Mathematics for Economists
Aalto BIZ
Spring 2021
Juuso Välimäki

Elementary analysis

The goal of these notes is to find sufficient conditions for the existence of a solution to constrained optimization problems in $\mathbb{R}^n$. We start by considering the notions of distance, convergence and continuity in a bit more detail.

Length and distance in $\mathbb{R}^n$

The only spaces that we will be interested in these notes are the various Cartesian products of the real line $\mathbb{R}$ denoted by $\mathbb{R}^n$. The exponent n is also called the dimension of the Euclidean space. Hence an element $\boldsymbol{x} \in \mathbb{R}^n$ is an ordered n-tuple $(x_1, ..., x_n)$ where each $x_i \in \mathbb{R}$.

Distance $d(\boldsymbol{x}, \boldsymbol{y})$ between two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ is usually based on the Euclidean norm or the length of a vector in $\boldsymbol{x} \in \mathbb{R}^n$ defined by

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

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

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

Proposition 1. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ denote points in $\mathbb{R}^n$. Then we have:
(a) $\|\boldsymbol{x}\| \geq 0$ and $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x}= \boldsymbol{0}$,
(b) $\|a\boldsymbol{x}\| = a\|\boldsymbol{x}\|$ for every real a,
(c) $\|\boldsymbol{x}-\boldsymbol{y}\| = \|\boldsymbol{y}-\boldsymbol{x}\|$,
(d) $\boldsymbol{x} \cdot \boldsymbol{y} \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\| \quad$ (Cauchy-Schwarz inequality),
(e) $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\| \quad$ (Triangle inequality).

Remark 1. 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.

This simple inequality is one of the most important results in all of mathematics. Equality holds if and only if $\boldsymbol{x} = \lambda \boldsymbol{y}$, i.e. $\boldsymbol{x}$ is proportional to $\boldsymbol{y}$. We have used this observation to argue that the gradient $\nabla f(\hat{\boldsymbol{x}})$ gives the direction of steepest ascent for a function f at point $\hat{\boldsymbol{x}}$.

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

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

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

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

Exercise In general, any function $\hat{d}(\boldsymbol{x}, \boldsymbol{y})$ satisfying (a), (c) and (e) in the above list is a distance. It is a good exercise to show that $\hat{d}(x, y) := \max_i |x_i - y_i|$ is a distance in this sense. Are all the other properties above also satisfied by this distance?

By the segment (a, b) we mean the set of all real number x such that $a < x < b$. By the interval $[a, b]$, we mean the set of all real numbers such that $a \leq x \leq b$.

For $\boldsymbol{x} \in \mathbb{R}^n$, we define analogs of intervals as follows. If $a_i < b_i$ for $i = 1, ..., n$, the set of all points $\boldsymbol{x} = (x_1, ..., x_n)$ in $\mathbb{R}^n$ whose coordinates satisfy $a_i \leq x_i \leq b_i$ for $(1 \leq i \leq n)$, is called an n-cell. If $\boldsymbol{x} \in \mathbb{R}^n$ and $\varepsilon > 0$, the open (or closed) neighborhood $B^{\varepsilon}(\boldsymbol{x})$ with center at $\boldsymbol{x}$ and radius ε is defined to be the set of all $\boldsymbol{y} \in \mathbb{R}^n$, such that $\|\boldsymbol{y} - \boldsymbol{x}\| < (\leq)\, \varepsilon$.

Open and closed sets

Definition 1. A point $\boldsymbol{x}$ is a limit point of the set $E \subset \mathbb{R}^n$ if every neighborhood of x contains a point $\boldsymbol{y} \in E$ with $\boldsymbol{y} \neq \boldsymbol{x}$.

We say that E is *closed* if every limit point of E is an element of E. A point $\boldsymbol{x}$ is an interior point of E if there is a neighborhood $B^{\varepsilon}(\boldsymbol{x})$ of $\boldsymbol{x}$ such that $B^{\varepsilon}(\boldsymbol{x}) \subset E$. 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 $\boldsymbol{x} \in \mathbb{R}^n$ such that $x \notin E$.

The set E is *bounded* if there is a real number M such that $\|\boldsymbol{x}\| < M$ for all $\boldsymbol{x} \in E$.

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 2. 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.

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 (Heine-Borel Theorem) starting from a more fundamental notion, but for practical matters, this is all we need.

Definition 2 (Compact sets)**.** A set $E \subset \mathbb{R}^n$ is called *compact* if it is closed and bounded.

Sequences and subsequences

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

Definition 4. A sequence $\{\boldsymbol{x}_n\}$ in $\mathbb{R}^n$ is said to converge if there is a point $\boldsymbol{x} \in \mathbb{R}^n$ with the following property: For every $\epsilon > 0$, there is an integer N such that $n \geq N$ implies that $d(\boldsymbol{x}_n, \boldsymbol{x}) < \epsilon$.

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

$$\lim_{n \to \infty} \boldsymbol{x}_n = \boldsymbol{x}.$$

Theorem 1. Let $\{\boldsymbol{x}_n\}$ be a sequence in $\mathbb{R}^n$.

(i) $\{\boldsymbol{x}_n\}$ converges to $\boldsymbol{x} \in \mathbb{R}^n$ if and only if every neighborhood of $\boldsymbol{x}$ contains all but finitely many of the terms of $\{x_n\}$.

(ii) If $\boldsymbol{x} \in \mathbb{R}^n, x' \in \mathbb{R}^n$, and if $\{\boldsymbol{x}_n\}$ converges to $\boldsymbol{x}$ and to $\boldsymbol{x}'$, then $\boldsymbol{x} = \boldsymbol{x}'$.

