \theta) \qquad\qquad G(\theta) = \{x \in \Gamma(\theta) \,|\, f(x, \theta) = v(\theta)\}$$

Then $v : \Theta \to X$ is continuous, and $G : \Theta \rightrightarrows X$ is nonempty and compact valued, and u.h.c.

Proof. The proof is divided in three parts. First it is proven that G is nonempty and compact valued, then that it is u.h.c. and finally that v is continuous.

i. G is nonempty valued and compact valued.

   (a) Let $\theta \in \Theta$, by hypothesis $\Gamma(\theta)$ is compact and nonempty. Since $f(\cdot, \theta)$ is continuous a maximum is attained on $\Gamma(\theta)$ by the extreme value theorem (Weierstrass). This proves that $G(\theta)$ is nonempty for arbitrary θ.

   (b) Let $\theta \in \Theta$, by hypothesis $\Gamma(\theta)$ is compact and nonempty. Since $G(\theta) \subseteq \Gamma(\theta)$ it follows that $G(\theta)$ is bounded, it is left to show closedness to establish compactness. Let $x_n \to x$ and $x_n \in G(\theta)$ for all n. Clearly $x_n \in \Gamma(\theta)$ for all n, since Γ is closed valued it follows that $x \in \Gamma(\theta)$, so its feasible. By definition of G we have $v(\theta) = f(x_n, \theta)$ for all n, since f is continuous we get $v(\theta) = \lim f(x_n, \theta) = f(x, \theta)$, then by definition $x \in G(\theta)$, which proves closedness.

ii. G is u.h.c.

   Consider $\theta \in \Theta$, a sequence in Θ such that $\theta_n \to \theta$ and a sequence in X such that $x_n \in G(\theta_n)$ for all n. Note that $x_n \in \Gamma(\theta_n)$. Since Γ is u.h.c. there exists a subsequence $x_{n_k} \to x \in \Gamma(\theta)$.

Now consider $z \in \Gamma(\theta)$. Since Γ is l.h.c. there exists a sequence in X such that $z_n \in \Gamma(\theta_n)$ and $z_n \to z$. In particular the subsequence $\{z_{n_k}\}$ also converges to z.

Since $x_n \in G(\theta_n)$ and $z_n \in \Gamma(\theta_n)$ it follows that $f(x_n, \theta_n) \geq f(z_n, \theta_n)$. Since f is *continuous in both arguments* we get by taking limits: $f(x, \theta) \geq f(z, \theta)$. Since the inequality holds for arbitrary $z \in \Gamma(\theta)$ we get the result: $x \in G(\theta)$. This proves u.h.c.

iii. v is continuous.

Let $\theta \in \Theta$ and $\theta_n \to \theta$ an arbitrary sequence converging to θ. Consider an arbitrary sequence in X such that $x_n \in G(\theta_n)$ for all n.

Let $\overline{v} = \limsup v(\theta_n)$. By proposition 2.9 there is a subsequence $\{\theta_{n_k}\}$ such that $v(\theta_{n_k}) \to \overline{v}$. Since G is u.h.c. there exists a subsequence of $\{x_{n_k}\}$ (call it $\left\{x_{n_{k_l}}\right\}$) converging to a point $x \in G(\theta)$. Then

$$\overline{v} = \lim v(\theta_{k_l}) = \lim f(x_{k_l}, \theta_{k_l}) = f(x, \theta) = v(\theta)$$

where the second equality follows from $x_{k_l} \in G(\theta_{k_l})$, the third one from f being continuous and the final one from $x \in G(\theta)$.

Let $\underline{v} = \liminf v(\theta_n)$ and by a similar argument we get $v(\theta) = \underline{v}$.

Since $v(\theta) = \liminf v(\theta_n) = \limsup v(\theta_n)$ we get $v(\theta) = \lim v(\theta_n)$ for arbitrary $\{\theta_n\}$ converging to θ. This proves continuity.

□

The Theorem of the Maximum (ToM) translates continuity of the primitives to u.h.c. of the solution (recall that since v is a function being u.h.c. implies continuity). In the following section application to economics are discussed but first two simpler examples (taken from Sundaram (1996)) are presented.

Example 15.1. Let $\Theta = [0, 1]$ and $X = [1, 2]$, also

$$f(x, \theta) = x^{\theta} \qquad \Gamma(\theta) = X$$

Clearly f is continuous on $\Theta \times X$ and Γ is continuous, compact and nonempty valued.

For $\theta > 0$ the function f is strictly increasing in x, then the solution is to pick the highest x possible. For $\theta = 0$ the function f is constant, then all $x \in X$ gives the same value and hence is optimal. This gives:

$$v(\theta) = \begin{cases} 1 & \text{if } \theta = 0 \\ 2^{\theta} & \text{if } \theta > 0 \end{cases} \qquad G(\theta) = \begin{cases} [1, 2] & \text{if } \theta = 0 \\ \{2\} & \text{if } \theta > 0 \end{cases}$$

Its easy to verify that v is continuous and that G is nonempty valued, compact valued and u.h.c.

Note that G fails to be l.h.c. at $\theta = 0$, choose $x = 1$ and let $\theta_n = 1 + 1/n$, then $\forall_n G(\theta_n) = \{2\}$, so any sequence $x_n \in G(\theta_n)$ is of the form $x_n = 2$ for all n. Then for any sequence $x_n \to 2 \neq x$, this disproves l.h.c.

Example 15.2. Let $\Theta = X = [0,1]$, in this example f is continuous but Γ fails to be l.h.c.

$$f(x,\theta) = x \qquad \Gamma(\theta) = \begin{cases} X & \text{if } \theta = 0 \\ \{0\} & \text{if } \theta > 0 \end{cases}$$

Clearly Γ is not l.h.c. at 0. The solution to the problem is:

$$v(\theta) = \begin{cases} 1 & \text{if } \theta = 0 \\ 0 & \text{if } \theta > 0 \end{cases} \qquad G(\theta) = \begin{cases} \{1\} & \text{if } \theta = 0 \\ \{0\} & \text{if } \theta > 0 \end{cases}$$

Note that v fails to be continuous and that G is neither u.h.c. nor l.h.c. at $\theta = 0$.

As with the optimization problems above convexity and quasi-convexity have implications over the conclusions of the ToM.

Theorem 15.2. ***(ToM under convexity)*** *Let $\Theta \subseteq \mathbb{R}^m$ and $X \subseteq \mathbb{R}^n$, let $f : \Theta \times X \to \mathbb{R}$ be a continuous function and $\Gamma : \Theta \rightrightarrows X$ a nonempty, compact valued, continuous correspondence. Define:*

$$v(\theta) = \max_{x \in \Gamma(\theta)} f(x,\theta) \qquad G(\theta) = \{x \in \Gamma(\theta) \,|\, f(x,\theta) = v(\theta)\}$$

i. If $f(\cdot,\theta)$ is concave in x for all θ and Γ is convex valued then G is convex valued.

ii. If $f(\cdot,\theta)$ is strictly concave in x for all θ and Γ is convex valued then G is single valued, hence a continuous function.

iii. If f is concave on $\Theta \times X$ and Γ has a convex graph then v is concave and G is convex valued.

iv. If f is strictly concave on $\Theta \times X$ and Γ has a convex graph then v is strictly concave and G is single valued, hence a continuous function.

Proof. Each part is proven

i. Consider $\theta \in \Theta$ and let $x_1, x_2 \in G(\theta)$ and $\lambda \in (0,1)$. Define $x = \lambda x_1 + (1-\lambda) x_2$. Since Γ is convex valued $x \in \Gamma(\theta)$. Since $f(\cdot,\theta)$ is concave we have:

$$\begin{aligned} f(x,\theta) &= f(\lambda x_1 + (1-\lambda) x_2, \theta) \\ &\geq \lambda f(x_1,\theta) + (1-\lambda) f(x_2,\theta) \\ &= \lambda v(\theta) + (1-\lambda) v(\theta) \\ &= v(\theta) \end{aligned}$$

But by the definition of v we have $v(\theta) \geq f(x,\theta)$ for all $x \in \Gamma(\theta)$, then $v(\theta) = f(x,\theta)$ which implies $x \in G(\theta)$, this proves that $G(\theta)$ is convex for arbitrary θ.

ii. The inequality above holds strictly if f is strictly concave. But $f(x,\theta) > v(\theta)$ contradicts $x_1, x_2 \in G(\theta)$ unless $x_1 = x_2$. This gives single-valuedness.

iii. Let $\theta_1, \theta_2 \in \Theta$ and define $\theta = \lambda\theta_1 + (1-\lambda)\,\theta_2$. Now choose $x_1 \in G(\theta_1)$ and $x_2 \in G(\theta_2)$ and define $x = \lambda x_1 + (1-\lambda)\,x_2$. Since Γ has a convex graph $x \in \Gamma(\theta)$. Since x is feasible but not necessarily optimal we have:

$$\begin{aligned} v(\theta) &\geq f(x,\theta) \\ &\geq \lambda f(x_1,\theta_1) + (1-\lambda)\, f(x_2,\theta_2) \\ &= \lambda v(\theta_1) + (1-\lambda)\, v(\theta_2) \end{aligned}$$

The second inequality follows from f being jointly concave. This establishes concavity of v.
G being convex valued follows as in (i).

iv. If f is jointly strictly concave the second inequality holds strictly proving strict concavity of v. G being convex valued follows as in (ii).

□

Theorem 15.3. ***(ToM under quasi-convexity)*** *Let $\Theta \subseteq \mathbb{R}^m$ and $X \subseteq \mathbb{R}^n$, let $f : \Theta \times X \to \mathbb{R}$ be a continuous function and $\Gamma : \Theta \rightrightarrows X$ a nonempty, compact valued, continuous correspondence. Define:*

$$v(\theta) = \max_{x\in\Gamma(\theta)} f(x,\theta) \qquad\qquad G(\theta) = \{x \in \Gamma(\theta) \,|\, f(x,\theta) = v(\theta)\}$$

i. If $f(\cdot,\theta)$ is quasi-concave in x for all θ and Γ is convex valued then G is convex valued.

ii. If $f(\cdot,\theta)$ is strictly quasi-concave in x for all θ and Γ is convex valued then G is single valued, hence a continuous function.

iii. If f is quasi-concave on $\Theta \times X$ and Γ has a convex graph then v is quasi-concave and G is quasi-convex valued.

iv. If f is strictly quasi-concave on $\Theta \times X$ and Γ has a convex graph then v is strictly quasi-concave and G is single valued, hence a continuous function.

Proof. The proof is an immediate modification of the proof of the Theorem.

i. Consider $\theta \in \Theta$ and let $x_1, x_2 \in G(\theta)$ and $\lambda \in (0,1)$. Define $x = \lambda x_1 + (1-\lambda)\,x_2$. Since Γ is convex valued $x \in \Gamma(\theta)$. Since $f(\cdot,\theta)$ is quasi-concave we have:

$$\begin{aligned} f(x,\theta) &= f(\lambda x_1 + (1-\lambda)\,x_2, \theta) \\ &\geq \min\{f(x_1,\theta), f(x_2,\theta)\} \\ &= v(\theta) \end{aligned}$$

But by the definition of v we have $v(\theta) \geq f(x,\theta)$ for all $x \in \Gamma(\theta)$, then $v(\theta) = f(x,\theta)$ which implies $x \in G(\theta)$, this proves that $G(\theta)$ is convex for arbitrary θ.

ii. By assumption Γ is convex valued, then proposition 13.3 applies since f is a strictly quasi-concave function maximized over a convex set. Then the set of maximizers is either empty or a singleton. Since G is nonempty valued it follows that it is a singleton for all θ.

iii. Let $\theta_1, \theta_2 \in \Theta$ and define $\theta = \lambda\theta_1 + (1-\lambda)\,\theta_2$. Now choose $x_1 \in G(\theta_1)$ and $x_2 \in G(\theta_2)$ and define $x = \lambda x_1 + (1-\lambda)\,x_2$. Since Γ has a convex graph $x \in \Gamma(\theta)$. Since x is feasible but not necessarily optimal we have:

$$\begin{aligned} v(\theta) &\geq f(x,\theta) \\ &\geq \min\{f(x_1,\theta_1), f(x_2,\theta_2)\} \\ &= \min\{v(\theta_1), v(\theta_2)\} \end{aligned}$$

The second inequality follows from f being jointly quasi-concave. This establishes concavity of v.
G being convex valued follows as in (i).

iv. If f is jointly strictly quasi-concave the second inequality holds strictly proving strict quasi-concavity of v. G being convex valued follows as in (ii).

□

Note that for the versions of the ToM under convexity or quasi-convexity the original assumptions of the Theorem are not needed, except when establishing that G is continuous (a consequences of being u.h.c and single valued). These extra versions can be then understood as separate results.

15.1 Applications

15.1.1 Budget set

One, useful, application of the material covered above is to determine properties of the budget correspondence, that indicates the feasible consumption bundles for a consumer given a price vector p and an endowment vector e. Suppose there are l goods, and that the agent has a fixed endowment of each good given by the vector $e \in \mathbb{R}^l_{++}$, the price of the goods is a vector $p \in \Delta$, where Δ is the n-dimensional open simplex. Define the budget set correspondence $B(\cdot, e) : \Delta \rightrightarrows \mathbb{R}^l_+$ by

$$B(p, e) = \left\{x \in \mathbb{R}^l_+ | p \cdot x \leq p \cdot e\right\}$$

Claim 15.1. $B(\cdot, e)$ is continuous on prices.

Proof. The claim is proved establishing u.h.c. and l.h.c. of B.

i. $B(\cdot, e)$ is upper hemi-continuous on prices.[4]

   Let $p \in \Delta$, $\{p_n\} \subset \Delta$ with $p_n \to p$ and $\{x_n\} \subset \mathbb{R}^l_+$ a sequence such that $x_n \in B(p_n, e)$.

   Since $p_n \to p \in \Delta$ there exists a closed ball, C, around p such that $C \subset \Delta$ and for n large enough $p_n \in C$. Let $\xi_i = \max\limits_{p \in C} \frac{p \cdot e}{p_i}$ for $i = 1, \ldots, l$. ξ_i is the maximum amount of x_i that can be bought in the neighborhood of p. Define $\xi = \max\{\xi_i\} + 1$, it is clear that for n large enough $x_n \in B_\xi(0)$, then $\{x_n\}$ is a bounded sequence, hence it admits a convergent subsequence $x_{n_k} \to x$.

   Since $x_{n_k} \in B(p_{n_k}, e)$ we have: $p_{n_k} \cdot x_{n_k} \leq p_{n_k} \cdot e$, since dot product is a continuous function taking limits we have $p \cdot x \leq p \cdot e$, which is $x \in B(p, e)$, proving u.h.c. of B.

ii. $B(\cdot, e)$ is lower hemi-continuous on prices.

   Let $p \in \Delta$, $\{p_n\} \subset \Delta$ with $p_n \to p$ and $x \in B(p, e)$. Define $\eta^i_n = \max\left\{0, \frac{p_n \cdot x - p_n \cdot e}{l p^i_n}\right\}$ and let $x_n = x - \eta_n$.

   Clearly $x_n \in B(p_n, e)$ since either $x \in B(p_n, e)$ or

$$p_n \cdot x_n = p_n \cdot x - \sum p^i_n \left(\frac{p_n \cdot x - p_n \cdot e}{l p^i_n}\right) = p_n \cdot x - (p_n \cdot x - p_n \cdot e) = p_n \cdot e$$

   then $p_n \cdot x_n \leq p_n \cdot e$.

   Moreover $x_n \to x$, since $x \in B(p, e)$ and $p_n \to p$ it follows that $p_n \cdot x - p_n \cdot e \to p \cdot x - p \cdot e \leq 0$, then $\eta_n = \max\{0, p_n \cdot x - p_n \cdot e\} \to 0$ which is $x_n \to x$. Then B is l.h.c.

   (a) Note that it wasn't checked if $x_n \geq 0$ for all n. This is not guaranteed by the construction above. With extra notation it can be guaranteed that $x^i_n \geq 0$.

□

[4] Note that by proposition 14.5 it is always a good idea to check if the set to which the correspondence maps is compact. If this is the case then one can establish u.h.c. by establishing closedness which is in general much easier.

15.1.2 Indirect utility and Marshalian demands

The consumer problem is often laid out without explicit endowments of the goods, instead the parameters are prices $p \in \mathbb{R}^l_{++}$ and a nominal income level $I \in \mathbb{R}_+$. The set of parameters is $\Theta = \mathbb{R}^l_{++} \times \mathbb{R}$. The indirect utility function and the Marshalian demand correspondence are:

$$v(p, I) = \max_{x \in B(p,I)} u(x) \qquad G(p, I) = \{x \in B(p, I) \,|u(x) = v(p, I)\}$$

where the budget set is given by the correspondence:

$$B(p, I) = \left\{x \in \mathbb{R}^l_+ | p \cdot x \leq I\right\}$$

I take as given that B is a nonempty, convex valued and continuous correspondence, and that u is a continuous function.

Claim 15.2. v and G have the following properties on Θ.

i. v is a continuous function on Θ and G is a nonempty, compact valued, u.h.c. correspondence.

ii. v is nondecreasing in I for fixed p and non-increasing in p for fixed I.

iii. v is jointly quasi-convex on (p, I).

iv. If u is (quasi) concave then v is (quasi) concave in I for fixed p.

v. If u is (quasi) concave then G is a convex valued correspondence.

vi. If u is strictly (quasi) concave then G is a continuous function.

Proof. Each part is proven separately.

i. Since u does not depend on p or I and it is continuous on X it is also continuous on $\Theta \times X$. Then the ToM applies, this gives the result.

ii. Let $p \geq p^{'}$ (in the vector sense) and fix $I > 0$. It is clear that $B(p, I) \subseteq B\left(p^{'}, I\right)$. Then bundle feasible at (p, I) is also feasible at $\left(p^{'}, I\right)$. It follows that $v(p, I) \leq v\left(p^{'}, I\right)$.

   Let $I \geq I^{'}$ and fix $p \geq 0$. As before $B(p, I) \supseteq B\left(p, I^{'}\right)$ which gives $v(p, I) \geq v\left(p, I^{'}\right)$.

iii. To show that v is quasi-convex in (p, I) let $p_1, p_2 \in \mathbb{R}^l_{++}$, $I_1, I_2 \in \mathbb{R}_+$ and $\lambda \in (0, 1)$. Define $p = \lambda p_1 + (1 - \lambda) p_2$ and $I = \lambda I_1 + (1 - \lambda) I_2$. Suppose for a contradiction that $v(p, I) > \max\{v(p_1, I_1), v(p_2, I_2)\}$ then there exists $x \in G(p, I) \subset B(p, I)$ such that

$$v(p, I) = u(x) > v(p_1, I_1) \geq u\left(x^{'}\right) \quad \forall_{x' \in B(p_1, I_1)}$$

   and:

$$v(p, I) = u(x) > v(p_2, I_2) \geq u\left(x^{'}\right) \quad \forall_{x' \in B(p_2, I_2)}$$

This implies $x \notin B(p_1, I_1)$ and $x \notin B(p_2, I_2)$. (or else x would have been chosen under (p_1, I_1) or (p_2, I_2)). Then:

$$p_1 \cdot x > I_1 \quad \wedge \quad p_2 \cdot x > I_2$$

But this implies, by multiplying by λ and $1 - \lambda$ and summing:

$$\begin{aligned} \lambda p_1 \cdot x + (1-\lambda) p_2 \cdot x &> \lambda I_1 + (1-\lambda) I_2 \\ p \cdot x &> I \end{aligned}$$

which contradicts $x \in B(p, I)$. Then $v(p, I) \leq \max\{v(p_1, I_1), v(p_2, I_2)\}$.

iv. To show that v is (quasi) concave in I for fixed p wee first show that $B(p, \cdot)$ has a convex graph. Let $I_1, I_2 \in \mathbb{R}_{++}$, $x_1 \in B(p, I_1)$, $x_2 \in B(p, I_2)$, $\lambda \in (0, 1)$. It follows that:

$$p \cdot x_1 \leq I_1 \quad p \cdot x_2 \leq I_2$$

Multiplying by λ and $(1 - \lambda)$ and summing, and using properties of dot product, we get:

$$p \cdot (\lambda x_1 + (1-\lambda) x_2) \leq \lambda I_1 + (1-\lambda) I_2$$

which is $\lambda x_1 + (1-\lambda) x_2 \in B(p, \lambda I_1 + (1-\lambda) I_2)$, which establishes the desired result.

Since u is (quasi) concave and B has a convex graph v is (quasi) concave by the ToM under convexity.

v. G is a convex valued correspondence.

The result follows by the ToM under (quasi) convexity, since u is concave (in particular since it does not depend on p or I) and B is convex valued (for any pair (p, I)).

vi. Since G is u.h.c. we only need it to be single valued for it to be a continuous function.

By the ToM under (quasi) convexity we get the result immediately.

$\square$

15.1.3 Nash equilibrium in normal form games

A normal form game is formed by:

i. A finite set of agents $I = \{1, \ldots, N\}$. A generic player is denoted i and the set of other players $-i$.

ii. For each player a finite action set A_i. Note $A = \times A_i$.

iii. For each player a payoff function $u^i : A \to \mathbb{R}$.

From the set of pure strategies of a player one can define the set of mixed strategies. $S_i = \Delta(A_i)$, a mixed strategy is a probability distribution over the set of possible actions A_i. Formally:

$$S_i = \Delta(A_i) = \left\{ s_i : A_i \to [0,1] \mid \sum_{a_i \in A_i} s_i(a_i) = 1 \right\}$$

Note that S_i is convex and compact. In fact S_i is the convex hull of A_i.

If players play mixed strategies they rank alternative strategies according to their expected payoffs, the expected payoffs are given by function $v^i : S_i \times S_{-i} \to \mathbb{R}$ which is:

$$\begin{aligned} v^i(t, s_{-i}) &= \sum_{a_i \in A_i} t(a_i) \left(\sum_{a_{-i} \in A_{-i}} \prod_{j \neq i} s_j(a_j) u^i(a_i, a_{-i}) \right) \\ &= \sum_{a \in A} \left(\left(t(a_i) \prod_{j \neq i} s_j(a_j) \right) u^i(a) \right) \end{aligned}$$

In a game where players play simultaneously in a noncooperative manner they have to answer optimally to a given strategy profile of the other players. The best response of a player to s_{-i} is given by:

$$\mathrm{BR}_i(s_i, s_{-i}) = \mathrm{BR}_i(s) = \left\{ t \in S_i | \forall_{r \in S_i} u^i(t, s_{-i}) \geq u^i(r, s_{-i}) \right\} = \operatorname*{argmax}_{t \in S_i} v^i(t, s_{-i})$$

Note that BR^i is the solution to the problem $V(s) = \max_{t \in S_i} v^i(t, s_{-i})$.

Since S_i is a fixed set it is also a constant correspondence with argument s, a strategy profile. It is then continuous as well as nonempty, compact and convex valued. Moreover v^i is continuous in s_{-i} and constant in s_i by construction, then it is continuous in s. v is also linear in t holding s_{-i} constant, then it is concave. It follows that the ToM under convexity applies, then the BR is a nonempty, compact and convex valued and u.h.c. correspondence for each player.

A Nash Equilibrium is defined as a strategy profile $s^\star \in S$ such that $s_i^\star \in \mathrm{BR}_i(s^\star)$ for all i. A way to think about it is to form a correspondence with the cartesian product of the individual BR correspondences, this is $\mathrm{BR} : S \to S$ defined as:

$$\mathrm{BR}(s) = \times \mathrm{BR}_i(s)$$

Note that BR is by construction a nonempty, compact and convex valued and u.h.c. correspondence.

A NE is then a fixed point of the correspondence BR. The following theorem will establish the existence of such fixed point.

Theorem 15.4. ***(Kakutani)*** *Let $S \subset \mathbb{R}^n$ be nonempty, compact and convex, and $\Gamma : S \rightrightarrows S$ be a nonempty valued, compact valued, convex valued and u.h.c. correspondence. Then Γ has a fixed point in S $(\exists_{\overline{x} \in S} \overline{x} \in \Gamma(\overline{x}))$.*

Remark 15.1. Since S is compact u.h.c is equivalent to Γ having a closed graph.

Note that S as defined above is nonempty, compact and convex, and that BR is nonempty valued, convex valued, compact valued and is u.h.c. then the Theorem applies proving the existence of a Nash Equilibrium in mixed strategies.

16 Monotonicity and Supermodularity

The material in this Section draws from chapter 10 of Sundaram (1996), but a more authoritative source is Topkis (1998). For continuity of these notes I follow Sundaram for most of the exposition and only make use of extra material as needed. Applications of these topics can be found in chapters 3 to 5 of Topkis (1998), chapter 12 of Fudenberg and Tirole (1991) and in the literature on Global Games (a good introduction to the literature can be found in Morris and Shin (2003)).

16.1 Monotonicity of the value function

While the ToM addresses the continuity of optimization problems another problem is to establish if the value of the problem and its solution are monotone in the parameters. The first question can be solved with the same tools used when addressing continuity. Establishing monotonicity of the set of optimizers is more challenging and will be dealt with later.

Theorem 16.1. ***(ToM under Monotonicity)*** *Let $\Theta \subseteq \mathbb{R}^m$ and $X \subseteq \mathbb{R}^n$, let $f : \Theta \times X \to \mathbb{R}$ be a continuous function and $\Gamma : \Theta \rightrightarrows X$ a nonempty, compact valued, continuous correspondence. Define:*

$$v(\theta) = \max_{x \in \Gamma(\theta)} f(x, \theta) \qquad\qquad G(\theta) = \{x \in \Gamma(\theta) \,|\, f(x, \theta) = v(\theta)\}$$

Let f and Γ be monotonic on θ in the sense that:

$$\theta \leq \theta^{'} \longrightarrow \Gamma(\theta) \subseteq \Gamma\left(\theta^{'}\right) \quad \wedge \quad \forall_x f(x, \theta) \leq f\left(x, \theta^{'}\right)$$

Then $v : \Theta \to X$ is continuous and non-decreasing on θ, and $G : \Theta \rightrightarrows X$ is nonempty and compact valued, and u.h.c.

Proof. Continuity of f along with the properties of G are established by the ToM. We say that v is non-decreasing on θ if:

$$\theta \leq \theta^{'} \longrightarrow v(\theta) \leq v\left(\theta^{'}\right)$$

This can be shown directly. Let $\theta \leq \theta^{'}$, then by definition, letting $x_\theta \in G(\theta)$ and $x_{\theta'} \in G\left(\theta^{'}\right)$ we have:

$$v(\theta) = f(x_\theta, \theta) \leq f\left(x_\theta, \theta^{'}\right) \leq f\left(x_{\theta'}, \theta^{'}\right) \leq v\left(\theta^{'}\right)$$

Where the second inequality follows from the monotonicity assumption on f and the second one from the monotonicity assumption on Γ, and properties of the supremum (or maximum). $\square$

Note that this result, as all the results in this section are mildly unsatisfactory since the relation $\geq$ is an incomplete relation in $\mathbb{R}^n$. Then for $\theta, \theta^{'} \in \Theta$ it can can be the case that neither $\theta \leq \theta^{'}$ or $\theta \geq \theta^{'}$. Also, note that the monotonicity assumption on Γ alone is not sufficient for the result. Consider the following example:

Example 16.1. Let $\Theta = [-1/2, 1/2]$ and $X = \mathbb{R}$. Let $f(x,\theta) = \frac{1}{\theta+1} + x$ and $\Gamma(\theta) = [-1/2, \theta]$. Clearly f is continuous in both θ and x, and $\Gamma(\theta)$ is always non-empty and compact, furthermore it is continuous, it also satisfies $\Gamma(\theta) \subseteq \Gamma\left(\theta'\right)$ for any $\theta \leq \theta'$. Yet f is strictly decreasing on θ.

