# Nonlinear Optimization

Christian Rossi

Academic Year 2024-2025

**Abstract**

The course aims to present the main methods for non-linear optimization, both continuous and discrete. Topics include: optimality conditions for both unconstrained and constrained problems, Lagrangian functions, and duality; gradient-based methods, Newton's methods, and step-size reduction techniques; recursive quadratic programming; and methods using penalty functions.

# Contents

CHAPTER 1

# Optimization

## 1.1 Introduction

Optimization is a powerful and widely used field of applied mathematics, playing a crucial role in solving real-world problems across various domains.

Given a set $X \subseteq \mathbb{R}^n$ and a function $f : X \to \mathbb{R}^n$ hat we aim to minimize, the goal is to find an optimal solution $x^* \in X$ such that:

$$f(\underline{x}^*) \leq f(\underline{x}) \qquad \forall \underline{x} \in X$$

Many decision-making problems cannot be effectively modeled using linear approaches due to their inherent nonlinearity.

**Example:**
We are given:

- $m$ warehouses, indexed by $i = 1, \ldots, m$, each with a capacity $p_i$ and a location constraint within an area $A_i \subseteq \mathbb{R}^2$.
- $n$ clients, indexed by $j = 1, \ldots, n$, each located at coordinates $(a_j, b_j)$ with a demand $d_j$.

Our goal is to determine the optimal warehouse locations and how to distribute the product to clients while minimizing transportation costs, ensuring that warehouse capacities and client demands are met. We assume that the total available supply is sufficient: $\sum_{i=1}^{m} p_i \geq \sum_{j=1}^{n} d_j$.

The decision variables for this problem are:

- $(x_i, y_i)$: the coordinates of of warehouse $i$ (for $1 \leq i \leq m$).
- $w_{ij}$: the quantity of product transported from warehouse $i$ to client $j$ (for $1 \leq i \leq m$, $1 \leq j \leq n$).
- $t_{ij}$: the distance between warehouse $i$ and client $j$, given by:

$$t_{ij} = \sqrt{(x_i - a_j)^2 + (y_i - b_j)^2}$$

We aim to minimize the total transportation cost:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} t_{ij} w_{ij}$$

Subject to the following constraints:

1. Warehouse capacity constraints:

$$\sum_{j=1}^{n} w_{ij} \leq p_i \qquad \forall i$$

2. Client demand satisfaction:

$$\sum_{i=1}^{m} w_{ij} \geq d_j \qquad \forall j$$

3. Warehouse location constraints:

$$(x_i, y_i) \in A_i \subseteq \mathbb{R}^2 \qquad \forall i$$

4. Non-negativity constraints:

$$w_{ij} \geq 0, t_{ij} \geq 0 \qquad \forall i, j$$

This formulation ensures that all client demands are met while keeping transportation costs minimal and adhering to warehouse capacities and location constraints.

**Example:**
In computerized tomography, we analyze a 3D volume $V \subseteq \mathbb{R}^3$ that is subdivided into $n$ small cubes, called voxels $V_j$. We assume that the matter density is constant within each voxel.

Our goal is to reconstruct a 2D slice of $V$, meaning we need to determine the density $x_j$ for each voxel $V_j$ based on measurements from $m$ X-ray beams. For the $i$-th beam:

- $a_{ij}$ represents the path length of the beam within voxel $V_j$.
- $I_0$ is the initial X-ray intensity at the source.
- $I_i$ is the intensity after passing through the volume.

According to the Beer-Lambert law, the total log-attenuation of the beam is linearly related to the density:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i = \log \frac{I_0}{I_i} \qquad i = 1, \ldots, m$$

We can formulate this as a linear system:

$$Ax = b, \quad x_j \geq 0 \qquad \forall j = 1, \ldots, n$$

However, this system is often infeasible due to measurement errors, non-uniformity of voxels, and other practical factors.

