Lecturer: Dabeen Lee

1 Outline

In this lecture, we study

- convex optimization,
- classes of convex optimization problems.

2 Convex optimization

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

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

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

Example 3.2. When $f(x) = (x-2)^2$ and $C = (3,5]$, the problem is feasible and bounded but unsolvable.

2.2 Formulation of convex optimization problems

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 3.3. 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 Classes of convex optimization problems

There are many interesting "classes" of convex optimization problems. In this lecture, we consider the following five classes of convex optimization problems.

- Linear programming (LP),
- Quadratic programming (QP),
- Semidefinite programming (SDP),

- Conic programming,
- Second-order cone programming (SOCP).

In fact, these problems are closely related, as the problem classes form a hierarchy described in Figure 3.1.

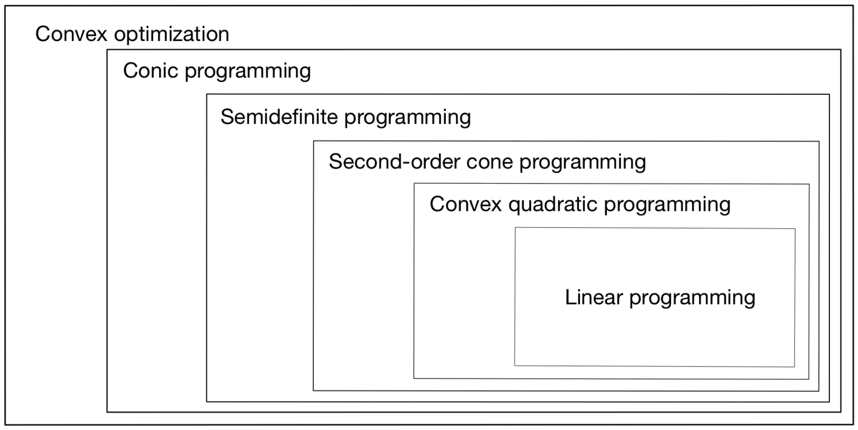

Figure 3.1: Hierarchy of classes of convex optimization problems

3.1 Linear programming

A linear program (LP) is an optimization problem of the following form.

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax \geq b \end{array} \tag{P}$$

where $c^\top x$ is the linear objective function and $Ax \geq b$ is the system of linear constraints. Note that the feasible region $P = \{x \in \mathbb{R}^d : Ax \geq b\}$ is a polyhedron, so the linear program is the problem of finding a point in a polyhedron that minimizes a linear function. Figure 3.2 describes a linear program with 2 variables.

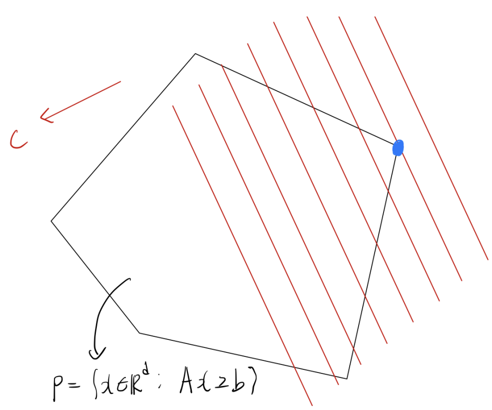

Figure 3.2: Geometric picture of a linear program

Example 3.4. Recall that the facility location problem can be modeled as follows:

$$\text{minimize} \quad \max_{i=1,\ldots,n} \|x - x^i\|_1$$

where $x^1, \ldots, x^n \in \mathbb{R}^d$ are the locations of households and we use the ℓ_1 norm for the norm $\|\cdot\|$. In fact, the problem is equivalent to a linear program. Introducing an auxiliary variable t, the problem is equivalent to

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & t \geq \max_{i=1,\ldots,n} \|x - x^i\|_1. \end{aligned}$$