Since f is strictly increasing on x for all θ, the value and solution to the problem of maximizing f subject to Γ are:

$$v(\theta) = \frac{1}{\theta+1} + \theta \qquad G(\theta) = \{\theta\}$$

It can be verified that v is strictly decreasing for $\theta < 0$ and strictly increasing for $\theta > 0$. For this consider the first derivative of v:

$$v'(\theta) = \frac{-1}{(\theta+1)^2} + 1$$

The derivative is nonnegative one for $\theta \geq 0$:

$$\begin{aligned} v'(\theta) &\geq 0 \\ 1 &\geq \frac{1}{(\theta+1)^2} \\ \theta &\geq 0 \end{aligned}$$

This proves that v is not non-decreasing on all of Θ.

On the other hand the monotonicity assumption on f is not necessary, as is made clear by the following example:

Example 16.2. Let $\Theta = [0,1]$ and $X = \mathbb{R}$. Let $f(x,\theta) = \frac{x}{\theta+1}$ and $\Gamma(\theta) = [0,\theta]$. Clearly f is continuous in both θ and x, and $\Gamma(\theta)$ is always non-empty and compact, furthermore it is continuous, it also satisfies $\Gamma(\theta) \subseteq \Gamma\left(\theta'\right)$ for any $\theta \leq \theta'$. Yet f is strictly decreasing on θ for fixed x.

Since f is strictly increasing on x for all θ, the value and solution to the problem of maximizing f subject to Γ are:

$$v(\theta) = \frac{\theta}{\theta+1} \qquad G(\theta) = \{\theta\}$$

Yet in this case v is strictly increasing for all θ. For this consider the first derivative of v:

$$v'(\theta) = \frac{1}{(\theta+1)^2} > 0$$

This proves that v is strictly increasing on all of Θ, even though f is decreasing on θ.

16.2 Lattices and supermodularity

Before stating the results on monotonicity of the solution let us define the greater than or equal relation on $\mathbb{R}^n$:

Definition 16.1. (Partial Order on $\mathbb{R}^n$) Let $x, y \in \mathbb{R}^n$ we say:

$$\begin{aligned} x = y &\quad \text{if } \forall_i \, x_i = y_i \\ x \geq y &\quad \text{if } \forall_i \, x_i \geq y_i \\ x > y &\quad \text{if } x \geq y \quad \wedge \quad x \neq y \\ x \gg y &\quad \text{if } \forall_i \, x_i > y_i \end{aligned}$$

Along with this relation we can define a certain type of sets and operators that will be used in what follows. These sets can be guaranteed to have a least and a greatest element, and the operators will help to introduce the concept of a supermodular function, these concepts are necessary for the results we seek. Supermodular functions over lattices are in fact what characterizes complementarity between inputs of a function. This complementarity is the key for proving monotonicity later.

Definition 16.2. (Lattice Operators) Let $x, y \in \mathbb{R}^n$. Lattice operators $\wedge$ and $\vee$ are defined as:

$$x \wedge y = \begin{bmatrix} \min[x_1, y_1] \\ \vdots \\ \min[x_n, y_n] \end{bmatrix} \qquad \text{and} \qquad x \vee y = \begin{bmatrix} \max[x_1, y_1] \\ \vdots \\ \max[x_n, y_n] \end{bmatrix}$$

Note that by construction $x \wedge y \leq x$ and $x \vee y \geq x$, with equality only if $x \leq y$ or $x \geq y$ respectively. In fact if $x \nleq y$ it follows that $x \wedge y < x$ and if $x \ngeq y$ then $x \vee y > x$.

We can now define a lattice as a set that is closed under the lattice operators:

Definition 16.3. (Lattice) A set $X \subseteq \mathbb{R}^n$ is a lattice is $\forall_{x,y \in X} x \wedge y \in X \quad \text{and} \quad x \vee y \in X$.

Lattices have many important properties, yet they won't be necessary for the results in the following section and are hence omitted. A detailed account of such properties is presented by Topkis (1998). One property that will be used is the fact that compact lattices admit a greatest and a least element:

Proposition 16.1. *Let X be a compact set. If X is a lattice then there exist elements $\overline{x}, \underline{x} \in X$ such that $\forall_{x \in X} \overline{x} \geq x \quad \wedge \quad \underline{x} \geq x$.*

Something that is worth mentioning is that among lattices it is possible to establish a set order:

Definition 16.4. (Strong Set Order) Let $A, B \subseteq \mathbb{R}^n$.

$$A \leq_{sso} B \iff \forall_{x \in A} \forall_{y \in B} x \wedge y \in A \quad \text{and} \quad x \vee y \in B$$

If A and B are singletons the strong set order is the usual inequality between vectors.

With this set order it is possible to define monotonicity for correspondences:

Definition 16.5. (Monotone non-decreasing or non-increasing correspondence) Let $\Theta \subseteq \mathbb{R}^m$ and $\varphi : \Theta \to \mathbb{R}^n$ a correspondence. φ is monotone non-decreasing in θ if:

$$\forall_{\theta,\theta'} \theta \leq \theta^{'} \longrightarrow \varphi\left(\theta\right) \leq_{sso} \varphi\left(\theta^{'}\right)$$

it is monotone non-increasing if instead $\varphi\left(\theta^{'}\right) \leq_{sso} \varphi\left(\theta\right)$.

Now we can define two (related) properties of functions that will be crucial for determining the monotonicity of the set of optimizers.

Definition 16.6. (Supermodular function) A function $f : S \to \mathbb{R}$ on a lattice S is supermodular on S if:

$$\forall_{s,s' \in S} f\left(s \vee s^{'}\right) + f\left(s \wedge s^{'}\right) \geq f\left(s\right) + f\left(s^{'}\right)$$

In applications $S = X \times \Theta$ and $s = (x, \theta)$.

Definition 16.7. (Non-decreasing (non-increasing) differences) Let $X \subseteq \mathbb{R}^n$, $\Theta \subseteq \mathbb{R}^m$ and $f : X \times \Theta \to \mathbb{R}$. $f\left(x,\theta\right)$ has non-decreasing differences in (x, θ) if for $x^{'} \geq x$ the difference $f\left(x^{'},\theta\right) - f\left(x,\theta\right)$ is non-decreasing in θ. It has non-increasing differences if $f\left(x^{'},\theta\right) - f\left(x,\theta\right)$ is non-increasing in θ. These conditions are equivalent to:

$$\forall_{x' \geq x} \forall_{\theta' \geq \theta} f\left(x^{'},\theta^{'}\right) - f\left(x,\theta^{'}\right) \geq (\leq) f\left(x^{'},\theta\right) - f\left(x,\theta\right)$$

The following proposition is useful to generate examples of supermodular functions, as it is to check if they are indeed supermodular. It turns out that all functions of real variable are supermodular and that supermodularity and increasing differences are equivalent for functions of two variables.

Proposition 16.2. *Let $f : S \to \mathbb{R}$ be a function:*

i. If $S \subseteq \mathbb{R}$ then f is supermodular.

ii. If $S \subseteq \mathbb{R}^n$ then f is supermodular if and only if it has non-decreasing differences.

iii. If $S = X \times \Theta$ and f is supermodular then f has non-decreasing differences.

iv. If $S = X \times \Theta$ and f is supermodular in (x, θ) then f is supermodular on x for fixed θ:

$$\forall_{x,x' \in X} \quad f\left(x \wedge x^{'},\theta\right) + f\left(x \vee x^{'},\theta\right) \geq f\left(x,\theta\right) + f\left(x^{'},\theta\right)$$

Proof.

i. Let $f : S \to \mathbb{R}$ and $s, s^{'} \in S \subseteq \mathbb{R}$, then wlog $s \leq s^{'}$, by completeness of the relation $\geq$ in the real numbers.
f is supermodular if and only if:

$$f\left(s \vee s^{'}\right) + f\left(s \wedge s^{'}\right) \geq f(s) + f\left(s^{'}\right)$$

By definition of the lattice operators we have: $s = s \wedge s$ and $s^{'} = s \vee s^{'}$, then:

$$f\left(s^{'}\right) + f(s) \geq f(s) + f\left(s^{'}\right)$$

which is always satisfied.

ii. The proof for $S \subseteq \mathbb{R}^2$ is presented below. The general proof is in Sundaram (1996, sec. 10.4, pg. 264).
Let $f : S \to \mathbb{R}$ and $S \subseteq \mathbb{R}^2$.

(a) Suppose f is supermodular and let $s = \left(x, \theta^{'}\right)$ and $s^{'} = \left(x^{'}, \theta\right)$ such that $x^{'} \geq x$ and $\theta^{'} \geq \theta$. By construction $s \vee s^{'} = \left(x^{'}, \theta^{'}\right)$ and $s \wedge s^{'} = (x, \theta)$. By supermodularity of f:

$$\begin{aligned} f\left(s \vee s^{'}\right) + f\left(s \wedge s^{'}\right) &\geq f(s) + f\left(s^{'}\right) \\ f\left(x^{'}, \theta^{'}\right) + f(x, \theta) &\geq f\left(x, \theta^{'}\right) + f\left(x^{'}, \theta\right) \end{aligned}$$

Rearranging we get:

$$f\left(x^{'}, \theta^{'}\right) - f\left(x, \theta^{'}\right) \geq f\left(x^{'}, \theta\right) - f(x, \theta)$$

which is the definition of increasing differences. Then f has increasing differences.

(b) Suppose f has increasing differences and let $s = (x, \theta)$ and $s^{'} = \left(x^{'}, \theta^{'}\right)$. Wlog $x \leq x^{'}$, by completeness of the relation $\geq$ in the real numbers. There are then two cases to check $\theta \leq \theta^{'}$ and $\theta > \theta^{'}$:

Case 1. Let $\theta \leq \theta^{'}$. Then $s \vee s^{'} = s^{'} = \left(x^{'}, \theta^{'}\right)$ and $s \wedge s^{'} = s = (x, \theta)$. In this the condition for supermodularity is satisfied seldom. f is supermodular if:

$$\begin{aligned} f\left(s \vee s^{'}\right) + f\left(s \wedge s^{'}\right) &\geq f(s) + f\left(s^{'}\right) \\ f\left(s^{'}\right) + f(s) &\geq f(s) + f\left(s^{'}\right) \end{aligned}$$

which is verified automatically.

Case 2. Let $\theta > \theta^{'}$. Then $s \vee s^{'} = \left(x^{'}, \theta\right)$ and $s \wedge s^{'} = \left(x, \theta^{'}\right)$. Since f has increasing differences we have, for $x \leq x^{'}$ and $\theta^{'} < \theta$:

$$\begin{aligned} f\left(x^{'}, \theta\right) - f(x, \theta) &\geq f\left(x^{'}, \theta^{'}\right) - f\left(x, \theta^{'}\right) \\ f\left(s \vee s^{'}\right) - f(s) &\geq f(s) - f\left(s \wedge s^{'}\right) \\ f\left(s \vee s^{'}\right) + f\left(s \wedge s^{'}\right) &\geq f(s) + f\left(s^{'}\right) \end{aligned}$$

which is the condition for supermodularity.

iii. Note that the proof of part (a) in the second part of the proposition does not use the fact that $S \subseteq \mathbb{R}^2$. The proof is valid for any $S = X \times \Theta$.

iv. This is immediate from taking the definition of supermodularity using $s = (x, \theta)$ and $s' = \left(x', \theta\right)$.

$\square$

Example 16.3. Let $X = \Theta = \mathbb{R}_+$ and $f(x, \theta) = x\theta$. We want to show that f is supermodular. We will do this in two ways. Directly and by means of the previous proposition.

i. Direct proof:
Let $(x, \theta), \left(x', \theta'\right) \in X \times \Theta$. Wlog $x \leq x'$.

(a) Suppose $\theta \leq \theta'$. Then $(x, \theta) \wedge \left(x', \theta'\right) = (x, \theta)$ and $(x, \theta) \vee \left(x', \theta'\right) = \left(x', \theta'\right)$. This implies:

$$f\left((x, \theta) \wedge \left(x', \theta'\right)\right) + f\left((x, \theta) \vee \left(x', \theta'\right)\right) = f(x, \theta) + f\left(x', \theta'\right)$$

Thus satisfying supermodularity trivially.

(b) Suppose $\theta > \theta'$. Then $(x, \theta) \wedge \left(x', \theta'\right) = \left(x, \theta'\right)$ and $(x, \theta) \vee \left(x', \theta'\right) = \left(x', \theta\right)$. This implies:

$$f\left((x, \theta) \wedge \left(x', \theta'\right)\right) + f\left((x, \theta) \vee \left(x', \theta'\right)\right) = f\left(x, \theta'\right) + f\left(x', \theta\right) = x\theta' + x'\theta$$

We want to establish that:

$$\begin{aligned} f\left((x, \theta) \wedge \left(x', \theta'\right)\right) + f\left((x, \theta) \vee \left(x', \theta'\right)\right) &\geq f(x, \theta) + f\left(x', \theta'\right) \\ x\theta' + x'\theta &\geq x\theta + x'\theta' \\ \left(x' - x\right)\left(\theta - \theta'\right) &\geq 0 \end{aligned}$$

which is verified since $x' \geq x$ and $\theta > \theta'$.

ii. Proposition:
Since f has its domain in $\mathbb{R}^2$ it suffices to show that f has non-decreasing differences. To see this let $x \leq x'$ and note:

$$f\left(x', \theta\right) - f(x, \theta) = \left(x' - x\right)\theta$$

Since $x' - x \geq 0$ this function is non-decreasing in θ. This completes the proof.

The following proposition characterizes supermodular functions that are twice continuously differentiable. Interestingly the only conditions are on the off-diagonal elements of the Hessian matrix. This is markedly different to what was found for convex and concave functions.

Proposition 16.3. *Let $f : S \to \mathbb{R}$ be twice continuously differentiable on $S \subseteq \mathbb{R}^n$ with S a lattice. f is supermodular in S if and only if*

$$\forall_{s \in S} \forall_{i \neq j} \frac{\partial^2 f}{\partial s_i \partial s_j}(s) \geq 0$$

16.3 Monotonicity of the set of optimizers

There are several results related to the monotonicity of the set of maximizers. The first one is referred to as Topkis' Theorem and establishes sufficient conditions for the solution of optimization problem to be a monotone increasing correspondence.

We first prove a simpler proposition that establishes that supermodularity translates into correspondences that are "lattice-valued":

Proposition 16.4. Consider the following maximization problem:

$$v(\theta) = \max_{x \in \Gamma(\theta)} f(x, \theta) \qquad G(\theta) = \{x \in \Gamma | f(x, \theta) = v(\theta)\}$$

If $\Gamma(\theta)$ is a compact lattice for all θ and f is continuous and supermodular in x then $G(\theta)$ is a compact lattice for all θ. Moreover $G(\theta)$ admits a greatest and a least element $\overline{x}(\theta)$ and $\underline{x}(\theta)$.

Proof. Let $\theta \in \Theta$ and $x, x^{'} \in G(\theta)$. Since f is supermodular in x we have:

$$f\left(x \wedge x^{'}, \theta\right) + f\left(x \vee x^{'}, \theta\right) \geq f(x, \theta) + f\left(x^{'}, \theta\right)$$

Since $\Gamma(\theta)$ is a lattice $x \wedge x^{'} \in \Gamma(\theta)$ and $x \vee x^{'} \in \Gamma(\theta)$. Then:

$$v(\theta) = f(x, \theta) \geq f\left(x \wedge x^{'}, \theta\right) \qquad \text{and} \qquad v(\theta) = f\left(x^{'}, \theta\right) \geq f\left(x \vee x^{'}, \theta\right)$$

Summing we have:

$$f(x, \theta) + f\left(x^{'}, \theta\right) \geq f\left(x \wedge x^{'}, \theta\right) + f\left(x \vee x^{'}, \theta\right)$$

Joining:

$$f\left(x \wedge x^{'}, \theta\right) + f\left(x \vee x^{'}, \theta\right) = f(x, \theta) + f\left(x^{'}, \theta\right) = 2v(\theta)$$

Now suppose for a contradiction that either $f\left(x \wedge x^{'}, \theta\right) \neq v(\theta)$ or $f\left(x \vee x^{'}, \theta\right) \neq v(\theta)$, since vi is defined as the sup it follows that one of them is strictly less, which contradicts the equality above. It must be then that they are both equal to $v(\theta)$. This gives: $x \wedge x^{'} \in G(\theta)$ and $x \vee x^{'} \in G(\theta)$.

Note that since $\Gamma(\theta)$ is compact and f is continuous in x it follows that G is compact. To see this is sufficient to establish that G is closed. Let $\{x_n\} \subseteq G(\theta)$ such that $x_n \to x$ then $f(x_n, \theta) \geq f(y, \theta)$ for all $y \in \Gamma(\theta)$. Since f is continuous taking limits gives $f(x, \theta) \geq f(y, \theta)$ for all $y \in \Gamma(\theta)$, then $x \in G(\theta)$.

Since $G(\theta)$ is a compact lattice it follows that it admits a greatest and a least element. $\square$

Theorem 16.2. (Topkis) Consider the following maximization problem:

$$v(\theta) = \max_{x \in \Gamma} f(x, \theta) \qquad G(\theta) = \{x \in \Gamma | f(x, \theta) = v(\theta)\}$$

If Γ is a compact lattice, f is supermodular in x and has non-decreasing (non-increasing) differences in (x, θ) then G is a monotone non-decreasing (non-increasing) correspondence.

Proof. Let $\theta, \theta \in \Theta$ and $x \in G(\theta)$ and $x^{'} \in G\left(\theta^{'}\right)$, we want to show that if $\theta \leq \theta^{'}$ then $x \wedge x^{'} \in G(\theta)$ and $x \vee x^{'} \in G\left(\theta^{'}\right)$.

Let $x \in G(\theta)$ and $x^{'} \in G\left(\theta^{'}\right)$, then:

$$\begin{aligned}
0 &\leq f\left(x^{'}, \theta^{'}\right) - f\left(x \vee x^{'}, \theta^{'}\right) && \text{optimality of } x^{'} \text{at } \theta^{'} \\
&\leq f\left(x \wedge x^{'}, \theta^{'}\right) - f\left(x^{'}, \theta^{'}\right) && \text{supermodularity in } x \\
&\leq f\left(x \wedge x^{'}, \theta\right) - f(x, \theta) && \text{increasing differences} \\
&\leq 0 && \text{optimality of } x \text{at } \theta
\end{aligned}$$

Then all this relations have to hold with equality. By the first line we get that $f\left(x \vee x^{'}, \theta^{'}\right) = f\left(x^{'}, \theta^{'}\right) = v\left(\theta^{'}\right)$ and by the last line that $f\left(x \wedge x^{'}, \theta\right) = f(x, \theta) = v(\theta)$. Then $x \vee x^{'} \in G\left(\theta^{'}\right)$ and $x \wedge x^{'} \in G(\theta)$.

Remark. Note that if f has instead non-increasing differences the proof goes without any major changes.

□

Corollary 16.1. The previous result extends to the case when $\Gamma : \Theta \rightrightarrows X$ is a correspondence provided that $\Gamma(\theta)$ is a compact lattice for all θ and Γ is monotone non-decreasing (non-increasing) in θ.

Proof. The proof of the theorem goes unchanged by noting that since $x \in G(\theta) \subseteq \Gamma(\theta)$ and $x^{'} \in G\left(\theta^{'}\right) \subseteq \Gamma\left(\theta^{'}\right)$ then by the monotonicity assumption $x \wedge x^{'} \in \Gamma(\theta)$ and $x \vee x^{'} \in \Gamma\left(\theta^{'}\right)$, ensuring that:

$$0 \leq f\left(x^{'}, \theta^{'}\right) - f\left(x \vee x^{'}, \theta^{'}\right) \qquad \text{and} \qquad f\left(x \wedge x^{'}, \theta\right) - f(x, \theta) \leq 0$$

The other two inequalities still follow from supermodularity and increasing differences. □

Finally the result of Topkis' Theorem can be strengthened:

Proposition 16.5. *Consider the maximization problem in Topkis' Theorem (Theorem 16.2) and let Γ and f satisfy the same assumptions. Then:*

i. Let $\overline{x}(\theta)$ and $\underline{x}(\theta)$ be the greatest and least elements of $G(\theta)$. They are both non-decreasing (non-increasing) functions:

$$\theta < \theta^{'} \longrightarrow \overline{x}(\theta) \leq \overline{x}\left(\theta^{'}\right) \quad \wedge \quad \underline{x}(\theta) \leq \underline{x}\left(\theta^{'}\right)$$

ii. If f satisfies strictly increasing (decreasing) differences in (x, θ) then if $\theta < \theta^{'}$ and $x \in G(\theta)$ and $x^{'} \in G\left(\theta^{'}\right)$ we have $x \leq x^{'}$.

Proof. Let $\theta < \theta^{'}$:

i. Let $x = \overline{x}(\theta)$ and $x^{'} = \overline{x}\left(\theta^{'}\right)$. Note that $x \vee x^{'} \geq x^{'}$ and that $x \vee x^{'} > x^{'}$ unless $x^{'} \geq x$. But $x \vee x^{'} > x^{'}$ contradicts $x^{'}$ being the greatest element of $G\left(\theta^{'}\right)$, since $x \vee x^{'} \in G\left(\theta^{'}\right)$ by Topkis' Theorem. This establishes monotonicity of $\overline{x}(\theta)$ since it must be that $\overline{x}(\theta) \leq \overline{x}\left(\theta^{'}\right)$ if $\theta < \theta^{'}$.
   Now let $x = \underline{x}(\theta)$ and $x^{'} = \underline{x}\left(\theta^{'}\right)$. Note that $x \wedge x^{'} \leq x$ and that $x \wedge x^{'} < x$ unless $x \leq x^{'}$. But $x \wedge x^{'} < x$ contradicts x being the least element of $G(\theta)$, since $x \wedge x^{'} \in G(\theta)$ by Topkis' Theorem. This establishes monotonicity of $\underline{x}(\theta)$ since it must be that $\underline{x}(\theta) \leq \underline{x}\left(\theta^{'}\right)$ if $\theta < \theta^{'}$.

ii. Let $x \in G(\theta)$ and $x^{'} \in G(\theta)$. Suppose for a contradiction that it does not hold that $x \leq x^{'}$. Then we have $x \wedge x^{'} < x$ and $x \vee x^{'} > x^{'}$. If f satisfies strictly increasing differences we have, since $\theta^{'} > \theta$:

$$f\left(x, \theta^{'}\right) - f\left(x \wedge x^{'}, \theta^{'}\right) > f(x, \theta) - f\left(x \wedge x^{'}, \theta\right)$$

   Note that this contradicts the sequence of equalities in Topkis' Theorem. Then it must be that $x \leq x^{'}$.

$\square$

16.4 Applications of Supermodularity

16.4.1 Spring 2009 - Q 2.1 [Werner] (Topkis - Input Taxes)

Consider a profit maximizing firm with single output and n inputs, with production function $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ assumed strictly increasing, continuous (but possibly non-differentiable), and $f(0) = 0$. Let $q \in \mathbb{R}_{++}$ be the price of output and $w \in \mathbb{R}^n_{++}$ be the vector of prices of inputs. The firm is taxed at rate $t > 0$ of its total cost. The firm's profit maximization problem is

$$\max_{x \geq 0} qf(x) - wx - t(wx)$$

. Let $x^\star(t)$ denote the profit maximizing vector of inputs (assumed unique) as function of tax rate t.

i. State a definition of production function f being supermodular. State a criterion for supermodularity of f under an additional assumption that f is twice differentiable.

   A function f is supermodular if $\forall x, y \in \mathbb{R}^n_+$

$$f(x \vee y) + f(x \wedge y) \geq f(x) + f(y)$$

   where $f(x \vee y) = f(\max(x_1, y_1), \ldots, \max(x_n, y_n))$ and $f(x \wedge y) = f(\min(x_1, y_1), \ldots, \min(x_n, y_n))$. Under an additional assumption that f is twice differentiable then a function is supermodular if and only if

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \geq 0$$

   for all $i \neq j$ and all $x \in \mathbb{R}^l_+$

ii. Show that if f is supermodular, then input demand $x^\star$ is a non-increasing function of t, that is, if $t' \geq t$, then $x^\star(t^\star) \leq x^\star(t)$. If you use a known mathematical theorem in your proof, make sure that you state that theorem clearly.

   To solve this problem we will use a corollary to Topkis' Theorem which states.

   Given a problem of the form $\max_{x \in S} F(x, y)$ with $x^*(y) = \arg\max_{x \in S} F(x, y)$ then x^* is a non-increasing function of t if

   (a) S is a lattice - The set $\mathbb{R}^n_+$ is a lattice, note that if x and y are in $\mathbb{R}^n_+$ then $x \wedge y$ and $x \vee y$ contain only non-negative elements and are thus in $\mathbb{R}^n_+$

   (b) F is supermodular in x (given y)- By assumption f is supermodular. Then note that linear functions are also supermodular because $x + y = x \wedge y + x \vee y$ since every element in both x and y is in one and only one of $x \wedge y$ and $x \vee y$. Then F is supermodular iff

$$\begin{aligned}
\begin{bmatrix} qf(x \vee y) - w(x \vee y) - t(w(x \vee y)) \\ -qf(x) + wx + t(wx) \end{bmatrix} &\geq \begin{bmatrix} qf(y) - wy - t(wy) \\ -qf(x \wedge y) + w(x \wedge y) + t(w(x \wedge y)) \end{bmatrix} \\
qf(x \vee y) - qf(x) &\geq qf(y) - qf(x \wedge y) \\
f(x \vee y) - f(x) &\geq f(y) - f(x \wedge y)
\end{aligned}$$

   which follows from the supermodularity of f. Thus F is supermodular.

(c) F has non-increasing differences in x, y - Let $x' \geq x$ and $t' \geq t$ then we want to verify

$$\begin{aligned}
qf\left(x'\right) - wx' - t'\left(wx'\right) - qf\left(x\right) + wx + t'\left(wx\right) &\leq qf\left(x'\right) - wx' - t\left(wx'\right) - qf\left(x\right) + w\left(x\right) + t \\
-t'\left(wx'\right) + t'\left(wx\right) &\leq -t\left(wx'\right) + t\left(wx\right) \\
t'\left(wx - wx'\right) &\leq t\left(wx - wx'\right) \\
t' &\geq t
\end{aligned}$$

The second to last line follows from the fact that $wx - wx'$ is a negative number because $x' \geq x$. This verifies that F has non-increasing differences in x, y.

Therefore, $x^{\star}\left(t\right)$ is a non-increasing function.

16.4.2 Fall 2009 - Q 1.1 [Werner] (Topkis - Production Price)

on function $f:\mathbb{R}_{+}^{n}\to\mathbb{R}_{+}$ assumed strictly increasing, continuous (but possibly non-differentiable), and $f(0)=0$. Let $q\in\mathbb{R}_{++}$ be the price of output and $w\in\mathbb{R}_{++}^{n}$ be the vector of prices of inputs. The firm's profit maximization problem is:

$$\max_{x\geq 0} qf(x)-w\cdot x$$

Let $x^{\star}(q)$ denote the profit maximizing vector of inputs (assumed unique) as function of output price q.

i. State a definition of production function f being supermodular. Show that the Cobb-Douglas production function $f(x)=\prod_{i=1}^{n}x_i^{\alpha_i}$, where $\alpha_i>0$ for all i, and $\sum_{i=1}^{n}\alpha_i<1$, is supermodular.

