Convex Optimization (II)

Lecture 7, Nonlinear Programming

National Taiwan University

November 1, 2016

Table of contents

Convex optimization problems in standard form

A convex optimization problem is one of the form

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

where $f_0, \ldots, f_m$ are convex functions. Compared with the general standard form problem, the convex problem has three additional requirements:

- the objective function must be convex,
- the inequality constraint functions must be convex,
- the equality constraint functions $h_i(x) = a_i^T x - b_i$ must be affine.

Convex optimization problems in standard form

- The feasible set of a convex optimization problem is convex, since it is the intersection of
  - the domain of the problem

$$D = \bigcap_{i=0}^{m} \mathbf{dom}\ f_i,$$

    (which is a convex set),
  - m (convex) sublevel sets $\{x \mid f_i(x) \leq 0\}$, and
  - p hyperplanes $\left\{x \mid a_i^T x = b_i\right\}$.
    - W.l.o.g., we assume that $a_i \neq 0$.
- In a convex optimization problem, we minimize a convex objective function over a convex set.

Quasiconvex Optimization Problems

- If f_0 is quasiconvex instead of convex, the problem

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

  is called a (standard form) quasiconvex optimization problem.
- Since the sublevel sets of a convex or quasiconvex function are convex, we conclude that for a convex or quasiconvex optimization problem the ϵ-suboptimal sets are convex.
- In particular, the optimal set is convex.

Concave maximization problems

- We also refer to

$$\begin{aligned} \text{maximize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & a_i^T x = b_i, i = 1, \dots, p, \end{aligned}$$

  as a convex optimization problem if the objective function f_0 is concave, and the inequality constraint functions $f_1, \dots, f_m$ are convex.
- This concave maximization problem is readily solved by minimizing the convex objective function $-f_0$.
  - All of the results, conclusions, and algorithms that we describe for the minimization problem are easily transposed to the maximization case.
- In a similar way the above maximization problem is called quasiconvex if f_0 is quasiconcave.

Definition of Convex Optimization Problem

A closer look

- Consider the example with $x \in \mathbf{R}^2$,

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

  which is in the standard form.
- This problem is not a convex optimization problem in standard form since the equality constraint function h_1 is not affine, and the inequality constraint function f_1 is not convex.
- Nevertheless the feasible set, which is $\{x \mid x_1 \leq 0, x_1 + x_2 = 0\}$, is convex.
- The problem, although not in a form of convex optimization problem, can be easily transformed to, and be shown to be equivalent to, a convex optimization problem.

Local and global optima (1/2)

- As an important property of convex optimization problems, any locally optimal point is also (globally) optimal.
- To see this, suppose that x is locally optimal for a convex optimization problem, i.e., x is feasible and

$$f_0(x) = \inf \left\{ f_0(z) \;\middle|\; z \text{ feasible}, ||z - x||_2 \leq R \right\},$$

  for some $R > 0$.
- Now suppose that x is not globally optimal, i.e., there is a feasible y such that $f_0(y) < f_0(x)$. Evidently $||y - x||_2 > R$, since otherwise $f_0(x) \leq f_0(y)$.

Local and global optima (2/2)

- Consider the point z given by

$$z = (1-\theta)x + \theta y, \theta = \frac{R}{2||y-x||_2}.$$

  Then we have $||z-x||_2 = R/2 < R$, and by convexity of the feasible set, z is feasible.
- By convexity of f_0 we have

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

  which leads to a contradiction. So, x is globally optimal.
- It is not true that locally optimal points of quasiconvex optimization problems are globally optimal (to be shown later).

An optimality criterion for differentiable f_0

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

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

- Let X denote the feasible set, i.e.,

$$X = \{x \mid f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p\}.$$

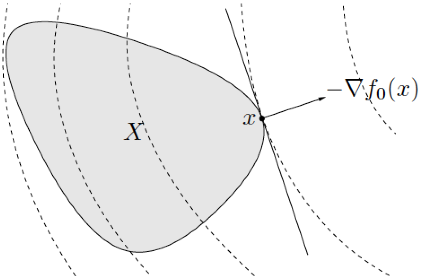
Then x is optimal if and only if $x \in X$ and

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

for all $y \in X$.

