Introduction to Convex Optimization

Lec 5: Convex Optimization Problems

Silin DU

Department of Management Science and Engineering
Tsinghua University
`dsl21@mails.tsinghua.edu.cn`

TSINGHUA UNIVERSITY
清华经管学院

August 20, 2022

Contents

Lecture Overview

In this lecture, we focus on several subclasses of convex optimization.

1. Convex functions.
2. Operations that preserve convexity.
3. Conjugate functions.
4. Quasiconvex functions.
5. Operations that preserve quasiconvexity.
6. Log-concave functions.
7. Convexity by generalized inequality.

We put some proofs in appendix.

Contents

Standard Form of an Optimization Problem

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\
& h_i(x) = 0, \quad i = 1, \dots, p
\end{array}
\tag{1}
$$

- $x \in \mathbb{R}^n$ is the optimization variable.
- $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the objective or cost function.
- $f_i(x) : \mathbb{R}^n \to \mathbb{R}, i = 1, ..., m$ are the inequality constraint functions.
- $h_i : \mathbb{R}^n \to \mathbb{R}, i = 1, ..., p$ are the equality constraint functions.
- The domain of the optimization problem

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom} f_i \cap \bigcap_{i=1}^{p} \mathbf{dom} h_i$$

  the domain of the optimization problem.
- Optimal value:

$$p^* = \inf \{ f_0(x) \mid f_i(x) \le 0, i = 1, ..., m, h_i(x) = 0, i = 1, ..., p \}$$

- $p^* = \infty$ if the problem is infeasible (no x satisfies the constraints). $p^* = -\infty$ is problem is unbounded below.

Optimal and Locally Optimal Points

- x is feasible if $x \in \mathbf{dom} f_0$ and it satisfies the constraints ($x \in \mathcal{D}$).
- A feasible x is optimal if $f_0(x) = p^*$; X_{opt} is the set of optimal points.
- x is locally optimal if there is an $R > 0$ such that x is optimal for

$$\begin{array}{ll} \text{minimize (over } z) & f_0(z) \\ \text{subject to} & f_i(z) \leq 0, \quad i = 1, \ldots, m, \quad h_i(z) = 0, \quad i = 1, \ldots, p \\ & \|z - x\|_2 \leq R \end{array}$$

Examples (with $n = 1, m = p = 0$)

- $f_0(x) = 1/x, \mathbf{dom} f_0 = \mathbb{R}_{++} : p^\star = 0$, no optimal point
- $f_0(x) = -\log x, \mathbf{dom} f_0 = \mathbb{R}_{++} : p^\star = -\infty$
- $f_0(x) = x \log x, \mathbf{dom} f_0 = \mathbb{R}_{++} : p^\star = -1/e, x = 1/e$ is optimal
- $f_0(x) = x^3 - 3x, p^\star = -\infty$, local optimum at $x = 1$

Standard Form Convex Optimization Problem I

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^\top x = b_i, \quad i = 1, \ldots, p \end{array} \tag{2}$$

Compared with the general standard form problem (Eq. 1), the convex problem has three additional requirements:

1. The objective function f_0 is convex.
2. The inequality constraint functions $f_1, ..., f_m$ must be convex.
3. The equality constraint functions $h_i(x)$ must be affine.

- If $f_0(x)$ is quasiconvex, then the problem is a quasiconvex optimization problem.
- Important Property: feasible set of a convex optimization problem is convex.
- Many problems can be reformulated into the convex optimization form.

Standard Form Convex Optimization Problem II

$$
\begin{array}{ll}
\text{minimize} & f_0(x) = x_1^2 + x_2^2 \\
\text{subject to} & f_1(x) = x_1 / \left(1 + x_2^2\right) \le 0 \\
& h_1(x) = (x_1 + x_2)^2 = 0
\end{array}
$$

- f_0 is convex; feasible set $\{(x_1, x_2) \mid x_1 = -x_2 \le 0\}$ is convex
- not a convex problem (according to our definition): f_1 is not convex, h_1 is not affine
- equivalent (but not identical) to the convex problem

$$
\begin{array}{ll}
\text{minimize} & x_1^2 + x_2^2 \\
\text{subject to} & x_1 \le 0 \\
& x_1 + x_2 = 0
\end{array}
$$

Local and Global Optimization Theorem

Theorem 1 (Local and Global Optimization Theorem)

Any local optimal solution of a convex optimization problem is also a global optimal solution.

- Suppose x is locally optimal, but there exists a feasible y with $f_0(y) \leq f_0(x)$
- x is locally optimal means there is an $R > 0$ such that

$$\forall z \text{ is feasible, } \|z - x\|_2 \leq R \Rightarrow f_0(z) \geq f_0(x)$$

- Consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$
- z is a convex combination of two feasible points, hence also feasible
- $\|z - x\|_x = R/2$ and

$$f_0(z) = f_0(\theta y + (1 - \theta)x) \leq \theta f_0(y) + (1 - \theta)f_0(x) < f_0(x)$$

  which contradicts the assumption that x is locally optimal.
- The first inequality is because of the convexity of f_0, and the second inequality is because of the assumption $f_0(y) < f_0(x)$.

