Optimization of functionals

D. González–Sánchez[1]

Fall 2025

[1]Secihti and Cinvestav-IPN, Mexico City, Mexico

Contents

Chapter 1

Basics of optimization in $\mathbb{R}^n$

1.1 Necessary conditions

Definition 1.1. *Let $X \subseteq \mathbb{R}^n$ and $f : X \to \mathbb{R}$. The point $\hat{x} \in X$ is a* ***local minimizer*** *of f if there is a ball $B_\delta = \{x \in \mathbb{R}^n \mid |x - \hat{x}| < \delta\}$ around $\hat{x}$ such that*

$$f(\hat{x}) \le f(x) \qquad \forall x \in B_\delta \cap X.$$

The proof of the following proposition follows from the latter definition.

Proposition 1.2. *Let X be an open subset of $\mathbb{R}^n$ and $f : X \to \mathbb{R}$. If $\hat{x} \in X$ is a local minimizer of f and there exists the directional derivative*

$$D_v^+ f(\hat{x}) := \lim_{t \downarrow 0} \frac{f(\hat{x} + tv) - f(\hat{x})}{t},$$

for some $v \in \mathbb{R}^n$, $v \neq 0$, then

$$D_v^+ f(\hat{x}) \ge 0.$$

If, in addition, there exists the two-sided directional derivative

$$D_v f(\hat{x}) := \lim_{t \to 0} \frac{f(\hat{x} + tv) - f(\hat{x})}{t},$$

then $D_v f(\hat{x}) = 0$.

Corollary 1.3. *Let X be an open subset of $\mathbb{R}^n$. If $\hat{x} \in X$ is a local minimizer of the differentiable function $f : X \to \mathbb{R}$, then*

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

1.2 Minimization of convex functions

The subset X of R^n is **convex** if for each $x, y \in X$ and $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda) y \in X.$$

Definition 1.4. Let X be a convex subset of R^n. The function $f : X \to \mathbb{R}$ is called

(a) **convex** if for every $x, y \in X$ and $\lambda \in [0, 1]$

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda) f(y); \tag{1.1}$$

(b) **strictly convex** if the inequality is strict for $x \neq y$ and $\lambda \in (0, 1)$.

Proposition 1.5. *Let X be a convex subset of $\mathbb{R}^n$. If $f : X \to \mathbb{R}$ is a convex function, then any local minimizer is a global minimizer.*

Proof. Let $\hat{x}$ be a local minimizer of f, thus

$$f(\hat{x}) \le f(y), \qquad \forall y \in X \cap U,$$

where U is some open subset of $\mathbb{R}^n$. If $x \in X$, then there is $y \in x \cap U$ and $0 < \lambda < 1$ such that

$$y = \lambda \hat{x} + (1 - \lambda)x.$$

Then

$$\begin{aligned} f(\hat{x}) &\le f(y) \\ &\le \lambda f(\hat{x}) + (1 - \lambda) f(x), \end{aligned}$$

that is, $(1 - \lambda) f(\hat{x}) \le (1 - \lambda) f(x)$. Therefore $f(\hat{x}) \le f(x)$ for each $x \in X$. □

Lemma 1.6. *Let $f : S \to \mathbb{R}$ be a convex function, $S \subseteq \mathbb{R}^n$ convex, and $a \in S$. Set, for $x \in S$,*

$$x_\lambda := \lambda x + (1 - \lambda)a, \qquad \lambda \in [0, 1]. \tag{1.2}$$

Then, for every $0 < \lambda < \lambda' \le 1$,

$$\frac{f(x_\lambda) - f(a)}{\lambda} \le \frac{f(x_{\lambda'}) - f(a)}{\lambda'}. \tag{1.3}$$

Further, if f is strictly convex and $x \neq a$, then

$$\frac{f(x_\lambda) - f(a)}{\lambda} < \frac{f(x_{\lambda'}) - f(a)}{\lambda'}, \qquad 0 < \lambda < \lambda' \le 1. \tag{1.4}$$

Proof. We only show the inequality when f is strictly convex, the other one is totally analogous. Pick $x, a \in S$ with $x \neq a$. Then, for $0 < \lambda < \lambda' \le 1$,

$$f\left(\frac{\lambda}{\lambda'} x_{\lambda'} + \left(1 - \frac{\lambda}{\lambda'}\right) a\right) < \frac{\lambda}{\lambda'} f(x_{\lambda'}) + \left(1 - \frac{\lambda}{\lambda'}\right) f(a)$$

since $0 < \lambda/\lambda' < 1$ and $x_{\lambda'} \neq a$, where $x_{\lambda'}$ is given by (1.2). Thus

$$\lambda'[f(x_\lambda) - f(a)] < \lambda[f(x_{\lambda'}) - f(a)], \qquad 0 < \lambda < \lambda' \le 1,$$

because $(x_{\lambda'})_{\frac{\lambda}{\lambda'}} = x_\lambda$, with the notation (1.2). □

Lemma 1.7. *Let $f : S \to \mathbb{R}$ be a C^1 function, where S is an open and convex subset of $\mathbb{R}^n$. The function f is convex in S if and only if*

$$f(x) - f(a) \geq \langle Df(a), x - a \rangle \quad \forall x, a \in S. \tag{1.5}$$

Likewise, f is strictly convex if and if the inequality is strict for every $x \neq a$.

Proof. Suppose that f is convex in S. Then for every $x, a \in S$ and $\lambda \in (0, 1]$

$$f(x) - f(a) \geq \frac{f(a + \lambda(x - a)) - f(a)}{\lambda}.$$

Letting $\lambda \to 0^+$, we obtain (1.5).

Conversely, let $x, a \in S$ and $\lambda \in [0, 1]$. Define $x_\lambda := \lambda x + (1 - \lambda)a$, then (1.5) yields

$$\begin{aligned} f(x) - f(x_\lambda) &\geq \langle Df(x_\lambda), x - x_\lambda \rangle, \\ f(a) - f(x_\lambda) &\geq \langle Df(x_\lambda), a - x_\lambda \rangle. \end{aligned}$$

Therefore

$$\lambda[f(x) - f(x_\lambda)] + (1 - \lambda)[f(a) - f(x_\lambda)] \geq \langle Df(x_\lambda), \lambda(x - x_\lambda) + (1 - \lambda)(a - x_\lambda) \rangle.$$

Since $\lambda(x - x_\lambda) + (1 - \lambda)(a - x_\lambda) = 0$, it follows that

$$\lambda f(x) + (1 - \lambda) f(a) \geq f(\lambda x + (1 - \lambda)a).$$

We now show the second equivalence. Suppose first that f is strictly convex and pick $x, a \in S$ with $x \neq a$. By (1.4), with $\lambda' = 1$,

$$\frac{f(a + \lambda(x - a)) - f(a)}{\lambda} < f(x) - f(a), \qquad 0 < \lambda < 1,$$

then

$$\begin{aligned} f(x) - f(a) &> \inf_{0<\lambda<1} \frac{f(a + \lambda(x - a)) - f(a)}{\lambda} \\ &= Df(a) \cdot (x - a). \end{aligned}$$