An optimality criterion for differentiable f_0

- The optimality criterion can be understood geometrically: If $\nabla f_0(x) \neq 0$, it means that $-\nabla f_0(x)$ defines a supporting hyperplane to the feasible set at x.



Proof of optimality condition

- The "if" part is obvious.
- For the "only if" part, suppose x is optimal, but the optimality condition $\nabla f_0(x)^T(y-x) \geq 0$ does not hold, i.e., for some $y \in X$ we have

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

- Consider the point $z(t) = ty + (1-t)x$, where $t \in [0,1]$ is a parameter. Since $z(t)$ is on the line segment between x and y, and the feasible set is convex, $z(t)$ is feasible. Note that

$$\left[\frac{d}{dt} f_0(z(t))\right]\Bigg|_{t=0} = \nabla f_0(x)^T(y-x) < 0,$$

so for small positive t, we have $f_0(z(t)) < f_0(x)$, which proves that x is not optimal.

Unconstrained problems

- For an unconstrained problem (i.e., $m = p = 0$), the optimality condition

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

reduces to the well known necessary and sufficient condition

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

for x to be optimal.

Unconstrained problems

- To see this, suppose x is optimal, which means here that $x \in \textbf{dom } f_0$, and for all feasible y we have $\nabla f_0(x)^T(y-x) \geq 0$. Since f_0 is differentiable, its domain is (by definition) open, so all y sufficiently close to x are feasible.
- Let us take $y = x - t\nabla f_0(x)$. Then for t small and positive, y is feasible, and so

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

from which we conclude $\nabla f_0(x) = 0$.
- If $\nabla f_0(x) = 0$ has no solutions, then there are no optimal points, possibly
    - the problem is unbounded below, or
    - the optimal value is finite, but not attained.
- On the other hand, $\nabla f_0(x) = 0$ can have multiple solutions.
    - In this case, each such solution is a minimizer of f_0.

Example – Unconstrained quadratic optimization.

- Consider the problem of minimizing the quadratic function
$$f_0(x) = (1/2)x^T Px + q^T x + r,$$
where $P \in \mathbf{S}^n_+$ (which makes f_0 convex).
- The necessary and sufficient condition for x to be a minimizer of f_0 is
$$\nabla f_0(x) = Px + q = 0.$$
- Several cases can occur, depending on whether this (linear) equation has no solutions, one solution, or many solutions.
  - 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, $x^* = -P^{-1}q$.
  - If P is singular, but $q \in \mathcal{R}(P)$, then the set of optimal points is the (affine) set $X_{opt} = -P^\dagger q + \mathcal{N}(P)$, where $P^\dagger$ denotes the pseudo-inverse of P.

Problems with equality constraints only (1/2)

- Consider the case where there are equality constraints but no inequality constraints, i.e.,

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax = b. \end{aligned}$$

  Here the feasible set is affine. We assume that it is nonempty.
- The optimality condition for a feasible x is that

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

  must hold for all y satisfying $Ay = b$.
- Since x is feasible, every feasible y has the form $y = x + v$ for some $v \in \mathcal{N}(A)$. The optimality condition can therefore be expressed as: $\nabla f_0(x)^T v \geq 0$ for all $v \in \mathcal{N}(A)$.

Problems with equality constraints only (2/2)

- If a linear function is nonnegative on a subspace, then it must be zero on the subspace, so it follows that $\nabla f_0(x)^T v = 0$ for all $v \in \mathcal{N}(A)$. In other words, $\nabla f_0(x) \perp \mathcal{N}(A)$.
- Using the fact that $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$, this optimality condition can be expressed as $\nabla f_0(x) \in \mathcal{R}(A^T)$, i.e., there exists a $\nu \in \mathbf{R}^p$ such that

$$\nabla f_0(x) + A^T \nu = 0.$$

Together with the requirement $Ax = b$ (i.e., that x is feasible), this is the classical Lagrange multiplier optimality condition.

Minimization over the nonnegative orthant (1/2)

- We consider the problem

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

where the only inequality constraints are nonnegativity constraints on the variables. The optimality condition is then

$$x \succeq 0, \quad \nabla f_0(x)^T(y - x) \geq 0 \text{ for all } y \succeq 0.$$

- The term $\nabla f_0(x)^T y$, which is a linear function of y, is unbounded below on $y \succeq 0$, unless we have $\nabla f_0(x) \succeq 0$.