Optimality Criterion for Differentiable f_0 I

Suppose that the objective f_0 in a convex optimization problem is differentiable, so that for all $x, y \in \textbf{dom} f_0$,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^\top (y - x)$$

Then x is optimal if and only if it is feasible ($x \in X$) and

$$\nabla f_0(x)^\top (y - x) \geq 0, \quad \text{for all feasible } y \tag{3}$$

If $\nabla f_0(x) \neq 0$, $-\nabla f_0(x)$ defines a supporting hyperplane to feasible set X at x; see Figure 1.

Optimality Criterion for Differentiable f_0 II

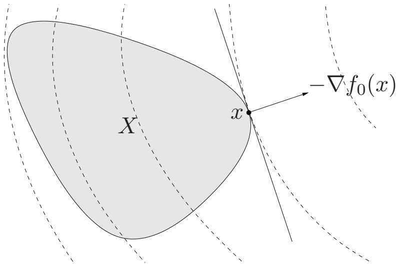


Fig. 1: The feasible set X is shown shaded. Some level curves of f_0 are shown as dashed lines. The point x is optimal: $-\nabla f_0(x)$ defines a supporting hyperplane (shown as a solid line) to X at x.

Proof. (By contradiction)

- Suppose $x \in X$ and satisfies Eq. 3. Then if $y \in X$ we have $f_0(y) \geq f_0(x)$, which shows x is optimal.
- Suppose x is optimal but Eq. 3 does not hold, i.e., for some $y \in X$ we have

$$\nabla f_0(x)^\top (y - x) < 0$$

- Consider $z(t) = ty + (1-t)x$, where $t \in [0,1]$ is a parameter. $z(t)$ is feasible since it is on the line segment between x and y.
-

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} f_0(z(t))\Big|_{t=0} &= \nabla f(z(t))^\top (y - x)\Big|_{t=0} \\ &= \nabla f(x)^\top (y - x) \leq 0 \end{aligned}$$

So $f_0(z(t)) < f_0(x)$ for t is small enough, which contradicts with x being optimal.

Next, we examine a few simple examples.

Unconstrainted Problems I

For an unconstrainted problem, the condition (Eq. 3) reduces to

$$\nabla f_0(x) = 0$$

for x to be optimal.

- Suppose x is optimal $\Rightarrow x \in \textbf{dom} f_0$ and for all feasible y we have $\nabla f_0(x)^\top (y - x) \geq 0$
- f_0 is differentiable, so all y sufficiently close to x are feasible.
- Take $y = x - t\nabla f_0(x)$ where $t \in \mathbb{R}$ is a parameter.
- For t small and positive, y is feasible, and so

$$\nabla f_0(x)^\top (y - x) = -t\|\nabla f_0(x)\|_2^2 \geq 0$$

for which we conclude $\nabla f_0(x) = 0$.

Unconstrainted Problems II

Unconstrainted quadratic optimization

- Consider the problem of minimizing the quadratic function

$$f_0(x) = (1/2)x^\top Px + q^\top x + r$$

  where $P \in \mathcal{S}_+^n$ (which makes f_0 convex). The necessary and sufficient condition for x to be a minimizer of f_0 is

$$\nabla f_0(x) = Px + q = 0.$$

- If $q \notin \mathcal{R}(P)$, then there is no solution. In this case f_0 is unbounded below.
- If $P \succ 0$ (which is the condition for f_0 to be strictly convex), then there is a unique minimizer, $x^\star = -P^{-1}q$.
- If P is singular, but $q \in \mathcal{R}(P)$, then the set of optimal points is the (affine) set $X_{\text{opt}} = -P^+q + \mathcal{N}(P)$, where P^+ denotes the pseudo-inverse of P.

Problems with Equality Constraints Only

Consider the probelm with equality constraints only, i.e.,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax = b \end{array}$$

x is optimal iff $\exists u$, such that $Ax = b, \nabla f_0(x) - A^\top u = 0$

- The optimality condition for a feasible x is that

  $$\nabla f_0(x)^\top (y - x) \geq 0$$

  hold for all y satisfying $Ay = b$.

- Since x is feasible, $A(x - y) = 0, (x - y) \in \mathcal{N}(A)$.
- $2x - y$ is also feasible $(A(2x - y) = b)$, so

  $$\nabla f_0(x)^\top (x - y) \geq 0$$

  which means $\nabla f_0(x)(x - y) = 0$ for all $(x - y) \in \mathcal{N}(A)$.

- In other words, $\nabla f_0(x) \perp \mathcal{N}(A)$. Therefore, $\nabla f_0(x) \in \mathcal{R}\left(A^\top\right)$. $(\mathcal{N}(A)^\perp = \mathcal{R}\left(A^\top\right))$
- $\nabla f_0(x) = A^\top u$ for some u.