(a) Let $x,y\in\mathbb{R}_{+}^{n}$ and define $x\wedge y=(\min(x_1,y_1),\ldots,\min(x_n,y_n))$ and $x\vee y=(\max(x_1,y_1),\ldots,\max(x_n,y_n))$. f is supermodular if for every $x,y\in\mathbb{R}_{+}^{n}$: $f(x\vee y)+f(x\wedge y)\geq f(x)+f(y)$. wlog let I_x be the set of indexes of the elements for which $x_i=\max(x_i,y_i)$ and I_y be the set for indexes for which $y_i=\max(x_i,y_i)\wedge y_i\neq x_i$.

$$\begin{aligned} f(x\vee y)-f(x) &\geq f(y)-f(x\wedge y)\\ \prod_{i\in I_x}x_i^{\alpha_i}\prod_{i\in I_y}y_i^{\alpha_i}-\prod_{i=1}^{n}x_i^{\alpha_i} &\geq \prod_{i=1}^{n}y_i^{\alpha_i}-\prod_{i\in I_x}y_i^{\alpha_i}\prod_{i\in I_y}x_i^{\alpha_i} \end{aligned}$$

where the product over an empty set is defined to be equal to 1.

$$\prod_{i\in I_x}x_i^{\alpha_i}\left(\prod_{i\in I_y}y_i^{\alpha_i}-\prod_{i\in I_y}x_i^{\alpha_i}\right)\geq\prod_{i\in I_x}y_i^{\alpha_i}\left(\prod_{i\in I_y}y_i^{\alpha_i}-\prod_{i\in I_y}x_i^{\alpha_i}\right)$$

By construction $\prod_{i\in I_y}y_i^{\alpha_i}>\prod_{i\in I_y}x_i^{\alpha_i}$ then:

$$\prod_{i\in I_x}x_i^{\alpha_i}\geq\prod_{i\in I_x}y_i^{\alpha_i}$$

which is verified by construction of the set I_x. This proves supermodularity of f. Alternatively a function $f\in C^2$ is supermodular if an only if: $\frac{\partial^2 f}{\partial x_i\partial x_j}(x)\geq 0$ for all x and all $i\neq j$. In the Cobb Douglas case:

$$\frac{\partial^2 f}{\partial x_i\partial x_j}(x)=\alpha_i\alpha_j x_i^{\alpha_i-1}x_j^{\alpha_j-1}\prod_{k\neq i\neq j}x_k^{\alpha^k}\geq 0$$

the inequality is verified since $x\in\mathbb{R}_{+}^{n}$ and $\alpha_i>0$ for all i.

ii. Show that if f is supermodular, then input demand $x^{\star}(q)$ is a nondecreasing function of q. If you use a known mathematical theorem in your proof, make sure that you state that theorem clearly.

(a) Theorem (Topkis): Let $\Theta \subset \mathbb{R}^m$ and $F : \mathbb{R}_+^n \times \Theta \to \mathbb{R}$ be a function and consider the problem:

$$\max_{x \in S} F(x, \theta) \qquad x^\star(\theta) = \left\{ x \in S | F(x, \theta) = \max_{x \in S} F(x, \theta) \right\}$$

If $S \subset \mathbb{R}_+^n$ is a lattice, F is supermodular in x for fixed θ and $F(x, \theta)$ has non-decreasing differences in (x, θ) then $x^\star(\theta)$ is monotone non-decreasing in θ. If $x^\star(\theta)$ is a singleton for every θ then $x^\star$ is a non-decreasing function of θ.

(b) Note that $\mathbb{R}_+^n$ is a lattice since for all $x, y \in \mathbb{R}_+^n$ it holds that $x \wedge y \in \mathbb{R}_+^n$ and $x \vee y \in \mathbb{R}_+^n$.

(c) Let $f : \mathbb{R}_+^n \to \mathbb{R}$ be supermodular in x and define $F(x, q) = qf(x) - w \cdot x$. Note that $F(x, q)$ is supermodular in x:

$$\begin{aligned} F(x \vee y, q) - F(x, q) &\geq F(y, q) - F(x \wedge y, q) \\ qf(x \vee y) - w \cdot (x \vee y) - qf(x) + w \cdot x &\geq qf(y) - w \cdot y - qf(x \wedge y) + w \cdot (x \wedge y) \\ q(f(x \vee y) + f(x \wedge y) - f(x) - f(x)) &\geq w \cdot (x - y + x \wedge y - x \vee y) \end{aligned}$$

By definition of the $\wedge$ and $\vee$ operators $x - y + x \wedge y - x \vee y = 0$, and by supermodularity of f $f(x \vee y) + f(x \wedge y) - f(x) - f(x) \geq 0$. Since $q \in \mathbb{R}_{++}$ this verifies the inequality and completes the proof.

(d) Note that $F(x, q)$ has non-decreasing differences in (x, q). Let $x_1 \geq x_2$ and $q_1 \geq q_2$ then it must be that:

$$\begin{aligned} F(x_1, q_1) - F(x_2, q_1) &\geq F(x_1, q_2) - F(x_2, q_2) \\ q_1 f(x_1) - w \cdot x_1 - q_1 f(x_2) + w \cdot x_2 &\geq q_2 f(x_1) - w \cdot x_1 - q_2 f(x_2) + w \cdot x_2 \\ q_1 (f(x_1) - f(x_2)) &\geq q_2 (f(x_1) - f(x_2)) \end{aligned}$$

Since f is a strictly increasing function $f(x_1) - f(x_2) \geq 0$ then the inequality is verified since $q_1 \geq q_2$.

(e) Since all the requirements of Topkis' theorem are satisfied, and assuming a unique solution, we have the result: $x^\star(q_1) \geq x^\star(q_2)$ for $q_1 \geq q_2$. That is, $x^\star(q)$ is a non-decreasing function.

Part V
Dynamic Programming

The topic of the following sections is how to state and solve (deterministic) dynamic programming problems. That is how to solve a Bellman equation of the form:

$$v(x) = \sup_{y \in \Gamma(x)} \{F(x, y) + \beta v(y)\}$$

where the solution is given by a function v satisfying the equation. The final objective is to establish conditions for a solution to exist and characterize the properties of such a solution. To do this some mathematical background has to be set up, this is done in Section 17 where the contraction mapping theorem is stated and proven, and sufficient conditions for an operator to be a contraction are established.

Once the basic tools are in place the problem at hand is to express usual sequential problems (stated in terms of infinite sums) in a recursive way, the equivalence between the two representations of the problem is established by optimality principle which is presented in Section 18, along with it the conditions for existence of a solution and the properties it can inherit from the objective function F and the correspondence Γ are listed.

All the exposition of the theoretical aspects follows (very) closely Section 3.2 and all of chapter 4 of Stokey et al. (1989).[5] Most proofs are relegated to the book since their treatment would require more time than the one the course has. Finally some applications of dynamic programming theory are presented. All of the applications are taken from previous preliminary exams in Macroeconomic Theory at the University of Minnesota.

[5]There is no reason to deviate.

17 Contraction Mapping Theorem

Three results are covered in this section that will be essential for studying dynamic programming (DP) problems. These results are the contraction mapping theorem, its corollary and the Blackwell sufficiency conditions. Before stating them recall the definition of a complete metric space and of a contraction mapping (or simply contraction) in a metric space:

Definition 17.1. A metric space is a pair (S, ρ) of a set and a metric (or distance) $\rho : S \times S \to \mathbb{R}$ such that for all $x, y, z \in S$:

i. $\rho(x, y) \geq 0$ and $\rho(x, y) = 0 \iff x = y$.

ii. $\rho(x, y) = \rho(y, x)$.

iii. $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$.

A metric space is furthermore complete if all Cauchy sequences in S converge to an element in S.

Definition 17.2. Let (S, ρ) be a metric space and $T : S \to S$ a function mapping S into itself. T is a contraction (with modulus β) if and only if there exists $\beta \in (0, 1)$ such that for all $x, y \in S$:

$$\rho(Tx, Ty) \leq \beta \rho(x, y)$$

The iterates of T are the mappings $\{T^n\}$ defined by $T^0 x = x$ and $T^n x = T(T^{n-1} x)$ fro $n = 1, 2, \ldots$.

The contraction mapping theorem establishes the existence and uniqueness of a fixed point in S for any contraction mapping, moreover it provides a simple algorithm to approximate the fixed point from any arbitrary point in the space. A fixed point is a point $x \in S$ such that $x = Tx$.

Theorem 17.1. ***(Contraction Mapping Theorem)*** *Let (S, ρ) be a complete metric space and $T : S \to S$ a contraction with modulus β, then:*

i. T has exactly one fixed point $v \in S$.

ii. For any $v_0 \in S$ and $n = 0, 1, \ldots$ it holds that:

$$\rho(T^n v_0, v) \leq \beta^n \rho(v_0, v)$$

Proof. The outline of the proof is to establish that the sequence $\{v_n\} \subset S$ with $v_n = T^n v_0$ is Cauchy and the use completeness of the space to argue that its limit is the fixed point of the mapping.

Let $v_0 \in S$ and define $v_{n+1} = T v_n$ so that $v_n = T^n v_0$. Since T is a contraction mapping:

$$\rho(v_2, v_1) = \rho(T v_1, T v_0) \leq \beta \rho(v_1, v_0)$$

By induction we get:

$$\rho(v_{n+1}, v_n) \leq \beta^n \rho(v_1, v_0)$$

Then for $m > n$ we get:

$$\begin{aligned}\rho\left(v_m, v_n\right) &\leq \rho\left(v_m, v_{m-1}\right)+\rho\left(v_{m-1}, v_{m-2}\right)+\ldots+\rho\left(v_{n+1}, v_n\right) \\ &\leq \left(\beta^{m-1}+\beta^{m-2}+\ldots+\beta^n\right) \rho\left(v_1, v_0\right) \\ &= \beta^n\left(\beta^{m-n-1}+\beta^{m-n-2}+\ldots+1\right) \rho\left(v_1, v_0\right) \\ &\leq \frac{\beta^n}{1-\beta} \rho\left(v_1, v_0\right)\end{aligned}$$

Since $\frac{\rho(v_1,v_0)}{1-\beta}$ is fixed, and finite, and $\beta^n \to 0$ its clear that for any $\epsilon > 0$ there exists N large enough for $\rho\left(v_m, v_n\right) \leq \epsilon$ for all $m, n \geq N$. Then $\{v_n\}$ is Cauchy and since S is complete there exists $v \in S$ such that $v_n \to v$.

Now we show that v is a fixed point of T. For all n and v_0:

$$\begin{aligned}\rho\left(Tv, v\right) &\leq \rho\left(Tv, T^n v\right)+\rho\left(T^n v, v\right) \\ &\leq \beta\rho\left(v, T^{n-1} v\right)+\rho\left(T^n v, v\right) \\ &= \beta\rho\left(v, v_{n-1}\right)+\rho\left(v_n, v\right)\end{aligned}$$

Since $v_n \to v$ it follows that $\rho\left(v, v_{n-1}\right) \to 0$ and $\rho\left(v_n, v\right) \to 0$. Since this is done for arbitrary n we get $\rho\left(Tv, v\right) \leq \epsilon$ for all $\epsilon > 0$ which implies $\rho\left(Tv, v\right) = 0$. By definition this is $Tv = v$, a fixed point.

To show uniqueness suppose for a contradiction that there is $v^{'} \neq v$ such that $Tv^{'} = v^{'}$, then:

$$0 \neq \rho\left(v^{'}, v\right) = \rho\left(Tv^{'}, Tv\right) \leq \beta\rho\left(v^{'}, v\right)$$

But this contradicts $\beta < 1$. Then v is the unique fixed point.

The second part of the theorem follows by induction, note that

$$\rho\left(Tv_0, v\right) = \rho\left(Tv_0, Tv\right) \leq \beta\rho\left(v_0, v\right)$$

and that for any $n \geq 1$:

$$\rho\left(T^n v_0, v\right) = \rho\left(T^n v_0, Tv\right) \leq \beta\rho\left(T^{n-1} v_0, v\right)$$

The result follows. □

The contraction mapping theorem is a very powerful and simple theorem, yet its results can be strengthened by further characterizing the fixed point. So far it has been established its existence in S and its uniqueness, the following corollary to the theorem allows to locate the fixed point in a given subset of S.

Corollary 17.1. *Let (S, ρ) be a complete metric space and $T : S \to S$ a contraction mapping with fixed point $v \in S$.*

i. If $S^{'} \subseteq S$ is closed and $T\left(S^{'}\right) \subseteq S^{'}$, then $v \in S^{'}$.

ii. If in addition there exists $S^{''} \subseteq S^{'}$ such that $T\left(S^{'}\right) \subseteq S^{''}$, then $v \in S^{''}$.

Proof. Let $v_0 \in S^{'}$, note that $\{T^n v_0\}$ is a sequence in $S^{'}$ and that $T^n v_0 \to v$, since $S^{'}$ is closed it follows that $v \in S^{'}$. If in addition $T\left(S^{'}\right) \subseteq S^{''}$ then it follows that $v = Tv \in S^{''}$. □

Finally a set of sufficiency conditions are established for a mapping on the space of bounded functions to be a contraction. In most economic applications these conditions are trivial to check.

Theorem 17.2. ***(Blackwell conditions)*** *Let $X \subseteq \mathbb{R}^l$ and $B(X)$ be the the space of bounded functions on X $(f : X \to \mathbb{R})$ with the sup-norm. Let $T : B(X) \to B(X)$, T is a contraction (with modulus β) if it satisfies the following two conditions:*

i. (monotonicity) Let $f, g \in B(X)$ and $f(x) \leq g(x)$ for all $x \in X$. Then $Tf(x) \leq Tg(x)$ for all $x \in X$.

ii. (discounting) There exists $\beta \in (0,1)$ such that $T(f + a)(x) \leq Tf(x) + \beta a$ for all $f \in B(X)$, $x \in X$ and $a \geq 0$.

Proof. If $f(x) \leq g(x)$ for all x we say that $f \leq g$.

Let $f, g \in B(X)$, by definition of the sup-norm $f(x) - g(x) \leq \|f - g\|$ for all $x \in X$, then $f(x) \leq g(x) + \|f - g\|$, using the notation defined at the beginning of the proof this is $f \leq g + \|f - g\|$, where $\|f - g\| > 0$ is a scalar. Then by hypothesis we have:

$$Tf \leq T(g + \|f - g\|) \leq Tg + \beta \|f - g\| \longrightarrow Tf - Tg \leq \beta \|f - g\|$$

But it also holds that $g(x) - f(x) \leq \|f - g\|$ which implies $Tg - Tf \leq \beta \|f - g\|$. Joining we have, for all $x \in X$:

$$|Tf(x) - Tg(x)| \leq \beta \|f - g\|$$

Taking sup we get:

$$\|Tf - Tg\| \leq \beta \|f - g\|$$

which establishes that T is a contraction. □

17.0.1 Extended Blackwell conditions

I also present a modified version of Blackwell's sufficiency conditions for vector valued functions. I first define the relevant set of functions.

Proposition 17.1. *Let $X \subset \mathbb{R}^n$ and $B(X) = \left\{f | f : X \to \mathbb{R} \wedge \exists_{M_f} \forall_{x \in X} |f(x)| \leq M_f\right\}$ the set of bounded functions defined on the set X. The space $S = B(X) \times B(X)$ equipped with the norm $\|f\| = \max\{\|f_1\|_\infty, \|f_2\|_\infty\} = \max\left\{\sup_{x \in X} |f_1(x)|, \sup_{x \in X} |f_2(x)|\right\}$ is a normed vector space. It is also a metric space with the metric $\rho(f, g) = \|f - g\|$.*

Proof. The proof proceeds by showing that $\|\cdot\|$ is a norm.

i. Clearly $\|f\| \geq 0$ and if $f(x) = 0$ for all $x \in X$ then $\|f\| = 0$. Finally:

$$\begin{aligned} \|f\| &= 0 \\ \max\left\{\sup_{x\in X}|f_1(x)|, \sup_{x\in X}|f_2(x)|\right\} &= 0 \end{aligned}$$

which happens if and only if $\sup_{x\in X}|f_1(x)| = 0$ and $\sup_{x\in X}|f_2(x)| = 0$. Again, this happens if and only if $f_1(x) = f_2(x) = 0$ for all $x \in X$. That is, if $f(x) = 0$ for all $x \in X$.

ii. $\|\alpha f\| = \max\left\{\sup_{x\in X}|\alpha f_1(x)|, \sup_{x\in X}|\alpha f_2(x)|\right\} = |\alpha| \max\left\{\sup_{x\in X}|f_1(x)|, \sup_{x\in X}|f_2(x)|\right\} = |\alpha| \|f\|$

iii. Triangle Inequality:

$$\begin{aligned} \|f+g\| &= \max\left\{\sup_{x\in X}|f_1(x) - g_1(x)|, \sup_{x\in X}|f_2(x) - g_2(x)|\right\} \\ &\leq \max\left\{\left(\sup_{x\in X}|f_1(x)| + \sup_{x\in X}|g_1(x)|\right), \left(\sup_{x\in X}|f_2(x)| + \sup_{x\in X}|g_2(x)|\right)\right\} \\ &\leq \max\left\{\sup_{x\in X}|f_1(x)|, \sup_{x\in X}|f_2(x)|\right\} + \max\left\{\sup_{x\in X}|g_1(x)|, \sup_{x\in X}|g_2(x)|\right\} \\ &= \|f\| + \|g\| \end{aligned}$$

The first inequality follows from properties of the absolute value and the second one from the inequality:

$$\sup_{x\in X}|f_i(x)| + \sup_{x\in X}|g_i(x)| \leq \max\left\{\sup_{x\in X}|f_1(x)| + \sup_{x\in X}|g_1(x)|, \sup_{x\in X}|f_2(x)| + \sup_{x\in X}|g_2(x)|\right\}$$

iv. Under the above three conditions $\|f\|$ is a norm.

v. Clearly the sum and scalar product of bounded functions is bounded.

□

Proposition 17.2. *Consider (S, ρ) with $S = B(X) \times B(X)$ and $\rho(f, g) = \|f - g\|$. (S, ρ) is a complete space.*

Proof. The proof starts by showing that a Cauchy sequence in S is formed by Cauchy sequences in $B(X)$. Then the completeness of $B(X)$ is used to establish the result.

i. Let $\{f_n\} \subset S$ be a Cauchy sequence and $\epsilon > 0$. There exists N such that $\forall_{n,m>N} \|f_n - f_m\| < \epsilon$ which is:

$$\max\left\{\sup_{x\in X}|f_{1n}(x) - f_{1m}(x)|, \sup_{x\in X}|f_{2n}(x) - f_{2m}(x)|\right\} < \epsilon$$

$$\sup_{x\in X}|f_{1n}(x) - f_{1m}(x)| < \epsilon \qquad \wedge \qquad \sup_{x\in X}|f_{2n}(x) - f_{2m}(x)| < \epsilon$$

This implies that the sequences $\{f_{1n}\} \subset B(X)$ and $\{f_{2n}\} \subset B(X)$ are Cauchy with respect to the sup-norm $\|f\|_\infty = \sup_{x\in X}|f(x)|$.

ii. Since the space $(B(X), \|f\|_\infty)$ is complete the above implies that the sequences $\{f_{1n}\}$ and $\{f_{2n}\}$ are convergent in the sup norm. $\exists_{f_1,f_2 \in B(X)} f_{1n} \to f_1 \wedge f_{2n} \to f_2$. Denote $f : S \to \mathbb{R}^2$ as $f(x) = [f_1(x)\ \ f_2(x)]^{'}$.

iii. Let $\epsilon > 0$. By convergence of $\{f_{1n}\}$ and $\{f_{2n}\}$ there exist numbers N_1 and N_2 such that:

$$\forall_{n \geq N_1} \|f_{1n} - f_1\|_\infty < \epsilon \qquad \wedge \qquad \forall_{n \geq N_2} \|f_{2n} - f_2\|_\infty < \epsilon$$

Then for $N = \max\{N_1, N_2\}$ it holds that:

$$\forall_{n \geq N} \|f_{1n} - f_1\|_\infty < \epsilon \qquad \wedge \qquad \|f_{2n} - f_2\|_\infty < \epsilon$$

which is:

$$\forall_{n \geq N} \sup_{x \in X} |f_{1n}(x) - f_1(x)| < \epsilon \qquad \wedge \qquad \sup_{x \in X} |f_{2n}(x) - f_2(x)| < \epsilon$$

implying then:

$$\forall_{n \geq N} \max\left\{\sup_{x \in X} |f_{1n}(x) - f_1(x)|, \sup_{x \in X} |f_{2n}(x) - f_2(x)|\right\} < \epsilon$$

which is:

$$\forall_{n \geq N} \|f_n - f\| < \epsilon$$

iv. The above proves that a Cauchy sequence converges on S over the given norm.

$\square$

Theorem 17.3. ***(Extended Blackwell)*** *Consider (S, ρ) with $S = B(X) \times B(X)$ and $\rho(f, g) = \|f - g\|$. Let $T : S \to S$ be an operator satisfying*

i. (Monotonicity) $f, g \in S$ and $f(x) \leq g(x)$, for all $x \in X$, implies $Tf(x) \leq Tg(x)$, for all $x \in X$, (where $f(x) \leq g(x)$ is taken in the vector sense, i.e. $f_1(x) \leq g_1(x)$ and $f_2(x) \leq g_2(x)$).

ii. (Discounting) there exists some $\beta \in (0, 1)$ such that $T(f + A) \leq Tf(x) + \beta A$ for $f \in S$, $A = [a\ a]^{'} \in \mathbb{R}^2_+$ and $x \in X$.

Then T is a contraction in S with modulus β.

Proof. The proof follows closely that of Blackwell's conditions

i. Let $f, g \in S$, and define $A = \left[\ \|f - g\| \quad \|f - g\|\ \right]^{'}$, it holds that:

$$\begin{aligned} f_1(x) - g_1(x) &\leq |f_1(x) - g_1(x)| \leq \sup_{x \in X} |f_1(x) - g_1(x)| \\ &\leq \max\left\{\sup_{x \in X} |f_1(x) - g_1(x)|, \sup_{x \in X} |f_2(x) - g_2(x)|\right\} = \|f - g\|_\infty \end{aligned}$$

By a similar argument $f_2(x) - g_2(x) \leq \|f - g\|_\infty$ then it holds that: $f(x) \leq g(x) + A$ for all $x \in X$

ii. By monotonicity and discounting:

$$Tf(x) \leq T(g+a)(x) \leq Tg(x) + \beta A$$

which holds for all $x \in X$.

iii. The same argument applies to show that $g(x) \leq f(x) + A$ and $Tg(x) \leq Tf(x) + \beta A$ for all $x \in X$.

iv. Joining:

$$Tf_i(x) - Tg_i(x) \leq \beta \|f-g\| \qquad \wedge \qquad Tg_i(x) - Tf_i(x) \leq \beta \|f-g\|$$

which implies:

$$|Tf_1(x) - Tg_1(x)| \leq \beta \|f-g\| \qquad \wedge \qquad |Tf_2(x) - Tg_2(x)| \leq \beta \|f-g\|$$

and then:

$$\sup_{x\in X} |Tf_1(x) - Tg_1(x)| \leq \beta \|f-g\| \qquad \wedge \qquad \sup_{x\in X} |Tf_2(x) - Tg_2(x)| \leq \beta \|f-g\|$$

v. Finally:

$$\begin{aligned} \|Tf - Tg\|_\infty &= \max\left\{\sup_{x\in X} |Tf_1(x) - Tg_1(x)|, \sup_{x\in X} |Tf_2(x) - Tg_2(x)|\right\} \\ &\leq \max\{\beta\|f-g\|, \beta\|f-g\|\} = \beta\|f-g\| \end{aligned}$$

This is the definition of T being a contraction with modulus β.

$\square$

18 The Bellman Equation

18.1 The neoclassical growth model

Recall from section 13.3 the finite horizon consumption savings model. The infinite horizon version of that model is the workhorse of modern macroeconomics and is known as the neoclassical growth model. There are two (related) ways of setting up the problem. One resembles the finite horizon problem already discussed, it is called sequence problem, the other form is to cast the problem as the solution to a functional equation, this dynamic programming approach has several advantages that will be presented in the next section.

As before, consider a discrete time, consumption-savings problem where the agent can either consume or save (invest) in capital that will be productive in the following period. The agent derives utility from consumption according to utility function u and discounts the future at a constant rate $\beta < 1$. Production only uses capital and the technology is described by a function f.

The problem of an agent endowed with k_0 units of capital is:

$$v(k_0) = \max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t) \qquad \text{s.t. } c_t + k_{t+1} \leq f(k_t) \quad c_t, k_t \geq 0 \quad k_0 \text{ given}$$

Provided that u is strictly increasing, a sustained assumption, we can eliminate consumption as before to get:

$$v(k_0) = \max_{\{k_{t+1}\}} \sum_{t=0}^{\infty} \beta^t u(f(k_t) - k_{t+1}) \qquad \text{s.t. } 0 \leq k_{t+1} \leq f(k_t) \quad k_0 \text{ given}$$

In the sequence problem, much like in the finite horizon problem before, the objective is to look for an infinite sequence that solves the problem and attains the maximum. This can prove to be too difficult in practice.

The dynamic programming problem takes a different approach. Instead of trying to solve the problem for all periods simultaneously the objective is to solve the problem one period at a time. That is, given the capital stock at the beginning of the period take an optimal investment decision for the next period. The problem is that, in order to make the decision, its necessary to know the extra value for the agent of the capital to be saved, we need a function that represents preferences over next period's capital.

The DP starts by assuming that we already know such a function. It is called a value function and is defined as v above. The value function is the maximum value given to the agent if she starts in a given period with initial capital k. Knowing v it is possible to cast the following problem:

$$\max_{0 \leq k_1 \leq f(k_0)} \{u(f(k_0) - k_1) + \beta v(k_1)\}$$

If we knew v the problem above could be solved. The solution to the problem is a policy function $g : \mathbb{R}_+ \to \mathbb{R}_+$ that gives the optimal capital next period given a capital level today. That is $k_1 = g(k_0)$.

It should be clear now that if $v(k_1)$ gives the maximum value starting in period 1 and the problem above maximizes that value and the value in period 0 (given by $u(f(k_0) - k_1)$) then the value of the whole problem is given by the maximum above. But that is the definition of v, then:

$$v\left(k_0\right)=\max_{0\leq k_1\leq f\left(k_0\right)}\left\{u\left(f\left(k_0\right)-k_1\right)+\beta v\left(k_1\right)\right\}$$

This is a functional equation, note that f and u are known functions, k_1 is a variable of choice and k_0 is given. Then this is an equation in the function v, the solution to this equation is the value function needed to solve the problem (to find the policy function).

In general solving functional equations is not easy, but this type of functional equation can be reinterpreted to both establish the existence of a solution and to obtain a method to find it.

Let u and f be bounded and continuous functions and define an operator $T: C\left(X\right)\rightarrow C\left(X\right)$ as:

$$Tv\left(k\right)=\max_{0\leq k^{'}\leq f\left(k\right)}\left\{u\left(f\left(k\right)-k^{'}\right)+\beta v\left(k^{'}\right)\right\}$$

boundedness of Tv is immediate for the sum of bounded is also bounded. Continuity is a consequence of the ToM, the objective function is continuous and since f is continuous and bounded the correspondence $\Gamma\left(k\right)=\left\{k^{'}|0\leq k^{'}\leq f\left(k\right)\right\}$ is continuous and compact valued.

Note that v, the solution to the functional equation is then a fixed point of the mapping T. It is left to verify that T is a contraction to establish the existence and uniqueness of the solution to the neoclassical growth model. It turns out that Blackwell's sufficient conditions are immediate:

i. (monotonicity) Let $v,w\in C\left(X\right)$ and $v\left(k\right)\leq w\left(k\right)$ for all k. Then:

$$Tv\left(k\right)=\max_{0\leq k^{'}\leq f\left(k\right)}\left\{u\left(f\left(k\right)-k^{'}\right)+\beta v\left(k^{'}\right)\right\}\leq\max_{0\leq k^{'}\leq f\left(k\right)}\left\{u\left(f\left(k\right)-k^{'}\right)+\beta w\left(k^{'}\right)\right\}=Tw\left(k\right)$$

ii. (discounting) Let $v\in C\left(X\right)$ and $a>0$. Then:

$$T\left(v+a\right)\left(k\right)=\max_{0\leq k^{'}\leq f\left(k\right)}\left\{u\left(f\left(k\right)-k^{'}\right)+\beta\left(v\left(k^{'}\right)+a\right)\right\}=\max_{0\leq k^{'}\leq f\left(k\right)}\left\{u\left(f\left(k\right)-k^{'}\right)+\beta v\left(k^{'}\right)\right\}$$

In particular:

$$T\left(v+a\right)\left(k\right)\leq Tv\left(k\right)+\beta a$$

It is possible to further characterize v and the policy function g, for that extra results are needed.