Minimization over the nonnegative orthant (2/2)

- The condition then reduces to $-\nabla f_0(x)^T x \geq 0$. But $x \succeq 0$ and $\nabla f_0(x) \succeq 0$, so we must have $\nabla f_0(x)^T x = 0$, i.e.,

$$\sum_{i=1}^{n} [\nabla f_0(x)]_i x_i = 0.$$

- Therefore, $[\nabla f_0(x)]_i x_i = 0$ for $i = 1, \ldots, n$. The optimality condition can therefore be expressed as

$$x \succeq 0, \quad \nabla f_0(x) \succeq 0, \quad x_i [\nabla f_0(x)]_i = 0, \quad i = 1, \ldots, n.$$

- The last condition is called **complementarity**, since it means that the set of indices corresponding to nonzero components of the vectors x and $\nabla f_0(x)$ are complementary (i.e., have empty intersection).

Quasiconvex optimization

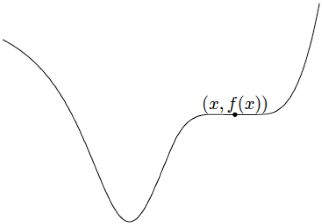
- Recall that a quasiconvex optimization problem has the standard form

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

  where the inequality constraint functions $f_1, \ldots, f_m$ are convex, and the objective f_0 is quasiconvex (instead of convex).
- Some basic differences between convex and quasiconvex optimization problems will be studied.
  - It would be shown that how solving a quasiconvex optimization problem can be reduced to solving a sequence of convex optimization problems.

Locally optimal solutions and optimality conditions

- The most important difference between convex and quasiconvex optimization is that a quasiconvex optimization problem can have locally optimal solutions that are not (globally) optimal.
- This phenomenon can be seen even in the simple case of unconstrained minimization of a quasiconvex function on $\mathbf{R}$.



Locally optimal solutions and optimality conditions

- Nevertheless, a variation of the optimality condition for convex problems ($\nabla f_0(x)^T(y-x) \geq 0$ for all $y \in X$) does hold for quasiconvex optimization problems with differentiable objective function.
- Let X denote the feasible set for the quasiconvex optimization problem described in a previous page.
- We first recognize that

$$f(y) \leq f(x) \Rightarrow \nabla f(x)^T(y-x) \leq 0$$

  for any quasiconvex differentiable function f.
- It then follows that x is optimal if

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

Quasiconvex optimization via convex feasibility problems

- One general approach to quasiconvex optimization relies on the representation of the sublevel sets of a quasiconvex function via a family of convex inequalities.
- Let $\phi_t : \mathbf{R}^n \to \mathbf{R}$, $t \in \mathbf{R}$, be a family of convex functions that satisfy

$$f_0(x) \leq t \iff \phi_t(x) \leq 0,$$

  and also, for each x, $\phi_t(x)$ is a nonincreasing function of t, i.e., $\phi_s(x) \leq \phi_t(x)$ whenever $s \geq t$.
- Let p^* denote the optimal value of the quasiconvex optimization problem. If the feasibility problem

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

  is feasible, then we have $p^* \leq t$. Otherwise, we have $p^* \geq t$.

Bisection for Quasiconvex Optimization (1/2)

Algorithm 4.1 Bisection method for quasiconvex optimization.

- given $l \leq p^*$, $u \geq p^*$, tolerance $\epsilon > 0$.
  repeat
  1. $t := (l + u)/2$.
  2. Solve the convex feasibility problem

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

  3. If the previous problem is feasible, $u := t$; else $l := t$.

  until $u - l \leq \epsilon$.

Bisection for Quasiconvex Optimization (2/2)

- The interval $[l, u]$ is guaranteed to contain p^*, i.e., we have $l \leq p^* \leq u$ at each step.
- In each iteration the interval is divided in two, i.e., bisected, so the length of the interval after k iterations is $2^{-k}(u - l)$, where $u - l$ is the length of the initial interval.
- It follows that exactly $\lceil \log_2((u - l)/\epsilon) \rceil$ iterations are required before the algorithm terminates.
- Each step involves solving the convex feasibility problem