Minimization over Nonnegative Orthant

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \preceq 0 \end{array}$$

- The optimality condition is

$$x \succeq 0, \nabla f_0(x)^\top (y - x) \geq 0 for all y \succeq 0$$

- $\nabla f_0(x)^\top y$ is unbounded below on $y \succeq 0$ unless $\nabla f_0(x) \succeq 0$
- The condition reduces to $-\nabla f_0(x)^\top x \geq 0$.
- Note that $x \succeq 0$ and $\nabla f_0(x) \succeq 0$. We must have $\nabla f_0(x)^\top x = 0$, i.e.,

$$\sum_{i=1}^{n} \left(\nabla f_0(x)\right)_i x_i = 0$$

- Since $\left(\nabla f_0(x)\right)_i \geq 0, x_i \geq 0$, then

$$\left(\nabla f_0(x)\right)_i x_i = 0, i = 1, ..., n$$

- x is optimal if and only if

$$x \in \mathbf{dom} f_0, \quad x \succeq 0, \quad \begin{cases} \nabla f_0(x)_i \geq 0, & x_i = 0 \\ \nabla f_0(x)_i = 0, & x_i > 0 \end{cases}$$

Contents

Equivalent Convex Problems I

Definition 1 (Equivalent Convex Problems)

Two problems are (informally) equivalent if the solution of one is readily obtained from the solution of the other, and vice versa.

Eliminating equality constraint

$$\begin{array}{lll} \text{minimize}_x & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, p \\ & Ax = b, & A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \end{array}$$

is equivalent to

$$\begin{array}{lll} \text{minimize}_z & f_0(Fz + x_0) & \\ \text{subject to} & f_i\left(Fz + x_0\right) \leq 0, & i = 1, \ldots, p, F \in \mathbb{R}^{n \times r}, r = \operatorname{rank}(F) \end{array}$$

where the range of F is the nullspace of A, i.e., $AF = 0$, and $Ax_0 = b$.

Equivalent Convex Problems II

- Introducing equality constraints

$$
\begin{array}{ll}
\text{minimize}_z & f_0\left(A_0 x+b_0\right) \\
\text{subject to} & f_i\left(A_i x+b_i\right) \leq 0, \quad i=1, \ldots, m
\end{array}
$$

is equivalent to

$$
\begin{array}{ll}
\text{minimize (over } x, y_i) & f_0\left(y_0\right) \\
\text{subject to} & f_i\left(y_i\right) \leq 0, \quad i=1, \ldots, m \\
& y_i=A_i x+b_i, \quad i=0,1, \ldots, m
\end{array}
$$

- Introducing slack variables for *linear inequalities*

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & a_i^{\top} x \leq b_i, \quad i=1, \ldots, m
\end{array}
$$

is equivalent to

$$
\begin{array}{ll}
\text{minimize(over } x, s) & f_0(x) \\
\text{subject to} & a_i^{\top} x+s_i=b_i, \quad i=1, \ldots, m \\
& s_i \geq 0, \quad i=1, \ldots m
\end{array}
$$

Equivalent Convex Problems III

- Epigraph Form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

is equivalent To

$$\begin{array}{ll} \text{minimize}_{x,t} & t \\ \text{subject to} & f_0(x) - t \leq 0 \\ & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

- Minimizing over some variables

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & f_i(x_1) \leq 0, \quad i = 1, \ldots, m \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x_1) \\ \text{subject to} & f_i(x_1) \leq 0, \quad i = 1, \ldots, m \end{array}$$

where $\tilde{f}_0\left(x_1\right) = \inf_{x_2} f_0\left(x_1, x_2\right)$

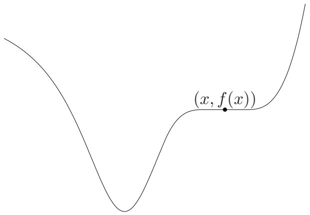
Contents

Standard Form

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
\tag{4}
$$

with $f_0 : \mathbb{R}^n \to \mathbb{R}$ quasiconvex, $f_1, ..., f_m$ convex.

- A quasiconvex optimization problem can have locally optimal solutions that are not (globally) optimal.
- Solving a quasiconvex optimization problem can be reduced to solving a sequence of convex optimization problems.



Optimality Condition

Let X be the feasible set for the quasiconvex optimization probelm (Eq. 4). It follows from the first-order condition for quasiconvexity that x is optimal if

$$x \in X, \quad \nabla f_0(x)^\top (y - x) > 0 \text{ for all } y \in X \setminus \{x\}$$

- The condition is only sufficient for optimality, which needs not hold for an optimal point.
- The condition requires the gradient $\nabla f_0(x) \neq 0$, whereas the condition in the convex case does not.

