Lecture 1: Preliminaries and some important results in elementary real analysis

Liu Peng
August 5, 2015

1 Real Line

Definition 1. *We use the following traditional notations:*
Natural numbers: $\mathbb{N} = \{1, 2, 3, \cdots\}$.
Integers: $\mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$.
Rational numbers: $\mathbb{Q} = \left\{x|x = \frac{p}{q}, p, q \in \mathbb{Z}, q \neq 0\right\}$.

This section will review the introduction of real line as an ordered field.

Definition 2. *A **field** is a set F with two operations, called* addition *and* multiplication*, which satisfy the following so-call "field axioms" (A), (M), and (D):*

(A) Axioms for addition

(A1) If $x \in F$ and $y \in F$, then their sum $x + y \in F$. (Closedness for addition.)

(A2) Addition is commutative: $x + y = y + x$ for all $x, y \in F$.

(A3) Addition is associative: $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$.

(A4) F contains an element 0 such that $0 + x = x$ for every $x \in F$. (Existence of identity elements.)

(A5) To every $x \in F$ corresponds an element $-x \in F$ such that $x + (-x) = 0$. (Existence of inverse elements.)

(M) Axioms for multiplication

(M1) If $x \in F$ and $y \in F$, then their product $x \cdot y \in F$. (Closedness for multiplication.)

(M2) Multiplication is commutative: $x \cdot y = y \cdot x$ for all $x, y \in F$.

(M3) Multiplication is associative: $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for all $x, y, z \in F$.

(M4) F contains an element $1 \neq 0$ such that $1 \cdot x = x$ for all $x \in F$. (Existence of identity elements.)

(M5) If $x \in F$ and $x \neq 0$ then their exists an element $1/x \in F$ such that $x \cdot (1/x) = 1$. (Existence of inverse elements.)

(D) The distributive law

$x \cdot (y + z) = x \cdot y + x \cdot z$ holds for all $x, y, z \in F$.

Question 1. *Is $(\mathbb{N}, +, \cdot)$ a field? $(\mathbb{Z}, +, \cdot)$? $(\mathbb{Q}, +, \cdot)$?*

Definition 3. *Let S be a set. An **order** on S is a relation, denoted by $<$, with the following two properties:*

*(i) **Completeness:** If $x \in S$ and $y \in S$ then one and only one of the statements $x < y$, $x = y$, $y < x$ is true.*

*(ii) **Transitivity:** If $x, y, z \in S$, then $x < y$ and $y < z$ imply $x < z$.*

The statement $x < y$ may be read as "x is smaller than y". The notation $\leqslant$ indicates that $x < y$ or $x = y$, without specifying which of these two is to hold. In other words, $x \leqslant y$ is the negation of $y < x$.

Definition 4. *An **ordered set** is a set S in which an order is defined.*

Example 1. *$\mathbb{Q}$ is an ordered set if $r < s$ is defined to mean that $s - r$ is a positive rational number.*

Definition 5. *Suppose S is an ordered set, $E \subset S$. If there exists a $\beta \in S$ such that $\beta \geqslant x$ for every $x \in E$, we say that E is bounded above, and call β an **upper bound** of E.*

Definition 6. *Suppose S is an ordered set, $E \subset S$, and E is bounded above. We call $\alpha \in S$ the **least upper bounded of** E or the **supremum of** E if (i) α is an upper bound of E, and (ii) if $\gamma < \alpha$ then γ is not an upper bound of E. We write $\gamma = \sup E$.*

Theorem 1. *There exists an ordered field $(\mathbb{R}, +, \cdot; <)$ which has the **least-upper-bound property**: for any $E \subset \mathbb{R}$, E is nonempty and bounded above, $\sup E$ exists in $\mathbb{R}$. Moreover, $\mathbb{R}$ contains $\mathbb{Q}$ as a subfield.*

Remark 1. *The ordered field $(\mathbb{Q}, +, \cdot; <)$ does not satisfy the least-upper-bound property. Here is a loose proof by providing a counter example.*

Notice that any decimal number with finitely many digits after the decimal point is a rational number. In addition, we know that $\sqrt{2}$ is not a rational number. Hence, written in decimal format, $\sqrt{2}$ must involve infinitely many digits after the decimal point: $\sqrt{2} = 1.41421356237\ldots$.

Let's define a sequence $\{a_n\}$ such that a_n rounds $\sqrt{2}$ to its n decimal places. It's clear that $a_n \in \mathbb{Q}$ and $a_n \to \sqrt{2} \notin \mathbb{Q}$: rational number set does not satisfy the least-upper-bound property.

2 Metric Space

Definition 7. *A set V is said to be a **metric space** if with any two elements x and y of V there is an associated real number $d(x, y) \in \mathbb{R}$, called the* distance *between x and y, such that*

*(a) **Positivity**: $d(x, y) \geqslant 0$, with equality iff $x = y$.*

*(b) **Symmetry**: $d(x, y) = d(y, x)$.*

*(c) **Triangle Inequality**: $d(x, z) \leqslant d(x, y) + d(y, z)$.*

*Any function with these three properties is called a distance function, or a **metric**.*

Example 2. *Let V be any nonempty set. A trivial way of making V a metric space is to use the metric $d : V \times V \to \mathbb{R}$, which is defined by*

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

*It's easy to check that (V, d) is indeed a metric space. Here d is called the **discrete metric** on V, and (V, d) is called a **discrete space**.*

Example 3. *Let $V = \mathbb{R}^n$. Define the metric d_p on V by*

$$d_p(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \qquad for \quad 1 \leqslant p < \infty \tag{2}$$

*d_p is called the **p−metric** on $\mathbb{R}^n$. When $n = 1$, x and y are both scalars and $d_p(x, y) = |x-y|$. When $p = 2$, the p-metric d_p on $\mathbb{R}^n$ is called the **Euclidian metric**. Accordingly, $(\mathbb{R}^2, d_2)$ is called the **n-dimensional Euclidian space**. The Euclidian metric is related to the Euclidian norm $||\cdot||$ through the identity $d(x, y) = ||x - y||$.* [1]

Throughout this and the following four lectures, when we refer to $\mathbb{R}^n$ without specifying a metric, you should understand that we view this set as the Euclidian space.

3 Limit and Complete Metric Space

A *sequence* in a set S is the specification of a point $x_k \in S$ for each integer $k \in \{1, 2, \cdots\}$. The sequence is usually written as $x_1, x_2, x_3, \cdots$ or, more compactly, simply as $\{x_k\}$.

Definition 8. *Let (V, d) be a metric space. A sequence of points $\{x_k\}$ in V is said to **converge to a limit** x, denoted $\lim_{k\to\infty} x_k = x$, if the distance $d(x_k, x)$ tends to zero as k goes to infinity, i.e., if for all $\varepsilon > 0$, there exists an integer $\delta(\varepsilon)$ such that for all $k > \delta(\varepsilon)$, we have $d(x_k, x) < \varepsilon$.*

We investigate some properties of limits.

Theorem 2. *A sequence can have at most one limit.*

Proof. Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$ converging to $x \in \mathbb{R}^n$. Suppose the statement is not true, i.e., the sequence converges also to $y \in \mathbb{R}^n$ and $x \neq y$.

By the definition of convergence, for any $\varepsilon > 0$, there exists a δ such that $d(x, x_k) < \varepsilon/2$ and $d(y, x_k) < \varepsilon/2$ hold at the same time for any $k > \delta$. Then triangle inequality gives

$$d(x, y) \leqslant d(x, x_k) + d(y, x_k) < \varepsilon \tag{3}$$

This contradicts with positivity, which gives $d(x, y) > 0$. (What is exactly the contradiction?) $\square$

A sequence $\{x_k\}$ in $\mathbb{R}^n$ is called **bounded** if there exists a real number M such that $||x_k|| \leqslant M$ for all k.

Theorem 3. *Every convergent sequence in $\mathbb{R}^n$ is bounded.*

[1] The **Euclidian norm** of an arbitrary $x \in \mathbb{R}^n$, denoted $||x||$, is defined as a function from $\mathbb{R}^n$ into $\mathbb{R}$:

$$||x|| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

.

Proof. Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$ converging to $x \in \mathbb{R}^n$. There exists a δ such that $||x_k - x|| < 1$ for all $k > \delta$. Let

$$M = \max\{||x_1||, ||x_2||, \cdots, ||x_\delta||, ||x|| + 1\}. \tag{4}$$

We have $||x_k|| \leqslant M$ for all k. (What guarantees the existence of M?) □

Theorem 4. *Let $\{x_k\}$ be a sequence in $\mathbb{R}$ converging to a limit x. Suppose that for every k, we have $a \leqslant x_k \leqslant b$. Then we have $a \leqslant x \leqslant b$.*

Proof. Left as an exercise. □

We now discuss briefly a major property that we often consider for a metric space, the property of **completeness**. A quick review of **Cauchy sequences** is a prerequisite for this, so we start with that.

Definition 9. *A sequence $\{x_k\}$ in a metric space V is called a **Cauchy sequence** if for any $\varepsilon > 0$, there exists an $M \in \mathbb{R}$ such that $d(x_m, x_n) < \varepsilon$ for all $m, n \geqslant M$.*

Theorem 5. *Let $\{x_k\}$ be a sequence in a metric space V.*

1. *If $\{x_k\}$ is convergent, then it is Cauchy.*

2. *If $\{x_k\}$ is Cauchy, then it is bounded, but it needs not converge in V*

Proof. 1: $d(x_m, x_n) \leqslant d(x_m, \lim x_k) + d(x_n, \lim x_k) \to 0$ as $m, n \to \infty$.

2: Consider a counter example: let $V = (0, 1]$, and the sequence be $\{1, \frac{1}{2}, \frac{1}{3}, \cdots\}$. This sequence is a Cauchy sequence, but it does not converge in this space. □

Definition 10. *A metric space V is said to be **complete** if every Cauchy sequence in V converges to a point in X.*

Example 4. *As we have seen, the sequence $\{1, \frac{1}{2}, \frac{1}{3}, \cdots\}$ makes metric space $(0, 1]$ incomplete.*

Example 5. *As shown in Remark 1, rational number set $\mathbb{Q}$ is incomplete. When the "gaps" between rational numbers are filled by irrational numbers, $\mathbb{R}$ is complete.*

4 Continuity and Differentiability

Definition 11. *Let $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^l$. Then, $f : S \to T$ is said to be **continuous at** $x \in S$ if for all sequences $\{x_k\}$ in S such that $\lim_{k\to\infty} x_k = x$, we have $\lim_{k\to\infty} f(x_k) = f(x)$.*

The function $f : S \to T$ is said to be **continuous** on S if it is continuous at all points in S.

Definition 12. *Let $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^l$. Then, $f : S \to T$ is said to be **differentiable at a point** $x \in S$ if there exists an $l \times n$ matrix A such that for all $\varepsilon > 0$, there is $\delta > 0$ such that $t \in S$ and $||t - x|| < \delta$ implies $||f(t) - f(x) - A(t - x)|| < \varepsilon ||t - x||$.*

Equivalently, f is differentiable at $x \in S$ if

$$\lim_{t\to x} \left(\frac{||f(t) - f(x) - A(t - x)||}{||t - x||} \right) = 0.$$

The matrix A in this case is called the **derivative of** f **at** x, and is denoted $Df(x)$. If f is differentiable at all points in S, then f is said to be differentiable on S. When f is differentiable on S, the derivative Df itself forms a function from S to $\mathbb{R}^{l\times n}$. If $Df : S \to \mathbb{R}^{l\times n}$ is a continuous function, then f is said to be **continuously differentiable** on S, and we call f is C^1.

The difference between differentiability and continuous differentiability is non-trivial. The following example shows that a function may be differentiable everywhere, but may still not be continuously differentiable.

Example 6. *Let $f : \mathbb{R} \to \mathbb{R}$ be given by*

$$f(x) = \begin{cases} 0, & if \quad x = 0 \\ x^2 sin(1/x^2), & if \quad x \neq 0 \end{cases} \tag{5}$$

For $x \neq 0$, we have

$$f'(x) = 2x\sin\left(\frac{1}{x^2}\right) - \left(\frac{2}{x}\right)\cos\left(\frac{1}{x^2}\right) \tag{6}$$

Since $|\sin(\cdot)| \leqslant 1$ and $|\cos(\cdot)| \leqslant 1$, but $(2/x) \to \infty$ as $x \to 0$, it is clear that the limit as $x \to 0$ of $f'(x)$ is not well defined. However, $f'(0)$ does exist! Indeed,

$$f'(0) = \lim_{x\to 0}\left(\frac{f(x) - f(0)}{x - 0}\right) = \lim_{x\to 0} x\sin\left(\frac{1}{x^2}\right) \tag{7}$$

Since $|\sin(1/x^2)| \leqslant 1$, we have $|x\sin(1/x^2)| \leqslant |x|$, so $x\sin(1/x^2) \to 0$ as $x \to 0$. This means $f'(0) = 0$. Thus, f is not C^1.

Most of functions in economics are continuous. In addition, most of them are real-valued, say utility functions and cost functions. Characterizing real-valued continuous function is hence of special interest for us. The intermediate value theorem gives part of the characterization.

Theorem 6. ***(Intermediate Value Theorem)*** *Let $D = [a, b]$ be an interval in $\mathbb{R}$ and let $f : D \to \mathbb{R}$ be a continuous function. If $f(a) < f(b)$, and if c is a real number such that $f(a) < c < f(b)$, then there exists $x \in (a, b)$ such that $f(x) = c$. A similar statement holds if $f(a) > f(b)$.*

Remark 2. *It might appear at first glance that the intermediate value theorem actually characterize continuous functions, i.e., that a function $f : [a, b] \to \mathbb{R}$ is continuous if and only if for any two points $x_1 < x_2$ and for any real number c lying between $f(x_1)$ and $f(x_2)$, there is $x \in (x_1, x_2)$ such that $f(x) = c$. The intermediate value theorem shows that the "only if" part is true. It is left as an exercise to show that the converse, namely the "if" part, is actually false.*

Theorem 7. ***Intermediate Value Theorem for the Derivative*** *Let $D = [a, b]$ be an interval in $\mathbb{R}$, and let $f : D \to \mathbb{R}$ be a function that is differentiable everywhere on D. If, and if c is a real number such hat $f'(a) < c < f'(b)$, then there is a point $x \in (a, b)$ such that $f'(c)$. A similar statement holds if $f'(b) < f'(a)$.*

It is very important to emphasize that the above Theorem does no assume that f is C^1. Indeed, if f were C^1, the result would be a trivial consequence of the Intermediate Value Theorem, since the derivative function f' would be a continuous function on $[a, b]$.