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

Quasiconvex Optimization Problem – An Example

- Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & ||Ax - b|| \leq \epsilon, \end{array}$$

  where $f_0(x) = \text{length}(x) = \min \{k \mid x_i = 0 \text{ for } i > k\}$. The problem variable is $x \in \mathbf{R}^n$; the problem parameters are $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $\epsilon > 0$.
- This is to find the minimum number of columns of A, taken in order, that can approximate the vector b within ϵ.
- It can be shown to be a quasiconvex optimization problem.
- The bisection algorithm can be applied by finding an appropriate family of functions $\phi_t(x)$ that satisfies

$$f_0(x) \leq t \iff \phi_t(x) \leq 0.$$

Linear Optimization Problems (1/2)

- When the objective and constraint functions are all affine, the problem is called a **linear program (LP)**. A general linear program has the form

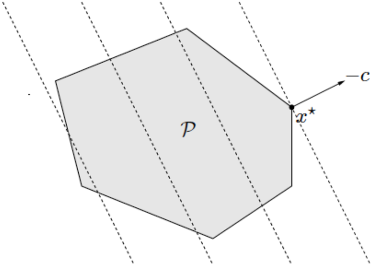
$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{subject to} \quad & Gx \preceq h \\ & Ax = b, \end{aligned}$$

where $G \in \mathbf{R}^{m \times n}$ and $A \in \mathbf{R}^{p \times n}$.

- Linear programs are a special case of convex optimization problems.
- It is common to omit the constant d in the objective function.

Linear Optimization Problems (2/2)

- We also refer to a maximization problem with affine objective and constraint functions as an LP since we can maximize an affine objective $c^T x + d$, by minimizing $-c^T x - d$ (which is still convex).
- The feasible set of an LP is a polyhedron $\mathcal{P}$; the problem is to minimize the affine function $c^T x + d$ over $\mathcal{P}$.



Standard and inequality forms of linear programs

- In a standard form LP the only inequalities are componentwise nonnegativity constraints $x \succeq 0$:

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

  - Some LP algorithms are developed specifically for standard form LP.

- If the LP has no equality constraints, it is called an inequality form LP, usually written as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b. \end{array}$$

Converting LPs to standard form

- In order to transform a general LP to a standard form LP, the first step is to introduce slack variables s_i for the inequalities, which results in

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx + s = h \\ & Ax = b \\ & s \succeq 0. \end{array}$$

- The second step is to express the variable x as $x = x^+ - x^-$, where $x^+, x^- \succeq 0$.
- This yields the problem

$$\begin{array}{ll} \text{minimize} & c^T x^+ - c^T x^- + d \\ \text{subject to} & Gx^+ - Gx^- + s = h \\ & Ax^+ - Ax^- = b \\ & x^+ \succeq 0, x^- \succeq 0, s \succeq 0, \end{array}$$

  which is an LP in standard form, with variables x^+, x^-, and s.
- How to convert LPs into an inequality form?

Examples of Linear Programming – Diet Problem

- A healthy diet contains m different nutrients in quantities at least equal to $b_1, \dots, b_m$. We can compose such a diet by choosing nonnegative quantities $x_1, \dots, x_n$ of n different foods.
- One unit quantity of food j contains an amount a_{ij} of nutrient i, and has a cost of c_j.
- We want to determine the cheapest diet that satisfies the nutritional requirements.
- This problem can be formulated as the LP

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

- Several variations on this problem can also be formulated as LPs.

Example – Chebyshev center of a polyhedron (1/2)

- We consider the problem of finding the largest Euclidean ball that lies in a polyhedron described by linear inequalities,

$$\mathcal{P} = \left\{ x \in \mathbf{R}^n \mid a_i^T x \leq b_i, i = 1, \dots, m \right\}.$$

- The center of the optimal ball is called the **Chebyshev center** of the polyhedron; it is the point deepest inside the polyhedron, i.e., farthest from the boundary;
- We represent the ball as

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

The variables in the problem are the center $x_c \in \mathbf{R}^n$ and the radius r; we wish to maximize r subject to the constraint $B \subseteq \mathcal{P}$.