18.2 A general framework and the principle of optimality

The problem to be studied in terms of infinite sequences is of the form:

$$v^{\star}(x_0) = \sup_{\{x_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t F(x_t, x_{t+1}) \qquad \text{s.t. } x_{t+1} \in \Gamma(x_t) \tag{18.1}$$

Corresponding to this problem is the following functional equation:

$$v(x) = \sup_{y \in \Gamma(x)} \{F(x, y) + \beta v(y)\} \tag{18.2}$$

Above X is the set of possible values for x, note that X is not necessarily an euclidean space, $\Gamma : X \rightrightarrows X$ is a correspondence that assigns feasible values of the choice variable and $F : \mathrm{Gr}(\Gamma) \to \mathbb{R}$ is a return or payoff function. $\beta > 0$ is a discount factor.

Some conditions have to be met for both problems to give the same solution, in the sense that $v(x) = v^{\star}(x)$ and that the optimal choice of one problem is the the same as the choice for the other. This equivalence between both problems is called the *principle of optimality*. After the validity of the principle has been established the properties of the solution to FE can be studied.

The conditions for the principle of optimality are stated below and the two propositions that constitute the principle are shown without proof.

It will be convenient to define the set of all possible feasible sequences for x, given an starting point x_0.

Definition 18.1. The set of all possible feasible sequences starting at $x_0 \in X$ is:

$$\Pi(x_0) = \{\{x_t\}_{t=0}^{\infty} | x_{t+1} \in \Gamma(x_t) \quad \wedge \quad x_0 \text{ given}\}$$

and $\underline{x} = (x_0, x_1, x_2, \ldots)$ is an element.

Assumption A.1: Γ is a nonempty valued correspondence.

Assumption A.2: For all $x_0 \in X$ and $\underline{x} \in \Pi(x_0)$ the following limit exists (although it might be infinite):

$$\lim_{n \to \infty} \sum_{t=0}^{n} \beta^t F(x_t, x_{t+1})$$

Remark. Assumption A.2 holds if F is bounded and $\beta \in (0, 1)$.

Note that under assumptions A.1 and A.2 $\Pi(x_0)$ is nonempty valued and problem (18.1) is well posed, moreover they are enough to guarantee that the function $v^{\star}$ satisfies equation (18.2).

Proposition 18.1. *Let X, Γ, F and β satisfy assumption A.1 and A.2, then $v^{\star}$ is a solution to the FE (18.2):*

$$v^{\star}(x) = \sup_{y \in \Gamma(x)} \{F(x, y) + \beta v^{\star}(y)\}$$

For $v^\star$ to be the only solution to the FE an extra condition is needed.

Proposition 18.2. *Let X, Γ, F and β satisfy assumption A.1 and A.2, if v is a solution to the FE (18.2) and for all $x_0 \in X$ and $\underline{x} \in \Pi(x_0)$ it holds that:*

$$\lim_{n\to\infty} \beta^n v(x_n) = 0$$

then $v = v^\star$.

The previous two propositions establish equivalence between the value of the two problems. It can also be shown that the optimizer of the SP problem also solves the FE in the following sense:

Proposition 18.3. *Let X, Γ, F and β satisfy assumption A.1 and A.2. Let $\underline{x}^\star \in \Pi(x_0)$ be a feasible plan that attains the supremum in (18.1), then:*

$$v^\star(x_t^\star) = F\left(x_t^\star, x_{t+1}^\star\right) + \beta v^\star\left(x_{t+1}^\star\right) \tag{18.3}$$

Again, under an extra boundedness condition a plan that solves the problem in (18.2) also solves the problem in the SP.

Proposition 18.4. *Let X, Γ, F and β satisfy assumption A.1 and A.2. Let $\underline{x}^\star \in \Pi(x_0)$ be a feasible plan that satisfies equation (18.3) and for which $\limsup \beta^t v^\star(x_t^\star) \le 0$, then $\underline{x}^\star$ attains the supremum in (18.1) for initial state x_0.*

Now we can define the optimal policy correspondence as:

$$G^\star(x) = \{y \in \Gamma(x) \,|\, v^\star(x) = F(x,y) + \beta v^\star(y)\}$$

We say that a plan $\underline{x}$ is generated by G if it satisfies $x_{t+1} \in G(x_t)$. The previous two propositions imply that any optimal plan of the sequence problem is generated by $G^\star$ and that if a plan is generated by $G^\star$ and satisfies the additional boundedness condition then it is also optimal.

Now we can concentrate in studying the properties of the DP in (18.2).

18.3 Bounded problems

Now we concentrate in establishing properties of the solution to the following problem:

$$v(x) = \max_{y \in \Gamma(x)} \{F(x,y) + \beta v(y)\} \tag{18.4}$$

$$G(x) = \{y \in \Gamma(x) \,|\, v(x) = F(x,y) + \beta v(y)\}$$

where v is the value function and G the policy correspondence.

Assumptions A.1 and A.2 have to be met for the implications of this sections to be valid on the original sequence problem. Additional assumptions are also imposed that ensure that the previous ones are met.

Assumption A.3: X is a a convex subset of $\mathbb{R}^l$ and Γ is a nonempty, compact valued and continuous correspondence.

Assumption A.4: The function $F : \mathrm{Gr}(\Gamma) \to \mathbb{R}$ is bounded and continuous and $\beta \in (0,1)$.

Since F is bounded and continuous it is natural to think that the solution to equation (18.4) lies in the set $C(X)$. What follows it to establish the existence of a solution by means of the contraction mapping theorem.

Define a mapping $T : C(X) \to C(X)$ as:

$$Tf(x) = \max_{y \in \Gamma(x)} \{F(x,y) + \beta f(y)\} \tag{18.5}$$

The solution to (18.4) is then a $v \in C(X)$ such that $v = Tv$. The following proposition establishes that T is a contraction from $C(X)$ into itself and also some properties of the policy correspondence G.

Proposition 18.5. *Let X, Γ, F and β satisfy assumption A.3 and A.4, and consider $C(X)$ the space of continuous bounded function on X along with the sup norm. Then:*

i. *T defined in (18.5) maps $C(X)$ into itself.*

ii. *T defined in (18.5) has a unique fixed point $v \in C(X)$, and for all $v_0 \in C(X)$*

$$\|T^n v_0 - v\| \leq \beta^n \|v_0 - v\|$$

iii. *Given v the optimal policy correspondence $G(x) = \{y \in \Gamma(x) \,|\, v(x) = F(x,y) + \beta v(y)\}$ is nonempty, compact valued and u.h.c.*

Proof. Each part is established separately.

i. Under A.3 and A.4 and given f continuous and bounded the function $F(x,y) + \beta f(y)$ is continuous in (x,y) and Γ satisfies all assumptions of the ToM, thus establishing that Tf is continuous.

   Since F and f are bounded then Tf is bounded as well. Note that there exists $M \geq 0$ such that $-M \leq F(x,y) + \beta f(y) \leq M$ for all (x,y), then for all x we have: $-M \leq \max_{y \in \Gamma(x)} \{F(x,y) + \beta f(y)\} \leq M$ which establishes boundedness of Tf.

   Then $Tf \in C(X)$ for any $f \in C(X)$.

ii. Blackwell conditions are met:

   (a) (monotonicity) Let $f, g \in C(X)$ and $f(x) \le g(x)$ for all x. Then:

$$Tf(x) = \max_{y\in\Gamma(x)} \{F(x,y) + \beta f(y)\} \le \max_{y\in\Gamma(x)} \{F(x,y) + \beta g(y)\} = Tg(x)$$

   (b) (discounting) Let $f \in C(X)$ and $a > 0$. Then:

$$T(f+a)(x) = \max_{y\in\Gamma(x)} \{F(x,y) + \beta(f(y)+a)\} = \max_{y\in\Gamma(x)} \{F(x,y) + \beta f(y)\} + \beta a$$

   In particular:

$$T(f+a)(x) \le Tf(x) + \beta a$$

   Then T is a contraction. By the contraction mapping theorem the result follows.

iii. The properties of G follow from the ToM which applies as shown before.

□

Additional assumption will help to characterize v and G better. The corollary of the contraction mapping theorem is the tool to be used now. First monotonicity can be inherited by the solution.

Assumption A.5: For all y $F(\cdot, y)$ is strictly increasing in its first l arguments.

Assumption A.6: Γ is monotone in the sense that if $x \le x^{'}$ the $\Gamma(x) \subseteq \Gamma\left(x^{'}\right)$.

Proposition 18.6. *Let X, Γ, F and β satisfy assumption A.3 to A.6, and let v be the unique solution to (18.4), then v is strictly increasing.*

Proof. Let $C^{'}(X) \subseteq C(X)$ be the set of bounded, continuous and non-decreasing functions and $C^{''}(X) \subseteq C^{'}(X)$ the set of strictly increasing functions. Clearly $C^{'}(X)$ is closed. By the corollary of the contraction mapping theorem it suffices to show that $T\left(C^{'}(X)\right) \subseteq C^{''}(X)$.

Let $f \in C^{'}(X)$ and consider $x < x^{'}$. We want to show that Tf is strictly increasing. This follows with A.5 and A.6:

$$Tf(x) = \max_{y\in\Gamma(x)} F(x,y) + \beta f(y) \le \max_{y\in\Gamma\left(x^{'}\right)} F(x,y) + \beta f(y) < \max_{y\in\Gamma\left(x^{'}\right)} F\left(x^{'}, y\right) + \beta f(y) = Tf\left(x^{'}\right)$$

where the first inequality follows from $\Gamma(x) \subseteq \Gamma\left(x^{'}\right)$, a larger choice set implies a higher than or equal maximum, the second inequality follows from F being strictly increasing. □

It is also possible to induce convexity as follows:

Assumption A.7: F is strictly concave in both arguments.

Assumption A.8: Γ has a convex graph.

Proposition 18.7. *Let X, Γ, F and β satisfy assumption A.3, A.4, A.7 and A.8, and let v be the unique solution to (18.4), then v is strictly concave and G is a continuous single valued function.*

Proof. Let $C^{'}(X)$ be the set of concave, bounded and continuous functions and $C^{''}(X)$ the set of strictly concave, bounded and continuous functions. Note that $C^{'}(X) \subseteq C(X)$ is closed and that $C^{''}(X) \subseteq C^{'}(X)$. To show that v is strictly concave we use the corollary of the contraction mapping theorem.

We want to show that for all $f \in C^{'}(X)$ it follows that $Tf \in C^{''}(X)$ where

$$Tf(x) = Tf(x) = \max_{y \in \Gamma(x)} F(x,y) + \beta f(y)$$

So let f be weakly concave on x, bounded and continuous. let $x_1, x_2 \in X$ and $\lambda \in (0,1)$ and define $x_\lambda = \lambda x_1 + (1-\lambda) x_2$. Let $y_i \in G(x_i) \subseteq \Gamma(x_i)$ and note that by A.8 $y_\lambda = \lambda y_1 + (1-\lambda) y_2 \in \Gamma(x_\lambda)$:

$$\begin{aligned}
Tf(x_\lambda) &= \max_{y \in \Gamma(x_\lambda)} F(x_\lambda, y) + \beta f(y) \\
&\geq F(x_\lambda, y_\lambda) + \beta f(y_\lambda) \\
&\geq F(x_\lambda, y_\lambda) + \lambda \beta f(y_1) + (1-\lambda) \beta f(y_2) \\
&> \lambda F(x_1, y_1) + (1-\lambda) F(x_2, y_2) + \lambda \beta f(y_1) + (1-\lambda) \beta f(y_2) \\
&= \lambda Tf(x_1) + (1-\lambda) Tf(x_2)
\end{aligned}$$

where the first inequality follows from $y_\lambda$being feasible at x_λ, the second one from f being concave and the third one from A.7. The final equality is obtained rearranging and recalling the optimality of y_1 and y_2 under x_1 and x_2 respectively. Joining results we get

$$Tf(x_\lambda) > \lambda Tf(x_1) + (1-\lambda) Tf(x_2)$$

Then the "image" of any concave function is a strictly concave function. This proves that $Tf \in C^{''}(X)$. Then by the corollary of the contraction mapping theorem the unique fix point of T belongs to $C^{''}$. That is, v is strictly concave.

Finally note that $G(x) = \underset{y \in \Gamma(x)}{\text{argmax}} F(x,y) + v(y)$. Since both F and v are strictly concave the single valuedness and continuity of G follow as an immediate consequence of the ToM under convexity part (ii). □

Finally there are conditions for v to be differentiable, allowing the use of first order conditions.

Assumption A.9: F is continuously differentiable on the interior of its domain, $\text{Gr}(A)$.

Proposition 18.8. *Let X, Γ, F and β satisfy assumption A.3, A.4, and A.7 to A.9, and let v be the unique solution to (18.4). If $x_0 \in IntX$ and $g(x_0) \in Int\Gamma(x_0)$ then v is continuously differentiable at x_0 with derivatives given by:*

$$v_i(x_0) = F_i(x_0, g(x_0))$$

Proof. Stokey et al. (1989, sec. 4.2, pp. 85). □

19 Applications[6]

19.1 Spring 2004 - Q2.2 [Chari] (Search and Unemployment)

Consider the following infinite horizon model. A worker who is employed begins each period with a wage, say w. The worker can either work at that wage or search and receive unemployment benefits b on the current period. If the worker chooses to work he is employed with wage w next period with probability δ or unemployed with probability $1-\delta$. An unemployed worker who searches receives a wage offer from a distribution $F(w)$. Wage offers are iid over time. The worker preferences are $\sum \beta^t u(c_t)$ where u is an increasing function. Assume no borrowing or lending.[7]

i. Set up the workers decision as a dynamic programming problem.

$$\begin{aligned} V^E(w) &= u(w) + \beta\delta V^U + \beta(1-\delta)V^E(w) \\ V^U &= u(b) + \beta \int \max\left\{V^E(\tilde{w}), V^U\right\} dF(\tilde{w}) \end{aligned}$$

The decision of a worker when facing a wage offer w is to accept it or reject it, the worker will accept if $V^E(w) > V^u$ and reject otherwise. Then the value of the worker is:

$$V(w) = \max\left[V^E(w), V^U\right]$$

ii. Show that the solution to the worker's problem is of the reservation wage form. Characterize the reservation wage. Show that the reservation wage is increasing in the unemployment benefits.

Let $C(W)$ be the set of continuous functions on W, note that since W is compact those function are also bounded. Let $\overline{C}(W)$ be the set of constant functions on W, these functions are by construction continuous and bounded. Define $\tilde{C}(W) = C(W) \times \overline{C}(W)$ and $T : \tilde{C}(W) \to \tilde{C}(W)$ as:

$$T\begin{pmatrix} V^E \\ V^U \end{pmatrix}(w) = \begin{bmatrix} u(w) + \beta\delta V^U + \beta(1-\delta)V^E(w) \\ u(b) + \beta\int \max\left\{V^E(\tilde{w}), V^U\right\} dF(\tilde{w}) \end{bmatrix}$$

note that since u is continuous and bounded over W it follows that for V^E and V^U continuous and bounded $TV^E(w) = u(w) + \beta\delta V^U + \beta(1-\delta)V^E(w)$ is also continuous and bounded also since b is constant and $\int \max\left\{V^E(\tilde{w}), V^U\right\} dF(\tilde{w})$ is independent of w it follows that $TV^U(w)$ is constant.

T is a contraction since it satisfies the extended Blackwell conditions.

(a) Monotonicity:

[6] This section takes material from solutions to some of the questions of Macroeconomic Theory preliminary exams at the University of Minnesota. The solutions were made with Dominic Smith.

[7] We furthermore assume that u is continuous and w has a bounded support $W = [0, \overline{w}]$.

Let $\left(V^E, V^U\right) \in \tilde{C}(W)$ and $\left(M^E, M^U\right) \in \tilde{C}(W)$ such that $M^U \geq V^U$ and for all $w \in W$ $M^E(w) \geq V^E(w)$. Clearly this implies $\max\left\{V^E(\tilde{w}), V^U\right\} \leq \max\left\{M^E(\tilde{w}), M^U\right\}$ for all $\tilde{w}$. Then:

$$\begin{aligned} T\begin{pmatrix} V^E \\ V^U \end{pmatrix}(w) &= \begin{bmatrix} u(w) + \beta\delta V^U + \beta(1-\delta) V^E(w) \\ u(b) + \beta \int \max\left\{V^E(\tilde{w}), V^U\right\} dF(\tilde{w}) \end{bmatrix} \\ &\leq \begin{bmatrix} u(w) + \beta\delta M^U + \beta(1-\delta) M^E(w) \\ u(b) + \beta \int \max\left\{M^E(\tilde{w}), M^U\right\} dF(\tilde{w}) \end{bmatrix} \\ &= T\begin{pmatrix} M^E \\ M^U \end{pmatrix}(w) \end{aligned}$$

(b) Discounting:
Let $a \in \mathbb{R}$. Then:

$$\begin{aligned} T\begin{pmatrix} V^E + a \\ V^U + a \end{pmatrix}(w) &= \begin{bmatrix} u(w) + \beta\delta\left(V^U + a\right) + \beta(1-\delta)\left(V^E(w) + a\right) \\ u(b) + \beta \int \max\left\{V^E(\tilde{w}) + a, V^U + a\right\} dF(\tilde{w}) \end{bmatrix} \\ &= \begin{bmatrix} u(w) + \beta\delta V^U + \beta(1-\delta) V^E(w) \\ u(b) + \beta \int \max\left\{V^E(\tilde{w}), V^U\right\} dF(\tilde{w}) \end{bmatrix} + \beta \begin{bmatrix} a \\ a \end{bmatrix} \\ &= T\begin{pmatrix} V^E \\ V^U \end{pmatrix}(w) + \beta \begin{bmatrix} a \\ a \end{bmatrix} \end{aligned}$$

Note that the set of increasing functions in $C(W)$ is a closed subset, call it $I(W)$, and that, for $\left(V^E, V^U\right) \in \tilde{C}(W)$m $T\begin{bmatrix} V^E \\ V^U \end{bmatrix} \in I(W) \times \overline{C}(W)$.

(a) Let $w < w^{'}$. Then $u(w) < u\left(w^{'}\right)$, and $V^E(w) \leq V^E\left(w^{'}\right)$

$$\begin{aligned} T\left[V^E\right](w) &= u(w) + \beta\delta V^U + \beta(1-\delta) V^E(w) \\ &\leq u(w) + \beta\delta V^U + \beta(1-\delta) V^E\left(w^{'}\right) \\ &< u\left(w^{'}\right) + \beta\delta V^U + \beta(1-\delta) V^E\left(w^{'}\right) \\ &= T\left[V^E\right]\left(w^{'}\right) \end{aligned}$$

Then $T\left[V^E\right]$ is strictly increasing.

It must be that $V^E(\bar{w}) > V^U$. Suppose the contrary, $V^E(\overline{w}) \leq V^U$, since V^E is strictly increasing it must be that $V^E(w) \leq V^U$ for all w. Then $\max\left\{V^E(\tilde{w}), V^U\right\} = V^U$ which gives $V^U = \frac{u(b)}{1-\beta}$. Also

$$\begin{aligned} V^E(\overline{w}) &= u(\overline{w}) + \beta\delta V^U + \beta(1-\delta) V^E(\overline{w}) \\ (1 - \beta(1-\delta)) V^E(\overline{w}) &= u(\overline{w}) + \beta\delta V^U \\ (1 - \beta(1-\delta)) V^U &\geq u(\overline{w}) + \beta\delta V^U \\ (1-\beta) V^U &\geq u(\overline{w}) \\ u(b) &\geq u(\overline{w}) \end{aligned}$$

For a contradiction suppose further that $b < \overline{w}$, so that the unemployment benefits cannot be higher than the maximum possible wage. With the extra assumption and the fact that u is strictly increasing the contradiction implies $V^E(\overline{w}) > V^U$.

Suppose that $V^E(0) \leq V^U$. Then by the intermediate value theorem there exists a wage $w^\star$ such that $V^E(w^\star) = V^U$. If the assumption does not hold then all wages are acceptable. $w^\star$ satisfies:

$$\begin{aligned} V^E(w^\star) &= u(w^\star) + \beta\delta V^U + \beta(1-\delta)V^E(w^\star) \\ V^U &= u(w^\star) + \beta\delta V^U + \beta(1-\delta)V^U \\ V^U &= \frac{u(w^\star)}{1-\beta} \end{aligned}$$

for $w \geq w^\star$:

$$\begin{aligned} V^E(w) &= u(w) + \beta\delta V^U + \beta(1-\delta)V^E(w) \\ V^E(w) &= \frac{1}{1-\beta(1-\delta)}\left(u(w) + \beta\delta\frac{u(w^\star)}{1-\beta}\right) \end{aligned}$$

and:

$$\begin{aligned} V^U &= u(b) + \beta F(w^\star)V^U + \beta\int_{w^\star}^{\overline{w}} V^E(\tilde{w})\,dF(\tilde{w}) \\ (1-\beta F(w^\star))\frac{u(w^\star)}{1-\beta} &= u(b) + \beta\int_{w^\star}^{\overline{w}} V^E(\tilde{w})\,dF(\tilde{w}) \end{aligned}$$

Replacing for $V^E(\cdot)$:

$$\beta\int_{w^\star}^{\overline{w}} V^E(\tilde{w})\,dF(\tilde{w}) = \frac{\beta}{1-\beta(1-\delta)}\int_{w^\star}^{\overline{w}} u(\tilde{w})\,dF(\tilde{w}) - \frac{\beta^2\delta(1-F(w^\star))}{1-\beta(1-\delta)}\frac{u(w^\star)}{1-\beta}$$

Joining:

$$\begin{aligned} u(b) &= \left[(1-\beta F(w^\star)) - \frac{\beta^2\delta(1-F(w^\star))}{1-\beta(1-\delta)}\right]\frac{u(w^\star)}{1-\beta} - \frac{\beta}{1-\beta(1-\delta)}\int_{w^\star}^{\overline{w}} u(\tilde{w})\,dF(\tilde{w}) \\ u(b) &= \left[\frac{1+\beta\delta-\beta F(w^\star)}{1-\beta(1-\delta)}\right]u(w^\star) - \frac{\beta}{1-\beta(1-\delta)}\int_{w^\star}^{\overline{w}} u(\tilde{w})\,dF(\tilde{w}) \\ u(b) &= \left[\frac{1+\beta\delta-\beta+\beta(1-F(w^\star))}{1-\beta(1-\delta)}\right]u(w^\star) - \frac{\beta}{1-\beta(1-\delta)}\int_{w^\star}^{\overline{w}} u(\tilde{w})\,dF(\tilde{w}) \\ u(b) &= u(w^\star) - \frac{\beta}{1-\beta(1-\delta)}\int_{w^\star}^{\overline{w}} u(\tilde{w})\,dF(\tilde{w}) + \frac{\beta}{1-\beta(1-\delta)}(1-F(w^\star))u(w^\star) \\ u(b) &= u(w^\star) - \frac{\beta}{1-\beta(1-\delta)}\int_{w^\star}^{\overline{w}} (u(\tilde{w}) - u(w^\star))\,dF(\tilde{w}) \end{aligned}$$

The LHS is constant the RHS is increasing in $w^\star$, to see that note that u is strictly increasing and that

$$G(w^\star) = \int_{w^\star}^{\overline{w}} (u(\tilde{w}) - u(w^\star))\,dF(\tilde{w})$$

is decreasing in $w^\star$. Let $w < w^{'}$ and consider

$$\begin{aligned}
G\left(w^{'}\right) - G(w) &= \int_{w^{'}}^{\overline{w}} \left(u(w) - u\left(w^{'}\right)\right) dF(\tilde{w}) - \int_{w}^{w^{'}} \left(u(\tilde{w}) - u(w)\right) dF(\tilde{w}) \\
&\leq \int_{w^{'}}^{\overline{w}} \left(u(w) - u\left(w^{'}\right)\right) dF(\tilde{w}) - \int_{w}^{w^{'}} \left(u\left(w^{'}\right) - u(w)\right) dF(\tilde{w}) \\
&= \int_{w}^{\overline{w}} \left(u(w) - u\left(w^{'}\right)\right) dF(\tilde{w}) \\
&= (1 - F(w)) \left(u(w) - u\left(w^{'}\right)\right) \\
&< 0
\end{aligned}$$

where the first inequality follows from $u(\tilde{w}) - u(w) \leq u\left(w^{'}\right) - u(w)$ since $\tilde{w} \leq w^{'}$, and the second inequality from $u(w) < u\left(w^{'}\right)$ since u is strictly increasing and $w < w^{'}$. Then $G\left(w^{'}\right) < G(w)$ which proves that the integral is strictly decreasing, hence its negative is strictly increasing.

Since the RHS side is increasing in $w^\star$ the reservation wage is an increasing function of b, a higher value of b increases the LHS of the equation which implies a higher value of $w^\star$ to increase the RHS.

iii. Calculate the probability that an unemployed worker is unemployed for N periods in a row as a function of the reservation wage and show that this probability is decreasing in N.

Since wage offers are iid the probability that a worker searching for a job receives N consecutive offers lower than its reservation wage is equal to the product of the probability that he receives an offer lower than $w^\star$, that is

$$\Pr\{w_1 < w^\star \wedge \ldots \wedge w_n \leq w^\star\} = [\Pr\{w \leq w^\star\}]^N = [F(w^\star)]^N$$

(a) How does this probability vary with unemployment benefits?

Since higher unemployment benefits imply a higher reservation wage and $F(\cdot)$ is an increasing function, a higher unemployment benefits increase the probability of remaining unemployed for N consecutive periods.

iv. Suppose now that there a large number of workers that all face the same problem. Define the unemployment rate as the fraction of all workers who are unemployed. How does the unemployment rate vary with b?

Assume there is a continuum of unit measure of workers. Let E_t and U_t be the fraction of workers employed and unemployed. The following conditions hold:

$$\begin{aligned}
E_{t+1} &= (1 - \delta) E_t + (1 - F(w^\star)) U_t \\
U_{t+1} &= \delta E_t + F(w^\star) U_t \\
E_t + U_t &= 1
\end{aligned}$$

In the steady state $E_{t+1} = E_t$ and $U_{t+1} = U_t$ which leads to $U = \delta\left(1 - U\right) + F\left(w^{\star}\right)U$ or

$$U = \frac{\delta}{\left(1 - \delta\right) - F\left(w^{\star}\right)}$$

Then since higher unemployment benefits increase the reservation wage they also increase $F\left(w^{\star}\right)$, which decreases the denominator, hence increasing the steady state rate of unemployment.

19.2 Spring 2008 - Q2.3 [Chari] (Search and Human Capital)

Consider an economy in which workers accumulate human capital while working.

A worker currently working at a job produces a single non-storable good according to h_t^{α} where h_t is human capital and $0 < \alpha < 1$.

Jobs disappear at the end of the period with probability δ. If the job does not disappear, the worker's human capital next period is $h_{t+1} = \gamma h_t$ where $\gamma > 1$. If the job disappears, the worker becomes unemployed. Unemployed workers receive one new job offer in each period.

