Convex Optimization

A supplementary note to Chapter 4 of *Convex Optimization* by S. Boyd and L. Vandenberghe

Optimization Lab.

IE department
Seoul National University

4th October 2009

Optimization:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \cdots, m, \\ & h_j(x) = 0, \quad j = 1, \cdots, p. \end{aligned}$$

- *Decision variables*: $x \in \mathbb{R}^n$
- *Objective function*: $f_0(x)$
- *Constraints* (inequality and equality): $f_i(x) \leq 0$ and $h_j(x) = 0$.
- A point $x \in \mathcal{D}$ is *feasible* if it satisfies all constraints; infeasible, otherwise where $\mathcal{D}$ is *domain of problem*, $\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{j=1}^{p} \operatorname{dom} h_j$.

- *Optimal value*: $p^* = \inf\{f_0(x) | f_i(x) \leq 0, \forall i, h_j(x) = 0, \forall j\}$.
- *Optimal solution*: A feasible x^* is optimal if $f_0(x^*) = p^*$. Denote by X_{opt} the set of optimal points.
- A feasible x is ϵ-*optimal* if $f_0(x) \leq p^* + \epsilon$.
- A feasible x is a *local optimum* if $\exists R > 0$ s.t. $f_0(x) = \inf\{f_0(z) | f_i(x) \leq 0,\ \forall i,\ h_j(x) = 0,\ \forall j,\ \|z - x\|_2 \leq R\}$.
- The *feasibility problem*: Find an x satisfying all the constraints.

Expressing problems in standard form

- Min-version

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \quad i = 1, \cdots, m, \\ & h_j(x) = 0, \quad j = 1, \cdots, p. \end{array}$$

- Max-version

$$\begin{array}{ll} \max & f_0(x) \\ \text{s.t.} & f_i(x) \geq 0, \quad i = 1, \cdots, m, \\ & h_j(x) = 0, \quad j = 1, \cdots, p. \end{array}$$

Equivalent problems

We say two problems are *equivalent* if, given a solution of one, we can efficiently find a solution of the other, and vice versa.

Example

Two problems are equivalent if $\alpha_i > 0$, $\forall i$, $\beta_j \neq 0$, $\forall j$:

$$\begin{array}{lll|lll} \text{min} & f_0(x) & & \text{min} & \alpha_0 f_0(x) & \\ \text{s.t.} & f_i(x) \le 0, & \forall i & \text{s.t.} & \alpha_i f_i(x) \le 0, & \forall i \\ & h_j(x) = 0, & \forall j & & \beta_j h_j(x) = 0, & \forall j \end{array}$$

Change of variables

Suppose $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one mapping. Define $\tilde{f}_i(x) = f_i(\phi(x))$ and $\tilde{h}_j(x) = h_j(\phi(x))$ Then the two problems are equivalent:

$$\begin{array}{ll} \min & f_0(x) \\ s.t. & f_i(x) \le 0, \ \ i = 1, \ldots, m \\ & h_i(x) = 0, \ \ i = 1, \ldots, p, \end{array} \qquad \begin{array}{ll} \min & \tilde{f}_0(x) \\ s.t. & \tilde{f}_i(x) \le 0, \ \ i = 1, \ldots, m \\ & \tilde{h}_i(x) = 0, \ \ i = 1, \ldots, p. \end{array}$$

Why?

Transformation of objective and constraints

Suppose $\psi_0 : \mathbb{R} \to \mathbb{R}$ is monotone increasing, $\psi_1, \dots, \psi_m : \mathbb{R} \to \mathbb{R}$ satisfy $\psi_i(u) \le 0$ iff $u \le 0$, and $\psi_{m+1}, \dots, \psi_{m+p} : \mathbb{R} \to \mathbb{R}$ satisfy $\psi_i(u) = 0$ iff $u = 0$. Define $\tilde{f}_i(x) = \psi_i(f_i(x))$, $i = 0, \dots, m$ and $\tilde{h}_i(x) = \psi_{m+i}(h_i(x))$, $i = 1, \dots, p$. Then following two are equivalent:

$$\begin{array}{ll} \min & f_0(x) \\ s.t. & f_i(x) \le 0, \;\; i = 1, \dots, m \\ & h_i(x) = 0, \;\; i = 1, \dots, p, \end{array} \qquad \begin{array}{ll} \min & \tilde{f}_0(x) \\ s.t. & \tilde{f}_i(x) \le 0, \;\; i = 1, \dots, m \\ & \tilde{h}_i(x) = 0, \;\; i = 1, \dots, p. \end{array}$$