Example – Chebyshev center of a polyhedron (2/2)

- We start by considering the simpler constraint that B lies in one halfspace $a_i^T x \leq b_i$, i.e.,

$$||u||_2 \leq r \Longrightarrow a_i^T(x_c + u) \leq b_i.$$

- Since

$$\sup\left\{a_i^T u \mid ||u||_2 \leq r\right\} = r||a_i||_2,$$

  we reach a linear inequality in x_c and r:

$$a_i^T x_c + r||a_i||_2 \leq b_i.$$

- Hence the Chebyshev center can be determined by solving the LP

$$\begin{array}{ll} \text{maximize} & r \\ \text{subject to} & a_i^T x_c + r||a_i||_2 \leq b_i, \quad i = 1, ..., m, \\ & r \geq 0 \end{array}$$

  with variables r and x_c.
- It can be shown that the constraint $r \geq 0$ is a redundant constraint.

Chebyshev inequalities (1/2)

- We consider a probability distribution for a discrete random variable x on a set $\{u_1, \dots, u_n\} \subseteq \mathbf{R}$ with n elements.
- We describe the distribution of x by a vector $p \in \mathbf{R}^n$, where $p_i = \mathbf{prob}(x = u_i)$, so p satisfies $p \succeq 0$ and $\mathbf{1}^T p = 1$. Conversely, if p satisfies $p \succeq 0$ and $\mathbf{1}^T p = 1$, then it defines a probability distribution for x.
- We assume that u_i are known and fixed, but the distribution p is not known.
- If f is any function of x, then $\mathbf{E} f = \sum_{i=1}^{n} p_i f(u_i)$ is a linear function of p.
- If $\mathcal{S}$ is any subset of $\mathbf{R}$, then

$$\mathbf{prob}(x \in \mathcal{S}) = \sum_{u_i \in \mathcal{S}} p_i$$

is a linear function of p.

Chebyshev inequalities (2/2)

- We assume to have the following prior knowledge:
  - We know upper and lower bounds on expected values of some functions of x, and probabilities of some subsets of $\mathbf{R}$.
  - It can be expressed as linear inequality constraints on p,

$$\alpha_i \leq a_i^T p \leq \beta_i, i = 1, ..., m.$$

- The problem is to give lower and upper bounds on $\mathbf{E} f_0(x) = a_0^T p$, where f_0 is some function of x.
- To find a lower bound we solve the LP

$$\begin{array}{ll} \text{minimize} & a_0^T p \\ \text{subject to} & p \succeq 0, \mathbf{1}^T p = 1 \\ & \alpha_i \leq a_i^T p \leq \beta_i, i = 1, ..., m, \end{array}$$

with variable p.

Piecewise-linear minimization

- Consider the unconstrained problem of minimizing the piecewise-linear, convex function

$$f(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i).$$

- This problem can be transformed to an equivalent LP by first forming the epigraph problem,

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \max\limits_{i=1,\ldots,m} (a_i^T x + b_i) \leq t, \end{array}$$

- Then, the inequality can be expressed as a set of m separate inequalities:

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

- This is an inequality-form LP, with variables x and t.

Linear-fractional programming

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

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

  where the objective function is given by

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

- The objective function is quasiconvex (and quasilinear) so linear-fractional programs are quasiconvex optimization problems.

Transforming to a linear program

- If the feasible set

$$\left\{x \mid Gx \preceq h, Ax = b, e^T x + f > 0\right\}$$

is nonempty, the linear-fractional program can be shown to be equivalent to a linear program

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

with variables y, z.

- Proof idea: Let

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

Solving Linear Programming Problems

- The Simplex method.
  - Developed by Dantzig in 1947.
  - One of the top 10 algorithms of the 20th century.
  - Usually very efficient for practical applications. Average-case performance: $\mathcal{O}(n^3)$.
  - Worst-case performance (though rarely happens): $\mathcal{O}(2^n)$.
- Interior-point methods
  - Developed since the late 70s' [Khachiyan1979, Karmarkar1984] with worst-case performance $\mathcal{O}(n^4)$, $\mathcal{O}(n^{3.5})$, respectively.
  - The average-case performance is still not better than the Simplex method.

Quadratic programming

Quadratic programming

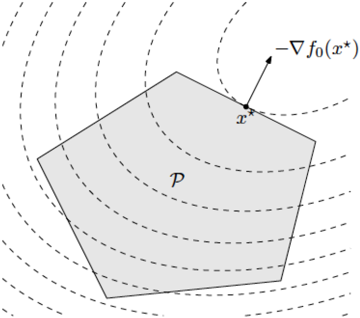
The convex optimization problem is called a **quadratic program (QP)** if the objective function is convex quadratic, and the constraint functions are affine, as expressed in the form

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T P x + q^T x + r \\ \text{subject to} & Gx \preceq h \\ & Ax = b, \end{array}$$

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

Quadratic optimization problems

- In a quadratic program (QP), we minimize a convex quadratic function over a polyhedron.