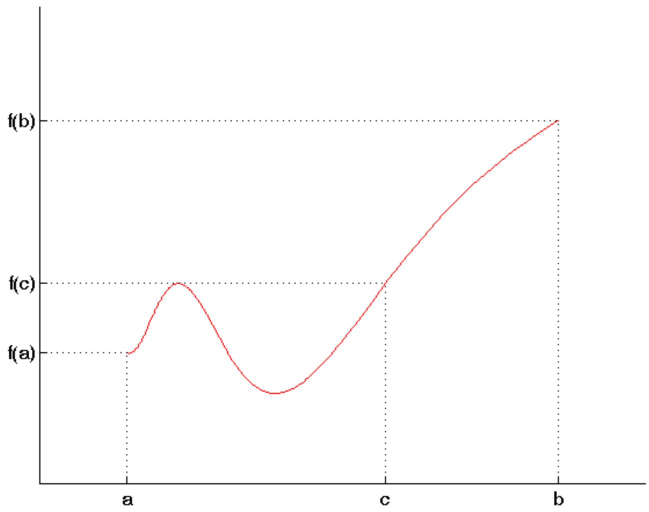

Figure 1: The Intermediate Value Theorem

Theorem 8. ***(Intermediate Value Theorem in $\mathbb{R}^n$)*** *Let $D \subset \mathbb{R}^n$ be a convex set, and let $f : D \to \mathbb{R}$ be continuous on D. Suppose that a and b are points in D such that $f(a) < f(b)$. Then, for any c such that $f(a) < c < f(b)$, there is $\hat{\lambda} \in (0, 1)$ such that $f((1 - \hat{\lambda})a + \hat{\lambda}b) = c$.*

5 Implicit Function Theorem

Theorem 9. ***(Implicit Function Theorem)*** *Let $F : S \subset \mathbb{R}^{m+n} \to \mathbb{R}^n$ be a C^1 function, where S is open. Let $(x^*, y^*) \in S$ such that $DF_y(x^*, y^*)$ is invertible, and let $F(x^*, y^*) = c$. Then, there is a neighborhood $U \subset \mathbb{R}^m$ of x^* and a C^1 function $g : U \to \mathbb{R}^n$ such that (i) $(x, g(x)) \in S$ for all $x \in U$, (ii) $g(x^*) = y^*$, and (iii) $F(x, g(x)) \equiv c$ for all $x \in U$. The derivative of g at any $x \in U$ may be ontained from the chain rule:*

$$Dg(x) = -(DF_y(x, y))^{-1} \cdot DF_x(x, y). \tag{8}$$

Let $m = n = 1$ and $F(x, y) = c$ on a neighborhood of (x^*, y^*), the total derivative is $\frac{\partial F}{\partial x}(x, y) \cdot dx + \frac{\partial F}{\partial y}(x, y) \cdot dy = 0$. Then we have $y'(x) = -\frac{\partial F/\partial x}{\partial F/\partial y}(x, y)$.

6 Weierstrass' Extreme Value Theorem

Theorem 10. ***Weierstrass' Extreme Value Theorem*** *Let $D \subset \mathbb{R}^n$ be compact, and let $f : D \to \mathbb{R}$ be a continuous function on D. Then f attains a maximum and minimum on D.*

Example 7. *Let $D = \mathbb{R}$, and $f(x) = x^3$ for all $x \in \mathbb{R}$. Then f is continuous, but D is not compact (it is closed, but not bounded). It's clear that f attains neither a maximum nor a minimum.*

Example 8. *Let $D = (0, 1)$ and $f(x) = x$ for all $x \in (0, 1)$. Then f is continuous, but D is again noncompact (it is bounded, but not closed). Again, f attains neither a maximum nor a minimum.*

Example 9. *Let $D = [0, 1]$, and let f be given by*

$$f(x) = \begin{cases} x, & if -1 < x < 1 \\ 0, & otherwise \end{cases} \tag{9}$$

Note that D is compact, but f fails to be continuous at just two points -1 and 1. In this case, $f(D)$ is the open interval $(-1, 1)$; consequently, f fails to attain either a maximum or a minimum on D.

Example 10. *Let $D = (0, \infty)$, and let $f : D \to \mathbb{R}$ be defined by*

$$f(x) = \begin{cases} 1, & if x \in \mathbb{Q} \\ 0, & otherwise \end{cases} \tag{10}$$

Then D is not compact (it is neither bounded nor closed), and f is discontinuous at every point in D. Nonetheless, f attains a maximum (at every rational number) and a minimum (at every irrational number).

Example 11. *Consider the utility maximization problem*

$$\begin{aligned} \max \quad & u(x) \\ s.t. \quad & x \in \{x \in \mathbb{R}^n_+ | p \cdot x \leqslant I\} \end{aligned} \tag{11}$$

where $p = \{p_1, \cdots, p_n\} \gg 0$ is a price vector and $I > 0$ denotes the income available for consumption.

Throughout the example, the utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ is assumed to be continuous on its domain. Then according to Weierstrass' Extreme Value Theorem, a solution to the maximization problem exists as long as the budget set $B(p, I) \equiv \{x \in \mathbb{R}^n_+ | p \cdot x \leqslant I\}$ is compact.

First, $B(p, I)$ being bounded is evident. (Verify!) The remaining is to check closedness. Suppose not, i.e., there exists a sequence $\{x_k\}$ in $B(p, I)$ converging to $x \notin B(p, I)$. Given $x \geqslant 0$, $x \notin B(p, I)$ implies $p \cdot x > I$. Then there exists $\varepsilon > 0$ such that $||x_k - x|| < \varepsilon$ and $p \cdot x_k > I$: contradiction.

Example 12. *Consider a problem of identifying a social optimal division of a given quantity of resources between two agents.*

$$\begin{aligned} \max \quad & \alpha u_1(x_1) + (1 - \alpha) u(x_2) \\ s.t. \quad & (x_1, x_2) \in F(w) \end{aligned} \tag{12}$$

where $f(w) \equiv \{(x_1, x_2) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ | x_1 + x_2 = w\}$ represents the set of possible divisions, and u_i is agent i's utility function, $i = 1, 2$.

The parameters of this problem are $\alpha \in (0,1)$ and $w \in \mathbb{R}^n_+$. It's easy to see that $F(w)$ is compact for any $w \in \mathbb{R}^n_+$. Moreover, the weighted utility function $\alpha u_1(x_1) + (1-\alpha)u(x_2)$ is continuous as a function of (x_1, x_2) whenever the underlying utility functions u_1 and u_2 are continuous on $\mathbb{R}^n_+$. Thus provided only that the utility functions u_1 and u_2 are continuous, the Weierstrass's Theorem assures us of the existence of a solution to this division problem.

Proof: (Weierstrass' Extreme Value Theorem).

Lemma 1. *If $f : D \to \mathbb{R}$ is continuous on D, and D is compact, then $f(D)$ is also compact.*

Proof. Pick any sequence $\{y_k\}$ in $f(D)$. The lemma will be proved if we can show that $\{y_k\}$ has a convergent subsequence, which converges to a point $y \in f(D)$.

For each y_k, there must be a $x_k \in D$ such that $y_k = f(x_k)$. Since D is compact, there must be a convergent subsequence $x_{m(k)}$ such that $\lim_{k\to\infty} x_{m(k)} = x$ and $x \in D$. Then f being continuous implies that $\lim_{x_{m(k)}\to x} f(x_{m(k)}) = f(x)$. Let $f(x_{m(k)}) = y_{m(k)}$ and $f(x) = y$, we have $\lim_{k\to\infty} y_{m(k)} = y$ and $y \in f(D)$. □

Lemma 2. *If $A \subset \mathbb{R}$ is compact, then* $\sup A \in A$ *and* $\inf A \in A$*, so the maximum and minimum of A are well defined.*

Proof. Since A is bounded, $\sup A \in \mathbb{R}$. For $k = 1, 2, \cdots$, let N_k represent the interval $(\sup A - \frac{1}{k}, \sup A]$, and let $A_k = N_k \cap A$. Then A_k must be nonempty for each k. Pick any point from A_k, and label it x_k.

We claim that $x_k \to \sup A$ as $k \to \infty$. This follows simply from the observation that since $x_k \in (\sup A - \frac{1}{k}, \sup A]$ for each k, it must be the case that $d(x_k, \sup A) < \frac{1}{k}$. Therefore, $d(x_k, \sup A) \to 0$ with k, establishing the claim.

But $x_k \in A_k \subset A$ for each k, and A is a closed set, so the limit of the sequence $\{x_k\}$ must be on A, establishing one part of the result. The other part of the result, $\inf A \in A$, can be established analogously. □

Then the theorem is an immediate consequence of these lemmata.

□

7 Exercise

1. Verify that the discrete metric is well-defined.

2. Verify that the p$-$metric on $\mathbb{R}^n$ is well-defined. (Hint: refer to the famous Minkowski's inequality.)

3. Check "$||x_k|| \leqslant M$ for all k" in the proof of Theorem 3.

4. Prove: The Euclidian metric d_2 is continuous on $\mathbb{R}^n$.

5. Show that the converse of the intermediate value theorem is false.

   Answer Consider the function $f : [0, 2] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 10, & x = 0 \\ \frac{1}{x}, & 0 < x \leqslant 1 \\ 2 - x, & 1 < x \leqslant 2 \end{cases} \tag{13}$$

   For any two points $x_1 < x_2$ and for any real number c lying between $f(x_1)$ and $f(x_2)$, there is $x \in (x_1, x_2)$ such that $f(x) = c$. But f is not continuous at $x = 0$.

6. Prove Theorem 4.

   Answer If $a = b$, $\{x_k\}$ is a constant sequence. Hence $x = a = b$. If instead $a < b$, we prove $a \leqslant x$. (The same argument gives $x \leqslant b$.) Suppose $x < a$, we have $\varepsilon \equiv a - x > 0$. Then there exist infinitely many x_k such that $|x_k - x| < \varepsilon$. Then $x_k < a$: contradiction.

7. Show that it is possible for two functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ to be discontinuous, but for their product $f \cdot g$ to be continuous. What about their composition $f \circ g$?

8. A consumer, who lives for two periods, has the utility function $v(c(1), c(2))$, where $c(t) \in \mathbb{R}^n$ denotes the consumer's consumption bundle in period t, $t = 1, 2$. The price vector in period t is given by $p(t) \in \mathbb{R}^n_+$, $p(t) = (p_1(t), \cdots, p_n(t))$. The consumer has an initial wealth of W_0, but has no other income. Any amount not spent in the first period can be saved and used for the second period. Savings earn interest at a rate $r \geqslant 0$. (Thus, a dollar saved in period 1 becomes $(1 + r)$ in period 2.)

   (a) Assume the usual nonnegativity constraints on consumption and set up the consumer's utility-maximization problem.

   (b) Show that the feasible set in this problem is compact if and only if $p(1) \gg 0$ and $p(2) \gg 0$.

9. A fishery earns a profit of $\pi(x)$ from catching and selling x units of fish. The firm owns a pool which currently has y_1 fish in it. If $x \in [0, y_1]$ fish are caught this period, the remaining $i = y_1 - x$ fish will frow to $f(i)$ fish by the begining of the next period, where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is the growth function for the fish population. The fishery wishes to set the volume of its catch in each of the next three periods so as to maximize the sum of its profits over this horizon. That is, it solves:

$$\begin{aligned} &\max \pi(x_1) + \pi(x_2) + \pi(x_3) \\ &s.t. x_1 \leqslant y_1 \\ &\quad\;\; x_2 \leqslant y_2 = f(y_1 - x_1) \\ &\quad\;\; x_3 \leqslant y_3 = f(y_2 - x_2) \end{aligned} \tag{14}$$

   and the nonnegativity constraints that $x_t \geqslant 0$, $t = 1, 2, 3$.

Show that if π and f are continuous on $\mathbb{R}_+$, then he Weierstrass' Theorem may be used to prove that a solution exists to this problem. (This is immediate if one can show that the continuity of f implies the compactness of the set of feasible triples (x_1, x_2, x_3).)

References

Ok, Efe A, *Real analysis with economic applications*, Vol. 10, Princeton University Press, 2007.

Rudin, Walter, *Principles of mathematical analysis*, Vol. 3, McGraw-Hill New York, 1964.

Sundaram, Rangarajan K, *A first course in optimization theory*, Cambridge university press, 1996.

