Mathematics for Economists: Lecture 7

Juuso Välimäki

Aalto University School of Business

Spring 2020

This lecture covers

1. Existence of solutions to optimization problems
2. Simplest constrained optimization
3. Optimization subject to an equality constraint: first-order conditions
4. Optimization subject to equality constraints: second-order condition

Existence of optimal choices: why care?

Let's start with a cautionary tale showing how silly characterizations of non-existent objects can look.

Example ('Proof' that 1 is the largest natural number)

Proof.

- Denote the largest natural number (i.e. strictly positive integer) by x.
- Since x is a natural number, also x^2 is a natural number.
- Since x is the largest natural number, we have:

$$x \geq x^2.$$

- Dividing both sides by x (a positive number since it is a natural number), we get

$$1 \geq x.$$

- Since all natural numbers are larger than or equal to 1, the claim follows.

Existence of optimal choices: how to show?

- There are numerous examples of 'results' characterizing objects that do not exist (in economics and elsewhere)
- In more complicated settings, it is not so easy to notice this
- How can we make sure our results are not like this
- Find sufficiently general conditions for optimal choices to exist.
- Our first task in this lecture is understand Weierstrass' theorem that does this.

Existence of optimal choices: Mathematical analysis

- We start by considering the notions of distance, convergence and continuity in a bit more detail.
- This will help us understand when optimization problems are well posed in the sense that they have optimal solutions.
- Distance $d(x, y)$ between two vectors x, y in $\mathbb{R}^n$ is usually based on the Euclidean norm or the length of a vector x in $\mathbb{R}^n$ defined by

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}. \tag{1}$$

- This is just the generalization of the Pythagorean theorem to an arbitrary dimension.
- A distance for $\mathbb{R}^n$ can be derived from this norm as

$$d(x, y) = \|x - y\|.$$

Existence of optimal choices: Mathematical analysis

Proposition

Let x and y denote points in $\mathbb{R}^n$. Then we have:

(*a*) $\|x\| \geq 0$ *and* $\|x\| = 0$ *if and only if* $x = \mathbf{0}$,

(*b*) $\|ax\| = a\,\|x\|$ *for every real a,*

(*c*) $\|x - y\| = \|y - x\|$,

(*d*) $x \cdot y \leq \|x\|\,\|y\|$ (*Cauchy-Schwarz inequality*) ,

(*e*) $\|x + y\| \leq \|x\| + \|y\|$ (*Triangle inequality*) .

Existence of optimal choices: Mathematical analysis

Remark

To see why the Cauchy-Schwarz inequality is true, consider the sum of squares

$$\sum_{i=1}^{n}(x_i + ty_i)^2.$$

This is a quadratic polynomial in t, and as a sum of squares, it is also non-negative. Hence its discriminant is non-positive, i.e.

$$(2\sum_{i=1}^{n} x_i y_i)^2 \leq 4(\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2).$$

Dividing both sides by 4 and taking square roots on both sides gives Cauchy-Schwarz inequality.

Existence of optimal choices: Mathematical analysis

Remark

From Cauchy-Schwarz, we get easily the triangle inequality:

$$\|x+y\|^2 = (x+y)\cdot(x+y) = \|x\|^2 + \|y\|^2 + 2x\cdot y$$

$$\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2.$$

The triangle inequality follows by taking square roots on both sides of the inequality. The inequality in the above expression results from the Cauchy-Schwarz inequality.

Existence of optimal choices: Mathematical analysis

Recall our definition of a neighborhood: for $x \in \mathbb{R}^n$ and $\varepsilon > 0$, the open (or closed) neighborhood $B^{\varepsilon}(x)$ with center at x and radius ε is defined to be the set of all $y \in \mathbb{R}^n$, such that $\|y - x\| < (\leq)\, r$.
In this subsection, we give some basic definitions on sets in $\mathbb{R}^n$.