To handle inconsistencies, we use a least squares formulation to minimize the reconstruction error:

$$\min \sum_{i=1}^{m} \left( b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2$$

Subject to:

$$x_j \geq 0 \qquad j = 1, \ldots, n$$

Since $n \gg m$ (many voxels, fewer beams), the problem may have multiple optimal solutions. To improve stability, we introduce a regularization term and minimize:

$$f(x) = \sum_{i=1}^{m} \left( b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2 + \delta \sum_{j=1}^{n} x_j$$

Here, $\delta > 0$ controls the strength of regularization.

The function $f(x)$ can include nonlinear terms to better account for material properties, image characteristics, or a stochastic model of attenuation. The number and directions of beams can also be optimized to improve reconstruction quality. In dynamic imaging, we can extend this to four dimensions to account for respiratory motion over time.

**Example:**
In this auction setting, bidders can place bids on combinations of discrete items, rather than bidding on individual items separately. Given:

- A set $N$ of $n$ bidders.
- A set $M$ of $m$ distinct items,
- For each subset $S \subseteq M$, bidder $j \in N$ is willing to pay $b_j(S)$ for that specific bundle $S$.

We assume that bidders may place higher bids for bundles than for individual items separately, i.e., there may be synergies in bundling items.

Our goal is to determine the allocation of items to bidders to maximize total revenue while ensuring that no item is allocated more than once. The decision variables will be:

- $b(S) = \max_{j \in N} b_j(S)$: the highest bid received for bundle.
- $x_S$: a binary variable indicating whether the highest bid for $S$ is accepted:

$$x_S = \begin{cases} 1 & \text{if the highest bid on } S \text{ is accepted} \\ 0 & \text{otherwise} \end{cases}$$

The formulation is the following:

$$\max \sum_{S \subseteq M} b(S) x_S$$

Subject to:

1. Each item can be allocated to at most one winning bundle:

$$\sum_{S \subseteq M | i \in S} x_S \leq 1 \qquad \forall i \in M$$

2. Each bundle is either selected or not:

$$x_S \in \{0,1\} \qquad \forall S \subseteq M$$

In essence, when $x_S = 1$, bundle $S$ is awarded to a bidder willing to pay the highest price, ensuring that the total revenue is maximized while maintaining a valid allocation of items.

This formulation has $2^{|M|}$ variables, making it computationally challenging for large sets of items. Efficient algorithms, such as branch-and-bound, linear programming relaxations, or heuristic methods, may be necessary for practical problem sizes.

## 1.2 Optimization problem

The general optimization problem is formulated as follows:

$$\begin{aligned} \min \ & f(\underline{x}) \\ \text{such that } & g_i(\underline{x}) \leq 0 \qquad i \leq i \leq m \\ & \underline{x} \in S \subseteq \mathbb{R}^n \end{aligned}$$

**Definition** (*Feasible region*)**.** The feasible region consists of all points that satisfy both the set constraints and the algebraic constraints:

$$X = S \cap \{\underline{x} \in \mathbb{R}^n \mid g_i(\underline{x}) \leq 0, 1 \leq i \leq m\}$$

Here, each constraint function $g_i : S \to \mathbb{R}$ defines a restriction on the feasible set.

**Definition** (*Objective function*)**.** The function to be minimized, known as the objective function, is given by:

$$f : X \to \mathbb{R}$$

It assigns a numerical value to each feasible solution $\underline{x}$, which we seek to minimize.

