A First Course in Optimization

Ishan Kapnadak

Autumn Semester 2022-23

Updated on: 2022-08-09

Abstract

Lecture Notes for the course EE 659: A First Course in Optimization taught in Spring 2022 by Prof. Vivek Borkar. Additional references include *A first course in optimization* by Rangarajan K. Sundaram, *Optimization by vector space methods* by David Luenberger, and *Nonlinear programming* by Dimitri P. Bertsekas

Contents

§1. Lecture 1

Definition 1.1: Open Ball

The *open ball* of radius ϵ centered around $\mathbf{x}_0 \in \mathbb{R}^d$ is defined as

$$B_\epsilon(\mathbf{x}_0) := \left\{\mathbf{x} \in \mathbb{R}^d \colon \|\mathbf{x} - \mathbf{x}_0\| < \epsilon\right\}.$$

Definition 1.2: Closed Ball

The *closed ball* of radius ϵ centered around $\mathbf{x}_0 \in \mathbb{R}^d$ is defined as

$$\overline{B}_\epsilon(\mathbf{x}_0) := \left\{\mathbf{x} \in \mathbb{R}^d \colon \|\mathbf{x} - \mathbf{x}_0\| \leq \epsilon\right\}.$$

Definition 1.3: Open and Closed Sets

A set $A \subset \mathbb{R}^d$ is said to be *open* if for all $\mathbf{x} \in A$, there exists an $\epsilon > 0$ such that $B_\epsilon(\mathbf{x}) \subset A$. A set A is said to be *closed* if A^{c} is open.

We have the following properties for open and closed sets.

1. Let $\mathcal{I}$ be an arbitrary index set. If A_α is open for each $\alpha \in \mathcal{I}$, then
$$\bigcup_{\alpha \in \mathcal{I}} A_\alpha$$
is open. In other words, open sets are closed under arbitrary unions.
2. Let $\mathcal{I}$ be a finite index set. If A_α is open for each $\alpha \in \mathcal{I}$, then
$$\bigcap_{\alpha \in \mathcal{I}} A_\alpha$$
is open. In other words, open sets are closed under finte intersections.
3. Let $\mathcal{I}$ be an arbitrary index set. If A_α is closed for each $\alpha \in \mathcal{I}$, then
$$\bigcap_{\alpha \in \mathcal{I}} A_\alpha$$
is closed. In other words, closed sets are closed under arbitrary intersections.
4. Let $\mathcal{I}$ be a finite index set. If A_α is closed for each $\alpha \in \mathcal{I}$, then
$$\bigcup_{\alpha \in \mathcal{I}} A_\alpha$$
is closed. In other words, closed sets are closed under finite unions.

Definition 1.4: Convergence of a sequence

Let $\{\mathbf{x}_n\}$ be a sequence in $\mathbb{R}^d$. Then, $\{\mathbf{x}_n\}$ converges to $\mathbf{x}^*$ (written $\mathbf{x}_n \to \mathbf{x}^*$) if for all $\epsilon > 0$, there exists an $n_0 \in \mathbb{N}$ such that

$$\mathbf{x}_n \in B_\epsilon(\mathbf{x}^*) \quad \forall n > n_0.$$

Equivalently, $\|\mathbf{x}_n - \mathbf{x}^*\| \to 0$.

Definition 1.5: Closure

Let $A \subset \mathbb{R}^d$. The *closure* of A (denoted $\overline{A}$) is the smallest closed set containing A. Equivalently, $\overline{A}$ is the intersection of all closed sets containing A.

Definition 1.6: Interior

Let $A \subset \mathbb{R}^d$. The *interior* of A (denoted A^{o}) is the largest open set contained in A. Equivalently, A^{o} is the union of all open sets contained in A.

Note that by definition, we have $A^{\mathrm{o}} \subset A \subset \overline{A}$.

Definition 1.7: Boundary

Let $A \subset \mathbb{R}^d$. The *boundary* of A is defined as

$$\partial A := \overline{A} \setminus A^{\mathrm{o}}.$$

Note that for a closed set, $\overline{A} = A$, and for an open set, $A^{\mathrm{o}} = A$.

