Optimization of functionals

D. González–Sánchez[1]

Fall 2025

[1]Conahcyt 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. A function $f : J \to \mathbb{R}$ is called

(a) **convex** if for every $x, y \in J$ 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 M be a convex subset of $\mathbb{R}^n$. If $f : M \to \mathbb{R}$ is a convex function, then any local minimizer is a global minimizer.*

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

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

where U is some open subset of X. If $x \in M$, then there is $y \in M \cap U$ 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 $x \in M$. □

Lemma 1.6. *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*

$$f(x) - f(a) \ge \langle Df(a), x - a \rangle \quad \forall x, a \in S. \tag{1.2}$$

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 $\lambda \in (0, 1]$

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

Letting $\lambda \to 0^+$, we obtain (1.2).

Conversely, let $x, a \in S$ and $\lambda \in [0, 1]$. Define $x_\lambda := \lambda x + (1-\lambda)a$, then (1.2) yields

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

Therefore

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

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 (**??**), 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, assume (**??**) holds for any pair of different points. 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 1.7 (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*

$$Df(x^*) \cdot (x - x^*) \geq 0 \quad \forall x \in X. \tag{1.3}$$

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 (1.3) holds, then by Proposition 1.6,

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

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

1.3 Lagrange multipliers

Theorem 1.8 (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*

$$\min_{z \in U} \{ f(z) \mid g(z) = 0 \} \tag{1.4}$$

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

$$Df(\hat{z}) = \hat{\lambda}^\top Dg(\hat{z}). \tag{1.5}$$

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 1.9. *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* (1.5) *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 (1.4).

Proof. It follows from Theorem 1.7. □

1.4 Inequality constraints

Let A be a nonempty convex subset of X and let $f_j : X \to \mathbb{R}$ be convex for $j = 0, 1, \dots, n$. Consider the **convex minimization problem**

$$\inf_{x\in A\cap F} f_0(x), \tag{1.6}$$

where

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

Remark 1.10. Let $f : X \to \mathbb{R} \cup \{\infty\}$ be a continuous convex function. Put

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

and

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

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

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

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

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

Indeed, let $x \in \text{int}(F)$ 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 $\text{int}(F) \subseteq G$, whenever $G \neq \emptyset$. ◇

Definition 1.11. *The problem* (1.6) *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 1.12 (Kuhn–Tucker)**.** *Suppose* $\overline{x} \in A \cap F$.

(a) *If* $\overline{x}$ *is a solution to the convex minimization problem* (1.6)*, then there exist nonnegative scalars* $\overline{\lambda}_0, \ldots, \overline{\lambda}_n$*, not all zero, such that*

$$\overline{\lambda}_j f_j(\overline{x}) = 0, \qquad 1 \leq j \leq n. \tag{1.7}$$

and

$$\mathcal{L}(\overline{x}, \overline{\lambda}_0, \ldots, \overline{\lambda}_n) = \min_{x \in A} \mathcal{L}(x, \overline{\lambda}_0, \ldots, \overline{\lambda}_n) \tag{1.8}$$

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

(b) *Assume that* (1.7) *and* (1.8) *hold with* $\overline{\lambda}_j \geq 0$*,* $1 \leq j \leq n$*, and* $\overline{\lambda}_0 = 1$*. Then* $\overline{x}$ *is a solution to problem* (1.6)*.*

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

$$f_0(x) - f_0(\overline{x}) < y_0, \; f_1(x) \leq y_1, \; \ldots \;, \; f_n(x) \leq y_n,$$

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

$$y_j > 0, \; 0 \leq j \leq n, \quad \Rightarrow \quad (y_0, \ldots, y_n) \in C. \tag{1.9}$$

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 the Separation theorem, there is a hyperplane that separates C and $\{0\}$, that is, for some $\overline{\lambda} = (\overline{\lambda}_0, \ldots, \overline{\lambda}_n) \neq 0$

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

From (1.9), we conclude that $\overline{\lambda}_j \geq 0$ for each j.

We now show (1.7). Suppose $f_k(\overline{x}) < 0$ for some $1 \leq k \leq n$. Put $y_k = f_k(\overline{x})$,

$$y_j = 0 \qquad j \geq 1,\ j \neq k,$$

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

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

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

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

Therefore (1.7) holds.

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

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

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

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

By letting $\varepsilon \downarrow 0$, we have

$$\mathcal{L}(x, \overline{\lambda}_0, \ldots, \overline{\lambda}_n) \geq \overline{\lambda}_0 f_0(\overline{x}).$$

Therefore (1.8) follows due to (1.7).

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

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

for some $x \in A$. This is a contradiction to (1.8), 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 (1.7) and (1.8),

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

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

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

Hint: Since $\operatorname{rank}(A) = n$*, 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

$$x \in \mathbb{R}^n, \ |x| = 1 \quad \Rightarrow \quad x^\top A x \geq \lambda_1. \tag{1.10}$$

(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

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

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

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

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 $-\nabla h(a)/|\nabla h(a)|$.

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 $\{e_k \mid k = 1, \ldots, n\}$ 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 $k = 1, \ldots, 2n$,

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

and $M := \max\{f(x_0 + d_k) \mid k = 1, \ldots, 2n\}$. 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 1.6 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.3), 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 1.6.

(b) It follows from (A.4) and Theorem 1.6.

□

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.9 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.9, 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.10, 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. □

A.3.1 Exercises

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

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

2.3 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?

2.4 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.

2.5 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 $x \in \mathbb{R}^n$.

2.6 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 $\rho \leq -1$.

2.7 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 $\lambda = \frac{x_2 - x_1}{x_3 - x_1}$.

2.8 If $x_1, \cdots, x_k$ are positive real numbers, show that

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

2.9 (Sydsæter et al. [**?**]) 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 $\alpha_1 + \ldots + \alpha_n < 1$.