For the converse, pick $x, a \in S$, with $x \neq a$. Then, for each $\lambda \in (0, 1)$,

$$\begin{aligned} f(x) - f(x_\lambda) &> Df(x_\lambda) \cdot (x - x_\lambda), \\ f(a) - f(x_\lambda) &> Df(x_\lambda) \cdot (a - x_\lambda), \end{aligned}$$

since $x_\lambda \neq a$. Hence, as above,

$$\lambda f(x) + (1 - \lambda) f(a) > f(\lambda x + (1 - \lambda)a), \quad \lambda \in (0, 1).$$

This completes the proof. □

Theorem 1.8 (First–order necessary and sufficient condition)**.** *Let X, U be sets in $\mathbb{R}^n$ such that $X \subseteq U$, X is convex, and U is open. Let $f : U \to \mathbb{R}$ be differentiable on U and convex on X. Then x^* is a global minimizer of f in X if and only if*

$$Df(x^*) \cdot (x - x^*) \geq 0 \quad \forall x \in X. \tag{1.6}$$

Proof. Suppose first that x^* is a minimizer of f and pick any $x \in X$. Since f is differentiable, there exists $D_v^+ f(x^*) = Df(x^*) \cdot v$, with $v = x - x^*$; by Proposition 1.2 $Df(x^*) \cdot (x - x^*) \geq 0$.

Conversely, if (1.6) holds, then by Proposition 1.7,

$$f(x) \geq f(x^*) + Df(x^*) \cdot (x - x^*) \geq f(x^*) \quad \forall x \in X.$$

Therefore x^* is a global minimizer of f in X. □

1.3 Lagrange multipliers

Theorem 1.9 (Lagrange)**.** *Let $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}^m$ be of class C^1, where U is an open subset of $\mathbb{R}^n$ and $m < n$. If $\hat{z}$ is a local minimizer to problem*

$$\min_{z \in U} \{f(z) \mid g(z) = 0\} \tag{1.7}$$

and $\operatorname{rank}(Dg(\hat{z})) = m$*, then there is a unique $\hat{\lambda} \in \mathbb{R}^m$ such that*

$$Df(\hat{z}) = \hat{\lambda}^\top Dg(\hat{z}). \tag{1.8}$$

Proof. Let us rewrite the optimization problem as

$$\min_{(x,y) \in U} \{f(x, y) \mid g(x, y) = 0\}$$

where $x \in \mathbb{R}^{n-m}$ and $y \in \mathbb{R}^m$. Since $\operatorname{rank}(Dg(\hat{x}, \hat{y})) = m$, where $(\hat{x}, \hat{y}) = \hat{z}$ is the given local minimizer, we can assume that the m rows of $D_y g(\hat{x}, \hat{y})$ are l.i.—otherwise the variables can be reordered. Then by the Implicit Function Theorem, there exists a local implicit C^1 function h such that $g(x, h(x)) = 0$, with $h(\hat{x}) = \hat{y}$, and

$$Dh(\hat{x}) = -[D_y g(\hat{x}, \hat{y})]^{-1} \cdot D_x g(\hat{x}, \hat{y}).$$

On the other hand, $\hat{x}$ is a local minimizer of the function $F(x) := f(x, h(x))$ and so $DF(\hat{x}) = 0$. By the Chain Rule, $D_x f(\hat{x}, h(\hat{x})) + D_y f(\hat{x}, h(\hat{x})) \cdot Dh(\hat{x}) = 0$, that is,

$$D_x f(\hat{x}, \hat{y}) = D_y f(\hat{x}, \hat{y}) \cdot [D_y g(\hat{x}, \hat{y})]^{-1} \cdot D_x g(\hat{x}, \hat{y}).$$

The result follows by defining $\hat{\lambda}^\top := D_y f(\hat{x}, \hat{y}) \cdot [D_y g(\hat{x}, \hat{y})]^{-1}$. □

Proposition 1.10. *Let $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}^m$ be differentiable, where U is an open and convex subset of $\mathbb{R}^n$. Suppose that $\hat{x}$ satisfies* (1.8) *for some $\hat{\lambda} \in \mathbb{R}^m$ and the function*

$$x \mapsto f(x) - \hat{\lambda}^\top g(x), \qquad x \in U,$$

is convex, then $\hat{x}$ is a global minimizer to problem (1.7).

Proof. It follows from Theorem 1.8. □

1.4 Inequality constraints

Let X denote a linear space an let A be a nonempty convex subset of X.

Suppose $f_j : X \to \mathbb{R}$ is convex, for $j = 0, 1, \ldots, n$. In this section, we consider the **convex minimization problem**

$$\inf_{x \in A \cap F} f_0(x), \tag{1.9}$$

where

$$F := \{x \in X \mid f_1(x) \leq 0, \ldots, f_n(x) \leq 0\}.$$

Remark 1.11. Let $f : X \to \mathbb{R}$ be a continuous convex function, where $X = \mathbb{R}^n$. Put

$$F = \{x \in X \mid f(x) \leq 0\}$$

and

$$G = \{x \in X \mid f(x) < 0\}.$$

Then G is open, because f is continuous, and $G \subseteq F$, hence

$$G \subseteq \mathrm{int}(F).$$

In general, $\mathrm{int}(F) \neq G$. Take, for instance, $f \equiv 0$. Nonetheless, if $G \neq \emptyset$, then

$$\mathrm{int}(F) = G.$$

Indeed, let $x \in \mathrm{int}(F)$ and $x_0 \in G$. Then there exists $0 < \varepsilon < 1$ such that

$$y := x + \varepsilon(x - x_0) \in F.$$

Observe that $f(y) \leq 0$, $f(x_0) < 0$, and

$$x = (1 - \lambda)y + \lambda x_0,$$

where $\lambda = \frac{\varepsilon}{1+\varepsilon} > 0$. Because f is convex, we have

$$f(x) \leq (1 - \lambda)f(y) + \lambda f(x_0) < 0$$

which proves that $x \in G$. Therefore $\mathrm{int}(F) \subseteq G$, whenever $G \neq \emptyset$. ◇

Definition 1.12. *The problem* (1.9) *is said to satisfy the* ***Slater's condition*** *if*

$$\{x \in A \mid f_1(x) < 0, \ldots, f_n(x) < 0\} \neq \emptyset.$$

In the following theorem, we use the **Lagrange function** $\mathcal{L} : X \times \mathbb{R}^{n+1} \to \mathbb{R}$ which is given by

$$\mathcal{L}(x, \lambda_0, \ldots, \lambda_n) := \lambda_0 f_0(x) + \ldots + \lambda_n f_n(x).$$

Theorem 1.13 (Kuhn–Tucker)**.** *Suppose* $\overline{x} \in A \cap F$.

(a) *If $\overline{x}$ is a solution to the convex minimization problem* (1.9)*, then there exist nonnegative scalars $\overline{\lambda}_0, \dots, \overline{\lambda}_n$, not all zero, such that*

$$\overline{\lambda}_j f_j(\overline{x}) = 0, \qquad 1 \le j \le n. \tag{1.10}$$

and

$$\mathcal{L}(\overline{x}, \overline{\lambda}_0, \dots, \overline{\lambda}_n) = \min_{x \in A} \mathcal{L}(x, \overline{\lambda}_0, \dots, \overline{\lambda}_n) \tag{1.11}$$

If, in addition, the Slater's condition holds, then $\overline{\lambda}_0 > 0$.

