In this lecture, we consider

- Operations preserving convexity,
- Introduction to convex optimization,
- Applications of convex optimization,
- Introduction to linear programming.

1 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.

1.1 Set operations

We first consider set operations that preserve convexity.

- Intersection: The intersection of any (possibly infinite) collection 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 $i = 1, \dots, k$, the Minkowski sum of them, defined by

$$C_1 + \cdots + C_k = \{x^1 + \cdots + x^k : \ x^i \in C_i \text{ for } i = 1, \dots, k\}$$

  is convex.

- Cartesian Product: Given convex sets $C_i \subseteq \mathbb{R}^{d_i}$ for $i = 1, \dots, k$, the Cartesian product of them, defined by

$$C_1 \times \cdots \times C_k = \{(x^1, \dots, x^k) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} : \ x^i \in C_i \text{ for } i = 1, \dots, k\}$$

  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

$$f^{-1}(C) = \{x : \ Ax + b \in C\}.$$

1.2 Function operations

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, Γ may be infinite.
- 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.
- Affine composition: Let $g : \mathbb{R}^p \to \mathbb{R}$ be a convex function, and take matrices $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$. Then $f : \mathbb{R}^d \to \mathbb{R}$ defined by $f(x) = g(Ax + b)$ is convex.
- Compositions: Let $h : \mathbb{R} \to \mathbb{R}$ be a univariate non-decreasing convex function, and let $g : \mathbb{R}^d \to \mathbb{R}$ be convex. Then $f = h \circ g$ is convex.

Example 2.1. 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. Let us show that both f_1 and f_2 are convex. Observe first 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. In summary, f_1 is convex regardless of whether C is convex or not, while f_2 is convex if the set C is convex.

2 Convex optimization problem

2.1 Basic optimization terminologies

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

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

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 2.2. 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:

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

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

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

Example 2.3. When $f(x) = (x-2)^2$ and $C = (3, 5]$, the problem is feasible and bounded but unsolvable.

2.2 Convex optimization

When the objective function f is convex 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

$$\text{minimize} \quad f(x) + I_C(x).$$

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

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

where

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

Exercise 2.4. If $g_1, \ldots, g_p$ are convex and $h_1, \ldots, h_q$ are affine functions, $C := \{x \in \mathbb{R}^d : \ g_i(x) \leq 0 \text{ for } i = 1, \ldots, p, \ h_i(x) = 0 \text{ for } j = 1, \ldots, q\}$ is a convex set.

Note that

$$\min_{x \in C} \quad f(x) \quad = \quad (-1) \times \max_{x \in C} \quad -f(x)$$

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.

3 Convex optimization applications

3.1 Portfolio optimization

Given d financial assets (stocks, bonds, etc), we want to allocate x_i fraction of our budget to asset $i \in [d]$. Hence, we impose condition

$$1^\top x = 1.$$

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 μ and covariance Σ. Then it follows that

$$\mathbb{E}\left[r^\top x\right] = \mu^\top x$$
$$\text{Var}\left[r^\top x\right] = x^\top \Sigma x$$

In words, the expected return, which is the expectation of the return $r^\top x$, is given by $\mu^\top x$. Moreover, the *risk* of my portfolio, which is often defined by the variance of the return $r^\top x$, is given by $x^\top \Sigma x$. 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

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

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.

Let us observe that the formulation above is a convex optimization problem. First, the feasible region is the intersection of a hyperplane

$$\{x \in \mathbb{R}^d : \ 1^\top x = 1\}$$

and a convex set C'. Therefore, the feasible region is a convex set. Moreover, it is known that any covariance matrix is positive semidefinite. This in turn implies that the objective function $\mu^\top x - \gamma x^\top \Sigma x$ is a concave function. Moreover, observe that

$$\begin{aligned}&\max \left\{ \mu^\top x - \gamma x^\top \Sigma x : \ 1^\top x = 1, \ x \in C' \right\} \\ &= -\min \left\{ -\mu^\top x + \gamma x^\top \Sigma x : \ 1^\top x = 1, \ x \in C' \right\}.\end{aligned}$$

As $\mu^\top x - \gamma x^\top \Sigma x$ is concave, it follows that the function $-\mu^\top x + \gamma x^\top \Sigma x$ is convex. Thus, the minimization problem is indeed a convex optimization problem. From now on, we may skip this process and simply say that concave maximization is equivalent to convex minimization.