Least-norm and least-norm-squared

$$\min\ \|Ax - b\|_2 \text{ versus } \min\ \|Ax - b\|_2^2.$$

Slack variables

- $f_i(x) \leq 0$ if and only if $\exists\ s_i \geq 0$ such that $f_i(x) + s_i = 0$; s_i is called a slack variable.
- The followings are equivalent:

$$\begin{array}{ll} \text{min} & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0,\ \ i = 1, \ldots, m \\ & h_i(x) = 0,\ \ i = 1, \ldots, p, \end{array}$$

$$\begin{array}{ll} \text{min} & f_0(x) \\ \text{s.t.} & f_i(x) + s_i = 0,\ \ i = 1, \ldots, m \\ & h_i(x) = 0,\ \ i = 1, \ldots, p \\ & s_i \geq 0,\ \ i = 1, \ldots, m. \end{array}$$

Introducing equality constraints

- The following two problems are equivalent:

$$
\begin{array}{lll}
\min & f_0(A_0x + b_0) & \\
\text{s.t.} & f_i(A_ix + b_i) \leq 0, & \forall i, \\
& h_j(x) = 0, & \forall j,
\end{array}
\qquad
\begin{array}{lll}
\min & f_0(y_0) & \\
\text{s.t.} & f_i(y_i) \leq 0, & \forall i, \\
& y_i = Ax_i + b_i & \forall i, \\
& h_j(x) = 0 & \forall j.
\end{array}
$$

Optimizing over some variables

Since $\inf_{x,y} f(x,y) = \inf_{x} \tilde{f}(x)$, where $\tilde{f}(x) = \inf_{y} f(x,y)$, following two are equivalent:

$$\begin{array}{lll} \min & f_0(x_1,x_2) & \\ \text{s.t.} & f_i(x_1) \le 0, & i=1,\cdots,m_1 \\ & g_j(x_2) \le 0, & j=1,\cdots,m_2 \end{array} \quad \Leftrightarrow \quad \begin{array}{lll} \min & \tilde{f}_0(x_1) & \\ \text{s.t.} & f_i(x_1) \le 0, & i=1,\cdots,m_1 \end{array}$$

where $\tilde{f}_0(x_1) = \inf\{f_0(x_1,z) | g_j(z) \le 0, j=1,\cdots,m_2\}$.

Example

Consider a strictly convex quadratic program constrained on some variables: $\min x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2$ s.t. $f_i(x_1) \le 0,\ i=1,\ldots,m$. Since, $\inf_{x_2} x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2 = x_1^T (P_{11} - P_{12} P_{22}^{-1} P_{12}^T) x_1$, we can obtain equivalent problem:

$$\min\ x_1^T (P_{11} - P_{12} P_{22}^{-1} P_{12}^T) x_1, \text{s.t. } f_i(x_1) \le 0,\ i=1,\ldots,m.$$

Epigraph problem form

- The following two problems are equivalent:

$$
\begin{array}{lll}
\min & f_0(x) & \\
\text{s.t.} & f_i(x) \le 0, & i = 1, \cdots, m \\
& h_j(x) = 0, & j = 1, \cdots, p
\end{array}
\qquad
\begin{array}{lll}
\min & t & \\
\text{s.t.} & f_0(x) - t \le 0, & i = 1, \cdots, m_1 \\
& f_i(x) \le 0, & i = 1, \cdots, m \\
& h_j(x) = 0, & j = 1, \cdots, p
\end{array}
$$

- Min-version

$$\begin{array}{llll} \min & f_0(x) & & f_0 \text{ is convex.} \\ \text{s.t.} & f_i(x) \leq 0, & i = 1, \ldots, m, & f_i \text{ are convex.} \\ & a_j^T x = b_j, & j = 1, \ldots, p. & \end{array}$$

- Max-version

$$\begin{array}{llll} \max & f_0(x) & & f_0 \text{ is concave.} \\ \text{s.t.} & f_i(x) \geq 0, & i = 1, \ldots, m, & f_i \text{ are concave.} \\ & a_j^T x = b_j, & j = 1, \ldots, p. & \end{array}$$

- Similarly, can define quasiconvex optimization problem: 'convex' ('concave') is replaced by 'quasi-convex' ('quasi-concave, resp.).