(b) *Assume that* (1.10) *and* (1.11) *hold with $\overline{\lambda}_j \ge 0$, $1 \le j \le n$, and $\overline{\lambda}_0 = 1$. Then $\overline{x}$ is a solution to problem* (1.9)*.*

Proof. (a) Let C be the set of elements $(y_0, y_1, \dots, y_n) \in \mathbb{R}^{n+1}$ that satisfy

$$f_0(x) - f_0(\overline{x}) < y_0,\ f_1(x) \le y_1,\ \dots\ ,\ f_n(x) \le y_n,$$

for some $x \in A$. Then C is convex, because A and the functions $f_0, \dots, f_n$ are convex. Since $\overline{x} \in A \cap F$,

$$y_j > 0,\ 0 \le j \le n, \quad \Rightarrow \quad (y_0, \dots, y_n) \in C. \tag{1.12}$$

In addition, $0 \notin C$. Indeed, if $0 \in C$, then there would exist $x' \in A$ such that $f_0(x') < f_0(\overline{x})$ and $x' \in F$. This is a contradiction because f attains its minimum at $\overline{x}$.

By Theorem A.6, there is a hyperplane that separates C and $\{0\}$, that is, for some $\overline{\lambda} = (\overline{\lambda}_0, \dots, \overline{\lambda}_n) \ne 0$

$$\langle \overline{\lambda} \mid y \rangle \ge 0 \qquad \forall y \in C.$$

From (1.12), we conclude that $\overline{\lambda}_j \ge 0$ for each j.

We now show (1.10). Suppose $f_k(\overline{x}) < 0$ for some $1 \le k \le n$. Put $y_k = f_k(\overline{x})$,

$$y_j = 0 \qquad j \ge 1,\ j \ne k,$$

and $y_0 = \varepsilon$, where $\varepsilon > 0$. Then $(y_0, y_1, \dots, y_n) \in C$, because $\overline{x} \in A \cap F$, and hence

$$\overline{\lambda}_0 \varepsilon + \overline{\lambda}_k f_k(\overline{x}) \ge 0.$$

By letting $\varepsilon \downarrow 0$, we have $\overline{\lambda}_k f_k(\overline{x}) \ge 0$ thus $\overline{\lambda}_k \le 0$. Since we had concluded that $\overline{\lambda}_k \ge 0$, we indeed have

$$f_k(\overline{x}) < 0 \quad \Rightarrow \quad \overline{\lambda}_k = 0.$$

Therefore (1.10) holds.

For each $x \in A$, put $z_j = f_j(x)$ for $1 \le j \le n$, and

$$z_0 = f_0(x) - f_0(\overline{x}) + \varepsilon,$$

where $\varepsilon > 0$. Then $(z_0, z_1, \dots, z_n) \in C$ and

$$\overline{\lambda}_0(f_0(x) - f_0(\overline{x}) + \varepsilon) + \overline{\lambda}_1 f_1(x) + \dots + \overline{\lambda}_n f_n(x) \ge 0$$

By letting $\varepsilon \downarrow 0$, we have
$$\mathcal{L}(x, \overline{\lambda}_0, \dots, \overline{\lambda}_n) \geq \overline{\lambda}_0 f_0(\overline{x}).$$

Therefore (1.11) follows due to (1.10).

Suppose now the Slater's condition holds. Recall that $\overline{\lambda}_0, \dots, \overline{\lambda}_n$ are nonnegative and not all zero. If $\overline{\lambda}_0 = 0$, then $\mathcal{L}(\overline{x}, \overline{\lambda}_0, \dots, \overline{\lambda}_n) = 0$ and

$$\mathcal{L}(x, \overline{\lambda}_0, \dots, \overline{\lambda}_n) < 0$$

for some $x \in A$. This is a contradiction to (1.11), then $\overline{\lambda}_0 > 0$.

(b) Let $x \in A \cap F$. In particular, $x \in F$ and, because $\overline{\lambda}_j \geq 0$, $1 \leq j \leq n$,

$$\sum_{j=1}^{n} \overline{\lambda}_j f_j(x) \leq 0.$$

Finally, due to (1.10) and (1.11),

$$\begin{aligned} f_0(\overline{x}) &= \mathcal{L}(\overline{x}, 1, \overline{\lambda}_1, \dots, \overline{\lambda}_n) \\ &\leq \mathcal{L}(x, 1, \overline{\lambda}_1, \dots, \overline{\lambda}_n) \\ &\leq f_0(x) \end{aligned}$$

for each $x \in A \cap F$.

□

Exercises

1.1 Let $f, g : S \to \mathbb{R}$ be convex functions, where $S \subseteq \mathbb{R}^n$ is convex. Show the following:

(a) If c is a nonnegative real number, then $f + cg$ is convex.

(b) If $F : \mathbb{R} \to \mathbb{R}$ is convex and increasing, then $F \circ f$ is convex.

(c) If $G : \mathbb{R} \to \mathbb{R}$ is concave and decreasing, then $G \circ g$ is concave.

1.2 Show that $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its **epigraph**

$$\{(x, y) \in \mathbb{R}^{n+1} \mid y \geq f(x)\}$$

is convex.

1.3 Prove that $f(x) = |x|$ is convex in $\mathbb{R}^n$. Is f strictly convex? What about $g(x) = |x|^2$?

1.4 Show that the set of minimizers (which could be empty) of any convex function is convex. Prove also that strictly convex functions have at most one global minimizer.

1.5 Let $f_n : \mathbb{R} \to \mathbb{R}$ be a convex function for each $n \in \mathbb{N}$. Prove the following assertions.

(a) If (f_n) converges to f (pointwise), then f is convex.

(b) If $F(x) := \sup_{n\geq 1} f_n(x)$ is finite for each $x \in J$, then F is convex.

1.6 **(Least squares)** Let $A \in \mathcal{M}_{m\times n}$, with $m > n$, and $b \in \mathbb{R}^m$. The system $Ax = b$ usually does not have a solution $x \in \mathbb{R}^n$, then an alternative is to find the *least-squares solution* $\hat{x}$—if it exists—, that is,

$$|A\hat{x} - b|^2 = \min_{x\in\mathbb{R}^n} |Ax - b|^2.$$

Assume rank$(A) = n$ and prove that there exists a unique global minimizer $\hat{x}$, given by

$$\hat{x} = (A^\top A)^{-1} A^\top b.$$

Hint: Since rank$(A) = n$, use the fact that $M^\top M$ is invertible.

1.7 Let $a \in \mathbb{R}^n$, $a \neq 0$. Use the Lagrange multipliers method to find the unique solution to the problem

$$\min_{x\in\mathbb{R}^n} \{a^\top x : |x|^2 = 1\}.$$

Hint: Use also the Cauchy-Schwarz inequality.

1.8 **(Spectral theorem)** Let $A \in \mathcal{M}_n(\mathbb{R})$ be a symmetric matrix.

(a) Use Lagrange multipliers to show that there exists $\lambda_1 \in \mathbb{R}$ and $u_1 \in \mathbb{R}^n$, $|u_1| = 1$, such that

$$Au_1 = \lambda_1 u_1$$

and

$$x \in \mathbb{R}^n,\ |x| = 1 \quad \Rightarrow \quad x^\top A x \geq \lambda_1. \tag{1.13}$$

(b) Show that λ_1 is the smallest eigenvalue of A.