Lecture 2: Separation Theorems and Unconstrained Optimization

Liu Peng

August 3, 2015

1 Separation Theorems

Definition 1. *Let $p \neq 0$ be a vector in $\mathbb{R}^n$, and let $a \in \mathbb{R}$. The set H defined by*

$$H = \{x \in \mathbb{R}^n | p \cdot x = a\} \tag{1}$$

*is called a **hyperplane** in $\mathbb{R}^n$, and will be denoted $H(p, a)$.*

Example 1. *A hyperplane in $\mathbb{R}^2$ is simply a straight line: if $p \in \mathbb{R}^2$ and $a \in \mathbb{R}$, the hyperplane $H(p, a)$ is simply the set of points (x_1, x_2) that satisfy $p_1x_1 + p_2x_2 = a$. Similarly, a hyperplane in $\mathbb{R}^3$ is a plane.*

Definition 2. *Two sets D and E are said to be **separated by the hyperplane** $H(p, a)$ in $\mathbb{R}^n$ if D and E lie on the opposite sides of $H(p, a)$, i.e., if we have*

$$\begin{aligned} p \cdot y \leqslant a, \textit{ for all } y \in D \\ p \cdot z \geqslant a, \textit{ for all } z \in E \end{aligned} \tag{2}$$

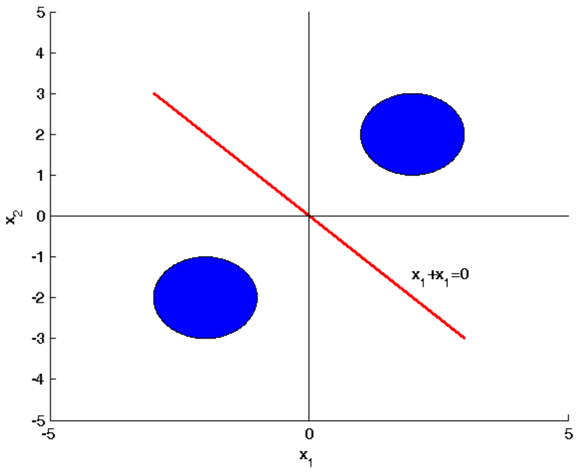

Figure 1: Separating hyperplane in $\mathbb{R}^2$.

Theorem 1 gives sufficient conditions for separating a point from a set.

Theorem 1. *Let D be a nonempty convex set in $\mathbb{R}^n$, and let x^* be a point in $\mathbb{R}^n$ that is not in D. Then, there is a hyperplane $H(p, a)$ in $\mathbb{R}^n$ with $p \neq 0$ which separates D and x^*. We may, if we desire, choose p to also satisfy $||p|| = 1$.*

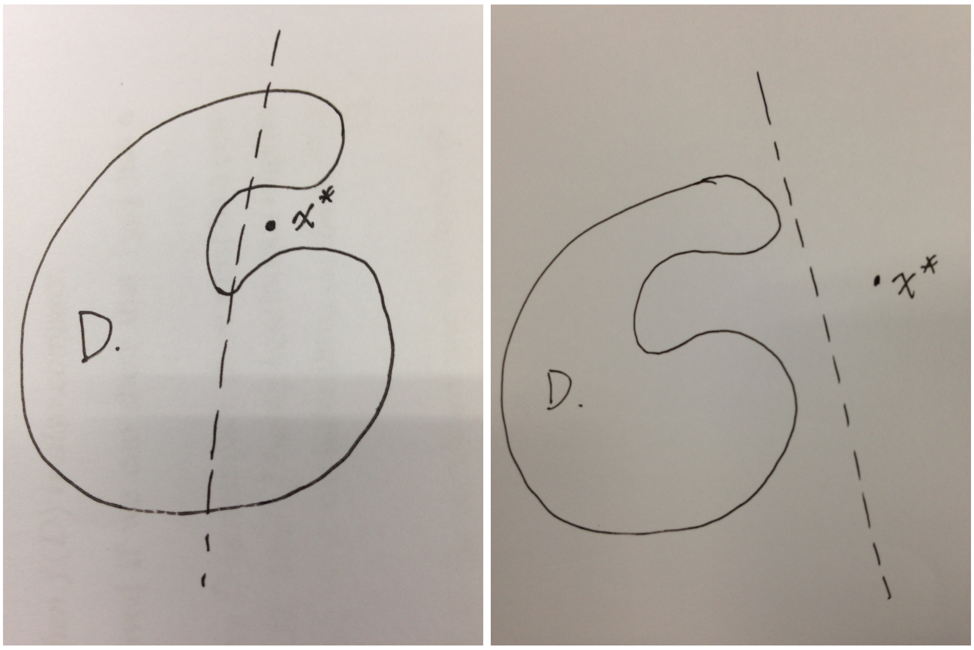

Figure 2: Being convex is a sufficient condition.

Proof. We prove the result for the case where D is closed. (For the case where D is open, refer to Sundaram(1996).)

Let $\bar{B}(x^*, r) = \{x \in \mathbb{R}^n \text{ s.t. } ||x - x^*|| \leqslant r\}$, i.e., $\bar{B}(x^*, r)$ is the closed ball with center x^* and radius r. Pick r sufficiently large so that $Y = \bar{B}(x^*, r) \cap D$ is nonempty. Since $\bar{B}(x^*, r)$ and D are both closed, so is Y. Since $\bar{B}(x^*, r)$ is bounded and $Y \subset \bar{B}(x^*, r)$, Y is also bounded. Therefore Y is compact. The metric function $f : Y \to \mathbb{R}$ defined by $f(y) = d_2(x^*, y)$ is clearly continuous on Y since the metric function d_2 is continuous. Thus, by the Weierstrass' Theorem, there exists a minimum y^* on Y, i.e., there exists y^* in Y such that

$$d_2(x^*, y^*) \leqslant d_2(x^*, y), \text{ for all } y \in Y. \tag{3}$$

If $y \in D \backslash Y$, then we have $y \notin \bar{B}(x^*, r)$ and hence $d_2(x^*, y) > r$. While, $d_2(x^*, y^*) \leqslant r$. Therefore, we have

$$d_2(x^*, y^*) \leqslant d_2(x^*, y), \text{ for all } y \in D. \tag{4}$$

Let $p = y^* - x^*$ and let $a = p \cdot y^*$. We will show that the hyperplane separates D and x^*.
Notice first that

$$\begin{aligned} p \cdot x^* &= (y^* - x^*) \cdot x^* \\ &= -(y^* - x^*) \cdot (y^* - x^*) + y^* \cdot (y^* - x^*) \\ &= -||p||^2 + a < a \end{aligned} \tag{5}$$

Next we show $p \cdot y \geqslant a$ for all $y \in D$. Given arbitrary $y \in D$, since D is convex, it is the

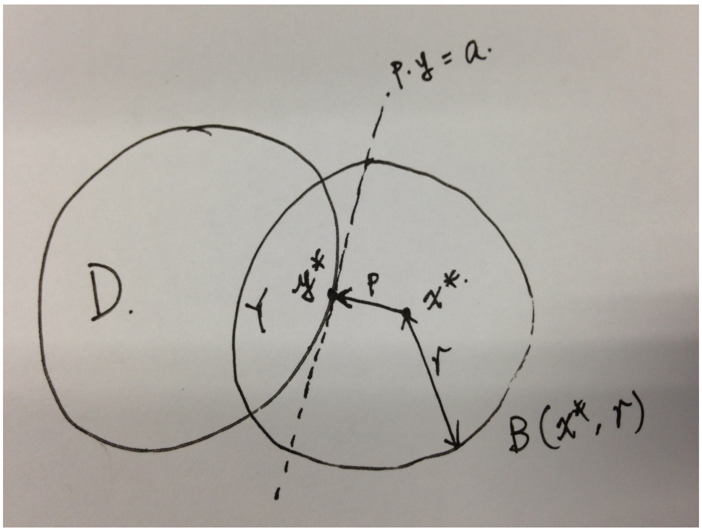

Figure 3: Proof of separating theorem when the set is closed.

case that for all $\lambda \in (0,1)$, the point $y(\lambda) = \lambda y + (1-\lambda)y^*$ is also in D. Then we have

$$\begin{aligned}
&d_2(x^*, y^*) \leqslant d_2(x^*, y(\lambda)) \\
\Leftrightarrow\ &||x^* - y^*||^2 \leqslant ||x^* - y(\lambda)||^2 \\
&\qquad = ||x^* - \lambda y - (1-\lambda)y^*||^2 \\
&\qquad = ||\lambda(x^* - y) + (1-\lambda)(x^* - y^*)||^2 \\
&\qquad = \lambda^2||x^* - y||^2 + 2\lambda(1-\lambda)(x^* - y)\cdot(x^* - y^*) \\
&\qquad\quad + (1-\lambda)^2||x^* - y^*||^2.
\end{aligned} \tag{6}$$

Rearranging terms, this gives us

$$0 \leqslant \lambda^2||x^* - y||^2 + 2\lambda(1-\lambda)(x^* - y)\cdot(x^* - y^*) - \lambda(2-\lambda)||x^* - y^*||^2. \tag{7}$$

Dividing both sides by $\lambda > 0$, we get

$$0 \leqslant \lambda||x^* - y||^2 + 2(1-\lambda)(x^* - y)\cdot(x^* - y^*) - (2-\lambda)||x^* - y^*||^2. \tag{8}$$

Taking limit as $\lambda \to 0$ and dividing the result by 2, we have

$$\begin{aligned}
0 &\leqslant (x^* - y)\cdot(x^* - y^*) - ||x^* - y^*||^2 \\
&= (x^* - y^*)\cdot(x^* - y - x^* + y^*) \\
&= (x^* - y^*)\cdot(x^* - y).
\end{aligned} \tag{9}$$

Rearranging terms again, we have what we want

$$p\cdot y = (y^* - x^*)\cdot y \geqslant (y^* - x^*)\cdot y^* = a. \tag{10}$$

Thus, we have shown that there is a hyperplane $H(p,a)$ that separates D and x^* when D is closed.

It remains to be shown that we may, without loss of generality, take $||p||$ to be unity. Suppose $||p|| \neq 1$. We will show the existence of $(\tilde{p}, \tilde{a})$ such that $H(\tilde{p}, \tilde{a})$ also separates D and x^*, which further satisfies $||\tilde{p}|| = 1$. Since $p \neq 0$, we have $||p|| > 0$. Define $\tilde{p} = p/||p||$ and $\tilde{a} = a/||p||$. Then we have $||\tilde{p}|| = 1$ and

$$\begin{aligned} \tilde{p} \cdot x^* &= \frac{p \cdot x^*}{||p||} < \frac{a}{||p||} = \tilde{a} \\ \tilde{p} \cdot y &= \frac{p \cdot y}{||p||} \geqslant \frac{a}{||p||} = \tilde{a} \end{aligned} \tag{11}$$

for all $y \in D$. This completes the proof of the theorem for the case where D is closed. □

We can easily generate the above theorem to the case where two sets are separated.

Theorem 2. *Let D and E be convex sets in $\mathbb{R}^n$ such that $D \cap E = \emptyset$. Then, there exists a hyperplane $H(p, a)$ in $\mathbb{R}^n$ which separates D and E. We may, if we desire, choose p to also satisfy $||p|| = 1$.*

2 Unconstrained Optimization

We now turn to a study of optimization theory under assumption of differentiability. Our principal objective here is to identify necessary conditions that the derivative of f must satisfy at an optimum.

Consider a real-valued function $f : D \subset \mathbb{R}^n \to \mathbb{R}$. We define the **interior** of D by

$$\text{int}D = \{x \in D | \text{ there is } r > 0 \text{ such that } B(x, r) \subset D\}. \tag{12}$$

A point x where f achieves a maximum will be called an unconstrained maximum of f if $x \in \text{int}D$. Unconstrained minima are defined analogously.

One observation is important before proceeding to the analysis. The concepts of maxima and minima are **global** concepts, i.e., they involve comparisons between the value of f at a particular point x and all other feasible points $z \in D$. On the other hand, the differentiability is a **local** property: the derivative of f at a point x tells us something about the behavior of f in a neighborhood of x, but nothing at all about the behavior of f elsewhere on D. This observation motivates the following definitions.

Definition 3. *A point $x \in D$ is a **local maximum** of f on D if there is $r > 0$ such that $f(x) \geqslant f(y)$ for all $y \in D \cap B(x, r)$.*

*A point $x \in D$ is an **unconstrained local maximum** of f on D if there is $r > 0$ such that $B(x, r) \subset D$, and $f(x) \geqslant f(y)$ for all $y \in B(x, r)$.*

2.1 First-Order Condition

Theorem 3. *Suppose $x^* \in D$ is an unconstrained local maximum of f on D and f is differentiable at x^*. Then $Df(x^*) = 0$. The same result is true if, instead, x^* is an unconstrained local minimum of f on D*

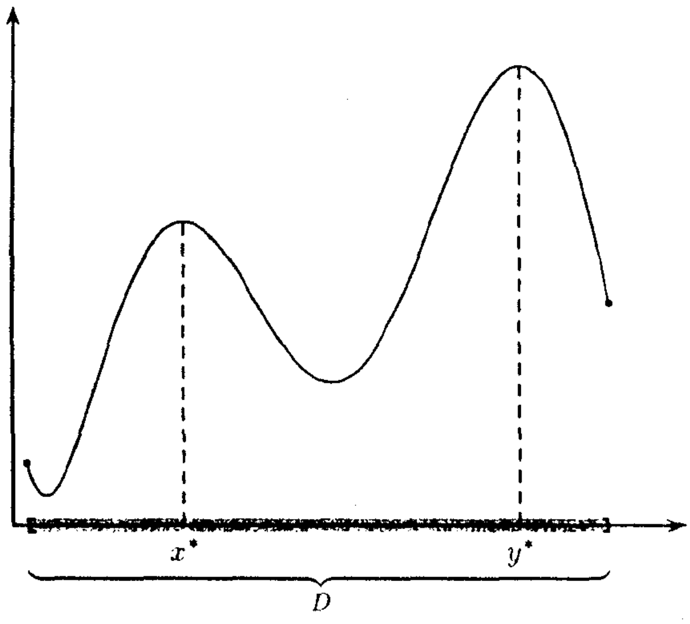

Figure 4: Global and Local Optima

Proof. See Sundaram (1996), section 4.5. □

Example 2. *Let $D = \mathbb{R}$, and let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$ for all $x \in \mathbb{R}$. Then, we have $f'(0) = 0$, but 0 is neither a local maximum nor a local minimum of f on $\mathbb{R}$. This is an example to show that the Theorem 3 provides only a necessary condition rather than a sufficient condition.*

2.2 Second-Order Conditions

The first-order conditions for unconstrained local optima do not distinguish between maxima and minima. To obtain such a distinction in the behavior of f at an optimum, we need to examine the behavior of the second derivative $D^2 f$. Usually, we call the points satisfying $Df(x) = 0$ **critical points**.