The offers are parameterized by a random variable z which is uniformly distributed between 0 and 1 and is i.i.d. over time. A worker who accepts a job at t has human capital $h_t = zh_{t-1}$. That is, part of the human capital $(1-z)\,h_{t-1}$ disappears forever. Workers are risk-neutral and maximize $\sum \beta^t c_t$.

i. Set up the worker's problem as a dynamic program.

$$\begin{aligned} V(h,z) &= \max\left\{V^E(zh), V^U(h)\right\} \\ V^E(h) &= h^{\alpha} + \beta\left[\delta E\left[V(h,z)\right] + (1-\delta)\,V^E(\gamma h)\right] \\ V^U(h) &= \beta E\left[V(h,z)\right] \end{aligned}$$

ii. Prove that unemployed workers have a reservation strategy of the form $z(h_{t-1})$ where they accept all offers greater than $z(h_{t-1})$ and reject all others.

(a) Let $C(H)$ be the space of continuous bounded functions defined on $H = \mathbb{R}_+$, that are also measurable with respect to the measure induced by the uniform distribution. Let $T: C(H)\times C(H) \to C(H)\times C(H)$ be a functional mapping pairs of functions in $C(X)$ to to pairs of functions in $C(X)$. Define

$$T\begin{bmatrix} V^E \\ V^U \end{bmatrix}(h) = \begin{bmatrix} h^{\alpha} + \beta\left[\delta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] + (1-\delta)\,V^E(\gamma h)\right] \\ \beta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] \end{bmatrix}$$

Note that the relevant space H is unbounded, yet the relevant functions are bounded provided that $\beta\gamma^{\alpha} < 1$. That is, provided that the maximum possible return (when the agent is employed and is never fired) is bounded. In that case the return is: $\sum \beta^t h_t^{\alpha} = \sum \beta^t\left(\gamma^t h_0\right)^{\alpha} = h_0^{\alpha}\sum\left(\beta\gamma^{\alpha}\right)^t$. Clearly if V^E and V^U are continuous and bounded on H so is their maximum, expected value and sum. (Note that since the mapping is defined over continuous bounded functions the measurability requirement is likely to be satisfied immediately).

(b) The mapping T satisfies the Extended Blackwell sufficient conditions for a contraction:

i. (Monotonicity) $f, g \in S$ and $f(x) \le g(x)$, for all $x \in X$, implies $Tf(x) \le Tg(x)$, for all $x \in X$, (where $f(x) \le g(x)$ is taken in the vector sense, i.e. $f_1(x) \le g_1(x)$ and $f_2(x) \le g_2(x)$).
Let $V^E, V^U, W^E, W^U \in C(H)$ such that for all $x \in H$ $V^E(h) \le W^E(h)$ and $V^U(h) \le W^U(h)$. Then:

$$\begin{aligned} TV^E(h) &= h^{\alpha} + \beta\left[\delta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] + (1-\delta)\,V^E(\gamma h)\right] \\ &\le h^{\alpha} + \beta\left[\delta E\left[\max\left\{W^E(zh), W^U(h)\right\}\right] + (1-\delta)\,W^E(\gamma h)\right] = TW^E(h) \end{aligned}$$

and

$$TV^U(h) = \beta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] \leq \beta E\left[\max\left\{W^E(zh), W^U(h)\right\}\right] = TW^U(h)$$

Thus proving monotonicity.

ii. (Discounting) there exists some $\beta \in (0,1)$ such that $T(f+A) \leq Tf(x) + \beta A$ for $f \in S$, $A = [a\ a]' \in \mathbb{R}^2_+$ and $x \in X$.
Let $V^E, V^U \in C(H)$ and $a \in \mathbb{R}$, then:

$$\begin{aligned}
T\begin{bmatrix} V^E + a \\ V^U + a \end{bmatrix}(h) &= \begin{bmatrix} h^\alpha + \beta\left[\delta E\left[\max\left\{V^E(zh) + a, V^U(h) + a\right\}\right] + (1-\delta)\left(V^E(\gamma h) + a\right) \\ \beta E\left[\max\left\{V^E(zh) + a, V^U(h) + a\right\}\right] \end{bmatrix} \\
&= \begin{bmatrix} h^\alpha + \beta\left[\delta E\left[\max\left\{V^E(zh), V^U(h)\right\} + a\right] + (1-\delta)\left(V^E(\gamma h)\right) + (1- \\ \beta E\left[\max\left\{V^E(zh), V^U(h)\right\} + a\right] \end{bmatrix} \\
&= \begin{bmatrix} h^\alpha + \beta\left[\delta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] + (1-\delta)\left(V^E(\gamma h)\right)\right] + \beta a \\ \beta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] + \beta a \end{bmatrix} \\
&= \begin{bmatrix} h^\alpha + \beta\left[\delta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] + (1-\delta)\left(V^E(\gamma h)\right)\right] + \beta a \\ \beta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] + \beta a \end{bmatrix} + \\
&= T\begin{bmatrix} V^E \\ V^U \end{bmatrix}(h) + \beta\begin{bmatrix} a \\ a \end{bmatrix}
\end{aligned}$$

Proving discounting.

iii. Then T is a contraction with respect to the norm $\|f\| = \max\{\|f_1\|_\infty, \|f_2\|_\infty\}$. By the contraction mapping theorem there exists a pair of functions $\left(V^E, V^U\right) \in C(H) \times C(H)$ such that they satisfy (jointly) the two functional equations above.

(c) Note that the set of non-decreasing functions is closed in $C(H)$, hence the set of non-decreasing functions in $C(X) \times C(X)$ is also closed (meaning that both of the functions in the pair are non-decreasing). Moreover, if $V^E, V^U \in C(H)$ are non-decreasing so are the pair formed by $T\begin{bmatrix} V^E \\ V^U \end{bmatrix}$. This follows from the maximum and expected value of two non-decreasing functions being non-decreasing, the sum of non-decreasing functions being non-decreasing and the fact that h^α is also non-decreasing.

By the corollary of the contraction mapping theorem it is established that the functions of the fixed point V^E and V^U are non-decreasing.

(d) Note that the set of homogenous of degree α functions (h.d.α) is closed in $C(H)$: let $f_n \to f$ and f_n be h.d.α for all n, then it holds that $f_n(\lambda h) = \lambda^\alpha f_n(h)$. Since $f_n \to f$ uniformly it also converges pointwise, then $\lim f_n(\lambda h) = \lambda^\alpha \lim f_n(h)$ which is $f(\lambda h) = \lambda^\alpha f(h)$, f is also h.d.α and then the set is closed.

Also, the mapping of h.d.α functions is also h.d.α. Let V^E and V^U be h.d.α, then:

$$\begin{aligned}
T\begin{bmatrix} V^E \\ V^U \end{bmatrix}(\lambda h) &= \begin{bmatrix} (\lambda h)^\alpha + \beta\left[\delta E\left[\max\left\{V^E(z\lambda h), V^U(\lambda h)\right\}\right] + (1-\delta)V^E(\gamma\lambda h)\right] \\ \beta E\left[\max\left\{V^E(z\lambda h), V^U(\lambda h)\right\}\right] \end{bmatrix} \\
&= \begin{bmatrix} \lambda^\alpha h^\alpha + \beta\left[\delta E\left[\max\left\{\lambda^\alpha V^E(zh), \lambda^\alpha V^U(h)\right\}\right] + (1-\delta)\lambda^\alpha V^E(\gamma h)\right] \\ \beta E\left[\max\left\{\lambda^\alpha V^E(zh), \lambda^\alpha V^U(h)\right\}\right] \end{bmatrix} \\
&= \begin{bmatrix} \lambda^\alpha\left(h^\alpha + \beta\left[\delta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] + (1-\delta)V^E(\gamma h)\right]\right) \\ \lambda^\alpha \beta E\left[\max\left\{V^E(zh), V^U(h)\right\}\right] \end{bmatrix} \\
&= \begin{bmatrix} \lambda^\alpha & 0 \\ 0 & \lambda^\alpha \end{bmatrix} T \begin{bmatrix} V^E \\ V^U \end{bmatrix}
\end{aligned}$$

By the corollary of the contraction mapping theorem it is established that the functions of the fixed point V^E and V^U are h.d.α.

(e) Under the above it is established that:

$$V^E(h) = h^\alpha V^E(1) \quad \wedge \quad V^U(h) = h^\alpha V^U(1)$$

which implies that $V^E(0) = V^U(0) = 0$, and that $V^E(h) > V^U(h)$ for all $h > 0$. Suppose the contrary, then $V^E(h) \leq V^U(h)$, then $V^E(1) \leq V^U(1)$, this implies that for all $z \in [0,1]$ $V^E(z) \leq V^U(1)$ and hence that:

$$V^U(1) = \beta \max\left\{V^E(z), V^U(1)\right\} = \beta V^U(1)$$

which is a contradiction unless $V^U(1) = 0$, in which case it has to be that $V^E(1) \leq 0$, yet by V^E being non-decreasing: $V^E(1) = 0$. This implies, by h.d.α that $V^E(h) = V^U(h) = 0$ for all h. This contradicts the V^E and V^Ubeing a fixed point for the mapping T. Take $h > 0$, under $V^E = V^U = 0$ one has from the first functional equation: $0 = h^\alpha$, a contradiction.

(f) An unemployed agent will accept a job offer for $z \in [0,1]$ such that: $V^E(zh) \geq V^U(h)$, which under $h.d.\alpha$ holds if and only if $z^\alpha V^E(1) \geq V^U(1)$, that is for $z \geq z^\star = \left(\frac{V^U(1)}{V^E(1)}\right)^{\frac{1}{\alpha}}$. Note that since $V^E(1) > V^U(1)$ the value $z^\star < 1$ and moreover is constant for all h.

iii. What can you say about $z(h_{t-1})$? Is it linear in h_{t-1}?

It is independent of h_{1-1}.

19.3 Spring 2012 - Q 1 [Kehoe] (Guess and Verify)

Infinitely lived consumers and dynamic programming

Consider an economy in which the representative consumer lives forever. There is a good in each period that can be consumed or saved as capital as well as labor. The consumer's utility function is

$$\sum_{t=0}^{\infty} \beta^t \log c_t$$

Here $0 < \beta < 1$. The consumer is endowed with 1 unit of labor in each period and with $\overline{k}_0$ units of capital in period 0. Feasible allocations satisfy

$$c_t + k_{t+1} \leq \theta k_t^{\alpha} l_t^{1-\alpha}$$

Here $\theta > 0$ and $0 < \alpha < 1$.

i. Formulate the problem of maximizing the representative consumer's utility subject to feasibility conditions as a dynamic programming problem. Write down the appropriate Bellman's equation.

   The Dynamic Programming version of this problem is for the consumer to solve

$$\begin{aligned} V(k) = & \max_{c,k',l} \{\log c + \beta V(k')\} \\ \text{s.t. } & c + k' \leq \theta k^{\alpha} l^{1-\alpha} \\ & c, k' \geq 0 \\ & 0 \leq l \leq 1 \end{aligned}$$

ii. Guess that the value function has the form $a_0 + a_1 \log k$. Solve the dynamic programming problem.

   The constraint will hold with equality because the utility function is strictly increasing in consumption, also production increases with labor and there is no disutility of it, hence there is a corner solution for labor indicating $l = 1$, so with the guess the problem becomes

$$a_0 + a_1 \log k = \log\left(\theta k^{\alpha} l^{1-\alpha} - k'\right) + \beta\left(a_0 + a_1 \log k'\right)$$

   Then the FOC is

$$\frac{1}{\theta k^{\alpha} l^{1-\alpha} - k'} = \frac{\beta a_1}{k'}$$

   solving for k'

$$\begin{aligned} k' &= \beta a_1 \left(\theta k^{\alpha} l^{1-\alpha} - k'\right) \\ &= \frac{\beta a_1 \left(\theta k^{\alpha} l^{1-\alpha}\right)}{1 + \beta a_1} \end{aligned}$$

   Then plugging this back into the value function you get

$$a_0 + a_1 \log k = \log\left(\theta k^{\alpha} l^{1-\alpha} - \frac{\beta a_1 \left(\theta k^{\alpha} l^{1-\alpha}\right)}{1 + \beta a_1}\right) + \beta\left(a_0 + a_1 \log\left(\frac{\beta a_1 \left(\theta k^{\alpha} l^{1-\alpha}\right)}{1 + \beta a_1}\right)\right)$$

Collection terms with k you get

$$
\begin{aligned}
a_1 \log k &= \alpha \log k + \beta a_1 \alpha \log k \\
a_1 \left(\log k - \beta\alpha \log k\right) &= \alpha \log k \\
a_1 &= \frac{\alpha}{1-\beta\alpha}
\end{aligned}
$$

which means the policy function is

$$
\begin{aligned}
k' &= \frac{\beta \frac{\alpha}{1-\beta\alpha}\left(\theta k^{\alpha} l^{1-\alpha}\right)}{1+\beta\frac{\alpha}{1-\beta\alpha}} = \beta\alpha\theta k^{\alpha} l^{1-\alpha} \\
l &= 1 \\
c &= \theta k^{\alpha} l^{1-\alpha} - \beta\alpha\theta k^{\alpha} l^{1-\alpha}
\end{aligned}
$$

19.4 Fall 2007 - Q2.1 [Jones] (Durable Goods)

Consider a single agent problem where each period, w total output is produced and can be divided into consumption of a perishable good, c_t and investment in a durable good, d_{xt}. The durable depreciates like a capital good, but is not directly productive. The stock of durables at any date, d_t, produces a flow of services that enters the utility function. Thus, the problem faced by the household with initial stock d_0 is:

$$\max_{c_t, d_t, d_{xt}} \sum_t \beta^t \left\{ u_1(c_t) + u_2(d_t) \right\}$$
$$\text{s.t.}$$
$$c_t + d_{xt} \leq w$$
$$d_{t+1} \leq (1-\delta) d_t + d_{xt}$$
$$c_t, d_t, d_{xt} \geq 0$$
$$d_0 \text{given}$$

where both u_1 and u_2 are strictly increasing and continuous. Note: you can ignore non-negativity constraints on investment, d_{xt} in this problem.

i. State a condition on either u_1 or u_2 (or both) such that you can write an equivalent problem in the following form:

$$\max_{\{d_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t F(d_t, d_{t+1})$$
$$\text{s.t.}$$
$$d_{t+1} \in \Gamma(d_t)$$
$$d_0 \text{given}$$

where $\Gamma(d) \in \mathbb{R}_+$. What is F? What is the correspondence Γ?

Under the conditions already given any solution to the problem requires that the two constraints are satisfied with equality. Suppose the contrary, that there is an optimum sequence $\{c_t, d_t, d_{xt}\}$ such that for some period $c_t + d_{xt} < w$ or $d_{t+1} < (1-\delta) d_t + d_{xt}$, then there exists an alternative allocation $\left\{c_t^{'}, d_t^{'}, d_{xt}\right\}$ such that $c_t^{'} = c_t + (w - c_t - d_t) > c_t$ and $d_{t+1}^{'} = d_{t+1} + ((1-\delta) d_t + d_{xt} - d_{t+1}) > d_{t+1}$. Since u_1 and u_2 are strictly increasing $u_1\left(c_t^{'}\right) > u_1(c_t)$ and $u_2\left(d_{t+1}^{'}\right) > u_2(d_{t+1})$, then $\sum_t \beta^t \left\{u_1\left(c_t^{'}\right) + u_2\left(d_t^{'}\right)\right\} > \sum_t \beta^t \left\{u_1(c_t) + u_2(d_t)\right\}$ which contradicts $\{c_t, d_t, d_{xt}\}$ being optimal.

Since the constraints hold with equality in any solution the problem is equivalent to:

$$\max_{d_t} \sum_t \beta^t \left\{ u_1(w + (1-\delta) d_t - d_{t+1}) + u_2(d_t) \right\}$$
$$\text{s.t.}$$
$$w + (1-\delta) d_t - d_{t+1} \geq 0$$
$$d_{t+1} \geq 0$$
$$d_0 \text{given}$$

Then $F(d_t, d_{t+1}) = u_1(w + (1-\delta) d_t - d_{t+1}) + u_2(d_t)$ and $\Gamma(d_t) = \{d_{t+1} | d_{t+1} \geq 0 \wedge w + (1-\delta) d_t - d$

ii. Write the Bellman equation for this problem.

$$v(d) = \max_{d' \in \Gamma(d)} \left[F\left(d, d'\right) + \beta v\left(d'\right) \right]$$

iii. State additional conditions on u_1 and u_2 such that the value function $v(d)$ is both strictly increasing and strictly concave. Prove these two properties.

(a) $v(d)$ strictly increasing. No further conditions are needed for this. Note that $F\left(\cdot, d'\right)$ is strictly increasing in d, this follows from u_1 and $w + (1-\delta)d - d'$ being strictly increasing in d and from u_2 being strictly increasing. Also note that for $d_1 < d_2$ it follows that $\Gamma(d_1) \subset \Gamma(d_2)$, to see this let $d' \in \Gamma(d_1)$, then $d' \geq 0$ and $w + (1-\delta)d_1 - d' \geq 0$ since $d_2 > d_1$ it follows that $w + (1-\delta)d_2 - d' \geq 0$ which implies $d' \in \Gamma(d_2)$.

Finally consider $v(d_1)$, since functions are continuous defined over a compact space (taking into account d_0 and the maximum sustainable level of durable goods $\overline{d} = \frac{w}{\delta}$) the solution to this problem exists and is attained by $d' \in G(d_1)$ where $G(\cdot)$ is the policy correspondence. It follows that:

$$v(d_1) = F\left(d_1, d'\right) + \beta v\left(d'\right)$$

for some $d' \in G(d_1)$. Since F is increasing in its first argument:

$$v(d_1) < F\left(d_2, d'\right) + \beta v\left(d'\right)$$

Since $d' \in G(d_1)$ it holds that $d' \in \Gamma(d_1)$ and hence $d' \in \Gamma(d_2)$. By definition of maximum it must be that:

$$v(d_1) < F\left(d_2, d'\right) + \beta v\left(d'\right) \leq \max_{d' \in \Gamma(d_2)} \left(F\left(d_2, d'\right) + \beta v\left(d'\right) \right) = v(2)$$

then: $v(d_1) < v(d_2)$.

(b) $v(d)$ strictly concave. Let u_1 and u_2 be strictly concave. Note that for all d_1, d_2 and $\theta \in (0,1)$ $\theta d_1' + (1-\theta)d_2' \in \Gamma(\theta d_1 + (1-\theta)d_2)$ where $d_1' \in \Gamma(d_1)$ and $d_2' \in \Gamma(d_2)$. It holds that since $d_1', d_2' \geq 0$ then $\theta d_1' + (1-\theta)d_2' \geq 0$. Moreover since $w + (1-\delta)d_1 - d_1' \geq 0$ and $w + (1-\delta)d_2 - d_2' \geq 0$, multiplying by θ and $1-\theta$ and summing inequalities gives: $w + (1-\delta)(\theta d_1 (1-\theta)d_2) - \left(\theta d_1' (1-\theta)d_2'\right) \geq 0$, this confirms the result.

Let $T : C(D) \to C(D)$ be an operator from the set of bounded continuous and weakly concave functions to itself, and $D = \left[0, \max\left(d_0, \frac{w}{\delta}\right)\right]$. Define $Tv(d) = \max_{d' \in \Gamma(d)} \left[F\left(d, d'\right) + \beta v\left(d'\right)\right]$. Note that if $w(d) \geq v(d)$ for all $d \in D$ it follows that

$$Tv(d) = \max_{d' \in \Gamma(d)} \left[F\left(d, d'\right) + \beta v\left(d'\right) \right] \leq \max_{d' \in \Gamma(d)} \left[F\left(d, d'\right) + \beta w\left(d'\right) \right] = Tw(d)$$

and that

$$T(v+a)(d) = \max_{d' \in \Gamma(d)} \left[F\left(d, d'\right) + \beta\left(v\left(d'\right) + a\right)\right] = Tv(d) + \beta a$$

Then T is a contraction by Blackwell's sufficient conditions, clearly the solution to the Bellman equation above is the unique fixed point of T. Moreover by the corollary of the contraction mapping theorem, since the image of the set of weakly concave functions is contained in the set of strictly concave functions $\left(C'(D)\right)$ (that is $T(C(D)) \subset C'(D)$) it follows that v (the fixed point) is strictly concave. Let v be weakly concave and $d_1, d_2 \in D$, $\theta \in (0,1)$ and $d_1' \in G(d_1)$, $d_2' \in G(d_2)$:

$$\begin{aligned}
Tv(d_\theta) = & \max_{d' \in \Gamma(\theta d_1 + (1-\theta) d_2)} \left[F\left(\theta d_1 + (1-\theta) d_2, d'\right) + \beta v\left(d'\right)\right] \\
& \geq F\left(\theta d_1 + (1-\theta) d_2, \theta d_1' + (1-\theta) d_2'\right) + \beta v\left(\theta d_1' + (1-\theta) d_2'\right) \\
& = u_1\left(w + (1-\delta)(\theta d_1 + (1-\theta) d_2) - \left(\theta d_1' + (1-\theta) d_2'\right)\right) + u_2(d_\theta) + \beta v\left(\theta d_1' + (1-\theta)\right. \\
& = u_1\left(\theta\left(w + (1-\delta) d_1 - d_1'\right) + (1-\theta)\left(w + (1-\delta) d_2 - d_2'\right)\right) + u_2(d_\theta) + \beta v\left(\theta d_1' + (1\right.
\end{aligned}$$

The inequality follows from $\theta d_1' + (1-\theta) d_2' \in \Gamma(\theta d_1 + (1-\theta) d_2)$. Using strict concavity of u_1 and u_2 and concavity of v one gets:

$$\begin{aligned}
Tv(d_\theta) > & \theta\left(u_1\left(w + (1-\delta) d_1 - d_1'\right) + u_2(d_1) + \beta v\left(d_1'\right)\right) \\
& + (1-\theta)\left(u_1\left(w + (1-\delta) d_2 - d_2'\right) + u_2(d_2) + \beta v\left(d_2'\right)\right) \\
= & \theta Tv(d_1) + (1-\theta) Tv(d_2)
\end{aligned}$$

Proving that Tv is strictly concave, hence the solution to the Bellman equation is too.

iv. For the remaining questions, assume that both u_1 and u_2 satisfy the Inada conditions and are continuously differentiable. State the envelope and the FOC for the functional equation problem in (b)

$$\begin{aligned}
F_2\left(d, d'\right) + \beta v'\left(d'\right) &= 0 \\
v'(d) &= F_1\left(d, d'\right)
\end{aligned}$$

where:

$$\begin{aligned}
F_2\left(d, d'\right) &= -u_1'\left(w + (1-\delta) d - d'\right) \\
F_1\left(d, d'\right) &= (1-\delta) u_1'\left(w + (1-\delta) d - d'\right) + u_2'(d)
\end{aligned}$$

v. Show that there is a unique steady state value of the stock, $d^\star$, such that if $d_0 = d^\star$, then $d_t = d^\star$ for all t. Show that $d^\star > 0$.

In a steady state:

$$u_1^{'}(w-\delta d^{\star}) = \beta\left((1-\delta)\, u_1^{'}(w-\delta d^{\star}) + u_2^{'}(d^{\star})\right)$$
$$\frac{1}{\beta}+\delta-1 = \frac{u_2^{'}(d^{\star})}{u_1^{'}(w-\delta d^{\star})}$$

Note that since u_1 and u_2 are strictly concave $u_1^{'}$ and $u_2^{'}$ are decreasing in their argument, then the numerator of the ratio above decreases with $d^{\star}$ and the denominator increases with $d^{\star}$. Then the ratio is unambiguously strictly decreasing in $d^{\star}$. Since the LHS is constant this leaves only one possible solution for the equation (the ratio can only cross $\frac{1}{\beta}+\delta-1$ once). Moreover since u_2 satisfies the inada conditions it cannot be that $d^{\star}=0$ since then $u_2^{'}(d^{\star})\to\infty$ which violates the equality, noting that $u_1^{'}(w)$ is well defined.

vi. Show that the policy functions for the solution, $c^{\star}(d)$ and $d'=g^{\star}(d)$ are increasing.

First note that since v is strictly concave the policy correspondence is single valued. The policy function for durable goods must satisfy:

$$-F_2(d,g(d)) = \beta v^{'}(g(d))$$
$$u_1^{'}(w+(1-\delta)\,d-g(d)) = \beta v^{'}(g(d))$$

Note that both $u_1^{'}$ and $v^{'}$ are decreasing in its argument. Consider $d_1<d_2$, suppose for a contradiction that $g(d_1)\geq g(d_2)$. From this we have: $\beta v^{'}(g(d_1))\leq\beta v^{'}(g(d_2))$ and hence:

$$u_1^{'}(w+(1-\delta)\,d_1-g(d_1))\leq u_1^{'}(w+(1-\delta)\,d_2-g(d_2))$$

Since $u_1^{'}$ is decreasing it must be that:

$$w+(1-\delta)\,d_1-g(d_1)\geq w+(1-\delta)\,d_2-g(d_2)$$
$$g(d_2)-g(d_1)\geq(1-\delta)(d_2-d_1)$$

Since $d_2>d_1$ this gives $g(d_2)>g(d_1)$ which is a contradiction. Hence g is strictly increasing in d.

For consumption, note that $c(d)=w+(1-\delta)\,d-g(d)$, then for $d_2>d_1$ one has $g(d_2)>g(d_1)$ and by strict concavity of v:

$$\beta v^{'}(g(d_1)) > \beta v^{'}(g(d_1))$$
$$u_1^{'}(c(d_1)) > u_1^{'}(c(d_2))$$

Since u_1 is strictly concave this implies $c(d_2)>c(d_1)$.

vii. Show that the system is globally stable. You can assume that the policy functions are differentiable for this part.

Since v is strictly concave it must be that:

$$\left[v^{'}(d_1)-v^{'}(d_2)\right][d_1-d_2]\leq 0$$

with equality only if $d_1 = d_2$. Since this is true for all d_1 and d_2 it is also true for d and $g(d)$ which gives:

$$\left[(1-\delta)\, u_1^{'}(w+(1-\delta)\, d-g(d))+u_2^{'}(d)-\frac{1}{\beta}u_1^{'}(w+(1-\delta)\, d-g(d))\right][d-g(d)] \leq 0$$
$$\left[\frac{u_2^{'}(d)}{u_1^{'}(w+(1-\delta)\, d-g(d))}-\left(\frac{1}{\beta}+\delta-1\right)\right][d-g(d)] \leq 0$$
$$\left[\frac{u_2^{'}(d)}{u_1^{'}(w+(1-\delta)\, d-g(d))}-\frac{u_2^{'}(d^\star)}{u_1^{'}(w+(1-\delta)\, d^\star-g(d^\star))}\right][d-g(d)] \leq 0$$

Now suppose that $d > d^\star$. Then $u_2^{'}(d) < u_2^{'}(d^\star)$ by strict concaveness of u_2. Suppose that $g(d) > d$ then it must be, for the inequality above to hold, that:

$$\frac{u_2^{'}(d)}{u_1^{'}(w+(1-\delta)\, d-g(d))} > \frac{u_2^{'}(d^\star)}{u_1^{'}(w+(1-\delta)\, d^\star-g(d^\star))}$$

in order for this to hold it must be that

$$u_1^{'}(w+(1-\delta)\, d-g(d)) < u_1^{'}(w+(1-\delta)\, d^\star-g(d^\star))$$

since u_1 is strictly concave this implies:

$$w+(1-\delta)\, d-g(d) > w+(1-\delta)\, d^\star-g(d^\star)$$
$$(1-\delta)\, d-g(d) > -\delta d^\star$$
$$d-g(d) > \delta(d-d^\star)$$

Since $d > d^\star$ this implies $d > g(d)$ which is a contradiction. Then it must be that $d > g(d)$. In the same way if $d < d^\star$ it must be that $d < g(d^\star)$.