Basically, minimizing t forces minimizing the original objective $\max_{i=1,\ldots,n} \|x - x^i\|_1$. Moreover, $t \geq \max_{i=1,\ldots,n} \|x - x^i\|_1$ is equivalent to imposing $t \geq \|x - x^i\|_1$ for $i = 1, \ldots, n$. The next step is to replace $t \geq \|x - x^i\|_1$ by a set of linear inequalities. Note that

$$\|x - x^i\|_1 = \sum_{j=1}^{d} |x_j - x_j^i|$$

where $x = (x_1, \ldots, x_d)^\top$ and $x^i = (x_1^i, \ldots, x_d^i)^\top$. By introducing an auxiliary variable s_{ij} for each absolute value term $|x_j - x_j^i|$, we replace $t \geq \sum_{j=1}^{d} |x_j - x_j^i|$ by $t \geq \sum_{j=1}^{d} s_{ij}$ and $s_{ij} \geq |x_j - x_j^i|$. Moreover, $s_{ij} \geq |x_j - x_j^i|$ is equivalent to $s_{ij} \geq x_j - x_j^i \geq -s_{ij}$. Therefore, we obtain

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & t \geq \sum_{j=1}^{d} s_{ij} & & \text{for } i = 1, \ldots, n, \\ & s_{ij} \geq x_j - x_j^i \geq -s_{ij} & & \text{for } i = 1, \ldots, n,\ j = 1, \ldots, d \end{aligned}$$

which is a linear program.

The dual linear program of (P) is given by

$$\begin{aligned} \text{maximize} \quad & b^\top y \\ \text{subject to} \quad & A^\top y = c, \\ & y \geq 0. \end{aligned} \tag{D}$$

Weak LP duality states that $\mathrm{OPT}(P) \geq \mathrm{OPT}(D)$, where $\mathrm{OPT}(P)$ and $\mathrm{OPT}(D)$ denote the optimal values of (P) and (D), respectively. *Strong LP duality* states that if (P) is feasible and bounded, then (D) is feasible and bounded and $\mathrm{OPT}(P) = \mathrm{OPT}(D)$.

3.2 Quadratic programming

A quadratic program (QP) is an optimization problem of the following form.

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^\top Q x + p^\top x \\ \text{subject to} \quad & Ax \geq b \end{aligned} \tag{QP}$$

The quadratic program is convex only if Q is positive semidefinite.

Example 3.5. Let us consider support vector machines. 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

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

The number of misclassifications can be used as penalty. Namely,

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

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

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

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

Here, $\|w\|_2^2 = w^\top w = w^\top I w$ where I is the identity matrix, and therefore, $\|w\|_2^2$ is a convex quadratic function. Moreover, the max terms in the objective can be replaced by adding some auxiliary variables. Note that the formulation is equivalent to

$$\begin{aligned} \text{minimize} \quad & \lambda w^\top w + \frac{1}{n} \sum_{i=1}^{n} t_i \\ \text{subject to} \quad & t_i \geq \max\{0, \ 1 - y_i(w^\top x_i - b)\} \text{ for } i = 1, \ldots, n. \end{aligned}$$

Next, we can rewrite the constraints as linear constraints as the following.

$$\begin{aligned} \text{minimize} \quad & \lambda w^\top w + \frac{1}{n} \sum_{i=1}^{n} t_i \\ \text{subject to} \quad & t_i \geq 1 - y_i(w^\top x_i - b) \quad \text{for } i = 1, \ldots, n, \\ & t_i \geq 0 \qquad\qquad\qquad\quad \text{for } i = 1, \ldots, n. \end{aligned}$$

Therefore, it is a convex quadratic program with a quadratic objective and linear constraints.

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

3.3 Semidefinite programming

A *semidefinite program* is an optimization problem of the following form. Let C and $A_1, \ldots, A_m$ be $d \times d$ matrices, and we have

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}(C^\top X) \\ \text{subject to} \quad & \operatorname{tr}(A_\ell^\top X) = b_\ell \text{ for } \ell = 1, \ldots, m \\ & X \succeq 0 \end{aligned} \tag{SDP}$$

where

$$\operatorname{tr}(C^\top X) = \sum_{i=1}^{d} \sum_{j=1}^{d} C_{ij} X_{ij} \quad \text{and} \quad \operatorname{tr}(A_\ell^\top X) = \sum_{i=1}^{d} \sum_{j=1}^{d} (A_\ell)_{ij} X_{ij}.$$