A preliminary definition first: A local maximum x of f on D is called a **strict local maximum** if there is $r > 0$ such that $f(x) > f(y)$ for all $y \in B(x, r) \cap D$, $y \neq x$.

Theorem 4. *Suppose f is a C^2 function on D, and x is an interior point.*

1. *If f has a local maximum at x, then $D^2 f(x)$ is negative semidefinite.*
2. *If f has a local minimum at x, then $D^2 f(x)$ is positive semidefinite.*
3. *If $Df(x) = 0$ and $D^f(x)$ is negative definite at some x, then x is a strict local maximum.*
4. *If $Df(x) = 0$ and $D^f(x)$ is positive definite at some x, then x is a strict local minimum.*

Proof. See Sundaram (1996), section 4.6. □

Notice that, as necessary conditions, part 1 and 2 identify only semidefiniteness of $D^2 f$ at the optimal points. But the sufficient conditions, part 3 and 4, require definiteness. Two natural questions could be: Can we strengthen the part 1 and 2 by identifying definiteness? and Can we strengthen the part 3 and 4 by requiring only semidefiniteness? The following two examples answer the questions negatively.

Example 3. *Let $D = \mathbb{R}$ and let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^4$ and $g(x) = -x^4$, respectively. Since $x^4 \geqslant 0$ everywhere, and $f(0) = g(0) = 0$, it is clear that 0 is a global minimum of f, and a global maximum of g. However, $f''(0) = g''(0) = 0$, so viewed as 1×1 matrices, $f''(0)$ is positive semidefinite, but not positive definite, while $g''(0)$ is negative semidefinite, but not negative definite.*

Example 4. *Let $D = \mathbb{R}$ and let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$. Then $f'(x) = 3x^2$ and $f''(x) = 6x$, so $f'(0) = f''(0) = 0$. Thus, viewed as a 1×1 matrix, $f''(0)$ is both positive semidefinite and negative semidefinite. If part 3 and 4 of Theorem 4 requires only semidefiniteness, $f''(0)$ will pass boh tests; however, 0 is neither a local maximum nor a local minimum.*

2.3 Using the First- and Second-Order Conditions

In application, we usually use first-order necessary conditions to identify all candidates for local optima and use the second-order sufficient conditions to check them one by one in order to decide which ones are local maxima, which ones are local minima, and which ones can not be justified to be local optima.

But notice that the usefulness of Theorems 3 and 4 is limited for at least two reasons. First, the conditions apply to only interior points. In many applications, optima happen on the boundaries. Second, using Theorems 3 and 4 gives only local optima rather tan global optima. The latter should be our final target.

The following two examples are used to show limitations of Theorems 3 and 4.

Example 5. *Consider the problem of maximizing $f(x) = 4x^3 - 5x^2 + 2x$ over the interval $[0, 1]$. Since $[0, 1]$ is compact, and f is continuous on this interval, the Weierstrass' Theorem shows that f has a maximum on this interval. There are two possibilities: either the maximum occurs at one of the boundary points 0 or 1, or it is an unconstrained maximum. In the latter case, it must meet the first-order conditions: $f'(x) = 12x^2 - 10x + 2 = 0$. The only points that satisfy this condition are $x = 1/2$ and $x = 1/3$. Evaluating f at the four points 0, $1/3$, $1/2$, and 1 shows that $x = 1$ is the point where f is maximized on $[0, 1]$.*

Example 6. *Let $D = \mathbb{R}$, and let $f : D \to \mathbb{R}$ be given by $f(x) = 2x - 3x^2$. It is easily checked that f is a C^2 function on $\mathbb{R}$, and that there are precisely two points at which $f'(x) = 0$: namely, at $x = 0$ and $x = 1$. Invoking the second-order conditions, we get $f''(0) = -6$, while $f''(0) = 6$. Thus, the point 0 is a strict local maximum, while the point 1 is a strict local minimum.*

However, there is nothing in the first- or second-order conditions that will help determine whether these points are global optima. In fact,they are not: global optima do not exist in this problem, since $\lim_{x \to +\infty} f(x) = +\infty$ and $\lim_{x \to -\infty} f(x) = -\infty$.

3 Exercise

1. Find and classify all critical points(local minimum, local maximum, neither) of each of the following functions. Are any of the local optima also global optima?

   (a) $f(x, y) = 2x^3 + xy^2 + 5x^2 + y^2$.

   (b) $f(x, y) = e^{2x}(x + y^2 + 2y)$.

   (c) $f(x, y) = xy(a - x - y)$.

   (d) $f(x, y) = x \sin y$.

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

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

   (g) $f(x, y) = \frac{x}{1+x^2+y^2}$.

   (h) $f(x, y) = (x^4/32) + x^2y^2 - x - y^2$.

References

Ok, Efe A, *Real analysis with economic applications*, Vol. 10, Princeton University Press, 2007.

Rudin, Walter, *Principles of mathematical analysis*, Vol. 3, McGraw-Hill New York, 1964.

Sundaram, Rangarajan K, *A first course in optimization theory*, Cambridge university press, 1996.

Lecture 3: Equality Constraints and the Theorem of Lagrange

Liu Peng
August 13, 2015

1 Constrained Optimization Problems

Many problems of interest in economic theory can be written in the form

$$\begin{aligned} \max \quad & f(x) \\ s.t. \quad & x \in \mathcal{D} \end{aligned} \tag{1}$$

Where $\mathcal{D} = U \cap \{x \in \mathbb{R}^n | g(x) = 0, h(x) \geqslant 0\}$, $U \subset \mathbb{R}^n$ is open, $g : \mathbb{R}^n \to \mathbb{R}^k$, and $h : \mathbb{R}^n \to \mathbb{R}^l$.

We call functions $g = (g_1, \cdots, g_k)$ **equality constraints** and functions $h = (h_1, \cdots, h_l)$ **inequality constraints**. In this lecture, we study the case where all the constraints are equality constraints, i.e., where the constraint set $\mathcal{D}$ can be represented as

$$\mathcal{D} = U \cap \{x \in \mathbb{R}^n | g(x) = 0\}. \tag{2}$$

2 First-Order Conditions and the Theorem of Lagrange

The Theorem of Lagrange provides a powerful characterization of local optima of equality-constrained optimization problems in terms of the behavior of the objective function f and the constraint functions g at these points. The conditions the theorem describes may be viewed as the first-order necessary conditions for local optima in these problems.

Theorem 1. ***The Theorem of Lagrange*** *Let $f : \mathbb{R}^n \to \mathbb{R}$, and $g_i : \mathbb{R}^n \to \mathbb{R}^k$ be C^1 functions, $i = 1, \cdots, k$. Suppose x^* is a local maximum or minimum of f on the set $\mathcal{D} = U \cap \{x \in \mathbb{R}^n | g(x) = 0\}$, where $U \subset \mathbb{R}^n$ is open. Suppose also that the rank $r(Dg(x^*)) = k$. Then, there exists a vector $\lambda^* = (\lambda_1^*, \cdots, \lambda_k^*) \in \mathbb{R}^k$ such that*

$$Df(x^*) + \sum_{i=1}^{k} \lambda_i^* Dg_i(x^*) = 0. \tag{3}$$

It must be stressed that the Theorem of Lagrange only provides necessary conditions for local optima x^*, and only for those local optima x^* which also meet the condition that $r(Dg(x^*)) = k$. [1] The next example shows clearly that these conditions are not sufficient, even if x meets the rank condition.

[1] A point satisfying this rank condition is usually called a non-degenerated point.

Example 1. *Let f and g be functions on $\mathbb{R}^2$ defined by $f(x,y) = x^3 + y^3$ and $g(x,y) = x - y$. Consider the equality-constrained optimization problem of maximizing and minimizing $f(x,y)$ over the set $\mathcal{D} = \{(x,y) \in \mathbb{R}^2 | g(x,y) = 0\}$.*

Let (x^*, y^*) be the point $(0,0)$, and let $\lambda^* = 0$. Then, $g(x^*, y^*) = 0$, so (x^*, y^*) is a feasible point. Moreover, since $Dg(x,y) = [1, -1]$ for any (x,y), it is clearly the case that $r(Dg(x,y)) = 1$. Finally, since $Df(x,y) = [3x^2, 3y^2]$, we have

$$Df(x^*, y^*) + \lambda^* Dg(x^*, y^*) = (0,0) + 0 \cdot (1,1) = (0,0). \tag{4}$$

Thus, if the conditions of the Theorem of Lagrange were also sufficient, then (x^*, y^*) would be either a local maximum of a local minimum of f on $\mathcal{D}$. It is, quite evidently, neither. For every $\varepsilon > 0$, we have $(-\varepsilon, -\varepsilon) \in \mathcal{D}$ and $f(-\varepsilon, -\varepsilon) = -2\varepsilon^3 < 0 = f(x^*, y^*)$; $(\varepsilon, \varepsilon) \in \mathcal{D}$ and $f(-\varepsilon, -\varepsilon) = 2\varepsilon^3 > 0 = f(x^*, y^*)$.

Proof. (Theorem of Lagrange) Assume, without loss of generality, that the $k \times k$ submatrix of $Dg(x^*)$ that has full rank is the $k \times k$ submatrix consisting of the first k rows and k columns. In addition, we will denote the first k coordinates of a vector x by w and the last $n-k$ coordinates by z, i.e., we write $x = (w, z)$. In this notation, we are given the data that (w^*, z^*) is a local maximum of f on $\mathcal{D} = U \cap \{(w,z) \in \mathcal{R}^n | g(w,z) = 0\}$ and that $r(Dg(w^*, z^*)) = k$. We are to prove that there exists λ^* such that

$$\begin{aligned} Df_w(w^*, z^*) + \lambda^* Dg_w(w^*, z^*) = 0 \\ Df_z(w^*, z^*) + \lambda^* Dg_z(w^*, z^*) = 0 \end{aligned} \tag{5}$$

Since $r(Dg(w^*, z^*)) = k$, the Implicit Function Theorem shows that there exists an open set V in $\mathbb{R}^{n-k}$ containing z^*, and a C^1 function $h : V \to \mathbb{R}^k$ such that $h(z^*) = w^*$ and $g(h(z), z) \equiv 0$ for all $z \in V$. Differentiating the identity $g(h(z), z) \equiv 0$ with respect to z by using chain rule, we ontain

$$Dg_w(h(z), z) Dh(z) + Dg_z(h(z), z) = 0. \tag{6}$$

At $z = z^*$, we have $h(z^*) = w^*$. Since $Dg_w(w^*, z^*)$ is invertible, we have

$$Dh(z^*) = -[Dg_w(w^*, z^*)]^{-1} Dg_z(w^*, z^*). \tag{7}$$

Now define

$$\lambda^* = -Df_w(w^*, z^*)[Dg_w(w^*, z^*)]^{-1}. \tag{8}$$

We will show that λ^* defined in this way meets the required conditions. Indeed, it follows from the definition of λ^* that

$$\begin{aligned} \lambda^* Dg_w(w^*, z^*) &= -Df_w(w^*, z^*)[Dg_w(w^*, z^*)]^{-1} Dg_w(w^*, z^*) \\ &= -Df_w(w^*, z^*), \end{aligned} \tag{9}$$

which is the same as

$$Df_w(w^*, z^*) + \lambda^* Dg_w(w^*, z^*) = 0. \tag{10}$$

Thus it remains to show that

$$Df_z(w^*, z^*) + \lambda^* Dg_z(w^*, z^*) = 0. \tag{11}$$

Define a function $F : V \to \mathbb{R}$ by $F(z) = f(h(z), z)$. Since f has a local maximum at (w^*, z^*), it is clear that F has a local maximum at z^*. Since V is an open set, z^* is an unconstrained local maximum of F, and the first-order conditions for an unconstrained maximum imply that $DF(z^*) = 0$, which is

$$Df_w(w^*, z^*)Dh(z^*) + Df_z(w^*, z^*) = 0. \tag{12}$$

Substituting for $Dh(z^*)$, we obtain

$$-Df_w(w^*, z^*)[Dg_w(w^*, z^*)]^{-1}Dg_z(w^*, z^*) + Df_z(w^*, z^*) = 0, \tag{13}$$

which, according to definition of λ^*, is the same as

$$Df_z(w^*, z^*) + \lambda^* Dg_z(w^*, z^*) = 0. \tag{14}$$

□

2.1 The Constraint Qualification

The condition in the Theorem of Lagrange that the rank of $Dg(x^*)$ be equal to the number of constraints k is called the **constraint qualification under equality constraints**. It plays a central role in the proof of the theorem; essentially, it ensures that $Dg(x^*)$ contains an invertible $k \times k$ submatrix, which may be used to define the vector λ^*.

More importantly, it turns out to be the case that if the constraint qualification is violated, then the conclusions of the theorem may also fail. That is, if x^* is a local optimum at which $r(Dg(x^*)) < k$, then there need not exist a vector λ^* such that $Df(x^*) + \sum_{i=1}^{k} \lambda_i^* Dg_i(x^*) = 0$. The following example, which involves seemingly well-behaved objective and constraint functions, illustrates this point.

Example 2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = -y$, and $g(x, y) = y^3 - x^2$, respectively. Consider the equality-constrained optimization problem*

$$\begin{aligned} \max \quad & f(x, y) \\ s.t. \quad & (x, y) \in \mathcal{D} = \{(x, y) \in \mathbb{R}^2 | g(x, y) = 0\}. \end{aligned} \tag{15}$$

Since $x^2 \geqslant 0$ for any real number x, and the constraint requires that $y^3 = x^2$, we must have $y \geqslant 0$ for any $(x, y) \in \mathcal{D}$; moreover, $y = 0$ if and only if $x = 0$. It easily follows that f attains a unique global maximum on $\mathcal{D}$ at the origin $(x, y) = (0, 0)$, so $r(Dg(0, 0)) = 0 < 1$. Thus, the constraint qualification is violated. Moreover, $Df(x, y) = (0, -1)$ at any (x, y), which means that there cannot exist any $\lambda \in \mathbb{R}$ such that $Df(0, 0) + \lambda Dg(0, 0) = (0, 0)$. Thus, the conclusions of the Theorem of Lagrange also fail.