Finally, since g is strictly increasing it must be that for all $d < d^\star$: $g(d) < d^\star$, and for all $d > d^\star$ that $g(d) > d^\star$. This along with the previous inequalities gives the stability result since d will converge monotonically to $d^\star$.

Part VI
Probability and Measure Theory

Consider an experiment that can have several (but finite) outcomes. For example trowing a die can turn out in getting any number from 1 to 6, or asking someone out can generate an affirmative response, a negative one or perhaps a maybe, or no response at all. A probability function is a function that assigns a value to each possible outcome while satisfying certain rules.

Its clear that since the outcomes here are finite, outcomes form a set $S = \{s_1, \ldots, s_n\}$ a probability is then a list $(\pi_1, \ldots, \pi_n)$ such that $\Pr(s_i) = \pi_i$:

i. $\pi_i \geq 0$ for all i.

ii. $\sum \pi_i = 1$.

Note that it is natural to define other outcomes that are formed by unions of the former ones, like getting an even number when trowing the die (the union of getting a two a four and a six) or getting a positive answer or a maybe when asking someone out. It is clear that the probability of these new outcomes is defined by the sum of probabilities of the original outcomes used to define them.

Formally we could say that for any set $A \subseteq S$ we define $I_A = \{i | s_i \in A\}$ and then a function $\mu : 2^S \to [0, 1]$ as:

$$\mu(A) = \Pr(A) = \sum_{i \in I_A} \pi_i$$

Furthermore we can define the expected value of a real valued function $f : S \to \mathbb{R}$ as $E[f] = \sum \mu(\{s_i\}) f(s_i)$.

This same discussion can be carried out if the possible outcomes are countably infinite, but it is difficult to generalize it otherwise. The objective now is to study which properties does this kind of function satisfy and how it is generalized to deal with cases where outcomes are arbitrary. The key to accomplish this objective is to realize that a probability is a function that maps sets into the interval $[0, 1]$, hence the study of functions that map sets into non-negative numbers will provide the necessary theory, these functions are called measures, for obvious reasons.

The following sections draw on the short exposition of measure theory contained in Chapter 7 of Stokey et al. (1989) and complements it with portions of Kolmogorov and Fomin (2012) (chapters 7 to 10). Both these references are introductory although they present all the relevant results. All the material is also covered in a more advanced manner in Kolmogorov and Fomin (1999).

As with previous sections the aim of the course is not to dwell in the mathematical details of the theory but rather present the most useful results for applications in economic theory, because of this many of the proofs will be omitted only including those that are either instructive of the way the theory is developed. Kolmogorov and Fomin (2012) is a good source for detailed (and easy to understand) proofs.

Finally Markov processes are defined following Stokey et al. (1989), chapter 8.

20 Measure

20.1 Measurable spaces (σ-algebras)

Before we define a measure recall that a measure has for domain a collection of sets. For a measure to have some desirable properties this collection of sets cannot be left unrestricted. It turns out that the appropriate family of sets to be consider is that of σ-algebra.

Definition 20.1. (σ-algebra) Let S be a set and $\mathcal{A} \subseteq 2^S$ a family of its subsets. $\mathcal{A}$ is a σ-algebra if and only if:

i. $\emptyset, S \in \mathcal{A}$.

ii. $A \in \mathcal{A}$ implies $A^c = S \backslash A \in \mathcal{A}$. We say that $\mathcal{A}$ is closed under complement.

iii. $A_n \in \mathcal{A}$ for $n = 1, \ldots$ implies $\cup A_n \in \mathcal{A}$. We say that $\mathcal{A}$ is closed under countable union.

   (a) Since $\cap A_n = (\cup A_n^c)^c$ we have that $\mathcal{A}$ is closed under countable intersection.

If $\mathcal{A}$ is only closed under finite union (or intersection) then $\mathcal{A}$ is an algebra.

A σ-algebra imposes certain consistency to the family of sets under consideration. The way to interpret it is that only subsets of the σ-algebra can be known, hence measured. Because of property (i) it is possible to know when one or all of the outcomes occurred. Also if there is an outcome that occurred it must be possible to determine if it didn't. Finally if it is possible to determine that some outcomes occurred individually it can also be determined if at least one or all of them were realized.

It is instructive to consider two simple examples of σ-algebras that arise from throwing a 4 sided die, then $S = \{1, 2, 3, 4\}$. One (trivial) σ-algebra is:

$$\mathcal{A} = \{\emptyset, S\}$$

Another one is the σ-algebra generated by the collection $\{\{1\}, \{2\}, \{3\}, \{4\}\}$, then:

$$\mathcal{A} = \left\{ \begin{array}{c} \{1\}, \{2\}, \{3\}, \{4\}, \{2,3,4\}, \{1,3,4\}, \{1,2,4\}, \{1,2,3\}, \\ \{1,2\}, \{2,3\}, \{3,4\}, \{1,3\}, \{1,4\}, \{2,4\}, \emptyset, S \end{array} \right\}$$

In this case $\mathcal{A} = 2^S$, but this is not necessarily true, imagine that one can only determine if an even number was thrown, then the outcomes are $\{\{1,3\}, \{2,4\}\}$, the σ-algebra is:

$$\mathcal{A} = \{\{1,3\}, \{2,4\}, \emptyset, S\}$$

When S has uncountably many elements this process cannot be exemplified as easily but one can always define the σ-algebra generated by a subset $\mathcal{A} \subseteq 2^S$ as the intersection of all σ-algebras that contain $\mathcal{A}$. Clearly the arbitrary intersection of σ-algebras is again a σ-algebra.

Now that we have defined a σ-algebra its possible to say what a measurable set and a measurable space are:

Definition 20.2. (Measurable Space) A pair $(S, \mathcal{A})$ where S is any set and $\mathcal{A}$ is a σ-algebra is called a measurable space. A set $A \in \mathcal{A}$ is called $\mathcal{A}$-measurable.

Note that we say that $A \subseteq S$ is measurable with respect to a σ-algebra $\mathcal{A}$ if its elements are identifiable, that is, if the outcomes represented in A can be told apart from other outcomes given the information in $\mathcal{A}$. For example the set $A = \{4\}$ is not measurable in the last example above, since its impossible to know if a 4 was the outcome of the throw.

A σ-algebra of special importance is the Borel σ-algebra.

Definition 20.3. (Borel σ-algebra) Let $S = \mathbb{R}$ and $\mathcal{A}$ be the set of open and half open intervals. The Borel algebra, noted by $\mathcal{B}$, is the σ-algebra generated by $\mathcal{A}$. A set $B \in \mathcal{B}$ is called a Borel set.

Note that the Borel algebra could have been defined equivalently with the closed and half closed intervals (use complement). In general one can define the Borel algebra for any metric space (S, ρ) as the smallest σ-algebra containing all the open balls. In the case of the Euclidean spaces it can also be generated with open rectangles.

What follows is to define the measure of a measurable set.

20.2 Measures

20.2.1 Measures in σ-algebras

Given a measurable space $(S, \mathcal{A})$ a measure is nothing but a function $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ with certain restrictions that guarantee its consistency:

Definition 20.4. (Measure) Let $(S, \mathcal{A})$ be a measurable space. A measure is an extended real-valued function $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ such that:

i. $\mu(\emptyset) = 0$

ii. $\mu(A) \geq 0$ for all $A \in \mathcal{A}$.

iii. μ is countably additive. If $\{A_n\}_{n=1}^{\infty}$ is a countable, disjoint sequence in $\mathcal{A}$, then:

$$\mu(\cup A_n) = \sum \mu(A_n)$$

If furthermore $\mu(S) < \infty$ then μ is said to be a finite measure, and if $\mu(S) = 1$ then μ is said to be a probability measure.

Definition 20.5. (Measure Space) A triple $(S, \mathcal{A}, \mu)$ where S is a set, $\mathcal{A}$ is a σ-algebra of its subsets and μ is a measure on $\mathcal{A}$ is called a measure space. The triple is called a probability space if μ is a probability measure.

An important concept is that of almost everywhere and almost surely. These are qualifiers for a given proposition that can be evaluated in sets of $\mathcal{A}$.

Definition 20.6. (Almost Everywhere and Almost Surely) Let $(S, \mathcal{A}, \mu)$ be a measure space. A proposition is said to hold almost everywhere (a.e.) or almost surely (a.s.) if there exists a set $A \in \mathcal{A}$ such that $\mu(A) = 0$ and the proposition holds in A^c.

An example of the use of a.e. or a.s. is when treating functions that are similar to each other. One can say that two functions are equivalent a.e. or that a function is continuous a.e. Then the functions f and g satisfy $f(x) = g(x)$ and $A = \{x | f(x) \neq f(y)\}$ satisfies $\mu(A) = 0$. In measure theory the behavior of functions a.e. is all that matters, then we can treat functions that have anomalies as long as those anomalies occur only in sets of measure zero.

There are some properties of a measure that are useful to keep in mind, a crucial one is used for Bayes law and the definition of conditional probability.

Proposition 20.1. *Let $(S, \mathcal{A}, \mu)$ be a measure space and $B \in \mathcal{A}$ a set. Define $\lambda : \mathcal{A} \to \overline{\mathbb{R}}$ as $\lambda(A) = \mu(A \cap B)$. Then λ is a measure on $(S, \mathcal{A})$. If in addition $\mu(B) < \infty$ then $\tilde{\lambda}$ defined as $\tilde{\lambda}(A) = {}^{\mu(A\cap B)}/_{\mu(B)}$ is a probability measure on $(S, \mathcal{A})$.*

Proof. First note that if $A, B \in \mathcal{A}$ then $A \cap B \in \mathcal{A}$, this follows from a σ-algebra being closed under countable intersection, by letting $A_1 = A$ and $A_n = B$ for $n \geq 2$ the result obtains. It is left to check the three properties of a measure:

i. $\lambda(\emptyset) = \mu(\emptyset \cap B) = \mu(\emptyset) = 0$.

ii. $\lambda(A) = \mu(A \cap B) \geq 0$.

iii. Let $\{A_n\}_{n=1}^{\infty}$ be a countable, disjoint sequence in $\mathcal{A}$, then note that the sequence $\{A_n \cap B\}_{n=1}^{\infty}$ is also disjoint and that:

$$\lambda(\cup A_n) = \mu((\cup A_n) \cap B) = \mu(\cup(A_n \cap B)) = \sum \mu(A_n \cap B) = \sum \lambda(A_n)$$

iv. If $\mu(B) < \infty$ then all the previous results hold for $\tilde{\lambda}$ by dividing everything by $\mu(B)$. Furthermore $\tilde{\lambda}(S) = \frac{\mu(S \cap B)}{\mu(B)} = \frac{\mu(B)}{\mu(B)} = 1$.

□

Another useful property is given by the following proposition, it reflects the intuitive property of measures being 'increasing':

Proposition 20.2. *Let $(S, \mathcal{A}, \mu)$ be a measure space and $A, B \in \mathcal{A}$ sets. If $A \subseteq B$ then $\mu(A) \leq \mu(B)$, if in addition μ is finite then $\mu(B \backslash A) = \mu(B) - \mu(A)$.*

Proof. Since $A \subseteq B$ there exits $C = B \backslash A = B \cap A^c$ such that $A \cup C = B$ and $A \cap C = \emptyset$. Then:

$$\mu(A) + \mu(C) = \mu(B)$$

since $\mu(C) \geq 0$ if follows that $\mu(A) \leq \mu(B)$. If μ is finite then all elements above are well defined and: $\mu(B \backslash A) = \mu(B) - \mu(A)$. □

The following property is widely used to establish properties of limits of functions, and of the Lebesgue integral:

Proposition 20.3. *Let $(S, \mathcal{A}, \mu)$ be a measure space:*

i. If $\{A_n\}$ is an increasing sequence in $\mathcal{A}$, that is, if $A_n \subseteq A_{n+1}$ for all n, then:

$$\mu(\cup A_n) = \lim \mu(A_n)$$

ii. If $\{B_n\}$ is an decreasing sequence in $\mathcal{A}$, that is, if $B_n \supseteq B_{n+1}$ for all n, then:

$$\mu(\cap B_n) = \lim \mu(B_n)$$

Proof. Stokey et al. (1989, Sec. 7.2). Satisfying these two properties makes a measure continuous. □

20.2.2 Measures in algebras and extensions

So far we have defined a measure on an σ-algebra, but a σ-algebra is usually a large collection of sets and defining a function on such a set while preserving the consistency required for a measure is not an easy task. An alternative is given by defining measures on algebras, which are smaller and less complicated collections of sets. It can be shown that these measures preserve all the desirable properties of the more complicated spaces, and also allow for an extension to σ-algebras, once the measure is properly constructed.

We start by defining a measure on an algebra.

Definition 20.7. (Measure) Let $(S, \mathcal{A})$ be a measurable space. A measure is an extended real-valued function $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ such that:

i. $\mu(\emptyset) = 0$

ii. $\mu(A) \geq 0$ for all $A \in \mathcal{A}$.

iii. If $\{A_n\}_{n=1}^{\infty}$ is a countable, disjoint sequence in $\mathcal{A}$, and $\cup A_n \in \mathcal{A}$, then:

$$\mu(\cup A_n) = \sum \mu(A_n)$$

If furthermore $\mu(S) < \infty$ then μ is said to be a finite measure, and if $\mu(S) = 1$ then μ is said to be a probability measure.

Note that condition (iii) also includes finite union of disjoint sets as a special case.

Definition 20.8. (σ-finite measure) Let S be a set, $\mathcal{A}$ an algebra of its subsets and μ a measure defined on $\mathcal{A}$. If there is a countable sequence of sets in $\mathcal{A}$, $\{A_n\}$, such that $\mu(A_n) \leq \infty$ and $S = \cup A_n$ then μ is σ-finite

It is now possible to extend the notion of this measure to a σ-algebra.

Theorem 20.1. (Caratheodory extension theorem) Let S be a set, $\mathcal{A}$ an algebra of its subsets and μ a measure defined on $\mathcal{A}$. Let $\mathcal{A}^\star$ be the smallest σ-algebra containing $\mathcal{A}$. There exists a measure $\mu^\star$ on $\mathcal{A}^\star$ such that $\mu^\star(A) = \mu(A)$ for all $A \in \mathcal{A}$.

The problem of uniqueness is also solved.

Theorem 20.2. (Hahn extension theorem) Let S be a set, $\mathcal{A}$ an algebra of its subsets, μ a measure defined on $\mathcal{A}$ and $\mathcal{A}^\star$ the minimal σ-algebra of $\mathcal{A}$. If μ is σ-finite then the extension $\mu^\star$ is unique.

To see how these theorems and the extension of a measure are used consider defining a measure on the Borel σ-algebra. It seems logical to define the measure of an interval $A = (a, b)$ as $\mu(A) = b - a$ if $b \geq a$ and $\mu(A) = 0$ otherwise (since the interval would be empty). Yet the Borel σ-algebra contains sets beyond simple intervals, and the countable union of intervals can give rise to weird sets. An answer to this problem is given by defining a measure on the Borel algebra, formed by all types of intervals and their finite unions. Defining a measure on this set seems straightforward:

i. $\mu(\emptyset) = 0$

ii. $\mu((a,b)) = \mu([a,b]) = \mu((a,b]) = \mu([a,b)) = b - a$

iii. $\mu((-\infty,\infty)) = \mu((-\infty,b]) = \mu([a,\infty)) = \infty$

iv. $\mu(\cup(a_n,b_n)) = \sum(b_n - a_n)$ if the intervals are disjoint.

The function μ can be verified to be a measure on the Borel algebra, and hence an extension to the Borel σ-algebra exists. If we restrict our attention to $S = [a,b]$ and the intervals contained in it we can define a σ-finite measure, obtaining uniqueness of the extension. This is how we can deal with complicated environments.

Once the measure is extended to the σ-algebra all the results obtained above apply.

20.2.3 Completion of a measure

One small detail is left to be checked. Sometimes there is a set $B \subseteq S$ such that $B \subseteq A \in \mathcal{A}$ and $\mu(A) = 0$, but if $B \notin \mathcal{A}$ then its measure is undefined, while it should be clearly zero. The completion of a σ-algebra to include these type of 'harmless' sets is what follows. Note that as before including sets or behaviors of measure zero is of no consequence.

Definition 20.9. (Completion of a σ-algebra) Let $(S,\mathcal{A},\mu)$ be a measure space. Define a collection $\mathcal{C}$ as:

$$\mathcal{C} = \{C \subset S | \exists_{A\in\mathcal{A}} \mu(A) = 0 \quad \wedge \quad C \subset A\}$$

The completion of σ-algebra $\mathcal{A}$ is:

$$\mathcal{A}' = \left\{B' \subseteq S \,|\, B' = (A \cup C_1) \setminus C_2 \qquad A \in \mathcal{A} \quad \wedge \quad C_1, C_2 \in \mathcal{C}\right\}$$

Note that by letting $C_1 = C_2 = \emptyset$ we get $\mathcal{A} \subseteq \mathcal{A}'$, $\mathcal{A}'$ includes all sets in 2^S that differ from a set in $\mathcal{A}$ by a set of measure 0.

Definition 20.10. (Completion of a measure) Let $(S,\mathcal{A},\mu)$ be a measure space and $\mathcal{A}'$ the completion of $\mathcal{A}$. $\mu(B') = \mu(B)$ for any $B' \in \mathcal{A}'$ that differs from $B \in \mathcal{A}$ by a set of measure 0.

The Caratheodory and Hahn extension theorems also apply for completions.

21 Measurable functions

A measurable function is a type of function for which it is possible to know (to measure) the conditions (the set) that originates certain outcomes. One can think of a function as mapping certain events in a given measure space to outcomes in another measure space. A function is measurable if the sets that induce a given outcome are measurable. Formally:

Definition 21.1. (Measurable function) Let $(S, \mathcal{A}, \mu)$ and $\left(S^{'}, \mathcal{A}^{'}, \mu^{'}\right)$ be measure spaces and $f : S \to S^{'}$ a function. f is measurable if and only if $f^{-1}\left(A^{'}\right) \in \mathcal{A}$ for all $A^{'} \in \mathcal{A}^{'}$.

A special case of notable importance is that of $\left(S^{'}, \mathcal{A}^{'}, \mu^{'}\right) = (\mathbb{R}, \mathcal{B}, \lambda)$, where λ is the Lebesgue measure on the plane. This are real valued functions. In this case the $\mathcal{B}$-measurable sets in $\mathbb{R}$ can be characterized in the following way:

Theorem 21.1. *Let $(S, \mathcal{A}, \mu)$ be a measure space and $f : S \to \mathbb{R}$. f is μ-measurable if and only if $f^{-1}((-\infty, c)) = \{x \in S | f(x) < c\} \in \mathcal{A}$ for all $c \in \mathbb{R}$.*

Proof. This theorem is stated as the definition of a real valued function f being μ-measurable in Stokey et al. (1989), but a formal proof is presented in Kolmogorov and Fomin (2012, Sec. 28, Thm. 1). It can also be stated with any of the inequalities $\geq, \leq, >, <$. □

Also when the measure space in question is a probability space one can characterize formally what a random variable is.

Definition 21.2. (Random variable) Let $(S, \mathcal{A}, \mu)$ be a probability space and $f : S \to \mathbb{R}$ a real valued function. f is a random variable if and only if f is measurable, that is, if and only if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$.

Generally it is very hard to find a function that is not measurable. The details of the example will depend on the spaces considered. For example if $f : \mathbb{R} \to \mathbb{R}$ and $\mathcal{A}$ is the set of all open (or closed) sets in $\mathbb{R}$ the definition of measurability is equivalent to that of continuity (the pre-image of an open set has to be open) and then all functions that are not continuous are not measurable. It is clear that more complete σ-algebras make more difficult to generate counterexamples. The following three results show how difficult it is to generate them:

Proposition 21.1. *Let $f : \mathbb{R} \to \mathbb{R}$.*

i. If f is continuous then f is measurable with respect to the Borel sets.

ii. If f is monotone then f is measurable with respect to the Borel sets.

Proof. Each case is proven:

i. Let f be continuous. Consider the set $f^{-1}((-\infty, c))$ for any $c \in \mathbb{R}$, note that $(-\infty, c)$ is open, since f is continuous then its pre-image is open, then it is a Borel set. Then its measurable.

ii. Let f be monotone increasing. Consider the set $f^{-1}((-\infty, c))$ for arbitrary $c \in \mathbb{R}$. Note that $f^{-1}((-\infty, c)) = (-\infty, a)$ or $f^{-1}((-\infty, c)) = (-\infty, a]$ or $f^{-1}((-\infty, c)) = (-\infty, \infty)$ or $f^{-1}((-\infty, c)) = \emptyset$ for some $a \in \mathbb{R}$. Monotonicity ensures that if $a \in f^{-1}((-\infty, c))$ and $b \leq a$ then $b \in f^{-1}((-\infty, c))$. Suppose its not, then there exists numbers $b \leq a$ such that $f(b) > c \geq f(a)$, contradicting monotonicity.

Note that all these sets are in $\mathcal{B}$, then f is $\mathcal{B}$-measurable.

□

Corollary 21.1. *The composition of measurable functions is measurable. In particular the composition of a continuous function with a measurable function is measurable.*

Proposition 21.2. *Let $S = \{s_1, s_2, \ldots\}$ be a countable set (potentially infinite) and $\mathcal{A} = 2^S$ a σ-algebra on S. Then all functions $f : S \to \mathbb{R}$ are measurable.*

Proof. The proof is immediate since the pre-image of a Borel set is a subset of S, then it belongs to $\mathcal{A} = 2^S$. □

In a more general way one can establish the measurability of a function by relating to a class of well behave 'simple' functions. The base for this class is the indicator function.

Definition 21.3. (Indicator Function) Let $(S, \mathcal{A})$ be a measurable space. An indicator function $\chi_A : S \to \mathbb{R}$ is:

$$\chi_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$$

Clearly χ_A is measurable if and only if $A \in \mathcal{A}$.

Definition 21.4. (Simple Function) Let $(S, \mathcal{A})$ be a measurable space. A simple function is a function that takes at most countably many values. When the function takes finitely many values it can be expressed as:

$$\phi(s) = \sum_{i=1}^{n} \alpha_i \chi_{A_i}(s)$$

where $\{A_i\}$ is a sequence of subsets of S and $\alpha_i \in \mathbb{R}$.

Characterizing the measurability of simple functions is slightly more complicated.

Proposition 21.3. *A simple function taking values $\{y_1, y_2, \ldots\}$is measurable if and only if the sets $A_i = \{s \in S | \phi(s) = y_n\}$ are measurable.*

Proof. Both directions are proven.

i. Let ϕ be measurable, and note that $\{y_n\} \in \mathcal{B}$, then its pre-image is measurable wrt $\mathcal{A}$.

ii. Let the sets be measurable, that is$A_i \in \mathcal{A}$, and consider $B \in \mathcal{B}$ a Borel set. Then:

$$\phi^{-1}(B) = \{s \in S | \phi(s) = y_i \in B\} = \bigcup_{y_i \in B} A_i$$

Since each $A_i \in \mathcal{A}_i$ and the union is taken over no more than countably many sets we have $\bigcup_{y_i \in B} A_i \in \mathcal{A}$ by definition of a σ-algebra. This proves measurability of $\phi^{-1}(B)$.

□

In what follows all simple functions will be considered measurable. The importance of simple functions is given by the applications of the following proposition.

Proposition 21.4. *Let $(S, \mathcal{A})$ be a measurable space and let $\{f_n\}$ be a sequence of measurable functions converging pointwise to f, that is $\lim f_n(s) = f(s)$ for all s. Then f is also measurable.*

Proof. The proof can be found in Stokey et al. (1989, Sec. 7.3) or in Kolmogorov and Fomin (2012, Sec. 28.1). □

Corollary 21.2. *If f is non-negative one can choose the sequence $\{f_n\}$ to be strictly increasing.*

Corollary 21.3. *If f is bounded one can choose the sequence $\{f_n\}$ to converge uniformly.*

The main application is the following result that gives a characterization of measurable functions in terms of simple functions:

Proposition 21.5. *A function f is measurable if and only if it an be represented as the limit of a uniformly converging sequence of measurable simple functions.*

Proof. The first direction is immediate from the previous proposition. If f is the limit of measurable functions then f is also measurable.

Let f be measurable. It is left to construct a converging sequence of simple functions that converges to f. wlog let $f(s) \geq 0$ for all s, then by the Archimedean principle there exists a non-negative integer m such that

$$\frac{m}{n} \leq f(s) < \frac{m+1}{n}$$

Let $f_n(s) = {}^{m}\!/_{n}$, since n is fixed and $m \in \mathbb{N} \cup \{0\}$ it follows that f_n can take at most countably many values, hence it is simple. f_n is also measurable since:

$$f_n^{-1}((-\infty, c)) = \{s \in S | f_n(s) \leq c\} = \left\{s \in S | f_n(s) \leq \frac{m^\star}{n}\right\} = \left\{s \in S | f_n(s) < \frac{m^\star + 1}{n}\right\}$$

For $m^\star$ chosen by the Archimedean principle. Note that the last set is $f^{-1}\left(\left(-\infty, \frac{m^\star+1}{n}\right)\right)$ which is measurable by assumption. Then f_n is measurable for all n.

Finally note that $f_n \to f$ uniformly since:

$$|f_n(s) - f(s)| \leq \left|\frac{m}{n} - \frac{m+1}{n}\right| = \frac{1}{n}$$

□

Other results will follow and are left stated without proof:

Proposition 21.6. *Let f, g be measurable functions and $\alpha \in \mathbb{R}$ then:*

i. $f + g$ *is measurable.*

ii. αf *is measurable.*

iii. fg *is measurable.*

iv. $1/f$ *is measurable provided that* $f(s) \neq 0$.

Finally continuity of functions is used to strengthen the intuition around measurability.

Proposition 21.7. *Let f, g be equivalent function defined on an interval E, that is they are equal a.e. If f and g are continuous then they coincide.*

Proof. Suppose not, then there exists $x \in E$ such that $f(x) \neq g(x)$. Let $\epsilon = |f(x) - g(x)|$, since f and g are continuous there exists δ such that for $x' \in B_\delta(x)$ it holds that $\left|f(x) - f\left(x'\right)\right| < \frac{\epsilon}{2}$ and $\left|g(x) - g\left(x'\right)\right| < \frac{\epsilon}{2}$. Then for all $x' \in B_\delta(x)$ it holds that $f\left(x'\right) \neq g\left(x'\right)$, but $B_\delta(x)$ has strictly positive measure, contradicting f and g being equivalent. □

Proposition 21.8. *A function f equivalent to a measurable function g is measurable.*

Proof. Since the functions are equivalent the sets $\{x|f(x) \leq c\}$ and $\{x|g(x) \leq c\}$ can differ in at most by a set of measure zero. Then if the second set is measurable so is the first one (taking into account the completion of the σ-algebra). This proves measurability. □

Corollary 21.4. *A function f equivalent to a continuous function is measurable.*

Proof. Immediate from continuous functions being measurable. □

This implies that if a function is continuos a.e. then it is measurable, again the behavior of functions in sets of measure zero carries no consequence. It turns out that this corollary can be strengthened. The result is powerful and is stated without a proof:

Theorem 21.2. ***(Luzin)*** *Let $f : [a, b] \to \mathbb{R}$ be a function. f is continuous if and only if for all $\epsilon > 0$ there exists a continuous function g such that $\mu\{x \in [a, b] \mid f(x) \neq g(x)\} < \epsilon$.*

This theorem shows that for the case of functions of real variable and real value measurability is equivalent to continuity, except on a set of arbitrarily small size. In other words a measurable function can be made continuous by altering its values on a set of arbitrarily small measure.

22 The Lebesgue integral

The Lebesgue integral is in at least two important ways a generalization of the Riemann integral and it serves a crucial purpose of defining what it means to take the expected value of a function with respect to a probability distribution. The first sense in which the Riemann integral is generalized is that the Lebesgue integral is defined over measurable function, a space that is much richer than that of Riemann integrable functions, the second sense is much more crucial: the Lebesgue integral is defined for functions that with domain in arbitrary sets, thus allowing to handle a more abstract and general class of functions.