(c) Show that exists $\lambda_2 \in \mathbb{R}$ and $u_2 \in \mathbb{R}^n$, $|u_2| = 1$, such that

$$Au_2 = \lambda_2 u_2$$

and

$$u_2^\top u_1 = 0.$$

Hint: Consider $W_1 = \{x \in \mathbb{R}^n \mid x^\top u_1 = 0\}$, verify that $Ax \in W_1$ for every $x \in W_1$, and find a minimizer u_2 of $x^\top Ax$ in some compact subset of W_1.

(d) Prove that there exist an orthonormal basis $\{u_1, \ldots, u_n\}$ of $\mathbb{R}^n$ and a vector $(\lambda_1, \ldots, \lambda_n)^\top$ such that

$$Au_j = \lambda_j u_j, \quad 1 \leq j \leq n.$$

1.9 Let $A \in \mathcal{M}_n(\mathbb{R})$ be a symmetric matrix. Prove the following:

(a) $\mathrm{tr}(A) := \sum_{j=1}^n A_{jj} = \sum_{j=1}^n \lambda_j$.

Hint: Recall Exercise 1.8(d) to show that $AU = U\Lambda$, where Λ is diagonal and the columns of U are eigenvectors.

(b) A is positive semidefinite if and only if its eigenvalues are nonnegative.

(c) A is positive definite if and only if its eigenvalues are positive.

1.10 Let $h : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Consider the problem

$$\inf\{h(x) \mid x \in \mathbb{R}^n\}.$$

In numerical analysis, a vector $v \in \mathbb{R}^n \setminus \{0\}$ is said to be a *descent direction* of f at a if $D_v h(a) < 0$. If $\nabla h(a) \neq 0$, then $-\nabla h(a)$ is called the *steepest-descent direction* of h at a. Justify these names by proving the following:

(a) If $D_v h(a) < 0$, then there exists $t_0 > 0$ such that

$$h(a + tv) < h(a) \qquad \forall t \in (0, t_0].$$

(b) There exists a solution to

$$\min_{v \in \mathbb{R}^n} \{D_v h(a) : |v|^2 = 1\},$$

and such a solution is given by $-\nabla h(a)/|\nabla h(a)|$.

Chapter 2

Semicontinuous functions

Definition 2.1. *Let (X, d) be a metric space. The function $f : X \to (-\infty, \infty]$ is* **lower semi-continuous** *(***lsc***) on X if*

$$\{x \in X \mid f(x) > \alpha\}$$

is open for every $\alpha \in \mathbb{R}$. Likewise, f is **upper semicontinuous** *(usc) if $-f$ is usc.*

Clearly, $f : X \to (-\infty, \infty]$ is lsc on X if and only if

$$\{x \in X \mid f(x) \leq \alpha\}$$

is closed for every $\alpha \in \mathbb{R}$.

In the following proposition, $X \times \mathbb{R}$ is endowed with the product topology. A characterization of lower semicontinuity is given by means of the **epigraph** of the function $f : X \to (-\infty, \infty]$

$$\operatorname{epi}(f) := \{(x, y) \in X \times \mathbb{R} \mid y \geq f(x)\}.$$

Proposition 2.2. *Let (X, d) be a metric space. The function $f : X \to (-\infty, \infty]$ is lsc if and only if* $\operatorname{epi}(f)$ *is closed.*

Proof. Assume first that f is lsc. Pick any sequence $(x_n, y_n) \in \operatorname{epi}(f)$, $n \in \mathbb{N}$, with

$$(x_n, y_n) \longrightarrow (\overline{x}, \overline{y}).$$

Let $\varepsilon > 0$. Then

$$y_n < \overline{y} + \varepsilon, \qquad n \geq N,$$

for some $N \in \mathbb{N}$, and hence

$$f(x_n) \leq y_n < \overline{y} + \varepsilon, \qquad n \geq N. \tag{2.1}$$

Since f is lsc, the set $\{x \in X \mid f(x) \leq \overline{y} + \varepsilon\}$ is closed, thus (2.1) yields

$$f(\overline{x}) \leq \overline{y} + \varepsilon.$$

By letting $\varepsilon \downarrow 0$, we have $(\overline{x}, \overline{y}) \in \operatorname{epi}(f)$. This proves that $\operatorname{epi}(f)$ is closed.

Conversely, assume epi(f) is closed. Let $\alpha \in \mathbb{R}$ and pick (x_n) any sequence in X with

$$f(x_n) \leq \alpha, \qquad n \in \mathbb{N},$$

and $x_n \to \overline{x}$. Notice that $(x_n, \alpha) \in \text{epi}(f)$ and $(x_n, \alpha) \to (\overline{x}, \alpha)$. Then $f(\overline{x}) \leq \alpha$ because epi(f) is closed. This proves that

$$\{x \in X \mid f(x) \leq \alpha\}, \qquad \alpha \in \mathbb{R},$$

is closed. □

Theorem 2.3. *Let (X, d) be a metric space. The function $f : X \to (-\infty, \infty]$ is lsc if and only if, for each $x \in X$ and any $x_k \to x$,*

$$\liminf_{k\to\infty} f(x_k) \geq f(x).$$

Proof. Assume f is lsc. Let $x \in X$ and $x_k \to x$. For each $\varepsilon > 0$, consider the real number

$$r_\varepsilon := \begin{cases} f(x) - \varepsilon & \text{if } f(x) \in \mathbb{R}, \\ 1/\varepsilon & \text{if } f(x) = \infty. \end{cases}$$

Notice that x is an element of the open set $\{y \in X \mid f(y) > r_\varepsilon\}$, thus there exists $\delta > 0$ such that

$$d(y, x) < \delta \implies f(y) > r_\varepsilon.$$

Since $x_k \to x$,

$$f(x_k) > r_\varepsilon \qquad \forall k \geqslant K,$$

for some $K \in \mathbb{N}$. Hence $\liminf_{k\to\infty} f(x_k) \geq r_\varepsilon$ and, by letting $\varepsilon \downarrow 0$, we obtain

$$\liminf_{k\to\infty} f(x_k) \geq f(x).$$

For the converse assertion, let $\alpha \in \mathbb{R}$. Pick any sequence (x_k) in X with

$$f(x_k) \leq \alpha, \qquad k \in \mathbb{N},$$

and $x_k \to x$. Observe that

$$\liminf_{k\to\infty} f(x_k) \leq \alpha.$$

On the other hand,

$$\liminf_{k\to\infty} f(x_k) \geq f(x).$$

Therefore $f(x) \leq \alpha$. This proves that

$$\{y \in X \mid f(y) \leq \alpha\}, \qquad \alpha \in \mathbb{R},$$

is closed. □

2.1 Existence of minimizers

Lemma 2.4. *Let (X, d) be a compact metric space. If $f : X \to (-\infty, \infty]$ is lsc, then f is bounded below, i.e., there exists $\alpha_0 \in \mathbb{R}$ such that*

$$f(x) \geq \alpha_0, \qquad \forall x \in X.$$

Proof. Since f is lsc, $\cup_{\alpha \in \mathbb{R}}\{x \in X \mid f(x) > \alpha\}$ is an open cover of X. Then the compactness of X yields the desired conclusion. □

Theorem 2.5. *Let (X, d) be a metric space and $f : X \to (-\infty, \infty]$ be an lsc function. If there exists $r \in \mathbb{R}$ such that*

$$S_r := \{x \in X \mid f(x) \leq r\}$$

is nonempty and compact, then the set of global minimizers is nonempty and compact.

