Convex Optimization (II)

Lecture 7, Convex Optimization

National Taiwan University

April 15, 2021

Table of contents

1. Quadratic optimization problems (§4.4)
    - Quadratic programming (QP)
    - Quadratically constrained quadratic programs (QCQP)
    - Second-order cone programming (SOCP)
    - More General Conic-Form Problems

Quadratic programming

Quadratic programming

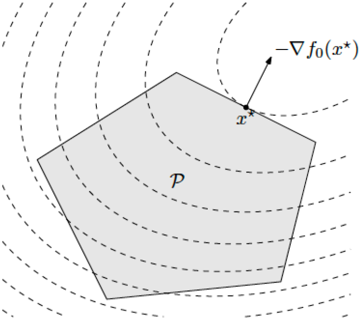
The convex optimization problem is called a **quadratic program (QP)** if the objective function is convex quadratic, and the constraint functions are affine, as expressed in the form

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

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

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

  where $P_i \in \mathbf{S}_+^n, i = 0, 1..., m$, 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 $i = 1, ..., m$. If $P_0 = 0$, it further reduces to LPs.

QP Examples

- Least-squares and regression
- Distance between polyhedra
- Bounding variance
- Chebyshev Inequality
- 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 a least-squares problems, with linear inequality constraints added, it 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.,

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

which is a QP.

Distance between polyhedra (1/2)

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

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

- 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.

Chebyshev inequalities (1/3)

- 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 (2/3)

- 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,

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

- 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

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

with variable p.

Chebyshev inequalities (3/3)

- The optimal value of this LP gives the lowest possible value of $\mathbf{E} f_0(X)$ for any distribution that is consistent with the prior information.
- Moreover, the bound is sharp: the optimal solution gives a distribution that is consistent with the prior information and achieves the lower bound.
- In a similar way, we can find the best upper bound by maximizing $a_0^T p$ subject to the same constraints.

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,

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

  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.
- One form of 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.

Markowitz portfolio optimization (3/3)

- Here we find the portfolio that minimizes the return variance (which is associated with the risk of the portfolio) subject to achieving a minimum acceptable mean return r_{min}, and satisfying the portfolio budget and no-shorting constraints.
- Another form can be to maximize the mean of the return subject to a constraint on the variance.

Second-order cone programming

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

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

  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
$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + ||P_i^T x||_2 \leq b_i, i = 1, \ldots, m. \end{array}$$
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 $a_i^T x \leq b_i$ should hold with a probability (or confidence) exceeding η , where $\eta \geq 0.5$, i.e., $\mathbf{prob}(a_i^T x \leq b_i) \geq \eta$.
- Letting $u = a_i^T x$, with σ^2 denoting its variance, this constraint can be written as

$$\mathbf{prob}\left(\frac{u - \bar{u}}{\sigma} \leq \frac{b_i - \bar{u}}{\sigma}\right) \geq \eta.$$

- 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

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

can be expressed as

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

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

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

can be expressed as the SOCP

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

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

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 $\beta \leq 1/2$), can be cast as an SOCP:

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

since $\Phi^{-1}(\beta) \leq 0$ under the assumption that $\beta \leq 1/2$.
  - 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.,

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

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

More General Conic-Form Problems – Beyond SOCP (1/2)

- Note that in the second-order cone problem

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

the second-order cone constraints $||A_i x + b_i||_2 \leq c_i^T x + d_i$ can be expressed as: $(A_i x + b_i, c_i^T x + d_i) \in K_i$ where

$$K_i = \left\{ (y, t) \in \mathbf{R}^{n_i + 1} \mid ||y||_2 \leq t \right\},$$

i.e., the second-order cone in $\mathbf{R}^{n_i+1}$.

- More generally, we can write the SOCP as

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & P_i x + q_i \in K_i \quad i = 1, \ldots, m \\ & Fx = g \end{array}$$

where $P_i = \begin{bmatrix} A_i \\ c_i^T \end{bmatrix} \in \mathbf{R}^{(n_i+1) \times n}$ and $q_i = \begin{bmatrix} b_i \\ d_i \end{bmatrix} \in \mathbf{R}^{(n_i+1)}$

More General Conic-Form Problems – Beyond SOCP (2/2)

- In the problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Px + q \in K \\ & Fx = g, \end{array}$$

  if the set K is chosen as a cone that satisfies certain properites, to be defined in a subsequent lecture, then it may lead to a more general convex optimization problem.
- For example, $K = \left\{(y, t) \in \mathbf{R}^{m+1} \mid ||y||_2 \leq t\right\}$, leads to an SOCP.
- If we choose $K = \mathbf{R}^m_+$, then it reduces to an LP.
- Suppose we choose $K = \mathbf{S}^m_+$, the positive semidefinite cone, and rewrite the problem as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \displaystyle\sum_{i=1}^{m} x_i P_i + Q \in K \\ & Fx = g, \end{array}$$

  where $P_i, Q \in \mathbf{S}^m$. Then it becomes a semidefinite program, to be studied in details sometime later.