Without loss of generality, we assume:

- The problem is a minimization problem, since maximization can be rewritten as:

$$\max_{\underline{x} \in X} f(\underline{x}) = \min_{\underline{x} \in X} -f(\underline{x})$$

- All algebraic constraints are inequality constraints, since equality constraints can be rewritten as two inequalities:

$$g(\underline{x}) = 0 \equiv \begin{cases} g(\underline{x}) \geq 0 \\ g(\underline{x}) \leq 0 \end{cases}$$

**Definition** (*Global optimum*)**.** A feasible solution $\underline{x}^* \in X$ is a global optimum if:

$$f(\underline{x}^*) \leq f(\underline{x}) \qquad \forall \underline{x} \in X$$

**Definition** (*Local optimum*)**.** A feasible solution $\overline{\underline{x}} \in X$ is a local optimum if there exists $\epsilon > 0$ such that:

$$f(\overline{\underline{x}}) \leq f(\underline{x}) \qquad \forall \underline{x} \in X \cap \mathcal{N}_\epsilon(\overline{\underline{x}})$$

Here, $\mathcal{N}_\epsilon(\overline{\underline{x}}) = \{\underline{x} \in X \mid \|\underline{x} - \overline{\underline{x}} \leq \epsilon\|\}$ is an epsilon-neighborhood around $\overline{\underline{x}}$.

For complex optimization problems, finding a global optimum is often computationally infeasible. Instead, we focus on obtaining good local optima within a reasonable computation time.

# CHAPTER 2

## Convex analysis

### 2.1 Introduction

In mathematical analysis and optimization, understanding the fundamental properties of sets in Euclidean space is crucial. This section introduces key definitions and properties that form the foundation of nonlinear optimization.

#### 2.1.1 Definitions

Let $S \subseteq \mathbb{R}^n$ be a subset of Euclidean space.

**Definition** (*Interior point*)**.** A point $\mathbf{x} \in S \subseteq \mathbb{R}^n$ is called an interior point of $S$ if there exists $\varepsilon > 0$ such that the open ball:

$$B_\varepsilon(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y} - \mathbf{x}\| < \varepsilon\} \subseteq S$$

**Definition** (*Boundary point*)**.** A point $\mathbf{x} \in \mathbb{R}^n$ is a boundary point of $S$ if, for every $\varepsilon > 0$, the open ball $B_\varepsilon(\mathbf{x})$ contains at least one point of $S$ and at least one point of $\mathbb{R}^n \setminus S$.

**Definition** (*Set interior*)**.** The interior of $S$, denoted $\mathrm{int}(S)$, is the set of all interior points of $S$.

**Definition** (*Set boundary*)**.** The boundary of $S$, denoted $\partial S$, is the set of all boundary points of $S$.

**Definition** (*Open set*)**.** A set $S \subseteq \mathbb{R}^n$ is said to be open if $S = \mathrm{int}(S)$.

**Definition** (*Closed set*)**.** A set $S \subseteq \mathbb{R}^n$ is said to be closed if its complement $\mathbb{R}^n \setminus S$ is open.

Intuitively, a closed set contains all its boundary points.

**Definition** (*Bounded set*)**.** A set $S \subseteq \mathbb{R}^n$ is bounded if there exists $M > 0$ such that $\|\mathbf{x}\| \leq M$ for all $\mathbf{x} \in S$.

**Definition** (*Compact set*)**.** A set is compact if it is both closed and bounded.

### 2.1.2 Properties

**Property 2.1.1.** A set $S \subseteq \mathbb{R}^n$ is closed if and only if every convergent sequence $\{\mathbf{x}_i\}_{i\in\mathbb{N}} \subseteq S$ has its limit in $\mathbf{x} \in S$.

**Property 2.1.2.** A set $S \subseteq \mathbb{R}^n$ is compact if and only if every sequence $\{\mathbf{x}_i\}_{i\in\mathbb{N}} \subseteq S$ has a convergent subsequence with its limit in $\mathbf{x} \in S$.

### 2.1.3 Optimal solutions existence

In general, for a function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ we can always determine a greatest lower bound (infimum) given by:

$$\inf_{\mathbf{x}\in S} f(\mathbf{x})$$

However, achieving the minimum value within $S$ depends on additional conditions.