Proposition 1.8

A set $A \subset \mathbb{R}^d$ is closed if and only if

$$\mathbf{x}_n \to \mathbf{x}^*, \mathbf{x}_n \in A \forall n \implies \mathbf{x}^* \in A.$$

Proof. Let $A \subset \mathbb{R}^d$ be closed, $\mathbf{x}_n \in A$ for all n, and $\mathbf{x}_n \to \mathbf{x}^*$. Assume to the contrary that $\mathbf{x}^* \notin A$. Then, $\mathbf{x}^* \in A^{\mathsf{c}}$, which is open by assumption. Thus, $\exists \epsilon > 0$ such that $B_\epsilon(\mathbf{x}^*) \subset A^{\mathsf{c}}$. This implies that $\mathbf{x}_n \notin B_\epsilon(\mathbf{x}^*)$ for all n, and thus $\mathbf{x}_n \not\to \mathbf{x}^*$, a contradiction. To prove the converse, assume that A is not closed. Thus, there exists a $\tilde{\mathbf{x}} \in \partial A$ such that $\tilde{\mathbf{x}} \notin A$. Then, for all $\epsilon > 0$, $B_\epsilon(\tilde{\mathbf{x}}) \cap A \neq \varnothing$. Let $\epsilon_n \downarrow 0$ and let $\mathbf{x}_n \in B_{\epsilon_n}(\tilde{\mathbf{x}}) \cap A$. Then, $\mathbf{x}_n \to \tilde{\mathbf{x}} \notin A$, a contradiction. $\square$

Definition 1.9: Limit Point

Let $\{\mathbf{x}_n\}$ be a sequence in $\mathbb{R}^d$. $\tilde{\mathbf{x}}$ is a *limit point* of $\{\mathbf{x}_n\}$ if there exists a subsequence $\{\mathbf{x}_{n_k}\}$ such that $\mathbf{x}_{n_k} \to \tilde{\mathbf{x}}$.

Proposition 1.10

$\{\mathbf{x}_n\}$ converges if and only if $\{\mathbf{x}_n\}$ has a unique limit point.

Definition 1.11: Supremum and Infimum

Let $A \subset \mathbb{R}$ be bounded. Then,

$$\begin{aligned} \sup A &:= \text{ smallest } x \in \mathbb{R} \cup \{+\infty\} \text{ such that } y \in A \implies y \leq x, \\ \inf A &:= \text{ largest } x \in \mathbb{R} \cup \{-\infty\} \text{ such that } y \in A \implies y \geq x. \end{aligned}$$

Definition 1.12: Cauchy Sequence

A sequence $\{\mathbf{x}_n\}$ is said to be *Cauchy* if $\lim\limits_{m,n\uparrow\infty} \|\mathbf{x}_m - \mathbf{x}_n\| = 0$.

Proposition 1.13

Cauchy sequences are bounded.

Proof. Let $\{\mathbf{x}_n\}$ be a Cauchy sequene and let $\epsilon > 0$. Pick N large enough such that

$$n, m > N \implies \|\mathbf{x}_m - \mathbf{x}_n\| < \epsilon.$$

We then have

$$\begin{aligned} &\mathbf{x}_n \in B_\epsilon(\mathbf{x}_m) \quad \forall n > N \\ &\implies \{\mathbf{x}_n \colon n > N\} \text{ is bounded} \\ &\implies \{\mathbf{x}_n\} \text{ is bounded.} \end{aligned}$$

□

Proposition 1.14

Cauchy sequences have at most one limit point.

Proof. Suppose $\{\mathbf{x}_n\}$ is a Cauchy sequence having two limit points, $\tilde{\mathbf{x}}$ and $\overline{\mathbf{x}}$. Then, there exist subsequences $\mathbf{x}_{\tilde{n}_k} \to \tilde{\mathbf{x}}$ and $\mathbf{x}_{\overline{n}_l} \to \overline{\mathbf{x}}$. We then have

$$\lim_{\tilde{n}_k, \overline{n}_l \uparrow \infty} \|\mathbf{x}_{\tilde{n}_k} - \mathbf{x}_{\overline{n}_l}\| = 0 \implies \tilde{\mathbf{x}} = \overline{\mathbf{x}}. \qquad \square$$

Definition 1.15: Complete Space

A metric space is *complete* if every Cauchy sequence converges.

§2. Lecture 2

Theorem 2.1: Bolzano-Weierstrass Theorem

Every bounded sequence]oin $\mathbb{R}^d$ has a convergent subsequence.