2.2 The Lagrangean Multipliers

The vector $\lambda^* = (\lambda_1^*, \cdots, \lambda_k^*)$ described in the Theorem of Lagrange is called the vector of **Lagrangean Multipliers** according to the local optimum x^*. The i-th multiplier λ_i^* measures, in a sense, the sensitivity of the value of the objective function at x^* to a small relaxation of the i-th constraint g_i. We demonstrate this interpretation of λ^* under some assumptions designed to simplify the exposition.

We begin with a clarification of the notion of the "relaxation" of a constraint. To this end, we will suppose in the rest of this subsection that he constraint function g are given in parametric form as

$$g(x; c) = g(x) + c, \tag{16}$$

where $c = (c_1, \cdots, c_k)$ is a vector of constants. This assumption enables a formal definition of the concept we need: a relaxation of the i-th constraint may now be thought of as an increase in the value of the constant c_i.

Now let C be some open set of feasible values of c. Suppose that for each $c \in C$, there is a global optimum, denoted $x^*(c)$, of f on the constrained set

$$\mathcal{D} = U \cap \{x \in \mathbb{R}^n | g(x; c) = 0\}. \tag{17}$$

Suppose further that, for each $c \in C$, the constraint qualification holds at $x^*(c)$, so there exists $\lambda^*(c) \in \mathbb{R}^k$ such that

$$Df(x^*(c)) + \sum_{i=1}^{k} \lambda_i^* Dg_i(x^*(c)) = 0. \tag{18}$$

Finally, suppose that $x^*(\cdot)$ is a differentiable function on C, that is, that the optimum changes smoothly with changes in the underlying parameters. Let

$$F(c) = f(x^*(c)) \tag{19}$$

be the value of the objective function at the optimum given the parameter c. Since f is C^1 and $x^*(\cdot)$ has been assumed to be differentiable on C, $F(\cdot)$ is also differentiable on C. We will show that

$$DF(c) = \lambda^*(c), \tag{20}$$

that is, that $\partial F(c)/\partial c_i = \lambda_i^*$. In words, this states precisely that λ_i^* represents the sensitivity of the objective function at x^* to a small relaxation in the i-th constraint.

First, note that from the definition of $F(\cdot)$, we have

$$DF(c) = Df(x^*(c))Dx^*(c), \tag{21}$$

where $Dx^*(c)$ is the $n \times k$ matrix whose (i, j)-th entry is $\partial x_i^*(c)/\partial c_j$. By the first-order condition at each c, we also have

$$Df(x^*(c)) = -\sum_{i=1}^{k} \lambda_i^*(c) Dg_i(x^*(c)). \tag{22}$$

By combining the last two expressions, we obtain

$$DF(c) = -\left(\sum_{i=1}^{k} \lambda_i^*(c) Dg_i(x^*(c))\right) Dx^*(c) = -\sum_{i=1}^{k} \lambda_i^*(c) Dg_i(x^*(c)) Dx^*(c). \tag{23}$$

Since $x^*(c)$ must be feasible at all c, it must be identically be true for each i and for all c that $g_i(x^*(c)) + c_i = 0$. Differentiating with respect to c, and rearranging, we obtain

$$Dg_i(x^*(c)) Dx^*(c) = -e_i, \tag{24}$$

where e_i is the i-th unit vector in $\mathbb{R}^k$, that is, the vector that has a 1 in the i-th place and zeros elsewhere. It follows that

$$DF(c) = -\sum_{i=1}^{k} \lambda_i^*(c)(-e_i) = \lambda^*(c), \tag{25}$$

and the proof is complete.

An economic interpretation of the result we have just derived bears mention. Since $\lambda^*(c) = \partial f(x^*(c))/\partial c$, a small relaxation in constraint i will raise the maximized value of the objective function by $\lambda_i^*(c)$. Therefore, $\lambda_i^*(c)$ also represents **the maximum amount the decision-maker will be willing to pay for a marginal relaxation of constraint** i, and is the marginal value or the **"shadow price"** of constraint i at c.

3 Second-Order Conditions

The Theorem of Lagrange gives us the first-order conditions for optimization problems with equality constraints. In this section, we describe **second-order conditions** for such problems. We will assume in this section that f and g are both C^2 functions. The following notation will come in handy: given any $\lambda \in \mathbb{R}^k$ define the function $L : \mathbb{R}^n \to \mathbb{R}$ by

$$L(x, \lambda) = f(x) + \sum_{i=1}^{k} \lambda_i g_i(x). \tag{26}$$

Note that the second derivative $D^2L(x, \lambda)$ of $L(x, \lambda)$ with respect to x is the $n \times n$ matrix defined by

$$D^2L(x, \lambda) = D^2 f(x) + \sum_{i=1}^{k} \lambda_i D^2 g_i(x). \tag{27}$$

Since f and g are both C^2 functions of x, so is $L(\cdot, \lambda)$ for any $\lambda \in \mathbb{R}^k$. Thus, $D^2L(x, \lambda)$ is a symmetric matrix and defines a quadratic form on $\mathbb{R}^n$.

Theorem 2. *Suppose there exist points $x^* \in \mathcal{D}$ and $\lambda^* \in \mathbb{R}^k$ such that $r(Dg(x^*)) = k$, and $Df(x^*) + \sum_{i=1}^{k} \lambda_i^* Dg_i(x^*) = 0$. Define*

$$\mathcal{Z}(x^*) = \{z \in \mathbb{R}^n | Dg(x^*)z = 0\}, \tag{28}$$

and let D^2L^* denote the $n \times n$ matrix $D^2L(x^*, \lambda^*) = D^2 f(x^*) + \sum_{i=1}^{k} \lambda_i^* D^2 g_i(x^*)$.

1. *If f has a local maximum on $\mathcal{D}$ at x^*, then $z' D^2 L^* z \leqslant 0$ for all $z \in \mathcal{Z}(x^*)$.*

2. *If f has a local minimum on $\mathcal{D}$ at x^*, then $z' D^2 L^* z \geqslant 0$ for all $z \in \mathcal{Z}(x^*)$.*

3. *If $z' D^2 L^* z < 0$ for all $z \in \mathcal{Z}(x^*)$ with $z \neq 0$, then x^* is a strict local maximum of f on $\mathcal{D}$.*

4. *If $z' D^2 L^* z > 0$ for all $z \in \mathcal{Z}(x^*)$ with $z \neq 0$, then x^* is a strict local minimum of f on $\mathcal{D}$.*

Proof. See Sundaram (1996), section 5.7. □

4 A "Cookbook" Procedure with an Example in Consumer Theory

Let an equality-constrained optimization problem of the form

$$\begin{aligned} \max \quad & f(x) \\ s.t. \quad & x \in \mathcal{D} \end{aligned} \tag{29}$$

Where $\mathcal{D} = U \cap \{x \in \mathbb{R}^n | g(x) = 0\}$, $U \subset \mathbb{R}^n$ is open, $g : \mathbb{R}^n \to \mathbb{R}^k$. Assume that functions f and g are both C^1. We describe here a "cookbook" procedure for using Theorem of Lagrange to solve this maximization problem. This procedure, usually called the **Lagrangean method**, involves three steps.

In the first step, we set up a function $L : \mathcal{D} \times \mathbb{R}^k \to \mathbb{R}$, called the **Lagrangean**, defined by

$$L(x, \lambda) = f(x) + \sum_{i=1}^{k} \lambda_i g_i(x). \tag{30}$$

The vector $\lambda = (\lambda_1, \cdots, \lambda_k)$ is called the vector of **Lagrange multipliers**.

As the second step, we find all critical points (x, λ) solving the equation system $DL(x, \lambda) = 0$ and $x \in U$. This equation system involves a system of $(n + k)$ equations in the $(n + k)$ unknowns:

$$\begin{aligned} \frac{\partial L}{\partial x_j}(x, \lambda) = 0, \quad & j = 1, \cdots, n \\ \frac{\partial L}{\partial \lambda_i}(x, \lambda) = 0, \quad & i = 1, \cdots, k \end{aligned} \tag{31}$$

As the third and the last step, what we do depends on whether the problem is numerical or theoretical (parameterized). For a numerical problem, we evaluate f at all critical points and then find the maximum by comparing the function values. For a parameterized problem we need first to verify whether f and g are both C^2. If this is true, then we need to employ the sufficient conditions in Theorem 2 to judge whether a critical point is a maximum or a minimum.

Next, we illustrate the method by solving a consumer's utility maximization problem.

Example 3. *The consumer has an income $I > 0$, and the price of commodity i is $p_i > 0$. The problem is to solve*

$$\begin{aligned} \max \quad & x_1 x_2 \\ s.t. \quad & I - p_1 x_1 - p_2 x_2 \geqslant 0 \\ & x_1 \geqslant 0, x_2 \geqslant 0 \end{aligned} \tag{32}$$

First, it's evident that the utility function is continuous on the budget set

$$\mathcal{B}(p, I) = \{(x_1, x_2) | I - p_1 x_1 - p_2 x_2 \geqslant 0, x_1 \geqslant 0, x_2 \geqslant 0\}, \tag{33}$$

and that the budget set is compact. Then, by the Weierstrass' Theorem, a solution (x_1^*, x_2^*) does exist.

In addition, notice that if either $x_1 = 0$ or $x_2 = 0$, $u(x_1, x_2) = 0$. But we can easily find a feasible consumption bundle which gives strictly positive utility, say $(x_1, x_2) = (I/2p_1, I/2p_2)$. Then we know that positivity constraints are not binding. Moreover, the optima must satisfy the equality constraint $I = p_1 x_1 + p_2 x_2$, since, if otherwise, utility can be increased.

Hence the problem is reformed in a standard equality-constraint maximization problem

$$\begin{aligned} \max \quad & x_1 x_2 \\ s.t. \quad & (x_1, x_2) \in \mathcal{B}^*(p, I), \end{aligned} \tag{34}$$

where the budget set $\mathcal{B}^*(p, I) = \mathbb{R}^2_{++} \cap \{(x_1, x_2) | p_1 x_1 + p_2 x_2 = I\}$. In notations we use in the previous sections, $U = \mathbb{R}^2_{++}$, and $g(x_1, x_2) = I - p_1 x_1 - p_2 x_2$.

Next, we can establish the Lagrangean and find the critical points. The Lagrangean is

$$L(x_1, x_2, \lambda) = x_1 x_2 + \lambda(I - p_1 x_1 - p_2 x_2). \tag{35}$$

The critical points are the solutions $(x_1^*, x_2^*, \lambda^*) \in \mathbb{R}^2_{++} \times \mathbb{R}$ to

$$\begin{aligned} x_2 - \lambda p_1 &= 0 \\ x_1 - \lambda p_2 &= 0 \\ I - p_1 x_1 - p_2 x_2 &= 0. \end{aligned} \tag{36}$$

The unique solution is $(x_1^*, x_2^*, \lambda^*) = (\frac{I}{2p_1}, \frac{I}{2p_2}, \frac{I}{2p_1 p_2})$.

Finally, we use Theorem 2 to judge whether $(x_1^*, x_2^*, \lambda^*)$ is a local maximum. Note that $Dg(x_1^*, x_2^*) = (-p_1, -p_2)$, so the set $\mathcal{Z}(x^*) = \{z \in \mathbb{R}^2 | Dg(x^*) z = 0\}$ is simply

$$\mathcal{Z}(x^*) = \{z \in \mathbb{R}^2 | z_1 = -\frac{p_2 z_2}{p_1}\}. \tag{37}$$

Defining $D^2 L^* = D^2 u(x^*) + \lambda^* D^2 g(x^*)$, we have

$$D^2 L^* = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda^* \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{38}$$

So for any $z \in \mathcal{Z}(x^*)$ with $z \neq 0$, we have $z' D^2 L^* z = -2 p_2 z_2^2 / p_1 < 0$. Thus $(x_1^*, x_2^*) = (\frac{I}{2p_1}, \frac{I}{2p_2})$ is a local maximum.

5 Exercise

1. Find the consumption bundle that maximizes the Cobb-Douglas Utility function.

$$\begin{aligned} \max \quad & x_1^a x_2^b \\ s.t. \quad & I - p_1 x_1 - p_2 x_2 \geqslant 0 \\ & x_1 \geqslant 0, x_2 \geqslant 0 \end{aligned} \tag{39}$$

   where $a > 0, b > 0, p_1 > 0, p_2 > 0, I > 0$.

References

Ok, Efe A, *Real analysis with economic applications*, Vol. 10, Princeton University Press, 2007.

Rudin, Walter, *Principles of mathematical analysis*, Vol. 3, McGraw-Hill New York, 1964.

Sundaram, Rangarajan K, *A first course in optimization theory*, Cambridge university press, 1996.

Lecture 4: Inequality Constraints and the Theorem of Kuhn and Tucker

Liu Peng
August 16, 2015

1 Constrained Optimization Problems with Inequality Constraints

The constraint set is now assumed to be

$$\mathcal{D} = U \cap \{x \in \mathbb{R}^n | h_i(x) \geqslant 0, i = 1, \cdots, l\} \tag{1}$$

where $U \subset \mathbb{R}^n$ is open, and $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \cdots, l$.

2 The Theorem of Kuhn and Tucker

In the statement of the theorem, we say that an inequality constraint $h_i(x) \geqslant 0$ is **effective** at a point x^* if the constraint holds with equality at x^*, that is, we have $h_i(x^*) = 0$. We will also use the expression $|E|$ to denote the **cardinality** of a finite set E, i.e., the number of elements in the set E.