Intuitively the Lebesgue integral is constructed in a similar way than the Riemann integral. To construct the latter one takes successively finer grids of the domain and evaluate the function at certain points, constructing step functions, one above the function and one below, then two sums are constructed and the value of the integral is defined as the (common) value of the limit of those sums as the length of the grid's spaces goes to zero.

The Lebesgue integral of a function $f : S \to \mathbb{R}_+$ is constructed by taking grids over the range of the function $\{y_i\}_{i=1}^n$ such that $0 = y_1 \leq \ldots \leq y_n$. Then one can define the sets $A_i = \{s \in S | y_i \leq f(s) < y_{i+1}\}$ and using the measure over S define $\lambda(A_i)$ and the sum $\sum y_i \lambda(A_i)$. The Lebesgue integral is then the limit of this sum as the values y_i are closer together.

The introduction before of simple functions makes sense when defining the Lebesgue integral. Its definition seems intuitive for this class of functions and Proposition 21.5 creates a bridge between them and the more general class of measurable functions, thus allowing to extend the Lebesgue integral to this broader family.

In what follows we restrict attention to non-negative, real valued functions.

Definition 22.1. (Lebesgue integral for simple functions) Let $(S, \mathcal{A}, \mu)$ be a measure space and $f : S \to \mathbb{R}_+$ a simple, μ-measurable function that takes no more than countably many values $\{y_1, y_2, \ldots\}$. The Lebesgue integral over the set $A \subseteq S$ is defined as:

$$\int_A f(s)\, d\mu = \sum_n y_n \mu(A_n) \tag{22.1}$$

where the sets A_n are defined as:

$$A_n = \{s \in A | f(s) = y_n\}$$

Note that these sets can be empty if there is no element of s in A for which f takes a given value. The Lebesgue integral is defined as long as the series in (22.1) is absolutely convergent. Note that if f takes finitely many values and μ is finite (or a probability measure) this condition is satisfied.

An example is given by the constant function, $f(s) = 1$ for all $s \in S$, then:

$$\int_A f(s)\, d\mu = \int_A d\mu = \mu(A)$$

It can be shown that the lebesgue integral satisfies some natural properties:

Proposition 22.1. *Let f and g be non-negative, measurable, simple and integrable functions on $(S, \mathcal{A}, \mu)$, a measure space, and $c \geq 0$ a constant. Then:*

i. $\int_A (f+g)(s)\, d\mu = \int_A f(s)\, d\mu + \int_A g(s)\, d\mu$

ii. $\int_A (cf)(s)\, d\mu = c \int_A f(s)\, d\mu$

iii. If f is bounded $|f(s)| \leq M$ a.e. then f is integrable and $\left|\int_A f(s)\, d\mu\right| \leq M\mu(A)$.

Proof. Kolmogorov and Fomin (2012, Sec. 29.1). □

Definition 22.2. (Lebesgue integral - Nonnegative functions) Let $(S, \mathcal{A}, \mu)$ be a measure space. A measurable function $f : S \to \mathbb{R}$ is said to be integrable on a set A if there exists a sequence $\{f_n\}$ of integrable simple functions converging uniformly to f on A. The Lebesgue integral is defined as:

$$\int_A f(s)\, d\mu = \lim \int_A f_n(s)\, d\mu \tag{22.2}$$

Note that this definition precludes the integral from being infinite, as shown in Kolmogorov and Fomin (2012, Sec. 29.1), the limit above exists provided that the functions f_n are integrable (recall that it was asked of the sum in (22.1) to be finite), moreover it is independent of the choice of sequence approximating f, this sequence can be furthermore be chosen to be strictly increasing (Stokey et al., 1989). A final thing is that the concept of the Lebesgue integral can be easily generalized to allow for infinite values, the definition in Stokey et al. (1989) allows for this.

What follows is a list of properties of the Lebesgue integral which should be familiar if there is any knowledge of the behavior of Riemann integrals. They are not of particular interest in this course.

Proposition 22.2. *Properties of the Lebesgue integral for non-negative measurable functions:*

i. $\int_A (f+g)(s)\, d\mu = \int_A f(s)\, d\mu + \int_A g(s)\, d\mu$

ii. $\int_A (cf)(s)\, d\mu = c \int_A f(s)\, d\mu$

iii. If g is measurable and integrable and f is bounded by g: $|f(s)| \leq g(s)$ a.e., then f is integrable and $\left|\int_A f(s)\, d\mu\right| \leq \int_A g(s)\, d\mu$.

   (a) If f is bounded and measurable then it is integrable.

iv. If $f \leq g$ a.e. then $\int f(s)\, d\mu \leq \int g(s)\, d\mu$.

v. If $A \subseteq B$ with $A, B \in \mathcal{A}$ then $\int_A f(s)\, d\mu \leq \int_B f(s)\, d\mu$

vi. Let $A = \cup A_n$ where $\{A_n\}$ is a finite or countable sequence of disjoint sets. If f is integrable on A then f is integrable on A_n for all n and:

$$\int_A f(s)\, d\mu = \sum_n \int_{A_n} f(s)\, d\mu$$

when the series on the right is absolutely convergent.

Finally it is noted that a non-negative integrable function induces a measure on a space, the following proposition makes this clear.

Proposition 22.3. *Let f be a non-negative, integrable function, then $\lambda : \mathcal{A} \to \mathbb{R}$ defined as:*

$$\lambda(A) = \int_A f(s)\, d\mu$$

is a measure on $(S, \mathcal{A})$.

Definition 22.3. (Lebesgue integral) Let $(S, \mathcal{A}, \mu)$ be a measure space. A measurable function $f : S \to \mathbb{R}$ is said to be integrable if the following two integrals are finite:

$$\int f^+(s)\, d\mu \qquad \int f^-(s)\, d\mu$$

where:

$$f^+(s) = \begin{cases} f(s) & \text{if } f(s) \geq 0 \\ 0 & \text{if } f(s) < 0 \end{cases} \qquad f^+(s) = \begin{cases} 0 & \text{if } f(s) \geq 0 \\ -f(s) & \text{if } f(s) < 0 \end{cases}$$

The integral of f is defined as:

$$\int f(s)\, d\mu = \int f^+(s)\, d\mu - \int f^-(s)\, d\mu \tag{22.3}$$

Recall that when $(S, \mathcal{A}, \mu)$ is a probability space the function f is called a random variable, the definitions above are then the definitions of the expected value of a random variable, this expected value exists when f is integrable, we have seen that a sufficient condition for this is to be bounded a.e. and the measure to be finite, this last condition is satisfied immediately by probability measures.

23 The Stieltjes integral

The Lebesgue-Stieltjes integral is a type of integral specially useful in probability theory, because of the resemblance between the Stieltjes measures and probability measures. To introduce the concept consider a real valued random variable that takes values on a closed interval $[a, b]$, this is for example the result of coin toss when catalogued as 0 or 1, the underlying probability space is formed by $S = \{H, T\}$, $\mathcal{A} = \{\emptyset, S, \{H\}, \{T\}\}$ and a probability measure on $\mathcal{A}$, a function $\mu : \mathcal{A} \to [0, 1]$ such that $\mu(\{H\}), \mu(\{T\}) \geq 0$, $\mu(S) = \mu(\{H\}) + \mu(\{T\}) = 1$ and $\mu(\emptyset) = 0$. The random variable is then a function $f : S \to \mathbb{R}$ such that $f(H) = 0$ and $f(T) = 1$. It seems natural to ask what is the probability that $f(s) = 1$, it is of course given by $\mu(T)$, in the same way can ask for the probability that $f(s) \leq c$ for any value c, the function that answers that question is called the cumulative distribution function. In this example we have:

$$F(c) = \Pr(f(s) \leq c) = \begin{cases} 0 & \text{if } c < 0 \\ \mu(H) & \text{if } 0 \leq c < 1 \\ 1 & \text{if } 1 \leq c \end{cases}$$

Since the measure μ is non-negative it is clear that F has to be a non-decreasing function, it is also continuous from the left, moreover it is possible to recover μ from knowledge of F:

$$\mu(H) = F(0) \qquad \mu(T) = 1 - F(0)$$

The Stieltjes measure is a general way of looking at this last step. It treats the problem of inducing a measure from a non-decreasing left continuous function. The application to probability theory is apparent since we deal with the CDF of a random variable, and not directly with its probability measure, as we saw before it is this latter object the one that defines the expected value.

23.1 The Stieltjes measure

Let $F : [a, b] \to \mathbb{R}$ be a non-decreasing and left-continuous function. Let $\mathcal{A}$ be an algebra of all subintervals of $[\alpha, \beta)$ (including open, closed and half-open intervals). Define a measure on $\mathcal{A}$ by:

$$\begin{aligned} m(\alpha, \beta) &= F(\beta) - F(\alpha + 0) \\ m[\alpha, \beta] &= F(\beta + 0) - F(\alpha) \\ m(\alpha, \beta] &= F(\beta + 0) - F(\alpha + 0) \\ m[\alpha, \beta) &= F(\beta) - F(\alpha) \end{aligned}$$

Now consider the Lebesgue extension of m, call it μ_F and the σ-algebra of all μ_F-measurable, call it $\mathcal{A}_F$. Note that $\mathcal{A}_F$ contains all subintervals of $[\alpha, \beta)$ and hence all the Borel sets of $[\alpha, \beta)$.

Definition 23.1. (Stieltjes measure) The measure μ_F described above is called the (Lebesgue-)Stieltjes measure and F its generating function.

This concept is easily extended to the whole real line. Some examples show the generality of this type of measure:

Example 23.1. Let $F(x) = x$, then the Stieltjes measure is nothing but the Lebesgue measure on the real line, that is, the extension of the concept of length of an interval.

Example 23.2. Let F be a jump function with discontinuity points $\{x_1, x_2, \ldots\}$ and corresponding jumps $\{h_1, h_2, \ldots\}$. The measure is of course:

$$m(\{x_n\}) = h_n \qquad m(\{x_1, x_2, \ldots\}^c) = 0$$

Then every subset of $[\alpha, \beta)$ is μ_F-measurable since their measure depends only on countable points. Any set A has measure given by:

$$\mu_F(A) = \sum_{x_n \in A} h_n$$

This number exists by assumption. A Stieltjes measure generated by a jump function is called a discrete measure. Note that all discrete random variables have CDF that are jump functions.

Example 23.3. Let F be an absolutely continuous non-decreasing function on $[\alpha, \beta)$. Absolutely continuous functions have a finite derivative a.e. let this derivative be $f = F^{'}$. Then the Stieltjes measure μ_F is defined for all Lebesgue measurable sets and:

$$\mu_F(A) = \int_A f(x)\, dx$$

clearly in this case $\mu_F(\{x\}) = 0$ since $\{x\}$ has Lebesgue measure 0.

The result follows from Lebesgue theorem:

Theorem 23.1. ***(Lebesgue)*** *If F is absolutely continuous on $[a, b]$ then the derivative $F^{'}$ is integrable on $[a, b]$ and:*

$$F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} F^{'}(x)\, dx$$

Proof. Kolmogorov and Fomin (2012, Sec. 33, Thm. 6). □

Applying this theorem here we get:

$$m(\alpha, \beta) = m[\alpha, \beta] = m(\alpha, \beta] = m[\alpha, \beta) = \int_{\alpha}^{\beta} f(x)\, dx$$

Since f is non-negative and integrable wrt all Lebesgue-measurable subsets of $[a, b]$ $\left(\mathcal{B}_{[a,b]}\right)$ we know by proposition (22.3) that

$$\mu_F(A) = \int_A f(x)\, dx$$

is a measure on $\left([a, b], \mathcal{B}_{[a,b]}\right)$ that coincides with m, since the extension is unique we get that μ_F is the Stieltjes measure we are looking for.

This type of measure is called absolutely continuous and is related to continuous random variables.

Now we can define the integral with respect to a Stieltjes measure:

Definition 23.2. (Lebesgue-Stieltjes integral) Let μ_F be Stieltjes measure with generating function F, and let g be a μ_F-measurable function, then the integral is defined as:

$$\int_a^b g(x)\,dF(x) = \int_{[a,b]} g(x)\,d\mu_F$$

If μ_F is discrete with $F(x) = \sum_{x_n \le x} h(x_n)$, then we have:

$$\int_a^b g(x)\,dF(x) = \sum_n g(x_n)\,h_n$$

If μ_F is absolutely continuous then:

$$\int_a^b g(x)\,dF(x) = \int_a^b g(x)\,f(x)\,dx$$

As hinted above in probability Stieltjes measures arise naturally. Let ξ be a random variable and define $F(x) = \Pr(\xi < x)$, then as noted above F is non-decreasing and continuous from the left, moreover $F(-\infty) = 0$ and $F(\infty) = 1$. The Lebesgue-Stieltjes measure allows us to define the expected value and variance of the random variable as:

$$E[\xi] = \int_{-\infty}^{\infty} x dF(x) \qquad V[\xi] = \int_{-\infty}^{\infty} (x - E[\xi])^2\,dF(x)$$

note that these definitions are valid for discrete and continuos random variables.

24 Markov Processes

As seen in Section 18 a great deal of problems can be expressed in a recursive setting, and the use of recursive methods can provide solution to problems that would otherwise be impossible to handle. When dealing with random variables the same topic arises, in particular one can think of a sequence made by the realizations of a random variable, since the sequence is ordered one can also think of each element of the sequence being realized sequentially, in this way its natural to consider the case in which one element of the sequence depends on the value of the previous element. More formally, when the distribution of one element of the sequence depends on the realization of the previous element. Markov processes are processes that behave in this way.

Since the objective is to introduce shocks to a dynamic program lets consider first the deterministic dynamic program of Section 18, characterized by the Bellman equation:

$$v(x) = \sup_{y \in \Gamma(x)} \{F(x,y) + \beta v(y)\} \tag{24.1}$$

The idea is to add a random variable whose realization z will affect the problem, z is a state of the problem and its drawn each period from a distribution characterized by the measure λ. Formally consider $(Z, \mathcal{Z}, \lambda)$ a probability space, then we can define the problem to be::

$$v(x,z) = \sup_{y \in \Gamma(x)} \left\{ F(x,y,z) + \beta \int v\left(y, z^{'}\right) d\lambda\left(dz^{'}\right) \right\} \tag{24.2}$$

Recall that λ maps sets of the σ-algebra $\mathcal{Z}$ to real numbers. The problem above can be solved using the results of Sections (20) to (23), but it is not general enough for our purposes since the distribution of z is fixed, and each draw is taken (each period) from the same distribution.

In general we want the distribution of $z^{'}$ to be influenced by the previous draw z, for this we need a special type of function, $Q : Z \times \mathcal{Z} \to \mathbb{R}$, such that for all $z \in Z$ it holds that $Q(z, \cdot)$ is a probability distribution for $z^{'}$. This is called a transition function and it allows to express the problem as:

$$v(x,z) = \sup_{y \in \Gamma(x)} \left\{ F(x,y,z) + \beta \int v\left(y, z^{'}\right) dQ\left(z, dz^{'}\right) \right\} \tag{24.3}$$

The objective is now to characterize transition functions and the properties of the process that they generate.

24.1 Transition functions

Definition 24.1. (Transition Function) Let $(Z, \mathcal{Z})$ be a measurable space. A transition function is a function $Q : Z \times \mathcal{Z} \to [0, 1]$ such that:

i. For each $z \in Z$ the function $Q(z, \cdot)$ is a probability measure on $(Z, \mathcal{Z})$.

ii. For each $A \in \mathcal{Z}$ the function $Q(\cdot, A)$ is a $\mathcal{Z}$-measurable function.

The interpretation is that for all current value of the random variable the transition function induces a probability measure for next period's value of the variable. Then $Q(a, A)$ is the probability that $z' \in A$ if the current value of the variable is a.

$$Q(a, A) = \Pr\left(z' \in A | z = a\right)$$

Any transition function defines two operators that will be of great importance later.

Definition 24.2. Let Q be a transition function on a measurable space $(Z, \mathcal{Z})$. Define $\mathcal{F}$ as the set of $\mathcal{Z}$-measurable functions and Λ the set of probability measures on $(Z, \mathcal{Z})$.

i. The Markov operator of Q is an operator T defined on the set of $\mathcal{Z}$-measurable functions:
$$Tf(z) = \int f\left(z'\right) Q\left(z, dz'\right)$$
for all $z \in Z$. T is the expected value of f in the next period if today's realization is z.

ii. The Adjoint operator of Q is $T^{\star}$ is an operator defined on probability measures on $(Z, \mathcal{Z})$:
$$T^{\star}\lambda(A) = \int Q(z, A)\, \lambda(dz)$$
for all $A \in \mathcal{Z}$. T gives the probability that $z' \in A$ if the current value of z is drawn from probability distribution λ.

These operators are important because they will allow to characterize the distribution of a sequence of random variables starting at some initial distribution. This is the objective when solving a stochastic dynamic programming problem. In order for T and $T^{\star}$ to be useful it is first necessary to check that they are sufficiently well behaved. The following propositions will establish that the operators can be used recursively and their proof will be instructive of how proofs go in measure theory.

Proposition 24.1. *Let $(Z, \mathcal{Z})$ be a measurable space and Q a transition function on that space with Markov operator T. Then $T : M^{+}(Z, \mathcal{Z}) \to M^{+}(Z, \mathcal{Z})$ where $M^{+}(Z, \mathcal{Z})$ is the space of nonnegative, extended variable $\mathcal{Z}$-measurable functions.*

Proof. The proof is done iteratively, first for indicator functions, then it is generalized to simple functions and then to arbitrary nonnegative measurable functions.

First note that in general for any $f \in M^{+}$ we have that Tf is a nonnegative function of extended real value, this follows immediately, then it is left to check that Tf is also measurable.

Case 1. Let $A \in \mathcal{Z}$ and $f = \chi_A$, note that χ_A is by construction measurable and nonnegative. Then:

$$Tf(z) = \int f\left(z^{'}\right) Q\left(z, dz^{'}\right) = \int \chi_A\left(z^{'}\right) Q\left(z, dz^{'}\right) = \int_A Q\left(z, dz^{'}\right) = Q(z, A)$$

Since $Q(\cdot, A)$ is measurable (as a function of z for fixed A) by definition we establish measurability of Tf.

Case 2. Let f be a simple function then there exists (finitely many) indicator functions such that: $f(z) = \sum_{i=1}^{n} a_i \chi_{A_i}(z)$. Then the Markov operator gives:

$$\begin{aligned} Tf(z) &= \int f\left(z^{'}\right) Q\left(z, dz^{'}\right) \\ &= \int \sum_{i=1}^{n} a_i \chi_{A_i}\left(z^{'}\right) Q\left(z, dz^{'}\right) \\ &= \sum_{i=1}^{n} a_i \int_A \chi_{A_i}\left(z^{'}\right) Q\left(z, dz^{'}\right) \\ &= \sum_{i=1}^{n} a_i T\chi_{A_i} \end{aligned}$$

By the previous case each $T\chi_{A_i}$ is measurable, then since sum and scalar product of measurable functions is also measurable we obtain that Tf is also measurable.

Case 3. Let f be an arbitrary nonnegative, extended real value, measurable function. By proposition 21.5 we know that since f is measurable and nonnegative it can be expressed as the limit of point-wise convergent sequence of simple functions. So for all z we have:

$$\begin{aligned} Tf(z) &= \int f\left(z^{'}\right) Q\left(z, dz^{'}\right) \\ &= \int \lim \phi_n\left(z^{'}\right) Q\left(z, dz^{'}\right) \\ &= \lim \int \phi_n\left(z^{'}\right) Q\left(z, dz^{'}\right) \\ &= \lim T\phi_n(z) \end{aligned}$$

where the third step of interchanging the limit and the integral follows from Lebesgue's Monotone Convergence theorem[8]. Finally the pointwise limit of measurable functions is a measurable function (Proposition 21.4), since $T\phi_n$ is measurable by the previous case the result obtains.

[8]The theorem states that if $\{f_n\}$ is a monotone increasing sequence of nonnegative measurable functions then that converges pointwise to f then $\int f d\mu = \lim \int f_n d\mu$. Recall from proposition 21.5 that the sequence $\{\phi_n\}$ of simple functions can be chosen to be monotone increasing.

□

Corollary 24.1. *Let $(Z, \mathcal{Z})$ be a measurable space and Q a transition function on that space with Markov operator T. Then $T : B(Z, \mathcal{Z}) \to B(Z, \mathcal{Z})$ where $B(Z, \mathcal{Z})$ is the space of bounded $\mathcal{Z}$-measurable functions.*

Proof. Let f be a bounded measurable function, then if $0 \leq f \leq m$ it holds that $0 \leq Tf \leq m$, since $Q(z, \cdot)$ is a probability measure. Then Tf is bounded. Measurability follows from the proposition above by applying it to $f = f^{+} - f^{-}$. □

Note that this allows us to apply iteratively the operator to a function since if $f \in B(Z, \mathcal{Z})$ then $Tf \in B(Z, \mathcal{Z})$, which allows to evaluate $T(Tf)$, and so on. It will also be important to apply the adjoint operator iteratively to a probability measure. The following proposition will enable us to do so.

Proposition 24.2. *Let $(Z, \mathcal{Z})$ be a measurable space and Q a transition function on that space with Adjoint operator $T^{\star}$. Then $T^{\star} : \Lambda(Z, \mathcal{Z}) \to \Lambda(Z, \mathcal{Z})$ where $\Lambda(Z, \mathcal{Z})$ is the space of probability measures on $(Z, \mathcal{Z})$.*

Proof. Let $\lambda \in \Lambda(Z, \mathcal{Z})$ and consider $T^{\star}\lambda(A) = \int Q(z, A)\, d\lambda(dz)$.

i. Since $Q(z, A) \geq 0$ for all (z, A)then $T^{\star}\lambda \geq 0$.

ii. $T^{\star}\lambda(\emptyset) = \int Q(z, \emptyset)\, \lambda(dz) = \int 0 \lambda(dz) = 0$, since $Q(z, \cdot)$ is a probability measure.

iii. $T^{\star}\lambda(Z) = \int Q(z, Z)\, \lambda(dz) = \int 1 \lambda(dz) = 1$, since $Q(z, \cdot)$ is a probability measure.

iv. It is left to show that $T^{\star}\lambda$ is countably additive. Let $\{A_i\} \subseteq \mathcal{Z}$ be a sequence of disjoint sets and $A = \cup A_i$, then:

$$\sum_{i=1}^{\infty} T^{\star}\lambda(A_i) = \sum_{i=1}^{\infty} \int Q(z, A_i)\, \lambda(dz_i) = \int \left(\sum_{i=1}^{\infty} Q(z, A_i) \right) \lambda(dz_i) = \int Q(z, A)\, \lambda(dz_i) = T^{\star}\lambda(A)$$

where $\sum Q(z, A_i) = Q(z, A)$ follows from Q being a σ-additive measure and interchange of the sum and the integral can be done because of the Lebesgue's monotone convergence theorem.

□

The following result establishes a duality between the Markov operator and its adjoint, in words it says that the expected value of a function tomorrow can be computed with either operator.

Proposition 24.3. *Let $(Z, \mathcal{Z})$ be a measurable space and Q a transition function on that space. Then for any function $f \in B(Z, \mathcal{Z})$ (or more generally $f \in M^{+}(Z, \mathcal{Z})$) it holds that:*

$$\int (Tf(z))\, \lambda(dz) = \int f\left(z'\right) T^{\star}\lambda\left(dz'\right) = \int \int f\left(z'\right) Q\left(z, dz'\right) \lambda(dz)$$

Then to obtain the expected value of function f tomorrow given a distribution λ of z today the order of integration does not matter.

Proof. Stokey et al. (1989, Sec. 8.1). $\square$

We can now define a sequence of probability measures over the sequence of random variables by iterating over Q with the Markov operator.

$$\begin{aligned} Q^{1}(z,A) &= Q(z,A) \\ &\vdots \\ Q^{n+1}(z,A) &= (TQ^{n}(\cdot,A))(z) = \int Q^{n}\left(z',A\right)Q\left(z,dz'\right) \end{aligned}$$

Then if a shock is drawn sequentially from Q the function $Q^{n}(z,A)$ will give the probability of going from initial point z to a value in set A in exactly n periods. Its easy to show that each Q^{n} is a transition function.

Finally its clear that starting from an initial probability (λ) the Adjoint operator can be used to define a sequence of probability measures $\{\lambda^{n}\}$ as $\lambda^{n} = T^{\star}\lambda^{n-1}$, we interpret λ as the distribution of the state z in the initial period and λ_{n} the (unconditional) distribution of z in the n^{th} period.

As a side note a transition function can have stronger properties that are of great use in stochastic dynamic programming:

Definition 24.3. (Feller Property) A transition function Q has the feller property if its Markov operator maps the set of continuous bounded function into itself. $T : C(Z) \to C(Z)$.

Definition 24.4. (Monotone transition functions) A transition function Q is said to be monotone if its Markov operator maps nondecreasing functions to nondecreasing functions.

24.2 Probability measures on spaces of sequences

The idea now is to study sequences of random variables and their probability distributions, this can be done using the transition function defined above.

The first task at hand is to define a probability distribution on a finite sequence of variables. For this let $(Z, \mathcal{Z})$ be a measurable space and for $t < \infty$ let $(Z^t, \mathcal{Z}^t) = (Z \times \ldots \times Z, \mathcal{Z} \times \ldots \times \mathcal{Z})$ be a product space. Now let Q be a transition function on $(Z, \mathcal{Z})$. A probability measure on the sequence given z_0, the initial value of the variable is:

Definition 24.5. (Probability measure on finite sequence) $\mu^t : Z \times \mathcal{Z}^t \to [0, 1]$ is the probability distribution for the finite sequence and its defined as:

$$\mu^t (z_0, B) = \int_{A_1} \ldots \int_{A_t} Q (z_{t-1}, dz_t) \cdot Q (z_{t-2}, dz_{t-1}) \cdots Q (z_0, dz_1)$$

where $B = A_1 \times \ldots \times A_t \in \mathcal{Z}^t$ is a rectangle in $\mathcal{Z}^t$. It can be shown that it is sufficient to define μ^t only for this type of set, since it can then be extended uniquely to measurable sets on $\mathcal{Z}^t$ by the Caratheodory and Hahn extension theorems.

The next task is to handle infinite sequences of realizations of z. To do this we need to be able to induce a σ-algebra $\mathcal{Z}^\infty$ on the set of infinite sequences and then a probability measure on that σ-algebra.

To do this define the set of finite-measurable rectangles. These sets establish outcomes for the variables for the first T periods, leaving unspecified what happens to the sequence afterwards.

Definition 24.6. (Finite-Measurable Rectangles)B is a finite measurable rectangle if its of the form:

$$B = A_1 \times \ldots \times A_T \times Z \times Z \times \ldots$$

for some finite T. Let $\mathcal{C}$ be the set of all finite measurable rectangles. Let $\mathcal{A}^\infty$ be the set of all finite unions of set in $\mathcal{C}$.

It can be shown that $\mathcal{A}^\infty$ is an algebra, then one can define $\mathcal{Z}^\infty$ to be the σ-algebra induced by $\mathcal{A}^\infty$. Then one can define a measure on finite-measurable rectangles $\mathcal{C}$ just as before, extend it to the algebra $\mathcal{A}^\infty$, and the extend the extension to $\mathcal{Z}^\infty$. This proves the existence of a measure for infinite sequences that coincides with our notion of measure for finite-measurable rectangles.

Now we can define what a stochastic process is:

Definition 24.7. (Stochastic Process) Let $(\Omega, \mathcal{F}, P)$ be a probability space. A stochastic process on $(\Omega, \mathcal{F}, P)$ is an increasing sequence of σ-algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}$, a measurable space $(Z, \mathcal{Z})$and a sequence of functions $\sigma_t : \Omega \to Z$ such that each σ_t is $\mathcal{F}_t$-measurable.