Proof. The restriction of f to S_r satisfies the hypotheses of Lemma 2.4, thus

$$l := \inf\{f(x) \mid x \in S_r\}$$

is a real number. For each $n \in \mathbb{N}$, there is $x_n \in S_r$ such that

$$l \leq f(x_n) < l + \frac{1}{n}.$$

Since S_r is compact, the sequence (x_n) has a convergent subsequence, say $x_{n_k} \to x_0$, with $x_0 \in S_r$. Then

$$f(x_0) \leq \liminf_{k \to \infty} f(x_{n_k}) \leq l$$

because f is lsc. This actually proves that $f(x_0) = l$. Observe that $f(x) > r$ for every $x \notin S_r$, then we have

$$f(x_0) \leq f(x) \qquad \forall x \in X.$$

Moreover,

$$\{\hat{x} \in X \mid f(\hat{x}) \leq l\}$$

is a closed subset of the compact set S_α, then the set of global minimizers is nonempty and compact. □

Let (X, d) be a metric space. If $f : X \to (-\infty, \infty]$ is not identically ∞, then f is a *proper function*.

Corollary 2.6. *Let (X, d) be a compact metric space. If $f : X \to (-\infty, \infty]$ is proper and lsc, then there exists $\hat{x} \in X$ such that*

$$f(\hat{x}) \leq f(x) \qquad \forall x \in X.$$

Definition 2.7. *Let $(X, \|\cdot\|)$ be a normed space. The function $f : X \to \mathbb{R}$ is* ***coercive*** *if*

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

Theorem 2.8. *Let $(X, \|\cdot\|)$ be a finite dimensional normed space (over $\mathbb{R}$ or $\mathbb{C}$). If $f : X \to \mathbb{R}$ is lsc and coercive, then f attains its global minimum in X.*

2.2 Ekeland's variational principle

Theorem 2.9 (Ekeland's variational principle)**.** *Let (X, d) be a complete metric space and let $f : X \to (-\infty, \infty]$ be a proper, bounded below, and lsc function. Assume that $\varepsilon > 0$ and $x_0 \in X$ satisfy*

$$f(x_0) \leqslant \varepsilon + \inf_{x \in X} f(x). \tag{2.2}$$

Then, for each $\lambda > 0$, there exists $\overline{x} \in X$ such that

$$f(\overline{x}) \leqslant f(x_0), \tag{2.3}$$

$$d(\overline{x}, x_0) \leqslant \lambda, \tag{2.4}$$

$$f(\overline{x}) < f(x) + \frac{\varepsilon}{\lambda} d(x, \overline{x}) \qquad \forall x \in X \setminus \{\overline{x}\}. \tag{2.5}$$

Proof. For each $x \in X$, consider the set

$$S(x) := \left\{ y \in X \,\middle|\, y \neq x,\ f(x) \geq f(y) + \frac{\varepsilon}{\lambda} d(y, x) \right\}.$$

Notice that $y \in S(x)$ implies $f(y) < f(x)$. On the other hand, if $S(\overline{x}) = \emptyset$ for some $\overline{x} \in X$, then $\overline{x}$ satisfies (2.5).

Step 1. *If $x' \in S(x_0)$, then $d(x', x_0) \leq \lambda$.*

Indeed,

$$\begin{aligned} \frac{\varepsilon}{\lambda} d(x', x_0) &\leq f(x_0) - f(x') \\ &\leq \varepsilon + \inf_X f - f(x') \\ &\leq \varepsilon, \end{aligned}$$

which yields the required inequality.

Step 2. *There exists a sequence (x_n) in X such that*

$$x_n \in S(x_{n-1}) \cup \{x_{n-1}\}, \qquad n \geq 1. \tag{2.6}$$

Furthermore, for each $n \geq 1$,

$$S(x_{n-1}) \neq \emptyset \quad \Rightarrow \quad x_n \in S(x_{n-1}), \tag{2.7}$$

and

$$x_n \neq x_{n-1} \quad \Rightarrow \quad f(x_n) < \varepsilon_n + \inf_{S(x_{n-1})} f, \tag{2.8}$$

where

$$\varepsilon_n := \frac{1}{2}[f(x_{n-1}) - \inf_{S(x_{n-1})} f]$$

We inductively define the sequence (x_n). Suppose that, for $n \geq 1$, x_{n-1} is known—recall that x_0 is given. If $S(x_{n-1}) = \emptyset$, then set $x_n = x_{n-1}$. Otherwise, $\inf_{S(x_{n-1})} f$ is a well-defined real number and ε_n is strictly positive, thus there is $x_n \in S(x_{n-1})$ such that

$$f(x_n) < \varepsilon_n + \inf_{S(x_{n-1})} f.$$

Step 3. *Suppose* (x_n) *satisfies* (2.7)*. If* $S(x_k) = \emptyset$ *for some* $k \geq 0$*, then there is* $\hat{k}$ *such that* $\overline{x} := x_{\hat{k}}$ *verifies* (2.3)*,* (2.4)*, and* (2.5)*.*

If $S(x_0) = \emptyset$, then $\overline{x} = x_0$ verifies the required inequalities. Suppose $S(x_0) \neq \emptyset$ but $S(x_k) = \emptyset$ for some $k \geq 1$. Let

$$\hat{k} = \min\{1 \leq j \leq k \mid S(x_j) = \emptyset\}.$$

Then $S(x_{\hat{k}}) = \emptyset$ and $\overline{x} = x_{\hat{k}}$ satisfies (2.5). Since $x_k \in S(x_{k-1})$, for $1 \leq k \leq \hat{k}$,

$$\frac{\varepsilon}{\lambda} d(x_0, x_1) \leq f(x_0) - f(x_1)$$
$$\vdots$$
$$\frac{\varepsilon}{\lambda} d(x_{\hat{k}-1}, x_{\hat{k}}) \leq f(x_{\hat{k}-1}) - f(x_{\hat{k}}).$$

By adding these inequalities up and using the triangle inequality, we have

$$\frac{\varepsilon}{\lambda} d(x_k, x_{\hat{k}}) \leq f(x_k) - f(x_{\hat{k}}), \qquad 0 \leq k < \hat{k}, \tag{2.9}$$

In particular, we see that $x_{\hat{k}} \in S(x_0)$. Thus $\overline{x} = x_{\hat{k}}$ satisfies (2.3) and, by Step 1, (2.4) also holds.