Proof. Suppose $d = 1$. Since $\{x_n\}$ is bounded, we have $x_n \in [a, b]$ for all n where $a, b \in \mathbb{R}$ and $a < b$. The idea is to keep halving the interval and pick a half interval containing infinitely many points. For example, consider the two half intervals $[a, \frac{a+b}{2}]$ and $[\frac{a+b}{2}, b]$. Since $\{x_n\}$ has infinitely many points, at least one of these two half intervals has infinitely many points. Call this half interval $[a_1, b_1]$ and repeat this argument again for $[a_1, b_1]$. This gives us a sequence $\{(a_n, b_n)\}$ satisfying

$$\begin{aligned} a_0 \leq a_1 \leq a_2 \leq \cdots b &\implies a_n \to a^* \\ b_0 \geq b_1 \geq b_2 \geq \cdots a &\implies b_n \to b^* \end{aligned}$$

where we define $a_0 := a$ and $b_0 := b$. Moreover, we have

$$|b_n - a_n| = \frac{b-a}{2^n} \to 0 \implies a^* = b^*.$$

Since there are infinitely many x_n's in $[a_k, b_k]$ for any k, pick $\tilde{x}_k \in [a_k, b_k] \cap \{x_n\}$ such that $\tilde{x}_k \neq \tilde{x}_j$ for $j < k$. Thus, $\tilde{x}_k \to a^* = b^*$. This can be generalised to $d > 1$ via induction and we leave this as an exercise to the reader. $\square$

Note that the above argument does not generalise to infinite dimensions. For example, consider the complete orthonormal space

$$\mathcal{L}_2[0, T] := \left\{ f\colon [0, T] \to \mathbb{R}\colon \int_0^T f^2(t)\,\mathrm{d}t < \infty \right\}$$

with inner product

$$\langle f, g \rangle := \int_0^T f(t)g(t)\,\mathrm{d}t.$$

Consider an orthonormal basis $\{e_n\}$ satisfying

$$\langle e_n, e_m \rangle = \begin{cases} 1 & \text{if } n = m, \\ 0 & \text{if } n \neq m. \end{cases}$$

Note that $\|e_n - e_m\| = \sqrt{2}$ whenever $n \neq m$ and thus $\{e_n\}$ has no convergent subsequence.

Proposition 2.2

Let $f\colon C \subset \mathbb{R}^d \to \mathbb{R}$ be bounded from below. Let $\beta = \inf_{\mathbf{x} \in C} f(\mathbf{x})$. Then, $\exists \{\mathbf{x}_n\} \in C$ such that $f(\mathbf{x}_n) \downarrow \beta$.

Theorem 2.3: Weierstrass Theorem

Let $C \subset \mathbb{R}^d$ be closed and bounded, and let $f\colon C \to \mathbb{R}$ be continuous. Then, f attains its minimum and maximum.

Proof. Let $\{\mathbf{x}_n\rangle \in C$ be such that $f(\mathbf{x}_n) \downarrow \beta := \inf_{\mathbf{x} \in C} f(\mathbf{x})$. By Bolzano-Weierstrass, $\exists \{\mathbf{x}_{n_k}\}$ such that $\mathbf{x}_{n_k} \to \mathbf{x}^*$. Since C is closed, $\mathbf{x}^* \in C$. Since f is continuous, $f(\mathbf{x}_{n_k}) \to f(\mathbf{x}^*) \implies f(\mathbf{x}^*) = \beta$. A similar argument holds for maximum. □

Corollary 2.1

Let $C \subset \mathbb{R}^d$ be closed and let $f\colon C \to \mathbb{R}$ satisfy

$$\lim_{\|\mathbf{x}\| \uparrow \infty} f(\mathbf{x}) = \infty.$$

Then, f attains its minimum on C.

Proof. Let $\{\mathbf{x}_n\}$ be such that $f(\mathbf{x}_n) \downarrow \beta := \inf_{\mathbf{x} \in C} f(\mathbf{x})$. Then, $\{\mathbf{x}_n\}$ is bounded, since otherwise $\exists \{\mathbf{x}_{n_k}\}$ such that $\|\mathbf{x}_{n_k}\| \uparrow \infty \implies f(\mathbf{x}_{n_k}) \to \infty \neq \beta$. The previous argument now follows through. □

Definition 2.4: Limit Supremum and Limit Infimum

Let $\{x_n\} \in \mathbb{R}$. We define

$$\begin{aligned}\limsup_{n\uparrow\infty} x_n &:= \lim_{n\uparrow\infty} \sup_{m\geq n} x_m = \inf_{n\geq 1} \sup_{m\geq n} x_m \\ \liminf_{n\uparrow\infty} x_n &:= \lim_{n\uparrow\infty} \inf_{m\geq n} x_m = \sup_{n\geq 1} \inf_{m\geq n} x_m\end{aligned}$$

We sometimes also denote the limit supremum as $\overline{\lim} x_n$ and the limit infimum as $\underline{\lim} x_n$.