**Theorem 2.1.1.** *Let $S \subseteq \mathbb{R}^n$ be a nonempty and compact set, and let $f : S \to \mathbb{R}$ be continuous. Then there exists $\mathbf{x}^* \in S$ such that:*

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \qquad \forall \mathbf{x} \in S$$

This result does not hold if $S$ is not closed, $S$ is not bounded, or $f(\mathbf{x})$ is not continuous on $S$. When a minimizing point $\mathbf{x}^* \in S$ exists, we denote it as $\min_{\mathbf{x}\in S} f(\mathbf{x})$.

**Definition** (*Cone*)**.** Given a set $S \subset \mathbb{R}^n$, the conic hull of $S$, denoted as $\text{cone}(S)$, is the set of all conic combinations of points in $S$, i.e., all points of the form:

$$\mathbf{x} = \sum_{i=1}^{m} \alpha_i \mathbf{x}_i \qquad \forall i, 1 \leq i \leq m$$

Here, $\mathbf{x}_1, \ldots, \mathbf{x}_m \in S$, and $\alpha_i \geq 0$.

**Definition** (*Affine subspace*)**.** The affine hull, denoted as $\text{aff}(S)$, is the smallest affine subspace containing $S$.

It coincides with the set of all affine combinations of points in $S$, i.e., all points of the form:

$$\mathbf{x} = \sum_{i=1}^{m} \alpha_i \mathbf{x}_i \qquad \forall i, 1 \leq i \leq m$$

Here, $\mathbf{x}_1, \ldots, \mathbf{x}_m \in S$, $\sum_{i=1}^{m} \alpha_i = 1$, and $\alpha_i \in \mathbb{R}$.

## 2.2 Convex analysis

**Definition** (*Convex set*)**.** A set $C \subset \mathbb{R}^n$ is convex if, for any two points $\mathbf{x}_1, \mathbf{x}_2 \in C$ and any $\alpha \in [0, 1]$, the following condition holds:

$$\alpha\mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2 \in C$$

This means that for any two points in the set, the entire line segment connecting them also lies within the set.

**Definition** (*Convex combination*)**.** A point $\mathbf{x} \in \mathbb{R}^n$ is a convex combination of points $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ if:

$$\mathbf{x} = \sum_{i=1}^{m} \alpha_i \mathbf{x}_i \qquad \forall i, 1 \leq i \leq m$$

Here, the coefficients satisfy: $\sum_{i=1}^{m} \alpha_i = 1$, and $\alpha_1 \geq 0$.

In other words, $\mathbf{x}$ is a weighted sum of the given points, with non-negative weights that sum to one.

**Property 2.2.1.** If $C_1, C - 2, \ldots, C_k$ are convex sets, then their intersection $\bigcap_{i=1}^{k} C_i$ is also convex.

**Definition** (*Polyedron*)**.** A polyhedron is the intersection of a finite number of closed half-spaces.

In the context of linear programming, the set of all optimal solutions forms a polyhedron.

**Definition** (*Convex hull*)**.** The convex hull of a set $S \subseteq \mathbb{R}^n$, denoted by $\text{conv}(S)$, is the smallest convex set that contains $S$.

Alternatively, it can be characterized as the set of all convex combinations of points in $S$.

**Definition** (*Extreme point*)**.** A point $x \in C$ is an extreme point of a convex set $C$ if it cannot be written as a convex combination of two other distinct points in $C$. That is, if:

$$\mathbf{x} = \alpha \mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2$$

Here, $\mathbf{x}_1, \mathbf{x}_2 \in C$ and $\alpha \in (0, 1)$.

Then it must be that $\mathbf{x}_1 = \mathbf{x}_2$, meaning $\mathbf{x}$ is not an interior point of any segment within $C$.