Step 4. *Suppose* (x_n) *satisfies* (2.7) *and* (2.8)*. If* $S(x_k) \neq \emptyset$ *for every* $k \geq 0$*, then* (x_n) *is convergent and* $\overline{x} = \lim_{n\to\infty} x_n$ *verifies* (2.3)*,* (2.4)*, and* (2.5)*.*

By property (2.7), we have

$$\frac{\varepsilon}{\lambda} d(x_k, x_n) \leq f(x_k) - f(x_n), \qquad k < n. \tag{2.10}$$

Then $f(x_n) < f(x_{n-1})$, for every $n \geq 1$, and hence $(f(x_n))$ converges—because it is a decreasing and bounded-below sequence. Moreover, since $(f(x_n))$ is a Cauchy sequence so is (x_n) because of (2.10). Since X is complete, there exists $\overline{x} = \lim_{n\to\infty} x_n$ and

$$\lim_{n\to\infty} f(x_n) \geq f(\overline{x})$$

due to the lower semicontinuity of f. On the other hand, fix k and let $n \to \infty$ in (2.10) to obtain

$$f(\overline{x}) + \frac{\varepsilon}{\lambda} d(x_k, \overline{x}) \leq f(x_k), \tag{2.11}$$

that is, $\overline{x} \in S(x_k)$ for each $k \geq 0$. In particular, $\overline{x} \in S(x_0)$, then $\overline{x}$ satisfies (2.3) and (2.4).

Suppose $\overline{x}$ does not satisfy (2.5), that is, there exists $x' \in S(\overline{x})$. Then

$$f(x') < f(\overline{x}) \tag{2.12}$$

and

$$f(x') + \frac{\varepsilon}{\lambda} d(x', \overline{x}) \leq f(\overline{x}).$$

The latter inequality along with (2.11) imply that $x' \in S(x_k)$ for every k. From (2.8),

$$2f(x_{k+1}) - f(x_k) < \inf_{S(x_k)} f \qquad \forall k,$$

hence $2f(x_{k+1}) - f(x_k) < f(x')$ and, by letting $k \to \infty$,

$$f(\overline{x}) \leq \lim_{k\to\infty} f(x_k) \leq f(x').$$

This inequality contradicts (2.12). We conclude that $\overline{x}$ indeed satisfies (2.5).

Therefore the sequence defined in Step 2 is convergent—by Steps 3 and 4—and its limit satisfies the theorem. □

Corollary 2.10. *Let (X, d) be a complete metric space and let $f : X \to (-\infty, \infty]$ be a proper, bounded below, and lsc function. For each $\varepsilon > 0$, there exists $\overline{x} \in X$ such that*

$$f(\overline{x}) < f(x) + \sqrt{\varepsilon} d(x, \overline{x}) \qquad \forall x \in X \setminus \{\overline{x}\}. \tag{2.13}$$

The following result, also known as Banach's Fixed Point Theorem, follows from Ekeland's variational principle (EVP).

Theorem 2.11 (Contraction mapping principle)**.** *Let (X, d) be a complete metric space and let $F : X \to X$ be a contraction mapping, that is, there is $0 < \beta < 1$ such that*

$$d(F(x), F(y)) \leq \beta d(x, y) \qquad \forall x, y \in X. \tag{2.14}$$

Then F has a unique fixed point $\overline{x}$, i.e., $F(\overline{x}) = \overline{x}$.

Proof. Let $f(x) := d(x, F(x))$, for each x in X, and $\varepsilon := (1 - \beta)^2/2$. Thus

$$\sqrt{\varepsilon} + \beta < 1. \tag{2.15}$$

By Corollary 2.10 to EVP, there exists $\overline{x}$ such that

$$d(\overline{x}, F(\overline{x})) < d(x, F(x)) + \sqrt{\varepsilon} d(x, \overline{x}) \qquad \forall x \neq \overline{x}.$$

Suppose $\overline{x} \neq F(\overline{x})$. Then, by the latter inequality and (2.14),

$$d(\overline{x}, F(\overline{x})) < (\beta + \sqrt{\varepsilon}) d(\overline{x}, F(\overline{x}))$$

which contradicts (2.15). Therefore $\overline{x} = F(\overline{x})$.

Concerning uniqueness, if $F(x') = x'$, then

$$d(\overline{x}, x') = d(F(\overline{x}), F(x')) \leq \beta d(\overline{x}, x')$$

and hence $d(\overline{x}, x')$. This proves the theorem. □

Exercises

2.1 Let $\mathcal{F}$ be a collection of lsc functions on the metric space (X, d). Show that

$$F(x) := \sup\{f(x) \mid f \in \mathcal{F}\}, \qquad x \in X,$$

is lsc.

Hint: Notice that $\{x \in X \mid F(x) > \alpha\} = \cup_{f\in\mathcal{F}}\{x \in X \mid f(x) > \alpha\}$.

2.2 Let $f, g : X \to (-\infty, \infty]$ be lsc functions on the metric space (X, d). If $r > 0$, then show that rf and $f + g$ are lsc.

2.3 Prove that $f : X \to \mathbb{R}$ is continuous if and only if f is both l.s.c. and u.s.c.

2.4 Let A be a subset of the metric space (X, d). Consider the function

$$I_A(x) = \begin{cases} 0 & \text{if } x \in A, \\ \infty & \text{if } x \notin A. \end{cases}$$

Show that I_A is lsc if and only if A is closed.

2.5 Let (X, d) be a metric space and $\emptyset \neq A \subseteq X$. Define

$$d(x, A) := \inf_{a\in A} d(x, a) \quad x \in X. \tag{2.16}$$

Prove the following:

(a) $d(x, A) \leq d(x, y) + d(y, A)$ for every $x, y \in X$,

(b) the function $d(\cdot, A) : X \to \mathbb{R}$ is uniformly continuous,

(c) if A is closed and $d(x, A) = 0$, then $x \in A$.

2.6 Let (X, d) be a metric space, $f : X \to \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$.

(a) Give an example of lsc functions f and g such that $g \circ f$ is not lsc.

(b) Suppose f is continuous and g is lsc. Prove that $g \circ f$ is lsc.

2.7 (**The Fundamental Theorem of Algebra** [3, 1, 5]). Let $p(z) = a_n z^n + \ldots + a_1 z + a_0$ be a polynomial with complex coefficients, $a_n \neq 0$ and $n \geq 1$. Define the function $f(z) := |p(z)|$ for each $z \in \mathbb{C}$.

(a) Show that f has a global minimizer.

Hint: Show that f is coercive.

(b) Find explicitly one (there could be more) global minimizer of f when (i) $p(z) = a_1 z + \ldots + a_n z^n$, that is $a_0 = 0$, and (ii) $p(z) = a_0 + a_k z^k$ with $a_k \neq 0$.

(c) Let $z_0 \in \mathbb{C}$. Explain why there exist complex numbers $c_0, c_1, \ldots, c_n$ such that

$$p(z) = c_0 + c_1(z - z_0) + \ldots + c_n(z - z_0)^n.$$

Hint: Write $p(z) = p((z - z_0) + z_0)$.

Further, prove that, for some $k = 1, \ldots, n$,

$$p(z) = c_0 + c_k(z - z_0)^k + (z - z_0)^{k+1} q(z),$$

where $c_k \neq 0$ and q is a polynomial.

(d) Let z_0 be a global minimizer of f, $t \in (0, 1)$, and $w \in \mathbb{C}$ satisfies $c_0 + c_k w^k = 0$. Suppose $f(z_0) > 0$, that is, $c_0 \neq 0$. Show that

$$f(z_0 + tw) \leq |c_0|(1 - t^k) + |tw|^{k+1}|q(z_0 + tw)|$$

and

$$t|w^{k+1} q(z_0 + tw)| < |c_0|$$

for some t small enough.