Local and global optima

Theorem

Any local optimum of convex optimization problems is also a global optimum.

Proof Let x be a local optimum: $\exists\ R > 0$ s.t. $f_0(x) \leq f_0(z)\ \forall$ feasible z such that $\|z - x\|_2 \leq R\}$. Suppose, on the contrary, $\exists\ z \in \mathcal{D}$ such that $f(z) < f(x)$. Then, $\exists y$ such that $\|y - x\|_2 < R$ and $y = \lambda x + (1 - \lambda)z$ for some $0 < \lambda < 1$. Since $f(y) \geq f(x) > f(z)$, $f(y) > \lambda f(x) + (1 - \lambda) f(z)$. A contradiction to convexity of f_0. $\square$

Remark

Not necess. true for quasiconvex minimization.

Theorem

For convex minimization with differentiable f_0, feasible x is optimal iff $\nabla f_0(x)^T(y-x) \geq 0$ for any feasible y.

Proof

"If." For any feasible y we have $f_0(y) \geq f_0(x) + \nabla f_0(x)^T(y-x) \geq f_0(x)$.
"Only if." Suppose not: $\exists y$ feasible such that $\nabla f_0(x)^T(y-x) < 0$.
For $\lambda \in [0,1]$, let $g(\lambda) = f_0((1-\lambda)x + \lambda y)$. Then,

$$\frac{d}{d\lambda} g(\lambda)\Big|_{\lambda=0} = \nabla f_0(x)^T(y-x) < 0,$$

which implies that for small enough $\lambda > 0$, we have $g(\lambda) < g(0)$. A contradiction to optimality of x. $\square$

Corollary

For unconstrained convex minimization, x is optimal iff x is feasible and $\nabla f_0(x) = 0$.

Corollary

For convex minimization with equality constraints $Ax = b$ only, x is optimal iff $\exists \lambda$ s.t. $A^T \lambda = \nabla f_0(x)$, where $\lambda \in \mathbb{R}^p$.

Proof

$$
\begin{aligned}
x \text{ optimal} \quad &\iff \quad \forall y \text{ s.t. } Ay = b, \nabla f_0(x)^T (y - x) \geq 0 \\
&\iff \quad \nabla f_0(x)^T z \geq 0, \forall z \in \mathcal{N}(A), \text{ null space of } A \\
&\iff \quad \nabla f_0(x)^T z = 0, \forall z \in \mathcal{N}(A) \\
&\iff \quad \nabla f_0(x) \perp \mathcal{N}(A) \iff \nabla f_0(x) \in \mathcal{R}(A), \text{ row space of } A \\
&\iff \quad \exists \lambda \text{ s.t. } A^T \lambda = \nabla f_0(x), \lambda \in \mathbb{R}^p. \square
\end{aligned}
$$

Example

Consider the problem

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & x \geq 0 \end{array}$$

- Optimality condition : $x \geq 0$, $\nabla f_0(x)^T(y - x) \geq 0$ for all $y \geq 0$.
- $\nabla f_0(x)^T y$ is unbounded below on $y \geq 0$ unless $\nabla f_0(x) \geq 0$, which implies $-\nabla f_0(x)^T x \geq 0$.
- But, as $x \geq 0$ and $\nabla f_0(x) \geq 0$, we get $\nabla f_0(x)^T x = 0$.
- Thus, optimality condition becomes

$$x \geq 0, \quad \nabla f_0(x) \geq 0, \; x_i(\nabla f_0(x))_i = 0, \; \text{complementary condition.}$$

$$
\begin{array}{lll}
\min & f_0(x) & f_0 \text{ is quasiconvex,} \\
\text{s.t.} & f_i(x) \leq 0, \ i = 1, \ldots, m, & f_i\text{'s are convex,} \\
& Ax = b. &
\end{array}
$$

1. It may have local optima that are not global optima.
2. A sufficient optimality condition: If x is feasible and $\nabla f_0(x)^T(y - x) > 0$, $\forall$ feasible $y \neq x$, then x is globally optimal. (Why?)