Theorem 1. ***(Theorem of Kuhn and Tucker)*** *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions, $i = 1, \cdots, l$. Suppose x^* is a local maximum of f on*

$$\mathcal{D} = U \cap \{x \in \mathbb{R}^n | h_i(x) \geqslant 0, i = 1, \cdots, l\}, \tag{2}$$

where $U \subset \mathbb{R}^n$ is open. Let $E \subset \{1, \cdots, l\}$ denote the set of effective constraints at x^*, and let $h_E = (h_i)_{i \in E}$. Suppose $r(Dh_E(x^*)) = |E|$. Then, there exists a vector $\lambda^* = (\lambda_1^*, \cdots, \lambda_l^*) \in \mathbb{R}^l$ such that the following conditions are met:

$$\begin{aligned} &[KT-1] \quad \lambda_i^* \geqslant 0 \quad \textit{and} \quad \lambda_i^* h_i(x^*) = 0 \quad i = 1, \cdots, l. \\ &[KT-2] \quad Df(x^*) + \sum_{i=1}^{l} \lambda_i^* Dh_i(x^*) = 0. \end{aligned} \tag{3}$$

Condition $[KT-1]$ in Theorem 1 is called the condition of **"complementary slackness"**. The terminology arises simply from the fact that we must have $\lambda_i^* = 0$ if $h_i(x^*) > 0$, and $h_i(x^*) = 0$ if $\lambda_i^* > 0$. That is, if one inequality if "slack", the other cannot be. As the remark following the Lagrange Theorem, we have to assert here that the Theorem of Kuhn and Tucker provides conditions that are only necessary for local optima, at which constraint qualification is met.

Proof. (Theorem of Kuhn and Tuker) To simplify notation in the proof, we will denote $|E|$ by k; we will also assume that the effective constraints at x^* are the first k constraints:

$$\begin{aligned} h_i(x^*) &= 0, \quad i = 1, \cdots, k \\ h_i(x^*) &> 0, \quad i = k+1, \cdots, l. \end{aligned} \tag{4}$$

There is no loss of generality in this assumption, since this can be done by renumbering the constraints.

For each $i \in \{1, \cdots, l\}$, define

$$V_i = \{x \in \mathbb{R}^n | h_i(x) > 0\}. \tag{5}$$

Let $V = \cap_{i=k+1}^{l} V_i$. By the continuity of h_i, V_i is open for each i, and so, therefore, is V. Now, let $D^* \subset D$ be the equality-constrained set defined through k equality constraints and given by

$$D^* = U \cap V \cap \{x \in \mathbb{R}^n | h_i(x) = 0, i = 1, \cdots, k\}. \tag{6}$$

By construction, we have $x^* \in D^*$. Since x^* is a local maximum of f on D, it is certainly a local maximum of f on D^*. Moreover, we have $r(Dh_E(x^*)) = k$, by hypothesis. Therefore, by the theorem of Lagrange, there exists a vector $\mu = (\mu_1, \cdots, \mu_k) \in \mathbb{R}^k$ such that

$$Df(x^*) + \sum_{i=1}^{k} \mu_i Dh_i(x^*) = 0. \tag{7}$$

Now define $\lambda \in \mathbb{R}^l$ by

$$\lambda_i = \begin{cases} \mu_i, & i = a, \cdots, k \\ 0, & i = k+1, \cdots, l. \end{cases} \tag{8}$$

We will show that the vector λ satisfies the properties stated in the theorem.

First, observe that for $i = k+1, \cdots, l$, we have $\lambda_i = 0$, and, therefore, $\lambda_i Dh_i(x^*) = 0$. Therefore,

$$\begin{aligned} Df(x^*) + \sum_{i=1}^{l} \lambda_i Dh_i(x^*) &= Df(x^*) + \sum_{i=1}^{k} \lambda_i Dh_i(x^*) \\ &= Df(x^*) + \sum_{i=1}^{k} \mu_i Dh_i(x^*) = 0. \end{aligned} \tag{9}$$

which establishes $[KT-2]$.

Now, for any $i \in \{1, \cdots, k\}$, we have $h_i(x^*) = 0$, so certainly it is the case that $\lambda_i h_i(x^*) = 0$ for $i \in \{1, \cdots, k\}$. For $i \in \{k+1, \cdots, l\}$, we have $\lambda_i = 0$, so it is also the case that $\lambda_i h_i(x^*) = 0$ for $i \in \{k+1, \cdots, l\}$.

It remains to show $\lambda_i \geqslant 0$ for all i. Since $\lambda_i = 0$ for $i \in \{k+1, \cdots, l\}$, we are required to show only $\lambda_i \geqslant 0$ for $i \in \{1, \cdots, k\}$. We will establish $\lambda_1 \geqslant 0$. A similar argument will establish $\lambda_i \geqslant 0$ for $i \in \{2, \cdots, k\}$.

Define for $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}$, the function $H = (H_1, \cdots, H_k)$ by

$$\begin{aligned} H_1(x, \eta) &= h_1(x) - \eta \\ H_i(x, \eta) &= h_i(x), \quad i = 2, \cdots, k. \end{aligned} \tag{10}$$

Notice that, since $h_E = (h_1, \cdots, h_k)$, we have $DH_x(x, \eta) = Dh_E(x)$ and we also have $DH_\eta(x, \eta) = (-1, 0, \cdots, 0)$ at any (x, η).

By the definition of H, we have $H(x^*, 0) = 0$. Moreover, $r(DH_x(x^*, 0)) = r(Dh_E(x^*)) = k$. Therefore, by the Implicit Function Theorem, there is a neighborhood of 0, and a C^1 function $\xi(0) = x^*$, and

$$H(\xi(\eta), \eta) = 0, \quad \eta \in N. \tag{11}$$

Differentiating this expression with respect to η and evaluating the derivative at $\eta(0) = x^*$, we have

$$DH_x(x^*, 0)D\xi(0) + DH_\eta(x^*, 0) = 0. \tag{12}$$

Since $DH_x(x, \eta) = Dh_E(x)$ and $DH_\eta(x, \eta) = (-1, 0, \cdots, 0)$ at any (x, η), this implies

$$Dh_E(x^*)D\xi(0) = (1, 0, \cdots, 0). \tag{13}$$

or, equivalently, that

$$\begin{aligned} Dh_1(x^*)D\xi(0) &= 1, \\ Dh_i(x^*)D\xi(0) &= 0, \quad i = 1, \cdots, k. \end{aligned} \tag{14}$$

Since $\lambda_i = 0$ for $i = k + 1, \cdots, l$, we now have

$$\begin{aligned} Df(x^*)D\xi(0) &= -\left(\sum_{i=1}^{l} \lambda_i Dh_i(x^*)\right) D\xi(0) \\ &= -\left(\sum_{i=1}^{k} \lambda_i Dh_i(x^*) D\xi(0)\right) \\ &= -\lambda_1. \end{aligned} \tag{15}$$

To complete the proof, we will show that $Df(x^*)D\xi(0) \leqslant 0$. To this end, we first show that there is an $\eta^* > 0$ such that for all $\eta \in [0, \eta^*)$, we must have $\xi(\eta) \in \mathcal{D}$, where $\mathcal{D}$ is the constraint set of the original problem.

If $\eta > 0$, then $H_i(\xi(\eta))$ for $i = 1, \cdots, k$, and from the definition of the functions H_i, this means

$$\begin{aligned} h_1(\xi(\eta)) &= \eta > 0, \\ h_i(\xi(\eta)) &= 0, \quad i = 2, \cdots, k. \end{aligned} \tag{16}$$

For $i = k + 1, \cdots, l$, we have $h_i(\xi(0)) = h_i(x^*) > 0$. Since both h_i and ξ are continuous, it follows that by choosing η sufficiently small, we can ensure that $h_i(\xi(\eta)) > 0$, for $i = k + 1, \cdots, l$.

Finally, by shrinking the value of η^* if need be, we can evidently ensure that $\xi(\eta) \in U$ for all $\eta \in [0, \eta^*)$, as claimed.

Now, since $\xi(0) = x^*$ is a local maximum, and $\xi(\eta)$ is in the feasible set for $\eta \in [0, \eta^*)$, it follows that for $\eta > 0$ and sufficiently close to zero, we must have $f(x^*) \geqslant f(\xi(\eta))$.

Therefore,

$$\left(\frac{f(\xi(0)) - f(\xi(\eta))}{\eta} \right) \leqslant 0 \tag{17}$$

for all $\eta > 0$ and sufficiently small. Taking limit as $\eta \to 0$, we obtain $Df(x^*)D\xi(0) \leqslant 0$. The proof is complete.

□

2.1 The Constraint Qualification

As with the analogous condition in the Theorem of Lagrange, the condition in the Theorem of Kuhn and Tuker that the rank of $Dh_E(x^*)$ be equal to $|E|$ is called the **constraint qualification**. This condition plays a central role in the proof of the theorem. Moreover, if the constraint qualification fails, the theorem itself could fail.

2.2 The Kuhn-Tuker Multipliers

The vector λ^* in the Theorem of Kuhn and Tuker is called the vector of **Kuhn-Tuker multipliers** corresponding to the local maximum x^*. As with the Lagrangean multipliers, the Kuhn-Tucker multipliers may also be thought of as measuring the sensitivity of the objective function at x^* to relaxations of the various constraints. Indeed, this interpretation is particularly intuitive in the context of inequality constraints. To wit, if $h_i(x^*) > 0$, then the i-th constraint is already slack, so relaxing it further will not help raise the value of the objective function in the maximization exercise, and λ_i^* must be zero. On the other hand, if $h_i(x^*) = 0$, then relaxing the i-th constraint may help increase the value of the maximization exercise, so we have $\lambda_i^* \geqslant 0$.

3 A "Cookbook" Procedure with a numerical example

Consider a problem of

$$\begin{aligned} \max \quad & f(x) \\ s.t. \quad & x \in \mathcal{D} = U \cap \{x | h(x) \geqslant 0\}. \end{aligned} \tag{18}$$

As the first step, we form the Lagrangean

$$L(x, \lambda) = f(x) + \sum_{i=1}^{l} \lambda_i h_i(x). \tag{19}$$

The second step is to find all solutions (x, λ) to the following set of equations:

$$\begin{gathered} \frac{\partial L}{\partial x_j}(x, \lambda) = 0, \quad j = 1, \cdots, n, \\ \frac{\partial L}{\partial \lambda_i}(x, \lambda) \geqslant 0, \quad \lambda_i \geqslant 0, \quad \lambda_i \frac{\partial L}{\partial \lambda_i}(x, \lambda) = 0, \quad i = 1, \cdots, l. \end{gathered} \tag{20}$$

All solutions to this system of equations are called ”critical points”.

At the last step, we compute the value of f at each x which is a part of a critical point and an element in U. Then, by comparing those values of f, we can find the optima.

Example 1. *Let $g(x, y) = 1 - x^2 - y^2$. Consider the problem of maximizing $f(x, y) = x^2 - y$ over the set*

$$\mathcal{D} = \{(x, y)|g(x, y) \geqslant 0\}. \tag{21}$$

We claim first here that there exists a maximum of f on $\mathcal{D}$, by the Weierstrass' extreme theorem. Note first that $Dg(x, y) = (-2x, -2y)$ at all points. Note also that at any point where the single constraint is effective, we must have either $x \neq 0$ or $y \neq 0$. Thus, the constraint qualification holds if the optimum occurs at a point (x, y) where $g(x, y) = 0$.

We establish the Lagrangean $L(x, y, \lambda) = x^2 - y + \lambda(1 - x^2 - y^2)$. The critical points are the solutions (x, y, λ) to

$$\begin{gathered} 2x - 2\lambda x = 0 \\ -1 - 2\lambda y = 0 \\ \lambda \geqslant 0, \quad (1 - x^2 - y^2) \geqslant 0, \quad \lambda(1 - x^2 - y^2) = 0. \end{gathered} \tag{22}$$

For the first equation to hold, we must have $x = 0$ or $\lambda = 1$. If $\lambda = 1$, then from the second equation, we must have $y = -\frac{1}{2}$, while from the third equation, we must have $x^1 + y^2 = 1$. This gives us two critical points of L, which differ only in the value of x:

$$(x, y, \lambda) = \left(\pm\frac{\sqrt{3}}{2}, -\frac{1}{2}, 1\right). \tag{23}$$

Note that at either of these two critical points, we have $f(x, y) = \frac{5}{4}$.

This leaves the case $x = 0$. If we also have $\lambda = 0$, then the second equation cannot be satisfied, so we must have $\lambda > 0$. This implies from the third equation that $x^2 + y^2 = 1$, so $y = \pm 1$. Since $y = 1$ is inconsistent with a positive value for λ from the second equation, the only possible critical point in this case is

$$(x, y, \lambda) = \left(0, -1, \frac{1}{2}\right). \tag{24}$$

At this critical point, we have $f(0, -1) = 1 < \frac{5}{4}$, which means this point cannot be a solution to the original maximization problem.

Since here does not exist any more critical point, we know that there are exactly two solutions to the optimization problem, namely the points $(x, y) = (\pm\frac{\sqrt{3}}{2}, -\frac{1}{2})$.

4 The General Case: Mixed Constraints

A constrained optimization problem with mixed constraints is a problem where the constraint set $\mathcal{D}$ has the form

$$\mathcal{D} = U \cap \{x \in \mathbb{R}^n | g(x) = 0, h(x) \geqslant 0\}, \tag{25}$$

where $U \subset \mathbb{R}^n$ is open, $g : \mathbb{R}^n \to \mathbb{R}^k$, and $h : \mathbb{R}^n \to \mathbb{R}^l$. For notational ease, define $\phi_i : \mathbb{R}^n \to \mathbb{R}^{k+l}$, where

$$\phi_i = \begin{cases} g_i, & i = 1, \cdots, k \\ h_{i-k}, & i = k+1, \cdots, k+l. \end{cases} \tag{26}$$

The following theorem is a simple consequence of combining the Theorem of Lagrange with the Theorem of Kuhn and Tuker.