Here, if we view matrix X as a $(d \times d)$-dimensional vector, then the objective and the equality constraints are "linear" in X. Hence, (SDP) is analogous to linear programming. Recall that we defined the linear programming (LP) dual of a given linear program. Likewise, we may define the notion of semidefinite programming (SDP) dual. The *dual* of (SDP) is

$$\begin{aligned} \text{maximize} \quad & \sum_{\ell=1}^{m} b_\ell y_\ell \\ \text{subject to} \quad & \sum_{\ell=1}^{m} y_\ell A_\ell \preceq C \end{aligned} \tag{dual-SDP}$$

where $\sum_{\ell=1}^{m} y_\ell A_\ell \preceq C$ means $C - \sum_{\ell=1}^{m} y_\ell A_\ell$ is positive semidefinite. If an optimization is in either form, we say that it is a semidefinite program.

We will study more about duality later in this course. We have dicussed LP duality, and in particular, we covered how to derive the dual of a linear program and learned duality theorems. The notion of duality extends to more general classes of convex programming problems. We will learn how to derive the dual of a given optimization problem, and we will define the associated weak and strong duality statements.

3.4 Conic programming

The linear program (P) can be rewritten as follows:

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax - b \in \mathbb{R}^n_+. \end{aligned}$$

Let us take a closer look at the nonnegative orthant $\mathbb{R}^n_+$. It satisfies the following properties.

1. $\mathbb{R}^n_+$ is a convex cone.
2. $\mathbb{R}^n_+$ is *pointed*, which means that if $v \in \mathbb{R}^n_+$ and $-v \in \mathbb{R}^n_+$, then it must be that $v = 0$.

In fact, $\mathbb{R}^n_+$ is not just a pointed convex cone. There are other important properties of $\mathbb{R}^n_+$.

3. $\mathbb{R}^n_+$ is *closed*, which means that for any convergent sequence $\{v^n\}_{n \in \mathbb{N}}$ contained in $\mathbb{R}^n_+$, its limit $\lim_{n \to \infty} v^n$ also belongs to $\mathbb{R}^n_+$.

4. $\mathbb{R}^n_+$ has a nonempty *interior*. Equivalently, $\mathbb{R}^n_+$ contains an *interior point*. A vector v is an interior point of a set K if there exists an open ball around v which is fully contained in K. Then the interior of a set K, denoted $\text{int}(K)$, is defined as the set of all its interior points. The interior of $\mathbb{R}^n_+$ is $\mathbb{R}^n_{++}$, the positive orthant.

In summary, the nonnegative orthant $\mathbb{R}^n_+$ is a pointed and closed convex cone with a nonempty interior. In fact, there are other closed convex cones that are pointed and have a nonempty interior. For example,

- The *Lorentz cone*.

$$\{(x_1, \ldots, x_{n-1}, x_n)^\top \in \mathbb{R}^n : \ \|(x_1, \ldots, x_{n-1})^\top\|_2 \leq x_n\}.$$

  Other equivalent names include the *second-order cone*, the *ice-cream cone*, and the *ℓ_2-norm cone*. Its interior is given by

$$\{(x_1, \ldots, x_{n-1}, x_n)^\top \in \mathbb{R}^n : \ \|(x_1, \ldots, x_{n-1})^\top\|_2 < x_n\}.$$

- The positive semidefinite cone.

$$\{S \in \mathbb{S}^d : \ x^\top S x \geq 0 \quad \text{for all } x \in \mathbb{R}^d\}.$$

  Its interior is the positive definite cone, the set of all positive definite matrices.

A *conic program* is an optimization problem defined with a pointed and closed convex cone K with a nonempty interior, as follows.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax - b \in K. \end{aligned} \tag{CP}$$

Again, when $K = \mathbb{R}^n_+$, the problem reduces to a linear program. As we use the arithmetic "$\geq$" to indicate that a vector belongs to $\mathbb{R}^n_+$, we use notation "$\geq_K$" to indicate that a vector belongs to cone K. Basically, $Ax - b \in K$ is equivalent to $Ax - b \geq_K 0$ and $Ax \geq_K b$.

Example 3.6. When K is the second-order cone, the conic program (CP) is referred to as a *second-order cone program*. When K is the positive semidefinite cone, (CP) is a semidefinite program.