(e) Prove the Fundamental Theorem of Algebra.

2.8 (**Baby EVP** [2]) Let X be a finite-dimensional vector space. Suppose $f : X \to \mathbb{R}$ is lsc and bounded below. Let $\varepsilon > 0$ and $x_0 \in X$ satisfy

$$f(x_0) \leq \inf f + \varepsilon. \tag{2.17}$$

Prove (without using Ekeland's variational principle!) that there exists $\overline{x} \in X$ such that

(i) $f(\overline{x}) \leq f(x_0)$,

(ii) $|\overline{x} - x_0| \leq \sqrt{\varepsilon}$, and

(iii) $f(\overline{x}) \leq f(x) + \sqrt{\varepsilon}|x - \overline{x}|$ for all $x \in X$.

In order to accomplish the proof, proceed as follows:

(a) Show that $g(x) = f(x) + \sqrt{\varepsilon}|x - x_0|$ has a global minimizer $\overline{x}$ in X.

Hint: Show that g is coercive and lsc.

(b) Use the inequality $g(\overline{x}) \leq g(x_0)$ to prove (i) and (ii).

(c) Finally, use (a) to prove (iii).

Hint: Notice that $|x - x_0| \leq |x - \overline{x}| + |\overline{x} - x_0|$.

Appendix A

Convexity in $\mathbb{R}^n$

A.1 Continuity of convex functions

Theorem A.1. *Let $f : S \to \mathbb{R}$ be a convex function. If x_0 is an interior point of S, then f is continuous at x_0.*

Proof. Let $\{e_k \mid k = 1, \dots, n\}$ be the canonical basis of $\mathbb{R}^n$. Since x_0 is an interior point of S, there exists $\varepsilon > 0$ such that $\overline{B}_1(x_0, \varepsilon) \subseteq S$. Define, for $k = 1, \dots, 2n$,

$$d_k = \begin{cases} \varepsilon e_{\frac{k+1}{2}} & \text{if } k \text{ is odd,} \\ -\varepsilon e_{\frac{k}{2}} & \text{if } k \text{ is even,} \end{cases}$$

and $M := \max\{f(x_0 + d_k) \mid k = 1, \dots, 2n\}$. Then

$$f(x) \leq M \quad \forall x \in \overline{B}_1(x_0, \varepsilon). \tag{A.1}$$

On the other hand, let $\{x_k\} \subseteq S$ be any sequence converging to x_0. Then there exists K such that $x_k \in B_1(x_0, \varepsilon)$ for every $k \geq K$. Furthermore,

$$x_k = \lambda_k x_0 + (1 - \lambda_k) y_k, \qquad k \geq K,$$

for some $\lambda_k \in [0, 1]$ and y_k such that $\|y_k - x_0\|_1 = \varepsilon$. Notice that

$$\lim_{k\to\infty} \|(1 - \lambda_k)(y_k - x_0)\| = 0,$$

that is, $\lim_{k\to\infty}(1 - \lambda_k) = 1$. Since f is convex,

$$f(x_k) \leq \lambda_k f(x_0) + (1 - \lambda_k) f(y_k), \quad k \geq K,$$

thus, by (A.1), $\overline{\lim}_{k\to\infty} f(x_k) \leq f(x_0)$.

On the other hand, the inequality $\underline{\lim}_{k\to\infty} f(x_k) \geq f(x_0)$ can be obtained by considering convex combinations of the form $x_0 = \theta_k x_k + (1 - \theta_k) z_k$ with $\|z_k - x_0\| = \varepsilon$. Therefore $\overline{\lim}_{k\to\infty} f(x_k) \leq f(x_0) \leq \underline{\lim}_{k\to\infty} f(x_k)$ implies the continuity of f at x_0. □

Corollary A.2. *Let $f : S \to \mathbb{R}$ be a convex function. If S is open, then f is continuous on S.*

A.2 Convex functions of class C^2

Theorem A.3. *Let* $f : S \to \mathbb{R}$ *be a* C^2 *function, where* $S \subseteq \mathbb{R}^n$ *is open and convex.*

(a) f *is convex in* S *if and only if* $D^2 f(x)$ *is positive semidefinite for every* $x \in S$.

(b) *If* $D^2 f(x)$ *is positive definite for every* $x \in S$, *then* f *is strictly convex.*

Proof. Let $x \in S$ and $h \in \mathbb{R}^n$.

(a) Suppose that f is convex on S and fix $x \in S$. Let $h \in \mathbb{R}^n$, $h \neq 0$. Then we can choose $N \in \mathbb{N}$ such that

$$x + n^{-1}h \in S \qquad \forall n \geq N.$$

Since S is convex, by Taylor theorem, there exists $\theta_n \in (0, 1)$ such that

$$f(x + n^{-1}h) = f(x) + n^{-1}Df(x)h + \frac{1}{2}n^{-2}h^\top D^2 f(x + \theta_n n^{-1}h)h, \qquad \forall n \geq N.$$

Theorem 1.7 implies

$$h^\top D^2 f(x + \theta_n n^{-1}h)h \geq 0, \qquad \forall n \geq N. \tag{A.2}$$

Notice that, when $n \to \infty$, $|\theta_n n^{-1}h| \to 0$ and hence

$$\lim_{n\to\infty} D^2 f(x + \theta_n n^{-1}h) = D^2 f(x)$$

because f is of class C^2. Then, by letting $n \to \infty$ in (A.2), it follows that

$$h^\top D^2 f(x)h \geq 0.$$

This proves that $D^2 f(x)$ is positive semidefinite at x.

Conversely, by Taylor theorem, with $h = x - a$,

$$f(x) = f(a) + Df(a) \cdot (x - a) + \frac{1}{2}(x - a)^\top D^2 f(a + \theta(x - a)) \cdot (x - a) \tag{A.3}$$

for some $\theta \in (0, 1)$. Since $D^2 f(\cdot)$ is positive definite, then $f(x) - f(a) \geq Df(a) \cdot (x - a)$ for each $x, a \in S$. Therefore f is convex by Theorem 1.7.

(b) It follows from (A.3) and Theorem 1.7.