- Quadratic programs include linear programs as a special case, by taking $P = 0$.

Quadratic optimization problems

- If the objective as well as the inequality constraint functions are convex and quadratic, as in

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, i = 1, \ldots, m \\ & Ax = b, \end{array}$$

  where $P_i \in \mathbf{S}_+^n, i = 0, 1 \ldots, m$, the problem is called a **quadratically constrained quadratic program (QCQP)**.
- In a QCQP, we minimize a convex quadratic function over a feasible region that is the intersection of ellipsoids (when $P_i \succ 0$).
- QCQPs include QPs as a special case, by taking $P_i = 0$, for $i = 1, \ldots, m$. If $P_0 = 0$, it further reduces to LPs.

QP Examples

- Least-squares and regression
- Distance between polyhedra
- Bounding variance
- Linear program with random cost
- Markowitz portfolio optimization

QP Examples – Least-squares and regression

- The problem of minimizing the convex quadratic function

$$||Ax - b||_2^2 = x^T A^T Ax - 2b^T Ax + b^T b$$

  is an (unconstrained) QP.
- It arises in many fields and has many names, e.g., regression analysis or least-squares approximation.
- This problem is simple enough to have the well known analytical solution $x = A^\dagger b$ ($A^\dagger$ is the pseudo-inverse of A).

Least-squares and regression with linear constraints

- For least-squares problems, with linear inequality constraints added, is called constrained regression or constrained least-squares, and there is no longer a simple analytical solution.
- As an example we can consider regression with lower and upper bounds on the variables, i.e.,

$$\begin{array}{ll} \text{minimize} & ||Ax - b||_2^2 \\ \text{subject to} & l_i \leq x_i \leq u_i, i = 1, ..., n, \end{array}$$

which is a QP.

Distance between polyhedra (1/2)

- We define the Euclidean distance between the polyhedra $P_1 = \{x | A_1 x \preceq b_1\}$ and $P_2 = \{x \mid A_2 x \preceq b_2\}$ in $\mathbf{R}^n$ as

$$\mathrm{dist}(P_1, P_2) = \inf \{||x_1 - x_2||_2 \mid x_1 \in P_1, x_2 \in P_2\}.$$

- If the polyhedra intersect, the distance is zero.
- To find the distance between P_1 and P_2, we can solve the QP

$$\begin{array}{ll} \text{minimize} & ||x_1 - x_2||_2^2 \\ \text{subject to} & A_1 x_1 \preceq b_1 \\ & A_2 x_2 \preceq b_2, \end{array}$$

with variables $x_1, x_2 \in \mathbf{R}^n$.

Distance between polyhedra (2/2)

- This problem is infeasible if and only if one of the polyhedra is empty.
- The optimal value is zero if and only if the polyhedra intersect, in which case the optimal point satisfies $x_1 = x_2 \in P_1 \cap P_2$.
- Otherwise the optimal x_1 and x_2 are the points in P_1 and P_2, respectively, that are closest to each other.

Bounding variance

- Consider again the Chebyshev inequalities example, where the variable is an unknown probability distribution given by $p \in \mathbf{R}^n$, about which we have some prior information.
- The variance of a random variable $f(x)$ is given by

$$\mathbf{E}f^2 - (\mathbf{E}f)^2 = \sum_{i=1}^{n} f_i^2 p_i - \left(\sum_{i=1}^{n} f_i p_i\right)^2,$$

  (where $f_i = f(u_i)$), which is a concave quadratic function of p.
- It follows that we can maximize the variance of $f(x)$, subject to the given prior information, by solving the QP

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{n} f_i^2 p_i - (\sum_{i=1}^{n} f_i p_i)^2 \\ \text{subject to} & p \succeq 0, \mathbf{1}^T p = 1 \\ & \alpha_i \leq a_i^T p \leq \beta_i, i = 1, ..., m. \end{array}$$

- The optimal value gives the maximum possible variance of $f(x)$, over all distributions that are consistent with the prior information; the optimal p gives a distribution that achieves this maximum variance.