Convex Representation of Sublevel Sets of f_0

If f_0 is quasiconvex, there exists a family of functions ϕ_t such that:

- $\phi_t(x)$ is convex in x for fixed t
- t-sublevel set of f_0 is 0-sublevel set of ϕ_t, i.e.,

$$f_0(x) \leq t \Longleftrightarrow \phi_t(x) \leq 0$$

For example, consider

$$f_0(x) = \frac{p(x)}{q(x)}$$

with p convex, q concave, and $p(x) \geq 0$, $q(x) > 0$ on $\mathbf{dom} f_0$

- It's easy to verify that $f_0(x)$ is quasiconvex. Note that $f_0(x) \geq 0$.

$$f_0(x) \leq t \Leftrightarrow \frac{p(x)}{q(x)} \leq t \Leftrightarrow p(x) - tq(x) \leq 0$$

  When $t \geq 0$, $\{x \mid p(x) - tq(x) \leq 0\}$ is convex.
- $\phi_t(x) = p(x) - tq(x)$ is convex in x for $t \geq 0$.
- $f_0(x) \leq t$ if and only if $\phi_t(x) \leq 0$.

Quasiconvex Optimization via Convex Feasibility Problems I

Let p^* denote the optimal value of the quasiconvex optimization problem (Eq. 4). If the following problem

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & \phi_t(x) \leq 0 \\ & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \tag{5}$$

is feasible, then $p^* \leq t$. Conversely, if the problem is infeasible, then $p^* \geq t$. We can solve a quasiconvex optimization problem using bisection, solving a convex feasibility problem at each step.

Quasiconvex Optimization via Convex Feasibility Problems II

Algorithm 1 Bisection method for quasiconvex optimization

Require: $l \leq p^*, u \geq p^*$, tolerance $\epsilon > 0$

```
repeat
    t := (l + u)/2
    Solve the convex feasiblity problem (Eq. 5) at t
    if feasible then
        u := t
    else
        l := t
until u − l ≤ ϵ
```

Complexity: requires exactly $\lceil \log_2((u-l)/\epsilon) \rceil$ iterations.

Contents

Linear Programming I

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

Standard form linear programming (LP)

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array}$$

Convert LP to standard forms

- Introduce slack variables s_i for the inequality constraints.
- Express the variable x as the difference of two nonnegative variables x^+ and x^-, i.e., $x = x^+ - x^-$

Linear Programming II

Diet problem: choose quantities $x_1, \ldots, x_n$ of n foods

- one unit of food j costs c_j. contains amount $a_{i,j}$ of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax \succeq b, \quad x \succeq 0 \end{array}$$

Piecewise-linear minimization

$$\text{minimize} \quad \max_{i=1,\ldots,m} \left(a_i^\top x + b_i \right)$$

equivalent to an LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^\top x + b_i \leq t, \quad i = 1, ..., m \end{array}$$

Chebyshev center of a polyhedron

Find the largest Euclidean ball that lies in a polyhedron

$$\mathcal{P} = \left\{x \in \mathbb{R}^n \mid a_i^\top x \leq b_i, i = 1, \ldots, m\right\}$$

The center of the optimal ball is called the Chebyshev center of the polyhedron. We represent the ball as

$$\mathcal{B} = \{x_c + u \mid \|u\|_2 \leq r\}$$

$\mathcal{B}$ in the halfspace $a_i^\top x \leq b_i$ if and only if

$$a_i^\top (x_c + u) \leq b_i, \quad \|u\|_2 \leq r$$

Note the dual norm of $\|\cdot\|_2$ is also Euclidean norm, i.e.,

$$\|a_i\|_2 = \sup\left\{a_i^\top x \mid \|x\|_2 \leq 1\right\}$$

Therefore, $\sup\left\{a_i^\top u \mid \|u\|_2 \leq r\right\} = r\|a_i\|_2$. We can solve the LP to get x_c, r.

$$\begin{array}{ll} \text{minimize} & r \\ \text{subject to} & a_i^\top x_c + r\|a_i\|_2 \leq b_i, \quad i = 1, \ldots, m \end{array}$$

Linear-Fractional Programming I

The problem of minimizing a ratio of affine functions over a polyhedron is called a linear-fractional program:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array} \tag{6}$$

where the objective function is given by

$$f_0(x) = \frac{c^\top x + d}{e^\top x + f}, \quad \mathbf{dom} f_0 = \left\{ x \mid e^\top x + f > 0 \right\}$$

- a quasiconvex optimization problem; can be solved by bisection
- also equivalent to the LP

$$\begin{array}{ll} \text{minimize} & c^\top y + dz \\ \text{subject to} & Gy \preceq hz \\ & Ay = bz \\ & e^\top y + fz = 1 \\ & z \geq 0 \end{array} \tag{7}$$

Linear-Fractional Programming II

To show the equivalence

- If x is feasible in Problem 6 then the pair

$$y = \frac{x}{e^\top x + f}, \quad z = \frac{1}{e^\top x + f}$$

  is feasible in Problem 7, with the same objective value $c^\top y + dz = f_0(x)$. It follows that the optimal value of Problem 6 is greater than or equal to the optimal value of Problem 7.
- If (y, z) is feasible in Problem 7, with $z \neq 0$, then $x = y/z$ is feasible in Problem 6, with the same objective value $f_0(x) = c^\top y + dz$.
- If (y, z) is feasible in Problem 7, with $z = 0$ and x_0 is feasible for Problem 6, then $x = x_0 + ty$ is feasible in Problem 6 for all $t \geq 0$.
- Moreover, $\lim_{t\to\infty} f_0(x_0 + ty) = c^\top y + dz$, so we can find feasible points in Problem 6 with objective values arbitrarily close to the objective value of (y, z).
- The optimal value of Problem 6 is less than or equal to the optimal value of Problem 7.

