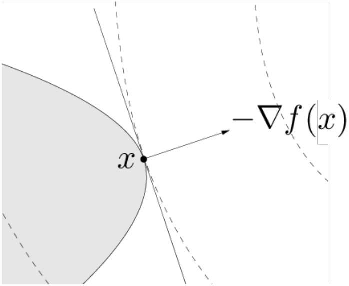
Canonical Problem Forms

Ryan Tibshirani
Convex Optimization 10-725/36-725

Last time: optimization basics

- Optimization terminology (e.g., criterion, constraints, feasible points, solutions)
- Properties and first-order optimality

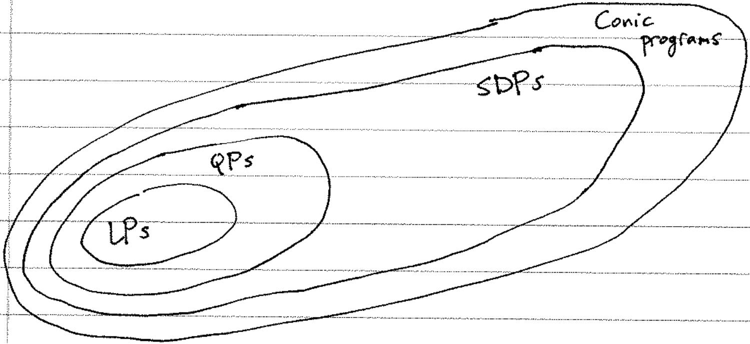


- Equivalent transformations (e.g., partial optimization, change of variables, eliminating equality constraints)

Outline

Today:

- Linear programs
- Quadratic programs
- Semidefinite programs
- Cone programs


Conic programs
SDPs
QPs
LPs

Linear program

A linear program or LP is an optimization problem of the form

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Dx \leq d \\ & Ax = b \end{aligned}$$

Observe that this is always a convex optimization problem

- First introduced by Kantorovich in the late 1930s and Dantzig in the 1940s
- Dantzig's simplex algorithm gives a direct (noniterative) solver for LPs (later in the course we'll see interior point methods)
- Fundamental problem in convex optimization. Many diverse applications, rich history

Example: diet problem

Find cheapest combination of foods that satisfies some nutritional requirements (useful for graduate students!)

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{subject to} \quad & Dx \geq d \\ & x \geq 0 \end{aligned}$$

Interpretation:

- c_j : per-unit cost of food j
- d_i : minimum required intake of nutrient i
- D_{ij} : content of nutrient i per unit of food j
- x_j : units of food j in the diet

Example: transportation problem

Ship commodities from given sources to destinations at minimum cost

$$
\begin{array}{ll}
\min_x & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\
\text{subject to} & \sum_{j=1}^{n} x_{ij} \leq s_i, \; i = 1, \ldots, m \\
& \sum_{i=1}^{m} x_{ij} \geq d_j, \; j = 1, \ldots, n, \;\; x \geq 0
\end{array}
$$

Interpretation:

- s_i : supply at source i
- d_j : demand at destination j
- c_{ij} : per-unit shipping cost from i to j
- x_{ij} : units shipped from i to j

Example: basis pursuit

Given $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$ with $p > n$. Suppose we seek the sparsest solution to underdetermined system of equations $X\beta = y$

Nonconvex formulation:

$$\begin{aligned} \min_{\beta} \quad & \|\beta\|_0 \\ \text{subject to} \quad & X\beta = y \end{aligned}$$

ℓ_1 approximation, often called basis pursuit:

$$\begin{aligned} \min_{\beta} \quad & \|\beta\|_1 \\ \text{subject to} \quad & X\beta = y \end{aligned}$$

Basis pursuit is a linear program. Reformulation:

$$\begin{array}{ll} \min_{\beta} & \|\beta\|_1 \\ \text{subject to} & X\beta = y \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \min_{\beta,z} & 1^T z \\ \text{subject to} & z \geq \beta \\ & z \geq -\beta \\ & X\beta = y \end{array}$$

(Check that this makes sense to you)

Example: Dantzig selector

Modification of previous problem, but allowing for $X\beta \approx y$ (not enforcing exact equality), the Dantzig selector:[1]

$$\begin{aligned} \min_{\beta} \quad & \|\beta\|_1 \\ \text{subject to} \quad & \|X^T(y - X\beta)\|_\infty \le \lambda \end{aligned}$$

Here $\lambda \ge 0$ is a tuning parameter

Again, this can be reformulated as a linear program (check this!)

