# 1 Outline

In this lecture, we consider

- Operations preserving convexity,
- Optimization terminologies,
- Introduction to convex optimization,
- Applications (Portfolio optimization, Uncertainty quantification, Support vector machine)

# 2 Operations preserving convexity

For many problems, it is important to recognize underlying convex structures. We can determine whether certain sets and functions are convex by understanding basic rules. Moreover, based on these rules, we can build complex convex sets and functions from simpler ones.

We first consider set operations that preserve convexity.

- Intersection: The intersection of any collection, possibly infinite, of covex sets is convex.
- Scaling: Given a convex set $C$ and $\alpha \in \mathbb{R}$,
$$\alpha C = \{\alpha x : \ x \in C\}.$$
- Minkowski sum: Given convex sets $C_i \subseteq \mathbb{R}^d$ for $i = 1, \ldots, k$, the sum of them, given by
$$C_1 + \cdots + C_k = \{x^1 + \cdots + x^k : \ x^i \in C_i \text{ for } i = 1, \ldots, k\}.$$
- Product: Given convex sets $C_i \subseteq \mathbb{R}^{d_i}$ for $i = 1, \ldots, k$, the product of them, given by
$$C_1 \times \cdots \times C_k = \{(x^1, \ldots, x^k) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} : \ x^i \in C_i \text{ for } i = 1, \ldots, k\}$$
is convex.
- Affine image: Given a convex set $C$ and matrices $A \in \mathbb{R}^{p\times d}$, $b \in \mathbb{R}^p$, we define an affine mapping $f(x) = Ax + b : \mathbb{R}^d \to \mathbb{R}^p$. Then
$$f(C) = \{Ax + b : \ x \in C\}.$$
- Inverse affine image: Given a convex set $C$ and matrices $A \in \mathbb{R}^{p\times d}$, $b \in \mathbb{R}^p$, we define an affine mapping $f(x) = Ax + b : \mathbb{R}^d \to \mathbb{R}^p$. Then
$$f^{-1}(C) = \{x : \ Ax + b \in C\}.$$

We next consider function operations preserving convexity.

- Nonnegative weighted sum: Let $f_1, \ldots, f_k : \mathbb{R}^d \to \mathbb{R}$ be convex functions. Then for any $\alpha_1, \ldots, \alpha_k \geq 0$,
$$\alpha_1 f_1 + \cdots + \alpha_k f_k$$
is convex.

- Maximum of arbitrary collection of convex functions: Let $\{f_\gamma\}_{\gamma \in \Gamma}$ be a collection of convex functions. Then $\max_{\gamma \in \Gamma} f_\gamma$ is also convex. Here, $\Gamma$ may be infinite.

- Affine composition: Let $g : \mathbb{R}^p \to \mathbb{R}$ be a convex function, and take matrices $A \in \mathbb{R}^{p \times d}$, $b \in$ ı. Then $f : \mathbb{R}^d \to \mathbb{R}$ defined by $f(x) = g(Ax + b)$ is convex.

- Minimizing out variables: Let $g(x, y)$ be convex function in $(x, y)$. Define $f$ by $f(x) = \inf_{y \in C} g(x, y)$ for some convex set $C$. Then $f$ is convex.

- Perspective function: Let $g(x)$ be a convex function. Then $f(x, t) = tg(x/t)$ is a convex function in $(x, t) \in \mathbb{R}^d \times \mathbb{R}_{++}$. Here, $f$ is called the perspective of $g$.

**Example 4.1.** Let $C$ be an arbitrary set of locations. Note that

$$f_1(x) = \max_{y \in C} ||x - y||$$

measures the longest distance from $x$ to a location in $C$, and

$$f_2(x) = \min_{y \in C} \|x - y\|$$

measures the shortest distance from $x$ to a location in $C$. Note that $g(x, y) = \|x - y\|$ is convex in $x$ and $y$. Then $f_1$ is convex as it is the pointwise maximum of some convex functions. Furthermore, if $C$ is convex, then $f_2$ is convex because it is a partial minimization of a convex function.

# 3 The problem of optimization

Given a function $f : \mathbb{R}^d \to \mathbb{R}$ and a set $C \subseteq \mathbb{R}^d$, we want to solve

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in C. \end{array} \tag{$P$}$$

Terminology 1:

- $x$ is called the *decision vector*, and the components of $x$ are called the *decision variables.* For example, decision variables capture how much to invest for a financial portfolio or where to build a hospital in a village.

- $f$ is called the *objective function* or *cost function.* For example, the cost of a production plan.

- $C$ is called the *domain*, *feasible region*, or *constraint set.* For example, production capacities, budget constraints.

Terminology 2:

- Any vector $x \in C$ is called a *feasible solution*

- We say that $(P)$ is *feasible* if $C \neq \emptyset$. Otherwise, $(P)$ is *infeasible*.
- If there exists $x \in C$ such that $f(x) \leq r$ for any $r \in \mathbb{R}$, then $(P)$ is *unbounded*.
- If there exists some $r \in \mathbb{R}$ such that $f(x) \geq r$ for all $x \in C$, then $(P)$ is *bounded*.

**Example 4.2.** When $C = \{(x_1, x_2) \in \mathbb{R}^2 : \ x_1^2 + x_2^2 \leq 1, \ x_1 + x_2 \geq 2\}$, then the problem is infeasible. When $f(x) = (x-2)^2$ and $C = [-1, 5]$, then the problem is feasible and bounded.