**Lemma 2.2.1.** *Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed, and convex set. Then, for any point $\mathbf{y} \notin C$, there exists a unique point $\mathbf{x}' \in C$ that is closest to $\mathbf{y}$. Moreover, $\mathbf{x}'$ is the closest point if and only if:*

$$(\mathbf{y} - \mathbf{x}')^T (\mathbf{x} - \mathbf{x}') \leq 0 \qquad \forall \mathbf{x} \in C$$

**Definition** (*Projection*)**.** The point $\mathbf{x}'$ is called the projection of $\mathbf{y}$ onto $C$.

**Definition** (*Supporting hyperplane*)**.** Let $S \subset \mathbb{R}^n$ be a nonempty set, and let $\overline{\mathbf{x}} \in \partial(S)$ (the boundary relative to the affine hull of $S$). A supporting hyperplane of $S$ at $\overline{\mathbf{x}}$ is defined as:

$$H = \{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{p}^T(\mathbf{x} - \overline{\mathbf{x}}) = 0\}$$

Such that either $S \subseteq H^-$ or $S \subseteq H^+$.

### 2.2.1 Separation theorem

**Theorem 2.2.2.** *Let $C \subset \mathbb{R}^n$ be a nonempty, closed, and convex set, and let $\mathbf{y} \notin C$. Then, there exists a vector $\mathbf{p} \in \mathbb{R}^n$ such that:*

$$\mathbf{p}^T \mathbf{x} < \mathbf{p}^T \mathbf{y} \qquad \forall \mathbf{x} \in C$$

*This means that there exists a hyperplane:*

$$H = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{p}^T \mathbf{x} = \beta\}$$

*With $\mathbf{p} \neq 0$ that separates $\mathbf{y}$ from $C$. That is:*

$$C \subseteq H^- = \left\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{p}^T \mathbf{x} \leq \beta\right\} \wedge \mathbf{y} \notin H^-$$

*Here, the condition $\mathbf{y} \notin H^-$ imples that $\mathbf{p}^T \mathbf{y} > \beta$.*

*Proof.* By the previous lemma, we know that for the closest point $\mathbf{x}' \in C$ to $\mathbf{y}$, we have:

$$(\mathbf{y} - \mathbf{x}')^T(\mathbf{x} - \mathbf{x}') \leq 0 \qquad \forall \mathbf{x} \in C$$

Defining $\mathbf{p} = (\mathbf{y} - \mathbf{x}') \neq 0$ and setting $\beta = (\mathbf{y} - \mathbf{x}')^T\mathbf{x}'$ we obtain:

$$\mathbf{p}^T\mathbf{x} \leq \beta \qquad \forall \mathbf{x} \in C$$

Moreover:

$$\mathbf{p}^T\mathbf{y} - \beta = (\mathbf{y} - \mathbf{x}')^T(\mathbf{y} - \mathbf{x}') = \|\mathbf{y} - \mathbf{x}'\|^2 > 0$$

Since $\mathbf{y} \notin C$. This proves the existence of a separating hyperplane. □

The separation theorem has three main consequences.

**Convex set as intersection** Any nonempty, closed convex set is the intersection of all closed half-spaces containing it.

**Existence of a supporting hyperplane** If $C$ is a nonempty convex set, then for every boundary point $\overline{\mathbf{x}} \in \partial(S)$, there exists at least one supporting hyperplane $H$ at $\overline{\mathbf{x}}$. That is, there exists $\mathbf{p} \neq 0$ such that:

$$\mathbf{p}^T(\mathbf{x} - \overline{\mathbf{x}}) \leq 0 \qquad \forall \mathbf{x} \in C$$

**Farkas lemma** Farkas' Lemma is a fundamental result in linear algebra and optimization, playing a key role in proving optimality conditions for nonlinear programming. It provides an alternative statement for the feasibility of a linear system, offering a way to certify infeasibility.