Definition

A point x is a limit point of the set $E \subset \mathbb{R}^n$ if every neighborhood of x contains a point $y \in E$ with $y \neq x$.
We say that E is *closed* if every limit point of E is an element of E. A point x is an interior point of E if there is a neighborhood $B^{\varepsilon}(x)$ of x such that $B^{\varepsilon}(x) \subset E$. We say that E is *open* if every point of E is an interior point.
The *complement* of E, denoted by E^c is the set of all points $x \in \mathbb{R}^n$ such that $x \notin E$.
The set E is *bounded* if there is a real number M such that $\|x\| < M$ for all $x \in E$.

Existence of optimal choices: Mathematical analysis

Exercise Is the empty set open or closed? Show that $A = \{x : a < x < b\}$ is an open set and that $A = \{x : a \leq x \leq b\}$ is a closed set. Show that the set $\{1, \frac{1}{2}, \frac{1}{3}, ...\}$ is neither open nor closed (hint: is 0 a limit point? Is it in the set?)

Proposition

A set $E \subset \mathbb{R}^n$ is open if and only if its complement is closed. A set $F \subset \mathbb{R}^n$ is closed if and only if its complement is open.

Existence of optimal choices: Mathematical analysis

A very important property for sets in mathematical analysis is called *compactness*.

we give here a definition of compactness for sets in $\mathbb{R}^n$ that should really be derived as a theorem starting from a more fundamental notion, but for practical matters, this is all we need.

Definition (Compact sets)

A set $E \subset \mathbb{R}^n$ is called *compact* if it is closed and bounded.

We discuss next the definition of sequences and their convergence:

Definition

If S is any set, a sequence in S is a function on the set $\mathbb{N} = \{1, 2, 3, ...\}$ of natural numbers and whose range is in S.

Definition

A sequence $\{x_n\}$ in $\mathbb{R}^n$ is said to converge if there is a point $x \in \mathbb{R}^n$ with the following property: For every $\epsilon > 0$, there is an integer N such that $n \geq N$ implies that $d(x_n, x) < \epsilon$.

We say that x_n converges to x, x is the limit of $\{x_n\}$ and we write $x_n \to x$,

$$\lim_{n\to\infty} x_n = x.$$

Theorem

Let $\{x_n\}$ *be a sequence in* $\mathbb{R}^n$*.*
(i) $\{x_n\}$ *converges to* $x \in \mathbb{R}^n$ *if and only if every neighborhood of* x *contains all but finitely many of the terms of* $\{x_n\}$ *.*
(ii) *If* $x \in \mathbb{R}^n, x' \in \mathbb{R}^n$*, and if* $\{x_n\}$ *converges to* x *and to* x'*, then* $x = x'$*.*
(iii) *If* $\{x_n\}$ *converges, then* $\{x_n\}$ *is bounded.*
(iv) *If* $E \subset \mathbb{R}^n$ *and* x *is a limit point of* E*, then there is a sequence* $\{x_n\}$ *in* E *such that* $x = \lim_{n\to\infty} x_n$*.*
(v) $x_n = (x_{1,n}, ..., x_{k,n}) \to x = (x_1, ..., x_k) \Leftrightarrow x_{i,n} \to x_i$ *for all* $i \in \{1, ..., k\}$*.*

The last part of the proposition claims that a sequence of vectors converges if and only if all of its coordinates converge.

Definition

Given a sequence $\{x_n\}$, consider an infinite sequence $\{n_k\}$ of positive integers, such that $n_1 < n_2 < \cdots$. Then the sequence $\{x_{n_i}\}$ is called a subsequence of $\{x_n\}$. If $\{x_{n_i}\}$ converges, its limit is called a subsequential limit of $\{x_n\}$.

Exercise Show that if $\{x_n\}$ converges to x, then all of its subsequences also converge to x.

Definition

A sequence $\{x_n\}$ is said to be a Cauchy sequence if for every $\epsilon > 0$ there is an integer N such that $d(x_n, x_m) < \epsilon$, if $n \geq N$ and $m \geq N$.

Real numbers are constructed in such a way that Cauchy sequences in $\mathbb{R}$ converge, i.e. have limits in $\mathbb{R}$. By part (v) of the previous theorem, the same is true for real vectors.

Theorem (Weierstrass)

Every bounded subset $E \subset \mathbb{R}^n$ with infinitely many elements has a limit point in $\mathbb{R}^n$.