Generalized Linear-Fractional Programming

A generalization of the linear-fractional program (6) is the generalized linearfractional program in which

$$f_0(x) = \max_{i=1,\ldots,r} \frac{c_i^\top x + d_i}{e_i^\top x + f_i}, \quad \mathbf{dom} f_0 = \left\{ x \mid e_i^\top x + f_i > 0, i = 1, ..., r \right\}$$

The objective function is the pointwise maximum of r quasiconvex functions, and therefore quasiconvex.

Von Neumann model of a growing economy

$$\begin{array}{ll} \text{maximize} & \min_{i=1,\ldots,n} x_i^+ / x_i \\ \text{subject to} & x^+ \succeq 0, Bx^+ \preceq Ax \end{array}$$

- $x, x^+ \in \mathbb{R}^n$: activity levels of n sectors, in current and next period.
- $(Ax)_i, (Bx)_i$: produced, consumed amounts of good i.
- x_i^+ / x_i: growth rate of sector i.
- allocate activity to maximize growth rate of lowest growing sector.

Quadratic Programming I

A convex optimization problem is called a quadratic program (QP) if the objective function is (convex) quadratic, and the constraint functions are affine.

$$
\begin{array}{ll}
\text{minimize} & (1/2)x^\top Px + q^\top x + r \\
\text{subject to} & Gx \preceq h \\
& Ax = b
\end{array}
$$

where $P \in \mathcal{S}_+^n, G \in \mathbb{R}^{m\times n},$ and $A \in \mathbb{R}^{p\times n}$.

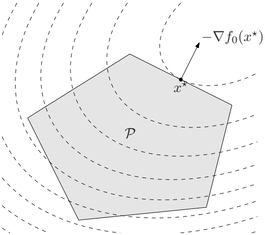


Fig. 2: Minimize a convex quadratic function over a polyhedron.

Quadratic Programming II

Least-squares

$$\text{minimize} \quad \|Ax - b\|_2^2$$

- optimal solution: $x^* = \left(A^\top A\right)^{-1} A^\top b$
- can add linear constraints, e.g., $l \preceq x \preceq u$

Linear program with random cost

$$\begin{array}{ll} \text{minimize} & \bar{c}^\top x + \gamma x^\top \Sigma x = \mathbf{E}\left(c^\top x\right) + \gamma \mathbf{E}\left(c^\top x\right) \\ \text{subject to} & Gx \preceq h, Ax = b \end{array}$$

- c is random vector with mean $\bar{c}$ and covariance Σ
- $c^\top x$ is a random variable with mean $\bar{c}^\top x$ and variance $x^\top \Sigma x$
- $\gamma > 0$ is risk aversion parameterl; controls the trade-off between expected cost and variance (risk)

Quadratic Constrained Quadratic Programming

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^\top P_0 x + q_o^\top x + r_0 \\ \text{subject to} & \frac{1}{2}x^\top P_i x + q_i^\top x + r_i \leq 0, \quad i = 1, ..., m \\ & Ax = b \end{array}$$

- $P_i \in \mathcal{S}_+^n, i = 0, 1, ..., m$; objective and constraints are convex quadratic
- If $P_1, ..., P_m \in \mathcal{S}_{++}^n$, feasible region is intersection of m ellipsoids and an affine set.

Second-Order Cone Programming I

$$\begin{array}{ll} \text{minimize} & f^\top x \\ \text{subject to} & \|A_i x + b_i\|_2 \leq c_i^\top x + d_i, \quad i = 1, \dots, m \\ & Fx = g \end{array}$$

where $x \in \mathbb{R}^n$ is the optimization variable, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

- We call a constraints of the form

$$\|Ax + b\|_2 \leq c^\top x + d$$

  where $A \in \mathbb{R}^{k \times n}$, a second-order cone constraint, since it is the same as requiring the affine function $(Ax + b, c^\top x + d)$ to lie in the second-order cone in $\mathbb{R}^{k+1}$.
- The second-order cone in $\mathbb{R}^{k+1}$ is defined as

$$C_k = \left\{ \begin{bmatrix} u \\ t \end{bmatrix} \mid u \in \mathbb{R}^k, t \in \mathbb{R}, \|u\|_2 \leq t \right\}$$

- For $n_i = 0$, SOCP reduces to an LP; if $c_i = 0$, it reduces to a QCQP.
- Second-order cone programs are more general than QCQPs and of LPs.

Second-Order Cone Programming II

Revisit the least-square problem.