3.2 Uncertainty quantification

Recall that given a portfolio x and the covariance matrix Σ of financial assets, the quadratic term

$$x^\top \Sigma x$$

models the risk of the portfolio x. However, in practice, it is difficult to predict the exact value of Σ. Instead, we estimate it and deduce an empirical estimation of it. Let us denote it as $\bar{\Sigma}$. Then

$$x^\top \bar{\Sigma} x$$

is an estimation of the actual risk term $x^\top \Sigma x$. Although $x^\top \bar{\Sigma} x$ may provide a proxy for the risk, underestimation of risk could result in a critical situation. Then the question is, given the empirical estimate of the risk term $x^\top \bar{\Sigma} x$, what would be the worst-case risk value under estimation noise?

How do we consider possible estimation noise? One common way to use some statistics tool such as

$$\left\| \underbrace{\Sigma}_{\text{true covariance}} - \underbrace{\bar{\Sigma}}_{\text{empirical covariance}} \right\|_{\text{nuc}} \leq \epsilon$$

which holds for some ϵ (with high probability) where ϵ depends on the number of data to obtaine the estimate $\bar{\Sigma}$. Here, $\|\cdot\|_{\text{nuc}}$ denotes what is called the *nuclear norm* over matrices. Then we may draw a ball around the empirical covariance matrix with respect to the nuclear norm. Formally, we consider

$$\begin{aligned}&\left\{ A \in \mathbb{R}^{d \times d} : \ \left\| A - \bar{\Sigma} \right\|_{\text{nuc}} \leq \epsilon \right\} \\ &= \left\{ \bar{\Sigma} + S \in \mathbb{R}^{d \times d} : \ \|S\|_{\text{nuc}} \leq \epsilon \right\}.\end{aligned}$$

Then it follows from the above statistical result that the true covariance matrix Σ belongs to the ball (with high probability). Then now we consider

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

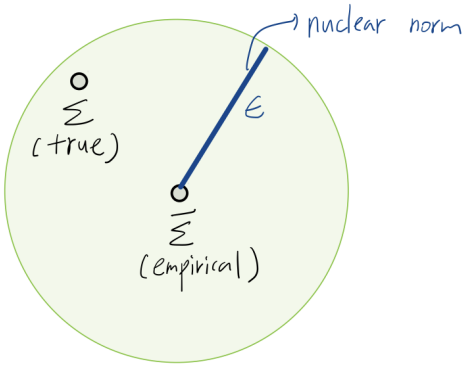

Figure 2.1: Nuclear-norm-ball around the empirical covariance matrix

Note that $\Sigma = \bar{\Sigma} + S$ for some matrix S with $\|S\|_{\text{nuc}} \leq \epsilon$. Moreover, we know that Σ is positive semidefinite. Therefore,

$$x^\top \Sigma x \leq \max \left\{ x^\top (\bar{\Sigma} + S) x : \ \bar{\Sigma} + S \succeq 0, \ \|S\|_{\text{nuc}} \leq \epsilon \right\}.$$

Hence, the optimum value provides an upper bound on the true risk value of portfolio x.

3.3 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. Here, the gap means the Euclidean distance between two consecutive hyperplanes

$$\{x \in \mathbb{R}^d : \ w^\top x = b\}, \quad \{x \in \mathbb{R}^d : \ w^\top x = b + 1\}.$$

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,

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

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

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

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

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 λ determines the trade-off between the margin size and the penalty.

3.4 LASSO (least absolute shrinkage and selection operator)

Based on n data points $(x_1, y_1), \ldots, (x_n, y_n)$, we want to find a linear rule

$$y = \beta^\top x$$

that best represents the relationship between x and y. The goal is to find β minimizing the "mean squared error", given by

$$\min_{\beta} \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta^\top x_i)^2 \quad = \quad \min_{\beta} \quad \frac{1}{n}\|y - X\beta\|_2^2$$

where the rows of X are $x_1^\top, \ldots, x_n^\top$.

However, there are issues such as highly collinear covariates and overfitting. Motivated by this, LASSO is a regression method that achieves variable selection and regularization. The LASSO problem is to solve

$$\begin{aligned} \text{minimize} \quad & \frac{1}{n}\|y - X\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \le t \end{aligned}$$

where t is a parameter determining the degree of regularization. Basically, the problem induces sparsity in β. The problem is often transformed into the *Lagrangian* form, given as follows.

$$\min_{\beta} \quad \frac{1}{n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

where λ is set to control the degree of regularization.

3.5 Facility location

Given the locations of n households $x_1, \ldots, x_n \in \mathbb{R}^d$, we wish to build a hospital covering the households. A desired location would minimize the longest distance to a household. The problem can be formulated as

$$\min_{x} \quad \max_{i=1,\ldots,n} \|x - x_i\|.$$