Definition 24.8. (Sample Path) Let $\omega \in \Omega$, then $(\sigma_1 (\omega), \sigma_2 (\omega), \ldots)$ is called the sample path of the stochastic process given ω.

Note that each σ_t is just a random variable that takes a value on Z given some event on Ω. In almost all cases we will have $(Z, \mathcal{Z}) = (\mathbb{R}, \mathcal{B})$. The selection of the probability space $(\Omega, \mathcal{F}, P)$ is also standard, since we are interested in the behavior of infinite sequences of the realizations of the random variable we can set $(\Omega, \mathcal{F}, P) = (Z^{\infty}, \mathcal{Z}^{\infty}, \mu(z_0, \cdot))$. The restriction that the σ-algebras are increasing follows from the draws being taken sequentially, this σ-algebras will be interpreted as possible histories, and any future history must include all of the possible previous histories from which it could have followed.

Given a stochastic process we can use probability measure P to induce measures on finite sets of sample paths.

Definition 24.9. (Probabilities on Paths) Let $C \in \mathcal{Z}^n$ we can define:

$$P_{t+1,\dots,t+n}(C) = P(\{\omega \in \Omega \mid (\sigma_{t+1}(\omega), \dots \sigma_{t+n}(\omega)) \in C\})$$

This is the probability that an event occurs and the sample path lies in C between periods $t+1$ and $t+n$.

Definition 24.10. (Stationary Stochastic Process) A stochastic process is said to be stationary if $P_{t+1,\dots,t+n}(C)$ is independent of t for all n and C. That is, if it does not matter the point in time where we start the sequence.

Definition 24.11. (Conditional probability) Let $P_{t+1,\dots,t+n}(C|a_{t-s}, \dots, a_{t-1}, a_t)$ be the conditional probability of the event $\{\omega \in \Omega \mid (\sigma_{t+1}(\omega), \dots \sigma_{t+n}(\omega)) \in C\}$ given that the event $\{\omega \in \Omega \mid \sigma_{\tau}(\omega) = a_{\tau}\}$ happened.

Now we can define what a Markov process is:

Definition 24.12. (Markov Process) A stochastic process is a Markov process if:

$$P_{t+1,\dots,t+n}(C|a_{t-s}, \dots, a_t) = P_{t+1,\dots t+n}(C|a_t)$$

for $t = 1, 2, \dots$, $n = 1, 2, \dots$, $s = 1, 2, \dots, t-1$ and $C \in \mathcal{Z}^n$.

The distribution of the path of a Markov process only depends on the last realization.

A general setting is easy to construct using a transition function Q. Let $(\Omega, \mathcal{F}, P) = (Z^{\infty}, \mathcal{Z}^{\infty}, \mu(z_0, \cdot))$ and for each T define $\mathcal{A}^T$ as the collection of all finite-measurable sets:

$$B = A_1 \times \dots \times A_T \times Z \times Z \times \dots$$

As before this forms an algebra, let $\mathcal{F}^T$ be the σ-algebra generated by $\mathcal{A}^T$. Clearly $\mathcal{F}^t \subseteq \mathcal{F}^{t+1}$. Then we can define the sequence of functions $\tilde{z}_t : \Omega \to Z$ as:

$$\tilde{z}_t(\omega) = \tilde{z}_t(a_1, a_2, \dots) = a_t$$

so that it selects the t^{th} realization of the sequence ω. These functions are clearly $\mathcal{F}^t$ measurable, since they don't contain information about future realizations of the variable.

The definition of P through Q can be used to verify that this process is a Markov process. Moreover it holds that:

$$P_{t+1}(C|a_{t-s}, \dots, a_t) = P_{t+1}(C|a_t) = Q(a_t, C)$$

for $C \in \mathcal{Z}$.

24.3 Markov chains

We now zoom into a special type of Markov process that is particularly useful in applications of dynamic programming. A Markov chain (or finite state Markov chain):

Definition 24.13. (Markov Chain) A Markov chain is a Markov process defined on a space $Z = \{z_1, \ldots, z_l\}$ with finite dimension (finitely many elements).

The relevance of Markov chains resides in two observations. First, they allow for a simple characterization of their transition function, as shown below. Second, most computational methods (and thus applications) of dynamic programming discretize the state space, effectively imposing that the space Z is finite.

Before characterizing the transition function of a Markov chain it is useful to note that the natural σ-algebra over Z is $\mathcal{Z} = 2^Z$ (the power set), and that the space of probabilities distributions over Z is formed by vectors $p \in \mathbb{R}^l$ such that $p_i \geq 0$ and $\sum_{i=1}^{l} p_i = 1$[9]. The transition function of the Markov process can be then characterized by a matrix:

Definition 24.14. (Markov Matrix / Stochastic Matrix) A square matrix $\Pi = [\pi_{ij}]$ of dimensions $l \times l$ is considered a Markov (or stochastic) matrix if $\pi_{ij} \geq 0$ for all i and j, and $\sum_{j=1}^{l} \pi_{ij} = 1$ for all i. Equivalently, if its rows are probability distributions on Z: $\pi_i \in \Delta^l$.

Note that the transition function of a Markov chain is a function $Q : Z \times \mathcal{Z} \to \mathbb{R}_+$ that gives the probability of a given set $A \in \mathcal{Z}$ given a current state z_i. We can then construct a Markov matrix by setting:

$$\pi_{ij} = Q\left(z_i, \{z_j\}\right) = \Pr\left(z_{t+1} = z_j | z_t = z_i\right)$$

So π_{ij} is interpreted as the probability that $z_{t+1} = z_j$ conditional on $z_t = z_i$. The row $\pi_i = (\pi_{i1}, \ldots, \pi_{il})$ is the conditional probability of z_{t+1}, given that $z_t = z_i$.

We can also go the other way, constructing a transition function Q from a Markov matrix Π. Let $A \in \mathcal{Z}$, since Z is finite $A = \{z_{a_1}, \ldots, z_{a_n}\}$ where a_1, …, a_n are $n \leq l$ indices. This gives:

$$Q\left(z_i, A\right) = \sum_{j=1}^{n} \pi_{ia_j}$$

We can now define the Markov operator and the adjoint Markov operator of a Markov chain making use of the Markov matrix (Markov!).

i. Recall that The Markov operator of Q is an operator T defined on the set of $\mathcal{Z}$-measurable functions:

$$Tf\left(z\right) = \int f\left(z'\right) Q\left(z, dz'\right)$$

[9] Formally $p \in \Delta^l$, where $\Delta^l = \left\{p \in \mathbb{R}^l_+ | \sum_{i=1}^{l} p_i = 1\right\}$ is the $l-1$ dimensional simplex. This same set is particularly useful in characterizing price systems in finite dimensional exchange economies.

for all $z \in Z$. T is the expected value of f in the next period if today's realization is z. For Markov chains the function f can be reduced to a row-vector $\vec{f} = (f(z_1), \ldots, f(z_l)) \in \mathbb{R}^l$, which reduces the integral to:

$$Tf(z_i) = \vec{f}\pi_i^{'}$$

more generally we have:

$$Tf = \vec{f}\Pi^{'}$$

the i^{th} element of Tf (which is an l-dimensional vector) corresponds to: $E[f(z_{t+1})|z_t = z_i]$.

ii. Recall that the adjoint operator of Q is $T^\star$ is an operator defined on probability measures on $(Z, \mathcal{Z})$:

$$T^\star\lambda(A) = \int Q(z, A)\lambda(dz)$$

for all $A \in \mathcal{Z}$. T gives the probability that $z^{'} \in A$ if the current value of z is drawn from probability distribution λ. Since the space is finite we can represent probabilities distributions as vectors in Δ^l. Let $p \in \Delta^l$ be an initial distribution on Z, we want to know the distribution on Z for the next period $(\hat{p})$:

$$T^\star p = \hat{p} = p\Pi \qquad \text{where: } \hat{p}_j = \sum_{i=1}^{l} p_i\pi_{ij}$$

The j^{th} element of $T^\star p$ (which is an l-dimensional row-vector) corresponds to the unconditional probability that $z_{t+1} = z_j$: $\Pr(z_{t+1} = z_j)$.

It shouldn't be a surprise that the Markov operator is characterized by $\Pi^{'}$ and the adjoint operator by its transpose Π[10].

As with general Markov processes there is a special interest in the limit behavior of the adjoint operator $\left(\lim\limits_{n\to\infty}\Pi^n p\right)$, in particular the existence and properties of an invariant distribution, that is $p^\star$ such that $p^\star = \Pi p^\star$ (generally $\lambda^\star = T\lambda^\star$). The problem of finding an invariant distribution is frequently cast as an eigenvector problem. Note that $p^\star$ is the eigenvector associated with any unit-eigenvalue of Π.

Another property that will be of interest is the presence of Ergodic sets. These are subsets of the space $E \subseteq Z$ that the process never leaves once it takes a value in them. Formally:

Definition 24.15. (Ergodic Set) A set $E \subseteq Z$ is ergodic if and only if $Q(z_i, E) = 1$ for all $z_i \in E$ and there does not exist a proper subset $E^{'} \subset E$ that is ergodic.

The ergodic sets are important because they tell us sections of the state space that are of interest. Only ergodic sets have positive mass in the invariant distribution.

Following SLP we now show 5 examples of the possible limit behavior of Markov chains. After the examples we state the main results on the existence and uniqueness of ergodic sets, invariant distributions, and the convergence of the sequences $\left\{\frac{1}{n}\sum\limits_{k=0}^{n}\Pi^k\right\}$ and $\{\Pi^n\}$. (clearly if the second sequence converges so does the first one).

[10] If vectors are assumed to be columns instead of rows then $Tf = \Pi f$ and $T^\star p = \Pi^{'} p$. The adjoint is characterized as the transpose of the Markov operator in any case.

Example 24.1. (Uniqueness of ergodic set, convergence of $\{\Pi^n\}$) Let $l = 2$ and consider the Markov matrix:

$$\Pi = \begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}$$

Clearly the only ergodic set is Z since one has positive probability of going to z_1 or z_2 starting in any state. Moreover:

$$\lim_{n\to\infty} \Pi^n = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

The invariant distribution is then $p^\star = (1/2, 1/2)$. Moreover $\lim (\Pi^n p_0) = p^\star$ for all $p_0 \in \Delta^2$.

Note: Convergence is easily defined in this setup since the limit is taken element wise.

Example 24.2. (Uniqueness of a ergodic set, convergence of $\{\Pi^n\}$) Let $l = 3$ and $\gamma \in (0, 1)$. Consider the Markov matrix:

$$\Pi = \begin{bmatrix} 1-\gamma & \gamma/2 & \gamma/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

There is a unique ergodic set $E = \{z_2, z_3\} \neq Z$. Note that the state z_1 is never reached again once you leave it. One can also show:

$$\Pi^n = \begin{bmatrix} (1-\gamma)^n & (1-(1-\gamma)^n)/2 & (1-(1-\gamma)^n)/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

clearly $\{\Pi^n\}$ converges:

$$\lim_{n\to\infty} \Pi^n = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

and the invariant distribution is $p^\star = (0, 1/2, 1/2)$.

Example 24.3. (Cyclical sets , convergence of $\left\{\frac{1}{n}\sum_{k=0}^{n}\Pi^k\right\}$) Consider an l-dimensional Markov chain and order its states into two subsets, the first one with k elements and the second one with $l-k$ elements. Suppose the Markov matrix has the form:

$$\Pi = \begin{bmatrix} 0 & \Pi_1 \\ \Pi_2 & 0 \end{bmatrix}$$

where the first matrix Π_1 is of dimension $k \times (l-k)$ and matrix Π_2 of dimension $(l-k)\times k$. Clearly if at one period one is in the first subset the next period one will be in the second subset, and vice-versa. So there are no proper subsets that form an ergodic subset, instead the process cycles from the first subset to the second period by period.

$$\Pi^{2n} = \begin{bmatrix} (\Pi_1\Pi_2)^n & 0 \\ 0 & (\Pi_2\Pi_1)^n \end{bmatrix} \qquad \Pi^{2n+1} = \begin{bmatrix} 0 & (\Pi_1\Pi_2)^n \Pi_1 \\ (\Pi_2\Pi_1)^n \Pi_2 & 0 \end{bmatrix}$$

In this example the sequence $\{\Pi^n\}$ does not converge but its odd and even elements do, then the sequence $\left\{\frac{1}{n}\sum_{k=0}^{n}\Pi^k\right\}$ does converge.

For example if $l=4$, $k=2$ and $\Pi_1=\Pi_2$:

$$\lim_{n\to\infty}\Pi^{2n}=\begin{bmatrix} 1/2 & 1/2 & 0 & 0\\ 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1/2 & 1/2\\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} \qquad \lim_{n\to\infty}\Pi^{2n+1}=\begin{bmatrix} 0 & 0 & 1/2 & 1/2\\ 0 & 0 & 1/2 & 1/2\\ 1/2 & 1/2 & 0 & 0\\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\Pi^k=\begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4\\ 1/4 & 1/4 & 1/4 & 1/4\\ 1/4 & 1/4 & 1/4 & 1/4\\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

An invariant distribution is found as one of the rows of the last limit: $p^\star=(1/4,1/4,1/4,1/4)$.

Example 24.4. (Two ergodic sets, Infinitely many invariant distributions) Consider an l-dimensional Markov chain and order its states into two subsets, the first one with k elements and the second one with $l-k$ elements. Suppose the Markov matrix has the form:

$$\Pi=\begin{bmatrix} \Pi_1 & 0\\ 0 & \Pi_2 \end{bmatrix}$$

where the first matrix Π_1 is of dimension $k\times(l-k)$ and matrix Π_2 of dimension $(l-k)\times k$. Clearly once the process enters the first subset it never leaves it. The same goes for the second subset. Then they are both ergodic. Also $\Pi^n=\begin{bmatrix} \Pi_1^n & 0\\ 0 & \Pi_2^n \end{bmatrix}$, this sequence converges if and only if $\{\Pi_1^n\}$ and $\{\Pi_2^n\}$ converge. Let $l=4$, $k=2$ and $\Pi_1=\Pi_2=\begin{bmatrix} 3/4 & 1/4\\ 1/4 & 3/4 \end{bmatrix}$, then:

$$\lim_{n\to\infty}\Pi^{n}=\begin{bmatrix} 1/2 & 1/2 & 0 & 0\\ 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1/2 & 1/2\\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

There are two invariant distributions:$p_1^\star=(1/2,1/2,0,0)$ and $p_2^\star=(0,0,1/2,1/2)$. But any convex combination of them is also an invariant distribution.

Example 24.5. (Two ergodic sets, Infinitely many invariant distributions) Let $l=3$, $\gamma\in(0,2)$ and $\alpha,\beta\geq 0$ such that $\alpha+\beta=1$. Consider the Markov matrix:

$$\Pi=\begin{bmatrix} 1-\gamma & \gamma\alpha & \gamma\beta\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}$$

As in the second example s_1 is a transient state (once you leave it you never come back), but there are now two ergodic sets $\{s_2\}$ and $\{s_3\}$. We also have:

$$\lim_{n\to\infty}\Pi^n=\lim_{n\to\infty}\begin{bmatrix} (1-\gamma)^n & (1-(1-\gamma)^n)\,\alpha & (1-(1-\gamma)^n)\,\beta\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}=\begin{bmatrix} 0 & \alpha & \beta\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}$$

The sequence $\{\Pi^n\}$ converges and there are two invariant distributions $p_1^\star = (0,1,0)$ and $p_2^\star = (0,0,1)$ given by the second and third rows of the limiting matrix. The first row is a convex combination of the limiting distribution.

Now we turn to the general results. The following theorem encompasses all the possible outcomes of a Markov chain. In particular, an ergodic set and a limit distribution always exist, but they need not be unique, and although the sequence $\{\Pi^n\}$ need not converge, the sequence $\left\{\frac{1}{n}\sum_{k=0}^{n}\Pi^k\right\}$ always converges, and its limit gives away the invariant distributions.

Theorem 24.1. *Let $Z = (z_1, \ldots, z_l)$ and denote the stochastic matrix by its elements: $\Pi = [\pi_{ij}]$. The powers of Π are also denoted by its elements $\Pi^n = \left[\pi_{ij}^{(n)}\right]$.*

i. Z can be partitioned into $M \geq 1$ ergodic sets and a transient set (an ergodic set always exists).

ii. The sequence $\left\{\frac{1}{n}\sum_{k=0}^{n}\Pi^k\right\}$ always converges to a Markov matrix $\overline{\Pi}$.

(a) For any $p_0 \in \Delta^l$ and $p_k = p_0\Pi^k$ it holds that: $\frac{1}{n}\sum_{k=0}^{n} p_k \to p_0\overline{\Pi}$.

iii. Each row of Q is an invariant distribution, and every invariant distribution is a convex combination of the rows of $\overline{\Pi}$ (so $p_0\overline{\Pi}$ is an invariant distribution for all $p_0 \in \Delta^l$).

We can strengthen these results by imposing extra structure on Π. We can get uniqueness of the ergodic set and the invariant distribution under a "reachability" condition (at least one state should be reachable in finite time starting from anywhere).

Theorem 24.2. *Let Z and Π as in Theorem 24.1. Π has a unique ergodic set if and only if there exists a state z_j such that for all $i \in (1, \ldots, l)$ there exist $n \geq 1$ such that $\pi_{ij}^{(n)} > 0$.*

Moreover, if this is the case Π has a unique invariant distribution $p^\star$ and all rows of $\overline{\Pi}$ are equal to $p^\star$ (so for any $p_0 \in \Delta^l$ we have $p_0\overline{\Pi} = p^\star$).

The previous result still does not rule out cyclicality in the ergodic set. We can get this under a "mixing" condition.

Theorem 24.3. *Let Z and Π as in Theorem 24.1. For $n = 1, 2, \ldots$ and $j = 1, \ldots, l$ define $\epsilon_j^{(n)} = \min_i \pi_{ij}^{(n)}$ and $\epsilon^{(n)} = \sum_{j=1}^{l}\epsilon_j^{(n)}$. Z has a unique ergodic set without cyclically moving subsets if and only if for some $N \geq 1$ it holds that $\epsilon^{(N)} > 0$.*

Moreover, if this is the case Π has a unique invariant distribution $p^\star$ and the sequence $\{\Pi^n\}$ converges (so for any $p_0 \in \Delta^l$ we have $\lim p_0\Pi^n = p^\star$).

Note that $\pi_{ij}^{(n)}$ is the probability that state j is reached from state i in n steps. $\epsilon_j^{(n)}$ measures the lowest such probability, since we don't know from which state we start from we need to know that the condition is satisfied for all states. This mixing property is stronger than the first one since we need there to be at least one column of non-zero elements, guaranteeing mixing towards one state (j) starting from any state (i), uniformly in time (the same time (N) for all the initial states).

24.4 Weak convergence of monotone Markov processes

In the previous section we obtained results characterizing the limiting behavior of Markov processes when the state space is finite. Unfortunately these results do not extend immediately to general Markov processes. The problem at hand is to establish when the sequence of distribution functions $\{\lambda_n\}$, constructed as $\lambda_n = T^{\star}\lambda_{n-1}$ with λ_0 given, converges. We must first define what it means for a sequence of distributions to converge. The simplest definition comes in the form of set-wise convergence (the equivalent of point-wise convergence for functions):

Definition 24.16. (Set-wise Convergence) Let $(Z, \mathcal{Z})$ be a measurable space and $\Lambda(Z, \mathcal{Z})$ the set of probability distributions. Consider a sequence $\{\lambda_n\} \subseteq \Lambda(Z, \mathcal{Z})$, we say that $\lambda_n \overset{s.w}{\to} \lambda \in \Lambda(Z, \mathcal{Z})$ if $\lambda_n(A) \to \lambda(A)$ for all $A \in \mathcal{Z}$.

This notion of convergence is intuitive but it turns out to be too strong for most applications. The following proposition shows why:

Definition 24.17. Let $(Z, \mathcal{Z})$ be a measurable space and $\Lambda(Z, \mathcal{Z})$ the set of probability distributions. Consider a sequence $\{\lambda_n\} \subseteq \Lambda(Z, \mathcal{Z})$. $\{\lambda_n\}$ converges set-wise to λ if and only if $\lim \int f(z)\, d\lambda_n = \int f(z)\, d\lambda$ for all bounded and measurable functions $f \in B(Z, \mathcal{Z})$.

Thus asking for set-wise convergence requires the expected value of a large class of functions to converge. A way to weaken this is to limit the space of functions for which convergence is required.

Definition 24.18. (Weak Convergence) Let (Z, ρ) be a metric space and $\mathcal{Z}$ the Borel set of Z. Define $\Lambda(Z, \mathcal{Z})$ as the set of probability distributions. Consider a sequence $\{\lambda_n\} \subseteq \Lambda(Z, \mathcal{Z})$, we say that $\{\lambda_n\}$ converges weakly to $\lambda \in \Lambda(Z, \mathcal{Z})$ if $\lim \int f(z)\, d\lambda_n = \int f(z)\, d\lambda$ for all bounded and continuous functions $f \in C(Z)$.

The main results we will obtain establish the existence of an invariant distribution under a continuity assumption on the Markov operator (the Feller property). We can then ensure uniqueness if the Markov operator is monotone and a mixing condition is satisfied, along with uniqueness we will obtain the weak convergence of $\{T^{\star n}\lambda_0\}$.

In what follows we consider $Z \subseteq \mathbb{R}^l$ for $l < \infty$, with $\mathcal{Z}$ the Borel σ-algebra of Z. The Markov process is characterized by its transition function Q, its Markov operator $T : B(Z, \mathcal{Z}) \to B(Z, \mathcal{Z})$ and its adjoint operator $T^{\star} : \Lambda(Z, \mathcal{Z}) \to \Lambda(Z, \mathcal{Z})$. We also define the inner product $\langle f, \lambda \rangle = \int f(z)\, d\lambda$.

We first expand on the Feller property through the following proposition:

Proposition 24.4. *The following three statements are equivalent:*

i. *(Feller property) If $f \in C(Z)$ then $Tf \in C(Z)$.*

ii. *If $z_n \to z$ then $Q(z_n, \cdot) \to Q(z, \cdot)$ (that is for all $A \in \mathcal{Z}$).*

iii. *If $\lambda_n \to \lambda$ then $T^{\star}\lambda_n \to \lambda$*

Then preserving continuity in conditional expected values (Tf is a conditional expected value) has equivalent statements in terms of the conditional distributions ($Q(z_n,\cdot)$) and unconditional distributions ($T^{\star}\lambda_n$). It turns out that continuity is enough to guarantee that an invariant distribution exists.

Theorem 24.4. *If $Z \subseteq \mathbb{R}^l$ is compact and Q satisfies the Feller property then an invariant distribution exists. That is, there is $\lambda^{\star} \in \Lambda(Z,\mathcal{Z})$ such that $\lambda^{\star} = T^{\star}\lambda^{\star}$.*

Yet, continuity is not enough to rule out the existence of many invariant distributions or of cycling sets. Monotonicity is needed for this. As before it is first useful to take a detour on what monotonicity (as in Definition 24.4) implies for distribution functions. We then have to impose an ordering of distribution functions to be able to talk about monotonicity.

Definition 24.19. (First Order Stochastic Dominance) A distribution μ (first order stochastically) dominates λ ($\mu \geq \lambda$) if $\int f(z)\,d\mu \geq \int f(z)\,d\lambda$ for all increasing, bounded and measurable function f.

In what follows we call a sequence $\{\lambda_n\}$ monotone if $\lambda_{n+1} \geq \lambda_n$ for all n, or if $\lambda_{n+1} \leq \lambda_n$ for all n. We can now establish the following result:

Proposition 24.5. *The following three statements are equivalent:*

i. *(Monotone property) If $f \in B(Z,\mathcal{Z})$ is weakly increasing then Tf is also weakly increasing.*

ii. *Let $\lambda,\mu \in \Lambda(Z,\mathcal{Z})$. If $\mu \geq \lambda$ then $T^{\star}\mu \geq T^{\star}\lambda$.*

iii. *If $z \geq z'$ then $Q(z,\cdot) \geq Q\left(z',\cdot\right)$ (that is for all $A \in \mathcal{Z}$).*

The last statement is particularly useful, since it translates monotonicity of the Markov operator directly into monotonicity of the transition function ("better" states lead to "better" distributions).

Now we introduce the final condition needed for the main result of this section. It is a mixing condition akin to that in Theorems 24.2 and 24.3, along with a restriction on the form of the set Z. To see why it is necessary to go SLP exercises 12.12 and 12.13.

Assumption. *The set $Z = [a,b]$ is a closed and bounded rectangle in $\mathbb{R}^l$ characterized by a and b*[11]*, and there exists $z \in Z$, $\epsilon > 0$ and $N \geq 1$ such that:*

$$Q^N(a,[c,b]) \geq \epsilon \qquad Q^N(b,[a,c]) \geq \epsilon$$

Under this assumption it is possible to reach the "upper" region of the rectangle, $[c,b]$, in finite time starting from the "lower" corner (a), and it is possible to reach the "lower" region of the rectangle, $[a,c]$, in finite time starting from the "upper" corner (b). It is possible to show that if one can move through the set from the corners it is possible to do it from anywhere (under a monotonicity assumption).

[11] A set $Z \subseteq \mathbb{R}^l$ is a closed and bounded rectangle if there are two vectors $a,b \in \mathbb{R}^l$ such that$a \leq b$ and $Z = [a_1,b_1] \times \ldots \times [a_l,b_l]$.

Proposition 24.6. *Let Q satisfy monotonicity and the previous assumption for some tuple (c, ϵ, N), then:*

$$Q^N(z, [c, b]) \geq \epsilon \qquad Q^N(z, [a, c]) \geq \epsilon \qquad \textit{for all } z \in Z$$

Finally we establish the convergence result.

Theorem 24.5. *Let $S = [a, b] \in \mathbb{R}^l$ be a rectangle and satisfy the assumption above. If Q is monotone and satisfies the Feller property, then Q has a unique invariant distribution $\lambda^\star$ and $T^{\star n}\lambda_0 \to \lambda^\star$ for all $\lambda_0 \in \Lambda(Z, \mathcal{Z})$.*

25 Laws of Large Numbers

References

Fudenberg, D. and Tirole, J. (1991). *Game Theory*. MIT Press.

Irigoyen, C., Rossi-Hansberg, E., and Wright, M. (2009). *Solutions Manual for 'Recursive Methods in Economic Dynamics'*. Harvard University Press.

Kolmogorov, A. and Fomin, S. (1999). *Elements of the Theory of Functions and Functional Analysis*. Number v. 1 in Dover books on mathematics. Dover.

Kolmogorov, A. and Fomin, S. (2012). *Introductory Real Analysis*. Dover Books on Mathematics. Dover Publications.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.

Milgrom, P. and Segal, I. (2002). Envelope Theorems for Arbitrary Choice Sets. *Econometrica*, 70(2):583–601.

Morris, S. and Shin, H. S. (2003). Global ggame: Theory and applications. In Dewatripont, M., Hansen, L., Turnovsky, S., and Congress, E. S. W., editors, *Advances in Economics and Econometrics: Theory and Applications : Eighth World Congress*, Econometric Society monographs. Cambridge University Press.

Rockafellar, R. (1997). *Convex Analysis*. Convex Analysis. Princeton University Press.

Stokey, N., Lucas, R., and Prescott, E. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

Sundaram, R. (1996). *A First Course in Optimization Theory*. Cambridge University Press.

Topkis, D. (1998). *Supermodularity and Complementarity*. Frontiers of Economic Research. Princeton University Press.

Wade, W. (2010). *An Introduction to Analysis*. Featured Titles for Real Analysis Series. Prentice Hall/Pearson.