Linear program with random cost (1/2)

- We consider an LP,

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Gx \preceq h \\ & Ax = b, \end{aligned}$$

  with variable $x \in \mathbf{R}^n$.
- We suppose that the cost function (vector) $c \in \mathbf{R}^n$ is random, with mean value $\bar{c}$ and covariance $E(c - \bar{c})(c - \bar{c})^T = \Sigma$.
  - For simplicity we assume that the other problem parameters are deterministic.
- For a given $x \in \mathbf{R}^n$, the cost $c^T x$ is a (scalar) random variable with mean $\mathbf{E} c^T x = \bar{c}^T x$ and variance

$$\mathbf{var}(c^T x) = \mathbf{E}(c^T x - \bar{c}^T x)^2 = x^T \Sigma x.$$

Linear program with random cost (2/2)

- In general there is a trade-off between small expected cost and small cost variance.
- One way to take variance into account is to minimize a linear combination of the expected value and the variance of the cost, i.e., $\mathbf{E}c^T x + \gamma \mathbf{var}(c^T x)$, which is called the risk-sensitive cost.
- The parameter $\gamma \geq 0$ is called the risk-aversion parameter, since it sets the relative values of cost variance and expected value. (For $\gamma > 0$, we are willing to trade off an increase in expected cost for a sufficiently large decrease in cost variance).
- To minimize the risk-sensitive cost we solve the QP

$$\begin{array}{ll} \text{minimize} & \bar{c}^T x + \gamma x^T \Sigma x \\ \text{subject to} & Gx \preceq h \\ & Ax = b. \end{array}$$

Markowitz portfolio optimization (1/2)

- We consider a classical portfolio problem with n assets or stocks held over a period of time.
- We let x_i denote the amount of asset i held throughout the period, with x_i in dollars, at the price at the beginning of the period.
- We let p_i denote the relative price change of asset i over the period.
- The overall return on the portfolio is $r = p^T x$ (dollars).
- The optimization variable is the portfolio vector $x \in \mathbf{R}^n$.

Markowitz portfolio optimization (2/2)

- We take a stochastic model for price changes: $p \in \mathbf{R}^n$ is a random vector, with known mean $\bar{p}$ and covariance Σ. Therefore with portfolio $x \in \mathbf{R}^n$, the return r is a (scalar) random variable with mean $\bar{p}^T x$ and variance $x^T \Sigma x$.
- The choice of portfolio x involves a trade-off between the mean of the return, and its variance.
- The classical portfolio optimization problem, introduced by Markowitz, is the QP

$$\begin{array}{ll} \text{minimize} & x^T \Sigma x \\ \text{subject to} & \bar{p}^T x \geq r_{min} \\ & \mathbf{1}^T x = 1, \\ & x \succeq 0, \end{array}$$

where x, the portfolio, is the variable.

Second-order cone programming

- A problem that is closely related to quadratic programming is the **second-order cone program (SOCP)**:

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

  where $x \in \mathbf{R}^n$ is the optimization variable, $A_i \in \mathbf{R}^{n_i \times n}$, and $F \in \mathbf{R}^{p \times n}$.
- We call a constraint of the form

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

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

Second-order cone programming

Second-order cone programming (SOCP)

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

- When $c_i = 0, i = 1, ..., m$, the SOCP is equivalent to a QCQP (which is obtained by squaring each of the constraints).
- Similarly, if $A_i = 0, i = 1, ..., m$, then the SOCP reduces to a (general) LP.
- Second-order cone programs are more general than QCQPs (and of course, LPs).

SOCP Examples – Robust linear programming (1/2)

- We consider a linear program in inequality form,

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

  in which there is some uncertainty or variation in the parameters c, a_i, b_i.
- As an example, we assume that c and b_i are fixed, and that a_i are known to lie in given ellipsoids:

$$a_i \in \mathcal{E}_i = \{\bar{a}_i + P_i u \mid ||u||_2 \leq 1\},$$

  where $P_i \in \mathbf{R}^{n \times n}$. (If P_i is singular we obtain 'flat' ellipsoids, of dimension rank P_i; $P_i = 0$ means that a_i is known perfectly.)
- We will require that the constraints be satisfied for all possible values of the parameters a_i, which leads us to the robust linear program