- Unconstrained:

$$\text{minimize} \quad \|Ax - b\|_2^2$$

- Adding constraints:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \quad \text{(Constrained QP)} \\ \text{subject to} & x \succeq 0 \end{array}$$

  equivalent to

$$\begin{array}{ll} \text{minimize} & t \qquad\qquad\qquad \text{(SOCP)} \\ \text{subject to} & \|Ax - b\|_2 \leq t \\ & x \succeq 0 \end{array}$$

Second-Order Cone Programming III

- Adding regularity constraints (Add penalty to large coefficients): (Ridge Regression)

$$
\begin{array}{lll}
\text{minimize} & \|Ax - b\|_2^2 & \text{(QCQP)} \\
\text{subject to} & \|x\|_2 \leq R_2 &
\end{array}
$$

equivalent to

$$
\begin{array}{lll}
\text{minimize} & t & \text{(SOCP)} \\
\text{subject to} & \|Ax - b\|_2 \leq t & \\
 & \|x\|_2 \leq R_2 &
\end{array}
$$

Second-Order Cone Programming IV

- LASSO

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \qquad \text{(QP)} \\ \text{subject to} & \|x\|_1 \le R_1 \end{array}$$

equivalent to

$$\begin{array}{ll} \text{minimize} & t \qquad \text{(SOCP)} \\ \text{subject to} & \|Ax - b\|_2 \le t \\ & \|x\|_1 \le R_1 \end{array}$$

We can transform l_1-norm constraints into linear constraints, e.g., $|x| \le 2$ can be transformed into $x \le 2$ and $x \ge 2$.

Robust Optimization I

The parameters in optimization problems are often uncertain, e.g., in an LP

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & a_i^\top x \leq b_i, \quad i = 1, ..., m \end{array}$$

There can be uncertainty in c, a_i, b_i.
Two common approaches to handling uncertainty (in a_i for simplicity)

- deterministic model: constraints must hold for all $a_i \in \mathcal{E}_i$

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & a_i^\top x \leq b_i \text{ for all } a_i \in \mathcal{E}_i, \quad i = 1, ..., m \end{array}$$

- stochastic model: a_i is random variable; constraints must hold with probability η

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \textbf{prob}\left(a_i^\top x \leq b_i\right) \geq \eta, \quad i = 1, ..., m \end{array}$$

Robust Optimization II

Deterministic approach via SOCP

- choose an ellipsoid as $\mathcal{E}_i$:

$$\mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \leq 1\} \quad \left(\bar{a}_i \in \mathbb{R}^n, P_i \in \mathbb{R}^{n \times n}\right)$$

  center is $\bar{a}_i$, semi-axes determined by singular values/vectors of P_i

- The robust linear constraint can be expressed as

$$\begin{aligned} \sup\left\{a_i^\top x \mid a_i \in \mathcal{E}_i\right\} &= \bar{a}_i^\top x + \sup\left\{u^\top P_i^\top x \mid \|u\|_2 \leq 1\right\} \\ &= \bar{a}_i^\top x + \|P_i^\top x\|_2 \leq b_i \qquad \text{(By the definition of dual norm)} \end{aligned}$$

- Robust LP

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & a_i^\top x \leq b_i \text{ for all } a_i \in \mathcal{E}_i, \quad i = 1, \dots, m \end{array}$$

  is equivalent to the SOCP

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \bar{a}_i^\top x + \|P_i^\top x\|_2 \leq b_i, \quad i = 1, \dots, m \end{array}$$

Robust Optimization III

Stochastic approach via SOCP

- Assume a_i is Guassian with mean $\bar{a}_i$, covariance Σ_i ($a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i)$)
- $a_i^\top x$ is Guassian r.v. with mean $\bar{a}_i^\top x$, variance $x^\top \Sigma_i x$; hence

$$\textbf{prob}\left(a_i^\top x \le b_i\right) = \Phi\left(\frac{b_i - \bar{a}_i^\top x}{\left\|\Sigma_i^{1/2} x\right\|_2}\right)$$

  where $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2} \mathrm{d}t$ is CDF of $\mathcal{N}(0,1)$.

- Robust LP

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \textbf{prob}\left(a_i^\top x \le b_i\right) \ge \eta, \quad i = 1, ..., m \end{array}$$

  with $\eta \ge 1/2$, is equivalent to the SOCP

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \bar{a}_i^\top x + \Phi^{-1}(\eta) \left\|\Sigma_i^{1/2} x\right\|_2 \le b_i, \quad i = 1, ..., m \end{array}$$

Generalized Inequality Constraints

Convex optimization problem with generalized inequality constraints

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \preceq_{K_i}, \quad i = 1, ..., m \\
 & Ax = b
\end{array}
\tag{8}
$$

where $f_0 : \mathbb{R} \to \mathbb{R}$ is convex and $f_i : \mathbb{R}^n \to \mathbb{R}^{K_i}$ are K_i-convex with respect to proper cone K_i

Many of the results for ordinary convex optimization problems hold for problems with generalized inequalities.

- The feasible set, any sublevel set, and the optimal set are convex.
- Any point that is locally optimal for Problem 8 is globally optimal.
- The optimality condition for differentiable f_0, given in Eq. 3, holds without any change.