**Lemma 2.2.3.** *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Exactly one of the following two statements is true:*

1. *There exists $\mathbf{x} \in \mathbb{R}^n$ such that:*

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad \forall \mathbf{x} \geq \mathbf{0}$$

2. *There exists $\mathbf{y} \in \mathbb{R}^m$ such that:*

$$\mathbf{y}^T\mathbf{A} \leq \mathbf{0}^T \qquad \mathbf{y}^T\mathbf{b} \geq \mathbf{0}^T$$

Thus, either $\mathbf{b}$ belongs to the cone of $\mathbf{A}$, or it is strictly separated by some hyperplane.

*Proof (forward direction).* Suppose there exists $\tilde{\mathbf{x}} \geq 0$ such that $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}$. Then, for any $\mathbf{y}$ satisfying $\mathbf{y}^T\mathbf{A} \leq \mathbf{0}$, we have:

$$\mathbf{y}^T\mathbf{b} = \mathbf{y}^T\mathbf{A}\tilde{\mathbf{x}} \leq \mathbf{0}$$

□

*Proof (backward direction).* Suppose $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq 0$ is infeasible, meaning $\mathbf{b} \notin \text{cone}(\mathbf{A})$. Since cone($\mathbf{A}$) is a nonempty, closed, convex set, and $\mathbf{b} \notin \text{cone}(A)$, by the separating hyperplane theorem, there exists $\mathbf{p} \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$ such that:

$$\mathbf{p}^T\mathbf{b} > \beta \quad \mathbf{p}^T\mathbf{z} \leq \beta \qquad \forall \mathbf{z} \in \text{cone}(\mathbf{A})$$

Since $\mathbf{0} \in \text{cone}(\mathbf{A})$, it follows that $\beta \geq 0$, implying $\mathbf{p}^T\mathbf{b} > 0$. Furthermore, since $\mathbf{z} \in \text{cone}(\mathbf{A})$, we get:

$$\mathbf{p}^T\mathbf{A}\mathbf{x} \leq \beta \qquad \forall \mathbf{x} \geq 0$$

Since $\mathbf{x} \geq \mathbf{0}$, this implies $\mathbf{p}^T\mathbf{A} \leq 0$. Thus, we have found $\mathbf{y} = \mathbf{p}$ satisfying the alternative system, proving the lemma. □

## 2.3 Convex functions

**Definition** (*Convex function*)**.** A function $f : C \to \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is convex if:

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \qquad \forall \mathbf{x}_1, \mathbf{x}_2 \in C \quad \forall \alpha \in [0,1]$$

**Definition** (*Strictly proper function*)**.** A function $f$ is strictly convex if the above inequality is strict for all distinct points $\mathbf{x}_1, \mathbf{x}_2 \in C$ and for any $\alpha \in (0,1)$:

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) < \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2)$$

**Definition** (*Concave function*)**.** A function $f$ is concave if $-f$ is convex, meaning:

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \geq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \qquad \forall \mathbf{x}_1, \mathbf{x}_2 \in C \quad \forall \alpha \in [0,1]$$

**Definition** (*Linear function*)**.** A function $f$ is linear if it is both convex and concave, which means:

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) = \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2)$$

**Definition** (*Epigraph*)**.** The epigraph of $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$, denoted as $\text{epi}(f)$, is the set:

$$\text{epi}(f) = \{(\mathbf{x}, y) \in S \times \mathbb{R} \mid f(\mathbf{x}) \leq y\}$$

It consists of all points lying on or above the graph of $f$.

**Definition** (*Domain*)**.** The domain of a convex function $f : C \to \mathbb{R}$ is given by:

$$\text{dom}(f) = \{\mathbf{x} \in C \mid f(\mathbf{x}) < +\infty\}$$

### 2.3.1 Properties

Let $C \subseteq \mathbb{R}^n$ be a nonempty convex set, and let $f : C \to \mathbb{R}$ be a convex function.