(iii) If $\{\boldsymbol{x}_n\}$ converges, then $\{\boldsymbol{x}_n\}$ is bounded.

(iv) If $E \subset \mathbb{R}^n$ and $\boldsymbol{x}$ is a limit point of E, then there is a sequence $\{\boldsymbol{x}_n \neq \boldsymbol{x}\}$ in E such that $\boldsymbol{x} = \lim_{n\to\infty} \boldsymbol{x}_n$.

(v) $\boldsymbol{x}_n = (x_{1,n}, ..., x_{k,n}) \to \boldsymbol{x} = (x_1, ..., x_k) \Leftrightarrow x_{i,n} \to x_i$ 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 5. Given a sequence $\{\boldsymbol{x}_n\}$, consider an infinite sequence $\{n_k\}$ of positive integers, such that $n_1 < n_2 < \cdots$. Then the sequence $\{\boldsymbol{x}_{n_i}\}$ is called a subsequence of $\{\boldsymbol{x}_n\}$. If $\{\boldsymbol{x}_{n_i}\}$ converges, its limit is called a subsequential limit of $\{\boldsymbol{x}_n\}$.

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

Definition 6. A sequence $\{\boldsymbol{x}_n\}$ is said to be a Cauchy sequence if for every $\epsilon > 0$ there is an integer N such that $d(\boldsymbol{x}_n, \boldsymbol{x}_m) < \epsilon$, 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 2. 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. The same construction generalizes easily to $\mathbb{R}^n$

An immediate consequence of this is the following theorem.

Theorem 3. (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 7. Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$. We write $f(\boldsymbol{x}) \to \hat{\boldsymbol{y}}$ as $\boldsymbol{x} \to \hat{\boldsymbol{x}}$, or

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

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

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

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

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

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

Proposition 3. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\hat{\boldsymbol{x}}$ if for every sequence $\{\boldsymbol{x}_n\}$ that converges to $\hat{\boldsymbol{x}}$, the sequence $\{f(\boldsymbol{x}_n)\}$ converges to $f(\hat{\boldsymbol{x}})$; in symbols,

$$\lim_{n \to \infty} f(\boldsymbol{x}_n) = f\left(\lim_{n \to \infty} \boldsymbol{x}_n\right).$$

A function is said to be continuous if it is continuous at all points in its domain. Continuity of a function f at a point $\hat{\boldsymbol{x}}$ is called a local property of f because it depends on the behavior of f only in the immediate vicinity of $\hat{\boldsymbol{x}}$. A property of f which concerns the whole domain of f is called a global property. Thus, continuity of f on its domain is a global property.

The following proposition gives yet another way of looking at continuity.

Proposition 4. 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.

Proposition 5. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$ be continuous functions, and let h be the composite function defined by

$$h(\boldsymbol{x}) = g(f(\boldsymbol{x})) \qquad \text{for } \boldsymbol{x} \in \mathbb{R}^n.$$

If f is continuous at $\hat{\boldsymbol{x}}$ and if g is continuous at $f(\hat{\boldsymbol{x}})$, then h is continuous at $\hat{\boldsymbol{x}}$.

Global properties of continuous functions

Definition 8. A function $f : E \to \mathbb{R}$ is said to be bounded if there is a real number M such that $|f(\boldsymbol{x})| \leq M$ for all $\boldsymbol{x} \in E$.

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:

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

Theorem 4 (Weierstrass' Theorem)**.** Suppose f is a continuous function on a compact set E, and

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

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

Proof. We show this for the supremum. The case for the infimum is analogous. Let $M = \sup_{\boldsymbol{x} \in E} f(\boldsymbol{x})$. Let $\{M_n\} \to M$ with $M_n < M$ for all n. By the definition of the supremum, there exists a sequence $\{\boldsymbol{x}_n\} \in E$ with $x_n \geq M_n$. Since E is compact, $\{\boldsymbol{x}_n\}$ has a convergent subsequence $\{\boldsymbol{x}_{n_k}\} \to x \in E$. Since $\{M_n\} \to M$, we also know that $\{M_{n_k}\} \to M$. By continuity of f,

$$M \geq f(\boldsymbol{x}) = \lim f(\boldsymbol{x}_{n_k}) \geq \lim M_{n_k} = M.$$

□

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.

Remark 2. To see that E must be closed and bounded and that f has to be continuous, consider the following examples where a single hypothesis (in brackets) of the theorem fails:

1. $f(x) = x$ and $E = \mathbb{R}$ (domain not bounded).
2. $f(x) = x$ and $E = \{x : 0 < x < 1\}$ (domain not closed).
3. $f(x) = x$ for $0 \leq x < 1, f(1) = 0$ and $E = \{x : 0 \leq x \leq 1\}$ (f not continuous).

Here are two more useful results. The first is a generalization of Weierstrass' theorem, the second is a generalization of the intermediate value theorem for functions of a single real variable.

Proposition 6. Let $f : X \to Y$ be a continuous function. the image $f(E)$ of any compact set $E \subset X$ is compact.

Proposition 7. Let $f : X \to Y$ be a continuous function. the image $f(E)$ of any connected set $E \subset X$ is connected.

Intervals (including the entire real line and the empty set) are the only connected sets on $\mathbb{R}$. It is surprisingly hard to give a general and easily verified definition of connected sets in $\mathbb{R}^n$, but for many applications of this theorem, it is enough to note that convex sets in $\mathbb{R}^n$ for any n are connected.