Note that lim sup and lim inf are always well-defined if we allow $\{\pm\infty\}$ as possibilities. This is because $\sup_{m\geq n} x_m$ is a non-increasing sequence and thus must converge (possibly to $-\infty$). Similarly, $\inf_{m\geq n} x_m$ is a non-decreasing sequence and thus must converge (possibly to $+\infty$). We also note that

1. $\limsup_{n\uparrow\infty} x_n \geq \liminf_{n\uparrow\infty} x_n$.
2. If $\limsup_{n\uparrow\infty} x_n = \liminf_{n\uparrow\infty} x_n = x^*$, then $x_n \to x^*$.

Definition 2.5: Lower and Upper Semicontinuous

$f\colon C \subset \mathbb{R}^d \to \mathbb{R}$ is said to be lower semicontinuous (l.s.c) if whenever $\mathbf{x}_n \to \mathbf{x}^*$ in C, then $\liminf_{n\uparrow\infty} f(\mathbf{x}_n) \geq f(\mathbf{x}^*)$.

$f\colon C \subset \mathbb{R}^d \to \mathbb{R}$ is said to be upper semicontinuous (u.s.c) if whenever $\mathbf{x}_n \to \mathbf{x}^*$ in C, then $\limsup_{n\uparrow\infty} f(\mathbf{x}_n) \leq f(\mathbf{x}^*)$.

Corollary 2.2

If $f\colon C \subset \mathbb{R}^d \to \mathbb{R}$ is lower semicontinuous, C is closed and bounded, then f attains its minimum.

Proof. Let $\{\mathbf{x}_n\} \in C$ be such that $f(\mathbf{x}_n) \downarrow \beta := \inf_{\mathbf{x}\in C} f(\mathbf{x})$. By Bolzano-Weierstrass, $\exists\,\{\mathbf{x}_{n_k}\}$ such that $\mathbf{x}_{n_k} \to \mathbf{x}^*$. Since C is closed, $\mathbf{x}^* \in C$. Then,

$$\beta = \lim_{n\uparrow\infty} f(\mathbf{x}_{n_k}) = \liminf_{n\uparrow\infty} f(\mathbf{x}_{n_k}) \geq f(\mathbf{x}^*) \geq \beta \implies f(\mathbf{x}^*) = \beta.$$

Similarly, an upper semicontinuous function attains its maximum on a closed and bounded domain. $\square$

Proposition 2.6

Let $g\colon C \times D \to \mathbb{R}$ where $C \subset \mathbb{R}^n$, $D \subset \mathbb{R}^m$. Define $f\colon C \to \mathbb{R}$ as

$$f(\mathbf{x}) := \sup_{\mathbf{y} \in D} g(\mathbf{x}, \mathbf{y}) \quad (\text{resp. } \inf_{\mathbf{y} \in D} g(\mathbf{x}, \mathbf{y}))$$

Suppose $f(\mathbf{x}) < \infty$ (resp. $f(\mathbf{x}) > -\infty$) for all $\mathbf{x} \in C$. If $g(\mathbf{x}, \mathbf{y})$ is continuous in $\mathbf{x}$ for all $\mathbf{y} \in D$, then f is lower semicontinuous (resp. upper semicontinuous).

Proof. Let $\mathbf{x}_n \to \mathbf{x}^*$ in C. Then,

$$\begin{aligned}
\liminf_{n \uparrow \infty} f(\mathbf{x}_n) &\geq \liminf_{n \uparrow \infty} g(\mathbf{x}_n, \mathbf{y}) \, \forall \mathbf{y} \in D \\
&= \lim_{n \uparrow \infty} g(\mathbf{x}_n, \mathbf{y}) \\
&= g(\mathbf{x}^*, \mathbf{y}).
\end{aligned}$$

Thus,

$$\begin{aligned}
& \liminf_{n \uparrow \infty} f(\mathbf{x}_n) \geq g(\mathbf{x}^*, \mathbf{y}) \quad \forall \mathbf{y} \in D \\
\implies & \liminf_{n \uparrow \infty} f(\mathbf{x}_n) \geq \sup_{\mathbf{y} \in D} g(\mathbf{x}^*, \mathbf{y}) = f(\mathbf{x}^*).
\end{aligned}$$

□