**Property 2.3.1.** For each $\beta \in \mathbb{R}$ (also $\beta \in +\infty$), the level sets

$$L_\beta = \{\mathbf{x} \in C \mid f(\mathbf{x}) \leq \beta\} \qquad L_\beta^+ = \{\mathbf{x} \in C \mid f(\mathbf{x}) < \beta\}$$

Are convex subsets of $\mathbb{R}^n$.

**Property 2.3.2.** A convex function $f$ is always continuous in the relative interior of its domain (relative to the affine hull of $C$).

**Property 2.3.3.** A function $f$ s convex if and only if its epigraph $\text{epi}(f)$ is a convex subset of $\mathbb{R}^{n+1}$.

### 2.3.2 Convex optimization

Consider the optimization problem:

$$\min_{\mathbf{x}\in C} f(\mathbf{x})$$

Here, $C \subseteq \mathbb{R}^n$ is convex and $f : C \to \mathbb{R}$ is a convex function.

**Proposition.** If $C$ and $f$ are convex, then any local minimum of $f$ on $C$ is also a global minimum.

*Proof.* Suppose $\mathbf{x}'$ is a local minimum. Assume for contradiction that there exists another point $\mathbf{x}^* \in C$ such that $f(\mathbf{x}^*) < f(\mathbf{x}')$. Since $f$ convex, for any $\alpha \in (0,1)$:

$$f(\alpha\mathbf{x}' + (1-\alpha)\mathbf{x}^*) \leq \alpha f(\mathbf{x}') + (1-\alpha)f(\mathbf{x}^*) < f(\mathbf{x}')$$

This contradicts the assumption that $\mathbf{x}'$ is a local minimum. Hence, any local minimum must also be a global minimum. □

**Proposition.** If $f$ is strictly convex on $C$, then there is at most one global minimum (provided $f$ is bounded below).

*Proof.* Suppose $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ are two distinct global minima. Since $C$ is convex, their midpoint belongs to $C$:

$$\frac{1}{2}\left(\mathbf{x}_1^* + \mathbf{x}_2^*\right) \in C$$

By strict convexity of $f$:

$$f\left(\frac{1}{2}\left(\mathbf{x}_1^* + \mathbf{x}_2^*\right)\right) < \frac{1}{2} f\left(\mathbf{x}_1^*\right) + \frac{1}{2} f\left(\mathbf{x}_2^*\right)$$

But both $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ cannot be global minima. □

**Proposition.** For a linear programming problem of the form:

$$P = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\} \neq \varnothing$$

Either there exists at least one optimal extreme point, or the objective function value is unbounded below over $P$.

### 2.3.3 Gradient-based characterization

**Proposition.** Let $f : C \to \mathbb{R}$ be a continuously differentiable function ($C^1$) on a nonempty, convex, and open set $C \subseteq \mathbb{R}^n$. Then, $f$ is convex if and only if:

$$f(\mathbf{x}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x} - \overline{\mathbf{x}}) \qquad \forall \mathbf{x}, \overline{\mathbf{x}} \in C \quad \mathbf{x} \neq \overline{\mathbf{x}}$$

This condition states that the function always lies above its first-order Taylor approximation, implying convexity.

**Proposition.** Let $f : C \to \mathbb{R}$ be a continuously differentiable function ($C^1$) on a nonempty, convex, and open set $C \subseteq \mathbb{R}^n$. Then, $f$ is strictly convex if and only if

$$f(\mathbf{x}) > f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x} - \overline{\mathbf{x}}) \qquad \forall \mathbf{x}, \overline{\mathbf{x}} \in C \quad \mathbf{x} \neq \overline{\mathbf{x}}$$

### 2.3.4 Hessian-based characterization

**Proposition.** Let $f : C \to \mathbb{R}$ be a twice continuously differentiable function ($C^2$) on a nonempty, convex, and open set $C \subseteq \mathbb{R}^n$. Then, $f$ is convex if and only if its Hessian matrix is positive semidefinite for all $\mathbf{x} \in C$:

$$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right) \succeq \mathbf{0} \qquad \forall \mathbf{x} \in C$$