Theorem 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$, and $\phi_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \cdots, k+l$ be C^1 functions. Suppose x^* maximizes f on*

$$\mathcal{D} = U \cap \{x \in \mathbb{R}^n | \phi_i(x) = 0, i = 1 \cdots, k, \phi_j(x) \geqslant 0, j = k+1, \cdots, k+l\}, \tag{27}$$

where $U \subset \mathbb{R}^n$ is open. Let $E \subset \{1, \cdots, k+l\}$ denote the set of effective constraints at x^*, and let $\phi_E = (\phi_i)_{i \in E}$. Suppose $r(D\phi_E(x^*)) = |E|$. Then, there exists $\lambda \in \mathbb{R}^{k+l}$ such that

1. *$\lambda_j \geqslant 0$ and $\lambda_j \phi_j(x^*) = 0$ for $j \in \{k+1, \cdots, k+l\}$.*
2. *$Df(x^*) + \sum_{i=1}^{k+l} D\phi_i(x^*) = 0$.*

5 Convexity and Optimization

Definition 1. *A set $D \in \mathbb{R}^n$ is called* ***convex*** *if the convex combination of any two points in D is itself in D, that is, if for all x and y in D and all $\lambda \in (0, 1)$, it is the case that $\lambda x + (1-\lambda)y \in D$.*

Definition 2. *A function $f : \mathcal{D} \to \mathbb{R}$ is* ***concave*** *on $\mathcal{D}$ if for all $x, y \in \mathcal{D}$ and $\lambda \in (0, 1)$, it is the case that*

$$f(\lambda x + (1-\lambda)y) \geqslant \lambda f(x) + (1-\lambda) f(y). \tag{28}$$

Similarly, $f : \mathcal{D} \to \mathbb{R}$ is ***convex*** *on $\mathcal{D}$ if for all $x, y \in \mathcal{D}$ and $\lambda \in (0, 1)$, it is the case that*

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y). \tag{29}$$

Strict concavity and strict convexity are defined, respectively, with the strict inequality holding.

5.1 Some General Observations

This subsection presents two results which indicate the importance of convexity for optimization theory. The first result establishes that in convex optimization problems, all local optima must also be global optima; and, therefore, that to find a global optimum in such problems, it always suffices to locate a local optimum. The second result shows that if a strictly convex optimization problem admits a solution, the solution must be unique.

Theorem 3. *Suppose $\mathcal{D} \subset \mathbb{R}^n$ is convex, and $f : \mathcal{D} \to \mathbb{R}$ is concave. Then,*

1. *Any local maximum of f is a global maximum of f.*

2. *The set* $\arg\max\{f(x)|x \in \mathcal{D}\}$ *of maximizers of* f *on* $\mathcal{D}$ *is either empty or convex.*

Proof. Suppose f admits a local maximum x that is not also a global maximum. Since x is a local maximum, there is $r > 0$ such that $f(x) \geqslant f(y)$ for all $y \in B(x,r) \cap \mathcal{D}$. Since x is not a global maximum, there is $z \in \mathcal{D}$ such that $f(z) > f(x)$. Since $\mathcal{D}$ is convex, $(\lambda x + (1-\lambda)z) \in \mathcal{D}$ for all $\lambda \in (0,1)$. Pick λ sufficiently close to unity so that $(\lambda x + (1-\lambda)z) \in B(x,r)$. By the concavity of f,

$$f(\lambda x + (1-\lambda)z) \geqslant \lambda f(x) + (1-\lambda)f(z) > f(x). \tag{30}$$

But $\lambda x + (1-\lambda)z$ is in $B(x,r)$ by construction, so $f(x) \geqslant f(\lambda x + (1-\lambda)z)$: contradiction, which establishes part 1.

To see part 2, suppose x_1 and x_2 are both maximizers of f on $\mathcal{D}$. Then, we have $f(x_1) = f(x_2)$. Further, for $\lambda \in (0,1)$, we have

$$f(\lambda x_1 + (1-\lambda)x_2) \geqslant \lambda f(x_1) + (1-\lambda)f(x_2) = f(x_1), \tag{31}$$

and this must hold with equality or x_1 and x_2 would not be maximizers. Thus, the set of maximizers must be convex, completing the proof. □

Theorem 4. *Suppose* $\mathcal{D} \subset \mathbb{R}^n$ *is convex, and* $f : \mathcal{D} \to \mathbb{R}$ *is strict concave. Then* $\arg\max\{f(x)|x \in \mathcal{D}\}$ *either is empty or contains a single point.*

Proof. Suppose the maximizer set is nonempty. We have already shown in the last theorem that the maximizer set is convex. Suppose this set contains two distinct points x and y. Pick any $\lambda \in (0,1)$ and let $z = \lambda x + (1-\lambda)y$. Then, the concavity of f gives

$$f(z) = f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda)f(y) = f(x), \tag{32}$$

which gives the contradiction. □

5.2 Convexity and Unconstrained Optimization

The following result shows that the first-order condition for unconstrained optima (i.e., the condition that $Df(x) = 0$) is both necessary and sufficient to identify global unconstrained maxima, when such maxima exist.

Theorem 5. *Let* $\mathcal{D} \subset \mathbb{R}^n$ *be convex, and* $f : \mathcal{D} \to \mathbb{R}$ *be a concave and differentiable function on* $\mathcal{D}$*. Then,* x *is an unconstrained maximum of* f *on* $\mathcal{D}$ *if and only if* $Df(x) = 0$.

Proof. We have known that the condition $Df(x) = 0$ must hold whenever x is any unconstrained local maximum. It must evidently also hold, therefore, if x is an unconstrained global maximum.

The inverse comes from a characterizing property of concavity. Suppose x and y are two points in $\mathcal{D}$. Function f being concavity implies

$$f(y) - f(x) \leqslant Df(x)(y-x). \tag{33}$$

If $Df(x) = 0$, this property is exactly the same as stating $f(x) \geqslant f(y)$. Since y is arbitrary, x is global maximum of f on $\mathcal{D}$. □

5.3 Convexity and the Theorem of Kuhn and Tuker

The following result states that the first-order conditions of the Theorem of Kuhn and Tuker are both necessary and sufficient conditions to identify optima of convex inequality-constrained optimization problems, provided a mild regularity condition is met.

Theorem 6. ***Theorem of Kuhn and Tucker under Convexity*** *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}$ be concave C^1 functions, $i = 1, \cdots, l$. Suppose there is some $\bar{x} \in U$ such that $h_i(\bar{x}) > 0$ for all $i \in 1 \cdots, l$. Suppose also that $U \subset \mathbb{R}^n$ is convex and open. Then x^* maximizes f on*

$$\mathcal{D} = \{x \in U | h_i(x) \geqslant 0, i = 1, \cdots, l\}, \tag{34}$$

if and only if there exists a vector $\lambda^* = (\lambda_1^*, \cdots, \lambda_l^*) \in \mathbb{R}^l$ such that the following conditions are met:

$$\begin{aligned} &[KT-1] \quad \lambda_i^* \geqslant 0 \quad \textit{and} \quad \lambda_i^* h_i(x^*) = 0 \quad i = 1, \cdots, l. \\ &[KT-2] \quad Df(x^*) + \sum_{i=1}^{l} \lambda_i^* Dh_i(x^*) = 0. \end{aligned} \tag{35}$$

Proof. See Sundaram (1996), page 194-198. □

The condition that there exist a point $\bar{x}$ at which $h_i(\bar{x}) > 0$ for all i is called **Slater's condition.** There are two points about this condition that bear stressing.

First, Slater's condition is used only in the proof that $[KT-1]$ and $[KT-2]$ are necessary at an optimum. It plays no role in proving sufficiency. That is, the condition $[KT-1]$ and $[KT-2]$ are sufficient to identify an optimum when f and the functions h_i are concave, regardless of whether Slater's condition is satisfied or not.

Second, the necessity of conditions $[KT-1]$ and $[KT-2]$ at any local maximum was established in Theorem 1, but under a different hypothesis, namely, that the rank condition described in the Theorem 1 held. Effectively, the necessity part of the Theorem 6 states that this rank condition can be replaced with the combination of Slater's condition and concave constraint functions. However, both parts of this combination are important: just as the necessity of $[KT-1]$ and $[KT-2]$ could fail if the rank condition is not met, the necessity of $[KT-1]$ and $[KT-2]$ could also fail if either Slater's condition or the concavity of the functions h_i fails.

6 Exercise

1. Solve the following maximization problem

$$\begin{aligned} \max \quad & \ln x + \ln y \\ s.t. \quad & x^2 + y^2 = 1 \\ & x, y \geqslant 0. \end{aligned} \tag{36}$$

References

Ok, Efe A, *Real analysis with economic applications*, Vol. 10, Princeton University Press, 2007.

Rudin, Walter, *Principles of mathematical analysis*, Vol. 3, McGraw-Hill New York, 1964.

Sundaram, Rangarajan K, *A first course in optimization theory*, Cambridge university press, 1996.

Lecture 5: Fixed Point Theorems

Liu Peng
August 16, 2015

Definition 1. *Let X be a metric space, and let $T : X \to X$ be a function that maps X into itself, a* ***fixed point*** *of T is an element $x \in X$ for which $T(x) = x$. In addition, we call a function mapping a set into itself* ***a self-map****.*

1 Contraction and Banach Fixed Point Theorem

Definition 2. *Let X be a metric space. A self-map Φ on X is said to be a* ***contraction*** *if there exists a real number $0 < K < 1$ such that*

$$d(\Phi(x), \Phi(y)) \leqslant Kd(x, y) \quad \textit{for all } x, y \in X. \tag{1}$$

In this case, the infimum of the set of all such K is called the ***contraction coefficient*** *of Φ.*

Why are contractions of interest? Because, when it is defined on a complete metric space, a contraction must map a point to itself, that is, it must have a fixed point. In fact more is true, it must have a unique fixed point. It is actually rather easy to give an intuition of this fact in the case of real functions - a geometric description is provided in Figure 1. The idea is to pick any point x in the domain of a given contraction $f : \mathbb{R} \to \mathbb{R}$, and observe the behavior of the sequence $(x, f(x), f(f(x)), \cdots)$. The contraction property tells us that this is a Cauchy sequence, and since $\mathbb{R}$ is complete, it must converge.

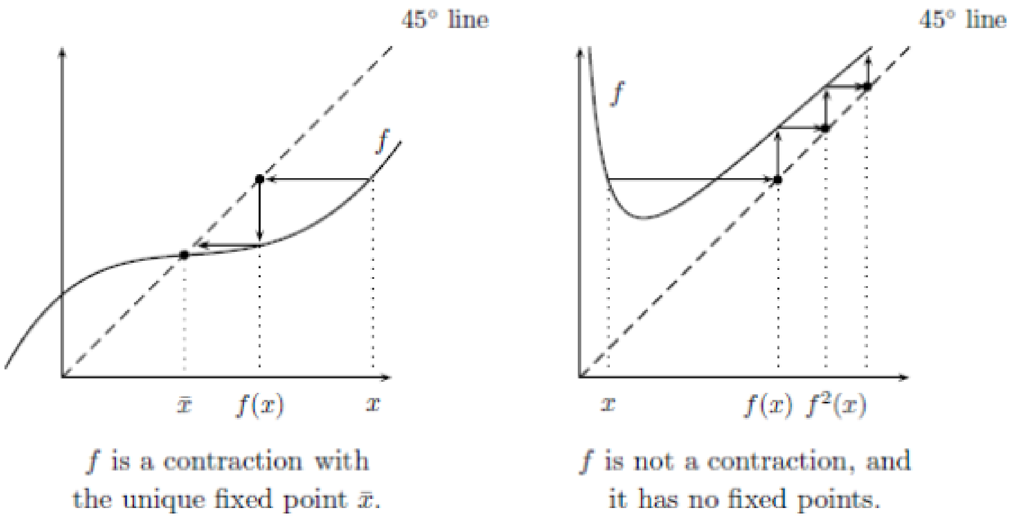

Figure 1: Contraction and Fixed Point

We now generalize this heuristic argument to the case of contraction defined on an arbitrary complete metric space.

Theorem 1. ***(The Banach Fixed Point Theorem)*** *Let X be a complete metric space. If $\Phi : X \to X$ is a contraction, then there exists a unique $x^* \in X$ such that $\Phi(x^*) = x^*$.*

Proof. Let us first prove the existence of a fixed point. Pick any $x_0 \in X$ and define $\{x_k\}$ recursively as $x_{k+1} = \Phi(x_k)$ for all $k = 0, 1, \cdots$. We claim that this sequence is Cauchy. To see this, let K be the contraction coefficient of Φ, and notice that, by the Principle of Mathematical Induction, we have $d(x_{k+1}, x_k) \leqslant K^k d(x_1, x_0)$ for all $k = 1, 2, \cdots$. Thus, for any $k > l$,

$$\begin{aligned} d(x_k, x_l) &\leqslant d(x_k, x_{k-1}) + \cdots + d(x_{l+1}, x_l) \\ &\leqslant (K^{k-1} + \cdots + K^l) d(x_1, x_0) \\ &= \frac{K^l(1 - K^{k-l})}{1 - K} d(x_1, x_0) \end{aligned} \tag{2}$$

so that $d(x_k, x_l) < \frac{K^l}{1-K} d(x_1, x_0)$. Then $\{x_k\}$ being Cauchy follows immediately from this inequality.

Since X is complete, $\{x_k\}$ must has a limit $x^* \in X$. Then, for any $\varepsilon > 0$, there exists M such that $d(x^*, x_m) < \frac{\varepsilon}{2}$ for all $m = M, M+1, \cdots$ and hence

$$\begin{aligned} d(\Phi(x^*), x^*) &\leqslant d(\Phi(x^*), x^{m+1}) + d(x^{m+1}, x^*) \\ &= d(\Phi(x^*), \Phi(x_m)) + d(x_{m+1}, x^*) \\ &\leqslant K d(x^*, x_m) + d(x_{m+1}, x^*) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2}. \end{aligned} \tag{3}$$

Since ε is arbitrarily small, we have $d(\Phi(x^*), x^*) = 0$, which implies $\Phi(x^*) = x^*$.