Conic Form Problem

Conic form problem: special case with affine objective and constraints

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Fx + g \preceq_K 0 \\ & Ax = b \end{array}$$

- It extends linear programming ($K = \mathbb{R}^m_+$) to nonpolyhedral cones.
- Conic form problem in standard form

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & x \succeq_K 0 \\ & Ax = b \end{array}$$

When K is S_+^k, the cone of positive semidefinite $k \times k$ matrices, the associated conic form problem is called a semidefinite program (SDP)

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b \end{array} \tag{9}$$

where $G, F_1, \cdots, F_n \in \mathcal{S}^k$, and $A \in \mathbb{R}^{p \times n}$.

- The inequality constraint is called linear matrix inequality (LMI)
- We can transform multiple LMI constraints into one LMI constraint. For example,

$$x_1 \hat{F}_1 + \cdots + x_n \hat{F}_n + \hat{G} \preceq 0, \quad x_1 \tilde{F}_1 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0$$

  is equivalent to single LMI

$$x_1 \begin{bmatrix} \hat{F}_1 & 0 \\ 0 & \tilde{F}_1 \end{bmatrix} + x_2 \begin{bmatrix} \hat{F}_2 & 0 \\ 0 & \tilde{F}_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{F}_n & 0 \\ 0 & \tilde{F}_n \end{bmatrix} + \begin{bmatrix} \hat{G} & 0 \\ 0 & \tilde{G} \end{bmatrix} \preceq 0$$

Semidefinite Programeamming II

Standard form of SDP

$$\begin{array}{ll} \text{minimize}_X & \operatorname{tr}(CX) \\ \text{subject to} & \operatorname{tr}(A_iX) = b_i, \quad i = 1, ..., p \\ & X \succeq 0 \end{array} \tag{10}$$

where $X, C, A_1, ..., A_p \in \mathcal{S}^n$.

- Note that $\operatorname{tr}(CX) = \sum_{i=1}^n \sum_{j=1}^n C_{ij}X_{ij}$ is a linear function of X.
- In an SDP that the variable is the matrix X, but it might be helpful to think of X as an array of n^2 numbers or simply as a vector in $\mathcal{S}^n$.
- Consider an example of an SDP for $n = 3$ and $p = 2$. Define the following matrices:

$$A_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 3 & 7 \\ 1 & 7 & 5 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 2 & 8 \\ 2 & 6 & 0 \\ 8 & 0 & 4 \end{pmatrix}, \quad \text{and} \quad C = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 9 & 0 \\ 3 & 0 & 7 \end{pmatrix}$$

and $b_1 = 11$ and $b_2 = 19$. Then the variable X will be the 3×3 symmetric matrix:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$$

Semidefinite Progreamming III

and so

$$\begin{aligned}\mathrm{tr}(CX) &= x_{11} + 2x_{12} + 3x_{13} + 2x_{21} + 9x_{22} + 0x_{23} + 3x_{31} + 0x_{32} + 7x_{33} \\ &= x_{11} + 4x_{12} + 6x_{13} + 9x_{22} + 0x_{23} + 7x_{33}\end{aligned}$$

since, in particular, X is symmetric. Therefore the SDP can be written as:

$$\begin{array}{ll}\text{minimize}_X & x_{11} + 4x_{12} + 6x_{13} + 9x_{22} + 0x_{23} + 7x_{33} \\ \text{subject to} & x_{11} + 0x_{12} + 2x_{13} + 3x_{22} + 14x_{23} + 5x_{33} = 11 \\ & 0x_{11} + 4x_{12} + 16x_{13} + 6x_{22} + 0x_{23} + 4x_{33} = 19 \\ & X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \succeq 0\end{array}$$

- Notice that SDP looks remarkably similar to a linear program. However, the standard LP constraint that x must lie in the nonnegative orthant is replaced by the constraint that the variable X must lie in the cone of positive semidefinite matrices.

Semidefinite Programeamming IV

SDP duality

Consider the standard form of SDP (Problem 10)

- The Lagrangian function is

$$L(X,\gamma,Y) = \mathrm{tr}(CX) + \sum_{i=1}^{p} \gamma_i \left(b_i - \mathrm{tr}(A_i X)\right) - \mathrm{tr}(XY), \quad \text{where } Y \succeq 0$$

- If $X = X^\top \in \mathbb{R}^{n\times n}, Y \in \mathbb{R}^{n\times n}$, then

$$\min_{Y\succeq 0} \mathrm{tr}(XY) = \begin{cases} 0, & X \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

- Dual problem

$$\begin{array}{lll} \max_{Y\succeq 0,\gamma} & g(Y,\gamma) & \\ \text{where} & g(Y,\gamma) & = \min_X L(X,\gamma,Y) \\ & & = \gamma^\top b + \min_X \mathrm{tr}\left[(C - \sum_{i=1}^p \gamma_i A_i - Y)X\right] \\ & & = \begin{cases} \gamma^\top b, & C - \sum_{i=1}^p \gamma_i A_i - Y = 0 \\ -\infty, & \text{otherwise} \end{cases} \end{array}$$

is equivalent to

$$
\begin{array}{ll}
\max & \gamma^\top b \\
\text{subject to} & C - \sum_{i=1}^{p} \gamma_i A_i - Y = 0 \\
 & Y \succeq 0
\end{array}
$$

$$\iff$$

$$
\begin{array}{ll}
\max & \gamma^\top b \\
\text{subject to} & \sum_{i=1}^{p} \gamma_i A_i \preceq C \quad \text{(LMI)}
\end{array}
$$

How to transform an SDP in LMI form into the standard form?