Quasiconvex optim. via sequence of convex feasibility

Let $\phi_t : \mathbb{R}^n \to \mathbb{R}, t \in \mathbb{R}$ such that

$$f_0(x) \le t \iff \phi_t(x) \le 0 \text{ and } \phi_s(x) \le \phi_t(x) \text{ for } s \ge t.$$

(For instance, $\phi_t(x) = f_0(x) - t$ satisfies conditions.) Then, optimal value p^* can be computed as follows:

$$\begin{array}{lll|l} \text{find} & x & & \text{If the problem is feasible, then } p^* \le t = UB. \\ \text{s.t.} & \phi_t(x) \le 0 & & \text{Otherwise, } p^* \ge t = LB. \\ & f_i(x) \le 0, & \forall i & t \leftarrow (UB + LB)/2 \\ & Ax = b & & \text{Repeat until } UB - LB \le \epsilon. \end{array}$$

Note that it requires $\lceil \log_2((UB - LB)/\epsilon) \rceil$ iterations for an ϵ-suboptimal solution.

- Quadratic program (QP) minimizes convex quadratic over polyhedron.

$$\begin{array}{ll} \min & \frac{1}{2}x^T Px + q^T x + r \\ \text{s.t.} & Gx \leq h \\ & Ax = b, \end{array}$$

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

- Quadratically constrained quadratic program (QCQP) minimizes convex quadratic over intersection of ellipsoids.

$$\begin{array}{ll} \min & \frac{1}{2}x^T P_0 x + q_0^T x + r_0 \\ \text{s.t.} & \frac{1}{2}x^T P_i x + q_i^T x + r_0 \leq 0,\ i = 1, \cdots, m \\ & Ax = b, \end{array}$$

where $P_i \in \mathbb{S}^n_+$, $i = 0, 1, \cdots, m$.

- QCQP $\supseteq$ QP $\supseteq$ LP.

Minimizing Euclidean norm of affine functions

- Least-squares and regression

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

Analytical solution is $x = A^\dagger b$. But, if linear inequality constraints are added, e.g.

$$\begin{array}{ll} \min & \|Ax - b\|_2^2 \\ \text{s.t.} & l_i \le x_i \le u_i, i = 1, \cdots, n, \end{array}$$

problem does not have analytical solution and solvable via QP.

- Distance between polyhedra For $\mathcal{P}_1 = \{x | A_1 x \le b_1\}, \mathcal{P}_2 = \{x | A_2 x \le b_2\}$,

$$\text{dist}(\mathcal{P}_1, \mathcal{P}_2) = \inf\{\|x_1 - x_2\|_2 | x_1 \in \mathcal{P}_1, x_2 \in \mathcal{P}_2\},$$

can be computed via following QP:

$$\begin{array}{ll} \min & \|x_1 - x_2\|_2^2 \\ \text{s.t.} & A_1 x \le b_1 (x_1 \in \mathcal{P}_1) \\ & A_2 x \le b_2 (x_2 \in \mathcal{P}_2). \end{array}$$

Linear program with random cost

Suppose that $c \in \mathbb{R}^n$ is random with mean $\bar{c}$ and covariance Σ. Then,

$$\mathrm{E}(c^T x) = \bar{c}^T x, \ \mathrm{Var}(c^T x) = \mathrm{E}(c^T x - \bar{c}^T x)^2 = x^T \Sigma x$$

A possible problem is to minimize a weighted sum of expected cost and uncertainty cost. To do so, we can consider γ, *risk-sensitive* cost so that the problem is formulated as follows:

$$\begin{array}{ll} \min & \bar{c}^T x + \gamma x^T \Sigma x \\ \text{s.t.} & Gx \leq h \\ & Ax = b. \end{array}$$

Markowitz portfolio optimization

- Let x_i be amount of asset or stock $i = 1, \ldots, n$; $x_i > 0 \leftrightarrow$ long position in asset i, $x_i < 0 \leftrightarrow$ short position in asset i.
- Let p_i be increase of price of i during a period; we assume p has mean vector $\bar{p}$ and covariance matrix Σ.
- Classical model minimizes risk guaranteeing a return with no shorting:

$$\begin{array}{ll} \min & x^T \Sigma x \\ \text{s.t.} & \bar{p}^T x \geq r_{\min}, \\ & \mathbf{1}^T x = 1, \; x \geq 0. \end{array}$$

- Various extensions are possible.

Homework

4.1, 4.2, 4.3, 4.5, 4.8, 4.13, 4.20, 4.21, 4.22