To prove the uniqueness. Observe that if $x \in X$ was another fixed point of Φ, we would then have $d(x, x^*) = d(\Phi(x), \Phi(x^*)) \leqslant K d(x, x^*)$, which is possible only if $x = x^*$. □

Example 1. ***(Cournot Equilibrium)*** *Two firms compete in a market by choosing their production levels, q_1 and q_2, respectively. Each firm responds to the other firm's production level by choosing its own level of output. Specifically*

$$\begin{aligned} q_1 &= r_1(q_2) = a_1 - b_1 q_2 \\ q_2 &= r_2(q_1) = a_2 - b_2 q_1 \end{aligned} \tag{4}$$

where a_1, a_2, b_1, b_2 are all positive and $q_i = 0$ if the above expression for q_i is negative. The function $r_i : \mathbb{R}_+ \to \mathbb{R}_+$ is called firm i-th reaction function. Define $r : \mathbb{R}_+^2 \to \mathbb{R}_+^2$ by $r(q_1, q_2) = (r_1(q_2), r_2(q_1))$.

The function r is a contraction if $b_1, b_2 < 1$:

$$\begin{aligned} d(r(q), r(q')) &= d((r_1(q_2), r_2(q_1)), (r_1(q_2'), r_2(q_1'))) \\ &= \sqrt{(r_1(q_2) - r_1(q_2'))^2 + (r_2(q_1) - r_2(q_1'))^2} \\ &= \sqrt{b_1^2(q_2 - q_2')^2 + b_2^2(q_1 - q_1')^2} \\ &\leqslant \max(b_1, b_2) d((q_1, q_2), (q_1', q_2')) \\ &= \max(b_1, b_2) d(q, q'). \end{aligned} \tag{5}$$

Therefore, we have an existence and uniqueness result for Cournot Equilibrium: r has a unique fixed point q^*, if $b_1, b_2 < 1$.

In addition, knowing that the unique fixed point exists, we can easily calculate out the equilibrium production from

$$\begin{aligned} q_1^* &= a_1 - b_1 q_2^* \\ q_2^* &= a_2 - b_2 q_1^*. \end{aligned} \tag{6}$$

It's easy to verify $q_1^* = \frac{a_1 - a_2 b_1}{1 - b_1 b_2}$ and $q_2^* = \frac{a_2 - a_1 b_2}{1 - b_1 b_2}$.

Remark 1. *Being a contraction is a sufficient condition for a self map to have a fixed point, rather than necessary. Consider a function $f : [0,1] \to [0,1]$ defined as follows*

$$f(x) = \begin{cases} 1, & 0 \leqslant x < \frac{1}{4} \\ \frac{3}{2} - 2x, & \frac{1}{4} \leqslant x < \frac{3}{4} \\ 0, & \frac{3}{4} \leqslant x \leqslant 1. \end{cases} \tag{7}$$

It is evident that this function is not a contraction: $d(f(\frac{1}{4}), f(\frac{3}{4})) = d(1, 0) = 1 > d(\frac{1}{4}, \frac{3}{4}) = \frac{1}{2}$. But it has a fixed point: $f(\frac{1}{2}) = \frac{1}{2}$.

1.1 The Method of Successive Approximations

For any "initial" point $x_0 \in X$, the sequence $x_k = \Phi(x_{k-1})$ that we constructed in the proof of the Banach Fixed Point Theorem not only converges to a fixed point x^*, it converges to the fixed point **monotonically**: each successive term is closer to x^* than its predecessor was. This is shown is the next theorem.

Theorem 2. *If Φ is a contraction on a complete metric space X and K is the corresponding contraction coefficient, then for any $x \in X$,*

$$d(\Phi^n(x), x^*) \leqslant K^n d(x, x^*), \tag{8}$$

where x^* is the unique fixed point and $\Phi^n(x) = \Phi^{n-1}(\Phi(x)) = \Phi^{n-2}(\Phi(\Phi(x))) = \cdots = \Phi(\Phi(\cdots \Phi(x)))$.

Proof.

$$\begin{aligned} d(\Phi^n(x), x^*) &= d(\Phi(\Phi^{n-1}(x)), \Phi(x^*)) \\ &\leqslant K d(\Phi^{n-1}(x), x^*) \\ &\leqslant \cdots \\ &\leqslant K^{n-1} d(\Phi(x), \Phi(x^*)) \\ &\leqslant K^n d(x, x^*). \end{aligned} \tag{9}$$

□

Example 2. *Suppose the response functions in out Cournot example become*

$$q_1 = r_1(q_2) = \frac{1}{2(1-q_2)}, \quad q_2 = r_2(q_1) = \frac{1}{2}e^{-q_1}. \tag{10}$$

Now it's not so easy to calculate the equilibrium production levels as the case where the response functions are both linear.

However, it's possible to show that the function $r : \mathbb{R}^2_+ \to \mathbb{R}^2_+$ defined in the last example is a contraction (so a Cournot equilibrium exists and is unique), and one can use the Method of Successive Approximations to compute an approximation to the equilibrium.

2 Brouwer Fixed Point Theorem

The example in Remark 1 above shows that the Banach Theorem seems somewhat limited. It seems intuitively clear that any C^1 function mapping the unit interval into itself will have a fixed point, but the Banach Theorem applies only to functions that satisfies $|f'(x)| \leqslant K$ for some $K < 1$.

The fixed point theorem due to Brouwer covers this case as well as a great many others that the Banach Theorem fails to cover because the relevant functions are not contractions.

Theorem 3. ***(Brouwer Fixed Point Theorem)*** *Let X be a nonempty, compact, convex subset of $\mathbb{R}^n$. Every continuous self map $f : X \to X$ has a fixed point.*

Remark 2. *Instead of proving the theorem rigorously, we show the case where $n = 1$. In $\mathbb{R}$, X being a nonempty, compact, and convex subset is equivalent to say that X is a closed interval, say $X = [0, 1]$. Hence, what's to be verified is that any continuous function $f : [0, 1] \to [0, 1]$ has a fixed point. To this end, define a new function on $[0, 1]$ as $g(x) = x - f(x)$. Evidently, we have $f(0) = 0 - f(0) \leqslant 0$ and $g(1) = 1 - g(1) \geqslant 0$. Then, by the Intermediate Value Theorem and the fact that g is continuous, there exists a point $x^* \in [0, 1]$ such that $g(x^*) = 0$, which is equivalently to say $f(x^*) = x^*$.*

Remark 3. *The Brouwer Theorem requires only that the function be continuous, not that it be a contraction, so there are lots of situations in which the Brouwer Theorem applies but the Banach Theorem doesn't, an illustration is in Remark 1. But conversely, the Banach Theorem doesn't require compactness or convexity, as Brouwer Theorem does. So there are also lots of situations where Banach Theorem applies but Brouwer Theorem does not.*

3 Kakutani Fixed Point Theorem

By a **correspondence** F from a nonempty set X into another nonempty set Y, we mean a map from X into $2^Y \backslash \{\emptyset\}$. Thus, for each $x \in X$, $F(x)$ is a nonempty subset of Y. We write $F : X \rightrightarrows Y$ to denote that F is a correspondence from X into Y. If $\cup_{x \in X} F(x) \subset X$, we refer to F as a self-correspondence.

Example 3. *For any $n \in \mathbb{N}$, $p \in \mathbb{R}^n_{++}$ and $I > 0$, define*

$$B(p, I) = \{x \in \mathbb{R}^n_+ : p \cdot x \leqslant I\}. \tag{11}$$

which is called the budget set of a consumer with income I at price p. If we treat (p, I) as variables, then it would be necessary to treat B as a correspondence. We have $B : \mathbb{R}^{n+1}_{++} \rightrightarrows \mathbb{R}^n_+$.

Example 4. *Consider a parameterized maximization problem*

$$\begin{aligned} \max \quad & u(x; \theta) \\ s.t. \quad & x \in \mathbb{R}^n_+ \cap \mathcal{D}(\theta). \end{aligned} \tag{12}$$

Fixing a value of θ, we have a collection of solutions. If we treat θ as a variable from some given set Θ, we can express the parameterized solution as a correspondence $x^* : \Theta \rightrightarrows \mathbb{R}^n_+$. A classical example is a Cobb-Douglas utility function with varying coefficients.

In most applications, the correspondences we deal with have some additional structures. Some important ones are summarized as follows.

Definition 3. *For any two metric spaces X and Y, a correspondence $F : X \rightrightarrows Y$ is said to be **convex-valued** if $F(x)$ is convex for each $x \in X$. Similarly, F is said to **compact-valued** if $F(x)$ is a compact subset of Y for each $x \in X$. In addition, F is said to **closed-valued** if $F(x)$ is a closed subset of Y for each $x \in X$.*

3.1 Continuity of Correspondences

Since correspondences are generalizations of functions, it seems reasonable to ask if we can extend the notion of "continuity" which we originally defined for functions, to the realm of correspondences. Of cause, we should be consistent with our original definition in the sense that the notion of "continuity" that we might define for a correspondence should reduce to ordinary continuity when that correspondence is single-valued. The following is a such definition.

Definition 4. *For any two metric spaces X and Y, a correspondence $F : X \rightrightarrows Y$ is said to be **upper hemi-continuous at** $x \in X$ if, for every open subset $O \subset Y$ with $F(x) \subset O$, there exists a $\delta > 0$ such that $F(B(x, \delta)) \subset O$, where $B(x, \delta) = \{y \in X \quad s.t. \quad ||y - x|| < \delta\}$. F is called upper hemi-continuous if it is upper hemi-continuous at all points in X.*

Intuitively speaking, upper hemicontinuity at x says that a small perturbation of x does not cause the image set $F(x)$ to suddenly get large. What we mean by this is illustrated in Figure 3.1, which depicts three correspondences mapping $\mathbb{R}_+$ into $\mathbb{R}_+$. According to the definition of upper hemicontinuity, Γ_1 and Γ_3 are upper hemicontinuous everywhere but x_1 and x_2. While, Γ_2 is upper hemicontinuous everywhere.

While upper hemi-continuity of a correspondence $F : X \rightrightarrows Y$ guarantees that the image set $F(x)$ of a point $x \in X$ does not "explode" due to a small perturbation of x, in some sense it allows for it to "implode". For instance, Γ_2 in Figure 3.1 is upper hemi-continuous at x_1, even though there is an intuitive sense in which Γ_2 is not continuous at this point because the "value" of Γ_2 changes dramatically when we perturb x_1 marginally. The following definition deals with this concern formally.

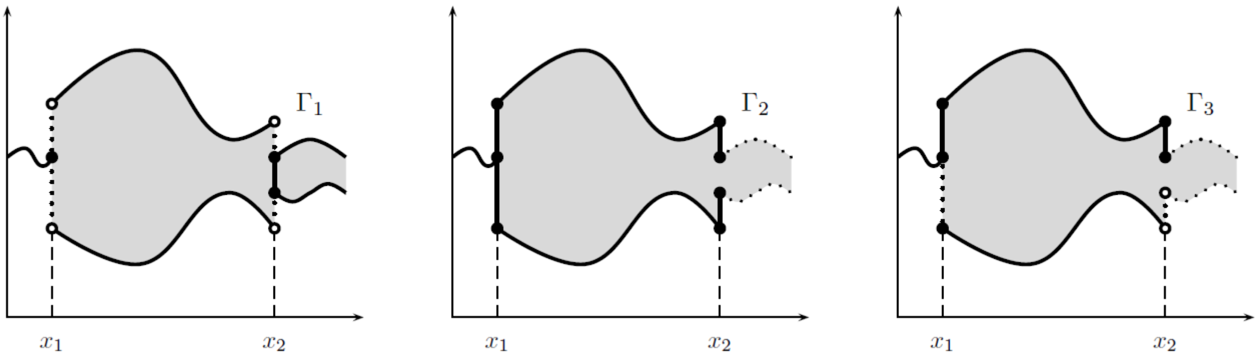

Figure 2: Upper and Lower Hemicontinuity

Definition 5. *For any two metric spaces X and Y, a correspondence $F : X \rightrightarrows Y$ is said to be* ***lower hemi-continuous at*** $x \in X$ *if, for every open subset $O \subset Y$ with $F(x) \cap O \neq \emptyset$, there exists a $\delta > 0$ such that*

$$F(y) \cap O \neq \emptyset \quad \textit{for all} \quad y \in \{x \in X : ||y - x|| < \delta\}. \tag{13}$$

F is called lower hemi-continuous on X if it is lower hemi-continuous at every point in X.

By applying the definition on the correspondences in Figure 3.1, we know that Γ_2 and Γ_3 are lower hemi-continuous everywhere except x_1 and x_2 and that Γ_1 is lower hemi-continuous.

A correspondence is said to be **continuous** when it is both upper and lower hemi-continuous. This behavior of such a correspondence is nicely regular in that small perturbations in its statement do not cause the image sets of nearby points to show drastic, upward or downward, alternations.

3.2 Kakutani Fixed Point Theorem

For a self-correspondence $F : X \rightrightarrows X$, **a fixed point** is defined as a point which satisfies $x^* \in F(x^*)$. With respect to the existence of a fixed point for a self-correspondence, The Kakutani Theorem gives positive answer under verifiable conditions.

Theorem 4. ***(Kakutani Fixed Point Theorem)*** *Let $X \subset \mathbb{R}^n$ be a compact convex set. Let $F : X \rightrightarrows X$ be an upper hemicontinuous convex valued correspondence. Then there exists a point $x^* \in X$ such that $x^* \in F(x^*)$.*

The Kakutani Theorem generalizes the Brouwer Theorem in a straightforward way. It is thus not surprising that this result finds wide applicability in the theory of games the competitive equilibrium.

Example 5. *Every game with finitely many strategies has a mixed-strategy Nash equilibrium.*

4 Exercise

1. In Brouwer Fixed Point Theorem, the sufficient conditions are X be closed, bounded, convex, and f be continuous. Provide one example for each of these four conditions to show their usefulness.

References

Ok, Efe A, *Real analysis with economic applications*, Vol. 10, Princeton University Press, 2007.

Rudin, Walter, *Principles of mathematical analysis*, Vol. 3, McGraw-Hill New York, 1964.

Sundaram, Rangarajan K, *A first course in optimization theory*, Cambridge university press, 1996.