□

A.3 Separation theorems

Definition A.4. *Let* $p \in \mathbb{R}^n \setminus \{0\}$ *and* $\beta \in \mathbb{R}$. *The* **hyperplane** *determined by* p *and* β *is the set*

$$H(p, \beta) := \{x \in \mathbb{R}^n \mid \langle p, x \rangle = \beta\}.$$

Theorem A.5. *Let* $C \subseteq \mathbb{R}^n$ *be a nonempty, convex, and closed set. If* $y \in \mathbb{R}^n \setminus C$, *then there exists a hyperplane* $H(p, \alpha)$, $p \neq 0$, *that separates* y *from* C, *that is,*

$$\langle p, y \rangle < \alpha \leq \langle p, c \rangle \quad \forall c \in C.$$

Furthermore, there exists a hyperplane $H(p, \beta)$, $p \neq 0$, *that strictly separates* y *from* C, *that is,*

$$\langle p, y \rangle < \beta < \langle p, c \rangle \quad \forall c \in C.$$

Proof. Since C is closed, there exists $c_0 \in C$ such that $0 < \|y - c_0\| \leq \|y - c\|$ for every $c \in C$. Define $p := c_0 - y$ and $\alpha := \langle p, c_0 \rangle$. Notice that $p \neq 0$ and

$$\langle p, y \rangle = \alpha - \|p\|^2 < \alpha.$$

For any $c \in C$ and $\lambda \in (0, 1]$, the point $c_\lambda := (1 - \lambda)c_0 + \lambda c$ belongs to C. Then

$$\begin{aligned} \|y - c_0\|^2 &\leq \|y - c_\lambda\|^2 \\ &= \|y - c_0 + \lambda(c_0 - c)\|^2 \\ &= \|y - c_0\|^2 + \lambda^2 \|c_0 - c\|^2 + 2\lambda \langle y - c_0, c_0 - c \rangle, \end{aligned}$$

which is equivalent to $2 \langle p, c_0 - c \rangle \leq \lambda \|c_0 - c\|^2$. By letting $\lambda \to 0$, we obtain

$$\langle p, y \rangle < \alpha \leq \langle p, c \rangle.$$

The second part of the theorem follows for any β in the interval $(\langle p, y \rangle, \alpha)$. □

Theorem A.6. *Let* $C \subseteq \mathbb{R}^n$ *be a nonempty and convex set. If* $y \notin C$, *then there exists a hyperplane* $H(p, \beta)$ *that separates* y *from* C, *that is,*

$$\langle p, y \rangle \leq \beta \leq \langle p, c \rangle \quad \forall c \in C.$$

Proof. Notice first that the closure $\overline{C}$ of C is also convex. Further, there exists $c_0 \in \overline{C}$ such that $\|y - c_0\| \leq \|y - c\|$ for every $c \in C$.

There are two cases for y, (i) $y \notin \overline{C}$ and (ii) y lies in the boundary of $\overline{C}$. Theorem A.5 implies the desired result for case (i). Assume (ii) y is a boundary point of $\overline{C}$, then there is a sequence $\{y_k\} \subseteq \mathbb{R}^n \setminus \overline{C}$ that converges to y. By Theorem A.5, there exists a hyperplane $H(\tilde{p}_k, \tilde{\beta}_k)$, $\tilde{p}_k \neq 0$, that separates $\overline{C}$ from y_k, $k \in \mathbb{N}$. Notice that $H(p_k, \beta_k)$, with

$$p_k := \frac{\tilde{p}_k}{\|\tilde{p}_k\|}, \qquad \beta_k := \frac{\tilde{\beta}_k}{\|\tilde{p}_k\|},$$

also separates $\overline{C}$ from y_k, for each $k \in \mathbb{N}$. Then we can pick a subsequence $\{p_{k_l}\}$ of $\{p_k\}$ such that $\lim_{k \to \infty} p_{k_l} = p$, for some p with $\|p\| = 1$. Therefore $H(p, \beta)$ separates y from C, where $\beta := \langle p, y \rangle$. □

Theorem A.7 (Separating hyperplane theorem)**.** *Let A and B be nonempty convex sets in $\mathbb{R}^n$ such that $A \cap B = \emptyset$. Then there exists a hyperplane that separates A and B.*

Proof. Let $D = A - B := \{x - y \mid x \in A,\ y \in B\}$. Then D is a convex set and $0 \notin D$. By Theorem A.6, there is a hyperplane $H(p, \alpha)$ such that

$$\langle p, x - y\rangle \leq \alpha \leq 0 \quad \forall x \in A,\ y \in B.$$

Define $\beta := \sup\{\langle p, x\rangle \mid x \in A\}$. Therefore the hyperplane $H(p,\beta)$ separates A and B. □

Exercises

A.1 Let $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ be convex sets. Prove that $A \times B$ is convex in $\mathbb{R}^{n+m}$.

A.2 Show any open ball $B_\varepsilon(x)$ in $\mathbb{R}^n$ is a convex set.

A.3 Let $A \subseteq \mathbb{R}^n$ be a convex set and denote by $\mathrm{int}(A)$ the set of its interior points. Is $\mathrm{int}(A)$ a convex set?

A.4 Show that the *simplex* $\{(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_+ \mid \sum_{j=1}^n \lambda_j = 1\}$ is convex and compact.

A.5 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. If $f(0) = 0$ and f is an *even function* ($f(x) = f(-x)$ for every $x \in \mathbb{R}^n$), show that $f(x) \geq 0$ for every $x \in \mathbb{R}^n$.

A.6 Let $f(x, y) = (x^{-\rho} + y^{-\rho})^{-1/\rho}$ for $(x, y) \in \mathbb{R}^2_{++}$ and $\rho \neq 0$. Show that f is

(a) concave if $\rho \geq -1$,

(b) convex if $\rho \leq -1$.

A.7 Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function. Show that $x_1 < x_2 < x_3$ implies

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

Hint: Consider the convex combination $x_2 = \lambda x_3 + (1 - \lambda)x_1$, where $\lambda = \frac{x_2 - x_1}{x_3 - x_1}$.

A.8 If $x_1, \cdots, x_k$ are positive real numbers, show that

$$\sqrt[k]{x_1 \cdots x_k} \leq \frac{x_1 + \ldots + x_k}{k}.$$

A.9 (Sydsæter et al. [4]) Consider the *Cobb–Douglas function*

$$f(x) = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$

defined on $\mathbb{R}^n_{++}$ for $\alpha_i > 0$ $(i = 1, 2, \ldots, n)$.

(a) Show that the kth leading principal minor of the Hessian $Hf(x)$ is

$$H_k f(x) = [f(x)]^k \frac{\alpha_1 \cdots \alpha_k}{(x_1 \cdots x_k)^2} \begin{vmatrix} \alpha_1 - 1 & \alpha_1 & \cdots & \alpha_1 \\ \alpha_2 & \alpha_2 - 1 & \cdots & \alpha_2 \\ & & \ddots & \\ \alpha_k & \alpha_k & \cdots & \alpha_k - 1 \end{vmatrix}.$$

(b) Show indeed that $H_k f(x) = [-f(x)]^k [1 - \sum_{i=1}^k \alpha_i] \frac{\alpha_1 \cdots \alpha_k}{(x_1 \cdots x_k)^2}$.

(c) Prove that f is strictly concave if $\alpha_1 + \ldots + \alpha_n < 1$.

Bibliography

[1] C. Fefferman. Classroom Notes: An Easy Proof of the Fundamental Theorem of Algebra. *Amer. Math. Monthly*, 74(7):854–855, 1967.

[2] J.-B. Hiriart-Urruty. A short proof of the variational principle for approximate solutions of a minimization problem. *Amer. Math. Monthly*, 90(3):206–207, 1983.

[3] J. E. Littlewood. Mathematical notes (14): “Every polynomial has a root.”. *J. London Math. Soc.*, 16:95–98, 1941.

[4] K. Sydsæter, P. Hammond, A. Seierstad, and A. Strøm. *Further mathematics for economic analysis*. Pearson Education, second edition, 2008.

[5] F. Terkelsen. The fundamental theorem of algebra. *Amer. Math. Monthly*, 83(8):647, 1976.