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

SOCP Examples – Robust linear programming (2/2)

- The robust linear constraint, $a_i^T x \leq b_i$ for all $a_i \in \mathcal{E}_i$, can be expressed as

$$\sup \left\{ a_i^T x \mid a_i \in \mathcal{E}_i \right\} \leq b_i.$$

- The lefthand side can be expressed as

$$\sup \left\{ a_i^T x \mid a_i \in \mathcal{E}_i \right\} = \bar{a}_i^T x + \sup \left\{ u^T P_i^T x \mid ||u||_2 \leq 1 \right\} = \bar{a}_i^T x + ||P_i^T x||_2.$$

- Thus, the robust LP can be expressed as the SOCP

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

  where the robust linear constraint becomes a second-order cone constraint.
- Note that the additional norm terms act as regularization terms; they prevent x from being large in directions with considerable uncertainty in the parameters a_i.

Linear programming with random constraints (1/2)

- We consider the aforementioned robust LP in a statistical framework.
- Suppose that the parameters a_i are independent Gaussian random vectors, with mean $\bar{a}_i$ and covariance Σ_i.
- We require that each constraint $a_i^T x \le b_i$ should hold with a probability (or confidence) exceeding η , where $\eta \ge 0.5$, i.e., $\textbf{prob}(a_i^T x \le b_i) \ge \eta$.
- Letting $u = a_i^T x$, with σ^2 denoting its variance, this constraint can be written as

$$\textbf{prob}\left(\frac{u-\bar{u}}{\sigma} \le \frac{b_i - \bar{u}}{\sigma}\right) \ge \eta.$$

- Since $(u-\bar{u})/\sigma$ is a zero mean unit variance Gaussian variable, the probability above is simply $\Phi((b_i - \bar{u})/\sigma)$, where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

is the cumulative distribution function of a zero mean unit variance Gaussian random variable.

Linear programming with random constraints (2/2)

- Thus the probability constraint

$$\textbf{prob}\left(\frac{u-\bar{u}}{\sigma} \le \frac{b_i-\bar{u}}{\sigma}\right) \ge \eta.$$

can be expressed as

$$\frac{b_i-\bar{u}}{\sigma} \ge \Phi^{-1}(\eta).$$

- With $u = a_i^T x$ and $\sigma = (x^T \Sigma_i x)^{1/2}$, the problem

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

can be expressed as the SOCP

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

where $\Phi^{-1}(\eta) \ge 0$ since $\eta \ge 1/2$.

Portfolio optimization with loss risk constraints (1/2)

- We consider again the classical Markowitz portfolio problem, and assume that the price change vector $p \in \mathbf{R}^n$ is a Gaussian random variable, with mean $\bar{p}$ and covariance Σ.
- Therefore the return r is a Gaussian random variable with mean $\bar{r} = \bar{p}^T x$ and variance $\sigma_r^2 = x^T \Sigma x$.
- Consider a loss risk constraint of the form $\mathbf{prob}(r \leq \alpha) \leq \beta$, where α is a given unwanted return level (e.g., a large loss) and β is a given maximum probability.
- This inequality is equivalent to

$$\bar{p}^T x + \Phi^{-1}(\beta) ||\Sigma^{1/2} x||_2 \geq \alpha$$

where Φ is the cumulative distribution function of a unit Gaussian random variable.

Portfolio optimization with loss risk constraints (2/2)

- The problem of maximizing the expected return subject to a bound on the loss risk (with $\beta \le 1/2$), can be cast as an SOCP:

$$\begin{array}{ll} \text{maximize} & \bar{p}^T x \\ \text{subject to} & \bar{p}^T x + \Phi^{-1}(\beta) \|\Sigma^{1/2} x\|_2 \ge \alpha \\ & x \succeq 0, \\ & \mathbf{1}^T x = 1 \end{array}$$

  since $\Phi^{-1}(\beta) \le 0$ under the assumption that $\beta \le 1/2$.
  - If $\beta > 1/2$, the loss risk constraint becomes nonconvex in x.
- There may be many extensions of this problem. For example, we can impose several loss risk constraints, i.e.,

$$\textbf{prob}(r \le \alpha_i) \le \beta_i, i = 1, ..., k,$$

  (where $\beta_i \le 1/2$).