Terminology 3:

- $\mathrm{OPT} := \min_{x \in C} f(x)$ is the *optimal value* of the optimization problem. Then

$$\mathrm{OPT} = \begin{cases} +\infty, & \text{if infeasible,} \\ -\infty, & \text{if feasible but unbounded,} \\ \text{finite,} & \text{if feasible and bounded.} \end{cases}$$

- A solution $x^* \in C$ such that $f(x^*) = \mathrm{OPT}$ is called an *optimal solution*.
- We say that $(P)$ is *solvable* if an optimal solution exists. If not, $(P)$ is *unsolvable*.

**Example 4.3.** When $f(x) = (x-2)^2$ and $C = (3, 5]$, the problem is feasible and bounded but unsolvable.

# 4 Convex optimization problem

When the objective function $f$ is a convex function and the feasible region is a convex set, then the optimization problem $(P)$ is referred to as a *convex optimization* or *convex minimization* problem. By using the indicator function for $C$, we can rewrite $(P)$ as

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & I_C(x) \leq 0 \end{aligned}$$

or

$$\min \quad f(x) + I_C(x).$$

Here, $f$, $I_C$, and $f + I_C$ are all convex. The standard form of a convex optimization problem is

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \dots, p, \\ & h_i(x) = 0, \quad i = 1, \dots, q \end{aligned} \tag{$P'$}$$

where

- the objective function $f$ is convex,
- the inequality constraint functions $g_1, \dots, g_p$ are convex, and
- the equality constraint functions $h_1, \dots, h_q$ are convex.

**Exercise 4.4.** If $g_1, \dots, g_p$ and $h_1, \dots, h_q$ are convex functions, $C := \{x \in \mathbb{R}^d : \ g_i(x) \leq 0 \text{ for } i = 1, \dots, p, \ h_i(x) = 0 \text{ for } j = 1, \dots, q\}$ is a convex set.

Note that

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

and $-f$ is concave when $f$ is convex. Therefore, the problem of maximizing a concave function over a convex domain is also a convex optimization problem.

### 4.1 Portfolio optimization

Given $d$ financial assets (stocks, bonds, etc), we want to allocate $x_i$ fraction of our budget to asset $i \in [d]$, i.e., $1^\top x = 1$. Here, $x_i < 0$ indicates a *short position*, which means borrowing shares, selling now, and returning the shares later, while $x_i \geq 0$ indicates a *long position*, buying shares now. Then $\|x\|_1 = \sum_{i=1}^d |x_i|$ means *leverage*.

Let $p_i$ be the initial price of asset $i$, and $p_i'$ be its price at the end of one period. Then the return of asset $i$ can be defined as $r_i := (p_i' - p_i)/p_i$. Moreover, the return of my portfolio can be measured by $r^\top x$. Here, $r$ is a random variable with mean $\mu$ and covariance $\Sigma$. Here, the variance term $x^\top \Sigma x$ is often used to measure the *risk* of my portfolio. By definition, the covariance matrix is positive semidefinite.

We want to find a portfolio that maximizes the expected return while guaranteeing a low risk. Then we consider

$$\begin{array}{ll} \text{maximize} & \mu^\top x - \gamma x^\top \Sigma x \\ \text{subject to} & 1^\top x = 1, \\ & x \in C' \end{array}$$

where $\gamma > 0$ is the risk aversion parameter. When $C' = \mathbb{R}^d_+$, we take long positions only. When $C' = \{x \in \mathbb{R}^d : \|x\|_1 \leq B\}$, then we allow short positions but there is a leverage limit.

### 4.2 Uncertainty quantification

Suppose we have chosen a portfolio $x$. To avoid high risk portfolios, can we measure the worst-case variance of the given portfolio? In practice, the covariance matrix $\Sigma$ is estimated through data, so it is subject to errors. Given a magnitude $\epsilon$ of potential errors, what is the highest risk of the portfolio?

$$\begin{array}{ll} \text{maximize} & x^\top (\Sigma + S) x \\ \text{subject to} & S \succeq 0, \\ & \|S\|_{\text{nuc}} \leq \epsilon \end{array}$$

where $\|S\|_{\text{nuc}}$ denotes the *nuclear norm* of $S$, defined as the sum of all eigenvalues of $S$. Is this problem convex?

### 4.3 Support vector machine

Given $n$ data $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i \in \{-1, 1\}$ are labels, we want to find a separating hyperplane

$$w^\top x = b$$

to classify data with $+1$ and data with $-1$. The goal is to find a separating hyperplane $w^\top x = b$ with the "gap" $(1/\|w\|_2)$ being maximized. Then the problem can be formulated as

$$\begin{array}{ll} \text{minimize} & \|w\|_2 \\ \text{subject to} & y_i(w^\top x_i - b) \geq 0, \ i = 1, \ldots, n. \end{array}$$

If this problem is feasible, then $x \to \text{sign}(w^\top x_i - b)$ is a valid classifier for the data set.

What if the data set is not entirely separable? What if no hyperplane separates the data without an error? In such cases, we force separation via a penalty term, instead of imposing hard constraints. The number of misclassifications can be used as penalty. Namely,

$$\sum_{i=1}^{n} 1(y_i \neq \mathrm{sign}(w^\top x_i - b)).$$

However, this is not convex. Instead, we apply the *hinge loss*[1], which is an upper bound on the number of misclassifications, given by

$$\sum_{i=1}^{n} \max\{0,\ 1 - y_i(w^\top x_i - b)\}.$$

Then we solve

$$\min_{w,b} \quad \lambda \|w\|_2^2 + \frac{1}{n}\sum_{i=1}^{n} \max\{0,\ 1 - y_i(w^\top x_i - b)\}$$

where $\lambda$ determines the trade-off between the margin size and the penalty.

[1]Here, $\max\{0, a\}$ is called the hinge function.