- We can write the equality constraint $Ax = b$ in Problem 9 as an LMI.
- We can rewrite $Ax = b$ as $\sum_{i=1}^{n} a_i x_i - b = 0$ where a_i is the i-th column of A.
- This can be written as two constraints: $\sum_{i=1}^{n} a_i x_i - b \geq 0$ and $\sum_{i=1}^{n} a_i x_i - b \leq 0$.

- Now we can simly add this to the LMI. So we get

$$x_1 F_1' + x_2 F_2' + \cdots + x_n F_n' + G' \preceq 0$$

where

$$F_i' = \begin{bmatrix} F_i & & \\ & \mathrm{diag}(a_i) & \\ & & \mathrm{diag}(-a_i) \end{bmatrix}, \quad G' = \begin{bmatrix} G & & \\ & \mathrm{diag}(-b) & \\ & & \mathrm{diag}(b) \end{bmatrix}$$

- At last, we write the dual to get

$$\begin{array}{ll} \max & \mathrm{tr}(G'X) \\ \text{subject to} & \mathrm{tr}(F_i'X) = c_i \\ & X \succeq 0 \end{array}$$

LP as SDP

LP

$$
\begin{array}{ll}
\min & c^\top x \\
\text{subject to} & Ax = b, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \\
& x \succeq 0
\end{array}
$$

is equivalent to SDP

$$
\begin{array}{ll}
\min & \operatorname{tr}(CX) \\
\text{subject to} & \operatorname{tr}(A_i X) = b_i, \quad i = 1, ..., m \\
& X \succeq 0
\end{array}
$$

where $C = \operatorname{diag}(c_1, c_2, ..., c_n), A_i = \operatorname{diag}(a_{i1}, a_{i2}, ..., a_{in}), X = \operatorname{diag}(x_1, x_2, ..., x_n)$

QCQP as SDP I

QCQP

$$\begin{array}{lll} \min_x & x^\top Q_0 x + q_0^\top x + c_0 & \\ \text{subject to} & x^\top Q_i x + q_i^\top x + c_i \leq 0, & i = 1, ..., m \\ & Q_i \succeq 0, i = 0, 1, .., m & \end{array}$$

is equivalent to

$$\begin{array}{lll} \min_{x,\theta} & \theta & \\ \text{subject to} & x^\top Q_0 x + q_0^\top x + c_0 - \theta \leq 0 & \\ & x^\top Q_i x + q_i^\top x + c_i \leq 0, & i = 1, ..., m \\ & Q_i \succeq 0, i = 0, 1, .., m & \end{array}$$

- We can factor each Q_i into $Q_i = M_i^\top M_i$ (Cholesky decomposition).
- By Schur complement, we have

$$\begin{bmatrix} I & M_i x \\ x^\top M_i^\top & -c_i - q_i^\top x \end{bmatrix} \succeq 0 \Longleftrightarrow I \succeq 0 \text{ and } -c_i - q_i^\top x - x^\top M_i^\top M_i x \geq 0$$

QCQP as SDP II

- Then we can write QCQP as

$$
\begin{array}{ll}
\min_{x,\theta} & \theta \\
\text{subject to} & \begin{bmatrix} I & M_0 x \\ x^\top M_0^\top & -c_0 - q_0^\top x + \theta \end{bmatrix} \succeq 0 \\
& \\
& \begin{bmatrix} I & M_i x \\ x^\top M_i^\top & -c_i - q_i^\top x \end{bmatrix} \succeq 0, \qquad i = 0, 1, .., m
\end{array}
$$

SOCP as SDP I

SOCP

$$\begin{array}{ll} \min & f^\top x \\ \text{subject to} & \|A_i x + b_i\|_2 \leq c_i^\top x + d_i, \quad i = 1, \ldots, m \end{array}$$

is equivalent to SDP

$$\begin{array}{ll} \min & f^\top x \\ \text{subject to} & \begin{bmatrix} \left(c_i^\top x + d_i\right) I & A_i x + b_i \\ \left(A_i x + b_i\right)^\top & c_i^\top x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m \end{array}$$

Note that

$$\begin{aligned} & \begin{bmatrix} \left(c_i^\top x + d_i\right) I & A_i x + b_i \\ \left(A_i x + b_i\right)^\top & c_i^\top x + d_i \end{bmatrix} \succeq 0 \\ \Longleftrightarrow \quad & c_i^\top x + d_i - \left(A_i x + b_i\right)^\top \left[\left(c_i^\top x + d_i\right) I\right]^{-1} A_i x + b_i \geq 0 \\ \Longleftrightarrow \quad & \|A_i x + b_i\|_2^2 \leq \left(c_i^\top x + d_i\right)^2 \end{aligned}$$

Application of SDP I

Max-eigenvalue minimization