Idea of proof for $\mathbb{R}$: Since E is bounded, it is contained in an interval $[-M, M]$ of length $2M$for some $M < \infty$.
Since E has infinitely many elements, either $[-M, 0]$ or $[0, M]$ or both have infinitely many elements.
Hence some interval of length M also contains infinitely many elements of E.
Continue this process of halving the interval to show that you can come up with a sequence of intervals of length $2^{-k}M$ containing infinitely many elements of E.
The midpoints of the sequences form a Cauchy sequence and hence they converge to a point $x \in \mathbb{R}$. This x is a limit point of E.

An immediate consequence of this is the following theorem.

Theorem

(Bolzano-Weierstrass Theorem)

Every bounded sequence in $\mathbb{R}^n$ *contains a convergent subsequence and every sequence in a compact set* $E \in \mathbb{R}^n$ *has a convergent subsequence whose limit is in* E.

Continuous Functions

Definition

Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$. We write $f(x) \to \hat{y}$ as $x \to \hat{x}$, or

$$\lim_{x \to \hat{x}} f(x) = \hat{y}, \tag{2}$$

if there is a point $y \in \mathbb{R}^m$ with the following property: For every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in B^\delta(\hat{x}) \Rightarrow f(x) \in B^\varepsilon(\hat{y}).$$

We say that f is *continuous at* $\hat{x}$ if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in B^\delta(\hat{x}) \Rightarrow f(x) \in B^\varepsilon(f(\hat{x})).$$

Another way of writing this is given in the following simple proposition.

Continuous Functions

Proposition

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ *is continuous at* $\hat{x}$ *if for every sequence* $\{x_n\}$ *that converges to* $\hat{x}$*, the sequence* $\{f(x_n)\}$ *converges to* $f(\hat{x})$*; in symbols,*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

A function is said to be continuous if it is continuous at all points in its domain.
The following proposition gives yet another way of looking at continuity.

Proposition

A function f is continuous if and only if the inverse image $f^{-1}(V)$ *is open (closed) for every open (closed) set V in Y.*

Global Properties of Continuous Functions

Definition

A function $f : E \to \mathbb{R}$ is said to be bounded if there is a real number M such that $|f(x)| \leq M$ for all $x \in E$.

Recall the definition of the least upper bound and greatest lower bound for a set A of real numbers.
We say that $\overline{a}$ is the least upper bound of A if for all $x \in A$, $x \leq \overline{a}$ and for all $a' < \overline{a}$, there is some $x \in A$ such that $x > a'$.
Similarly, we say that $\underline{a}$ is the greatest lower bound of A if for all $x \in A$, $x \geq \underline{a}$ and for all $a' > \underline{a}$, there is some $x \in A$ such that $x < a'$.
We write:

$$\overline{a} := \sup A, \underline{a} := \inf A.$$

Global Properties of Continuous Functions

Theorem (Weierstrass' Theorem)

Suppose f is a continuous function on a compact set E, and

$$M = \sup_{x \in E} f(x), \qquad m = \inf_{x \in E} f(x).$$

Then there exists a point $\overline{x}, \underline{x} \in E$ such that $f(\overline{x}) = M$ and $f(\underline{x}) = m$.

Global Properties of Continuous Functions

Proof.

We show this for the supremum. The case for the infimum is analogous. Let $M = \sup_{x \in E} f(x)$. Let $\{m_n\} \to M$ with $m_n < M$ for all n. Then By the definition of supremum, there must be a sequence $\{x_n\} \in E$ with $x_n \geq m_n$. Since E is compact, $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\} \to x \in E$. Since $\{m_n\} \to M$, we also know that $\{m_{n_k}\} \to M$. By continuity of f,

$$M \geq f(x) = \lim f(x_{n_k}) \geq \lim m_{n_k} = M.$$

□

This theorem ensures that our maximization and minimization problems have solutions as long as the objective function is continuous and the feasible set is compact.