This means that for all $\mathbf{y} \in \mathbb{R}^n$:

$$\mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} \geq \mathbf{0}$$

**Proposition.** Let $f : C \to \mathbb{R}$ be a twice continuously differentiable function ($C^2$) on a nonempty, convex, and open set $C \subseteq \mathbb{R}^n$. Then, $f$ is convex if and only if its Hessian matrix is positive definite for all $\mathbf{x} \in C$:

$$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right) \succ \mathbf{0} \qquad \forall \mathbf{x} \in C$$

This means that for all $\mathbf{y} \in \mathbb{R}^n$:

$$\mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} > \mathbf{0} \qquad \forall \mathbf{y} \neq \mathbf{0}$$

This is a sufficient condition for strict convexity but not necessary.

### 2.3.5 Subgradient of convex functions

n general, convex (or concave) functions are not necessarily differentiable everywhere. To handle nondifferentiable points, we extend the concept of the gradient for $C^1$ functions to piecewise $C^1$ functions through the notion of subgradients.

**Definition** (*Subgradient*)**.** Let $C \subseteq \mathbb{R}^n$ and $f : C \to \mathbb{R}$ be a convex function. A vector $\boldsymbol{\gamma} \in \mathbb{R}^n$ is called a subgradient of $f$ at $\mathbf{x} \in C$ if:

$$f(\mathbf{x}) \geq f(\mathbf{x}) + \boldsymbol{\gamma}^T(\mathbf{x} - \overline{\mathbf{x}}) \qquad \forall \mathbf{x} \in C$$

**Definition** (*Subdifferential*)**.** The subdifferential 0f $f$ at $\underline{\mathbf{x}}$, denoted as $\partial f(\mathbf{x})$, is the set of all subgradients of $f$ at $\mathbf{x}$:

$$\partial f(\mathbf{x}) = \{\boldsymbol{\gamma} \in \mathbb{R}^n \mid f(\mathbf{x}) \geq \boldsymbol{\gamma}^T(\mathbf{x} - \overline{\mathbf{x}}) \quad \forall \mathbf{x} \in C\}$$

If $f$ is differentiable at $\overline{\mathbf{x}}$, then $\partial f(\overline{\mathbf{x}})$ contains only one element: the usual gradient $\nabla f(\overline{\mathbf{x}})$. If $f$ is not differentiable, then $\partial f(\overline{\mathbf{x}})$ is a set of possible slopes that support $f$ at $\overline{\mathbf{x}}$.

Let $C \subseteq \mathbb{R}^n$ be convex, and let $f : C \to \mathbb{R}$ be a convex function.

**Property 2.3.4.** For every interior point $\overline{\mathbf{x}} \in \mathrm{int}(C)$, there exists at least one subgradient. In particular, there exists a vector $\boldsymbol{\gamma} \in \mathbb{R}^n$ such that the hyperplane:

$$H = \{(\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid y = f(\overline{\mathbf{x}}) + \boldsymbol{\gamma}^T(\mathbf{x} - \overline{\mathbf{x}})\}$$

is a supporting hyperplane of the epigraph of $f$ at $(\overline{\mathbf{x}}, f(\overline{\mathbf{x}}))$.

The existence of at least one subgradient at every interior point of $C$ is a necessary and sufficient condition for $f$ to be convex on $\text{int}(C)$.

**Property 2.3.5.** For every $\mathbf{x} \in C$, the subdifferential $\partial f(\mathbf{x})$ is a nonempty, convex, closed and bounded set.

**Property 2.3.6.** A point $\mathbf{x}^*$ is a global minimum of $f$ on $C$ if and only if:

$$\mathbf{0} \in \partial f(\mathbf{x}^*)$$

That is, the zero vector must be included in the subdifferential at $\mathbf{x}^*$.