Global Properties of Continuous Functions

Remark

To see that E must be closed and bounded and that f has to be continuous, consider the following examples:

1. $f(x) = x$ *and* $E = \mathbb{R}$.
2. $f(x) = x$ *and* $E = \{x : 0 < x < 1\}$.
3. $f(x) = x$ *for* $0 \leq x < 1, f(1) = 0$ *and* $E = \{x : 0 \leq x \leq 1\}$.

Constrained Optimization: Example

We start with a simplest example of constrained optimization to set up expectations for the more general case to follow.

Example

Consider finding the maximum for $f(x) = 3 + 2x - x^2$ on the feasible set $F = \{x : -\infty < a \leq x \leq b < \infty$.
Since f is continuous and the feasible set F is compact. Therefore Weierstrass' theorem guarantees the existence of a maximizer, i.e. an $x \in F$ such that for all $y \in F$, we have $f(x) \geq f(y)$.
Notice that f is strictly increasing for $x < 1$ and strictly decreasing for $x > 1$.
If $a \leq 1 \leq b$, then the function is maximized at its critical point $x = 1$.
We say that a direction $(x - x_0)$ is feasible from $x_0 \in F$ if for a small Δ, we have $x_0 + \Delta(x - x_0) \in F$.

Constrained Optimization: Example continued

Linear approximation by the derivative gives:

$$f(x_0 + \Delta(x - x_0)) - f(x_0) = f'(x_0)\Delta(x - x_0).$$

If we have a maximum at x_0, then for all feasible direction

$$f'(x_0)\Delta(x - x_0) \leq 0.$$

- If $a < x_0 < b$, then we must have $f'(x_0) = 0$ since both directions $x > x_0, x < x_0$ are feasible.
- If $f'(x_0) > 0$, then $x > x_0$ cannot be feasible if x_0 is a maximum. Therefore $x_0 = b$ if x_0 is the optimal choice and $f'(x_0) > 0$. Similarly, if $f'(x_0) < 0$ and x_0 is the optimum, then $x_0 = a$.

Optimization with a single equality constraint

- Local considerations: Let $f : \mathbb{R}^n \to \mathbb{R}$ be the objective function to be maximized
- Suppose the constraints take the form $g(x) = g(x_1, ..., x_n) = 0$.
- In other words, $F = \{x : g(x) = 0\}$. We write the maximization problem often as:

$$\max_x f(x)$$
$$\text{subject to } g(x) = 0.$$

Optimization with a single equality constraint

- A solution to this problem finds point $\hat{x}$ such that $f(\hat{x}) \geq f(x)$ for all $x \in F$.
- What can we say about such an $\hat{x}$?
- At this point, we do not know if it exists. If it exists, and f is differentiable, then for small Δ,

$$f(\hat{x} + \Delta(x - \hat{x})) - f(\hat{x}) = Df(\hat{x})(x - \hat{x})\Delta \leq 0$$

  for all feasible directions $(x - \hat{x})$.
- But how do we know which directions are feasible?

Optimization with a single equality constraint

- Assume that the function g defining the constraint is also differentiable.
- To find the feasible directions, we go back to implicit function theorem.
- If $\hat{x} \in F$ and $\frac{\partial g}{\partial x_i}(\hat{x}) \neq 0$ for some $i \in \{1, ..., n\}$, then we can find a write $x_i = h(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) =: h(x_{-i})$ in a neighborhood of $\hat{x}_{-i}$ so that

$$g(h(x_{-i}), x_{-i}) = 0.$$

- Notice that it is not possible to use the implicit function theorem if at a critical point of the constraint function.
- Therefore, we must assume that $Dg(\hat{x}) \neq 0$.
- We call this the constraint qualification and we will see different versions of this in more complex situations with multiple constraints.

Optimization with a single equality constraint

- Since the function g is at constant value in the feasible set, we have for all feasible directions $(x - \hat{x})$:

$$\nabla g(\hat{x})(x - \hat{x}) = 0.$$

- Notice also that if $(x - \hat{x})$ is feasible, then also $-(x - \hat{x})$ is feasible. From the linear approximation above, this means that for all feasible directions,

$$Df(\hat{x})(x - \hat{x}) = 0.$$

- But therefore we have shown that at optimum $\hat{x}$,

$$\nabla f(\hat{x}) = \mu \nabla g(\hat{x}).$$

- We have the following necessary condition for a constrained optimum at $\hat{x}$:
  1. the gradient of the objective function must be a scalar multiple of the gradient of the constraint function at the optimum
  2. the choice must be feasible, i.e. $g(\hat{x}) = 0$
  3. we have assumed constraint qualification at optimum

Lagrangean function

- The previous discussion motivates the following function that incorporates the constraints into an augmented objective function called the Lagrangean function.
- For a constrained optimization problem, we define the following function of $n+1$ variables:

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

- We call the new variable μ the Lagrange multiplier. We will give it a good economic interpretation later in the course.
- We are interested in the critical points of this augmented function. Therefore we look for $(\hat{x}, \hat{\mu})$ such that

$$\frac{\partial \mathcal{L}}{\partial x_i}(\hat{x}, \hat{\mu}) = \frac{\partial f}{\partial x_i}(\hat{x}) - \hat{\mu}\frac{\partial g}{\partial x_i}(\hat{x}) = 0 \text{ for all } i,$$

$$\frac{\partial \mathcal{L}}{\partial \mu}(\hat{x}, \hat{\mu}) = g(\hat{x}) = 0.$$

Lagrangean function

- As argued above, these are the first-order conditions for the constrained optimization problem.
- In order to know if we have found a local maximum or a minimum, we need to look at the second-order Taylor -approximations and the definiteness of the Hessian matrix at the critical point.
- As before, write the second-order Taylor approximation to $f : \mathbb{R}^n \to \mathbb{R}$ at $\hat{x}$ as:

$$f(x) = f(\hat{x}) + Df(\hat{x})(x - \hat{x}) + (x - \hat{x}) \cdot D^2(\hat{x})(x - \hat{x}).$$

- If $\hat{x}$ is a maximum, then for all feasible directions $(x - \hat{x})$, we have

$$i) Df(\hat{x})(x - \hat{x}) = 0,$$

$$ii) (x - \hat{x}) \cdot D^2(\hat{x})(x - \hat{x}) \leq 0.$$

Lagrangean function

- Since the feasible directions are give by vectors $(x - \hat{x})$ such that

$$\nabla g(\hat{x}) \cdot (x - \hat{x}) = 0,$$

  the condition for having a local maximum at $\hat{x}$ is equivalent to checking the negative definiteness of the bordered Hessian where we need
  1. The Lagrangean $\mathcal{L}$
  2. The Equality constraint h
- To get the bordered Hessian, start with the derivative of the Lagrangean with respect to the choice variables x at the critical point $\hat{x}$: $D_x^2\mathcal{L}(\hat{x})$ and 'border' it with the derivative of the constraint function (to capture the restriction to feasible directions).

$$D^2\mathcal{L} = \begin{bmatrix} 0 & Dg(\hat{x}) \\ [Dg(\hat{x})]^T & D_x^2\mathcal{L}(\hat{x}) \end{bmatrix}$$

Lagrangean function

- In the special case where we have only two choice variables, I let the variables be x, y for notational ease, we need to examine

$$D^2\mathcal{L} = \begin{bmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{bmatrix}$$

Lagrangean function

- How do we determine the negative definiteness of the bordered Hessian?
  1. Leading principal minors must alternate in sign [1]
  2. $\det D^2\mathcal{L}(\hat{x})$ must have the same sign as $(-1)^n$.
- How many principal minors to examine?
  - You need to check the sign of the last $(n-1)$ leading principal minors
  - For completeness, I state here that with more constraints, you need to border the Hessian with the derivatives of all binding constraints. If you have k such constraints, then you need to examine the sign of $(n-k)$ leading principal minors.
- Bordered Hessians are a bit of a nightmare for me. They are tedious to compute and they tell nothing of significance in the end. We will see later how we can bypass them to a large extent by concentrating on a subset of problems where first-order conditions also turn out to be sufficient.

[1] Recall that a leading principal minor of k^{th} order is obtained from a matrix A by deleting its last k rows and columns.