[1] Candes and Tao (2007), "The Dantzig selector: statistical estimation when p is much larger than n"

Standard form

A linear program is said to be in standard form when it is written as

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Any linear program can be rewritten in standard form (check this!)

Convex quadratic program

A convex quadratic program or QP is an optimization problem of the form

$$\begin{aligned} \min_x \quad & c^T x + \frac{1}{2} x^T Q x \\ \text{subject to} \quad & Dx \leq d \\ & Ax = b \end{aligned}$$

where $Q \succeq 0$, i.e., positive semidefinite

Note that this problem is not convex when $Q \not\succeq 0$. When we say quadratic program or QP from now on, we implicitly assume that $Q \succeq 0$

Example: portfolio optimization

Construct a financial portfolio, trading off performance and risk:

$$\begin{aligned} \max_x \quad & \mu^T x - \frac{\gamma}{2} x^T Q x \\ \text{subject to} \quad & 1^T x = 1 \\ & x \geq 0 \end{aligned}$$

Interpretation:

- μ : expected assets' returns
- Q : covariance matrix of assets' returns
- γ : risk aversion
- x : portfolio holdings (percentages)

Example: support vector machines

Given $y \in \{-1,1\}^n$, $X \in \mathbb{R}^{n \times p}$ with rows $x_1, \ldots x_n$, recall the support vector machine or SVM problem:

$$\begin{aligned} \min_{\beta, \beta_0, \xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{subject to} \quad & \xi_i \geq 0, \ i = 1, \ldots n \\ & y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i, \ i = 1, \ldots n \end{aligned}$$

This is a quadratic program

Example: lasso

Given $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$, recall the lasso problem:

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p} \quad & \|y - X\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \leq s \end{aligned}$$

Here $s \geq 0$ is a tuning parameter. Indeed, this can be reformulated as a quadratic program (check this!)

Alternative way to parametrize the lasso problem (called Lagrange, or penalized form):

$$\min_{\beta \in \mathbb{R}^p} \; \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

Now $\lambda \geq 0$ is a tuning parameter. And again, this can be rewritten as a quadratic program (check this!)

Standard form

A quadratic program is in standard form if it is written as

$$\begin{aligned} \min_{x} \quad & c^T x + \frac{1}{2} x^T Q x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Any quadratic program can be rewritten in standard form

Motivation for semidefinite programs

Consider linear programming again:

$$
\begin{aligned}
\min_{x} \quad & c^T x \\
\text{subject to} \quad & Dx \leq d \\
& Ax = b
\end{aligned}
$$

Can generalize by changing $\leq$ to different (partial) order. Recall:

- $\mathbb{S}^n$ is space of $n \times n$ symmetric matrices
- $\mathbb{S}^n_+$ is the space of positive semidefinite matrices, i.e.,

$$\mathbb{S}^n_+ = \{X \in \mathbb{S}^n : u^T X u \geq 0 \text{ for all } u \in \mathbb{R}^n\}$$

- $\mathbb{S}^n_{++}$ is the space of positive definite matrices, i.e.,

$$\mathbb{S}^n_{++} = \big\{X \in \mathbb{S}^n : u^T X u > 0 \text{ for all } u \in \mathbb{R}^n \setminus \{0\}\big\}$$

Facts about $\mathbb{S}^n$, $\mathbb{S}^n_+$, $\mathbb{S}^n_{++}$

- Basic linear algebra facts:

$$X \in \mathbb{S}^n \implies \lambda(X) \in \mathbb{R}^n$$
$$X \in \mathbb{S}^n_+ \iff \lambda(X) \in \mathbb{R}^n_+$$
$$X \in \mathbb{S}^n_{++} \iff \lambda(X) \in \mathbb{R}^n_{++}$$

- We can define an inner product over $\mathbb{S}^n$: given $X, Y \in \mathbb{S}^n$,

$$\langle X, Y \rangle = \operatorname{tr}(XY)$$

  Will also denote this by $X \bullet Y$

- We can define a partial ordering over $\mathbb{S}^n$ (called the Loewner ordering): given $X, Y \in \mathbb{S}^n$,

$$X \succeq Y \iff X - Y \in \mathbb{S}^n_+$$

Semidefinite program

A semidefinite program or SDP is an optimization problem of the form

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & x_1 F_1 + \ldots + x_n F_n \preceq F_0 \\ & Ax = b \end{aligned}$$

Here $F_j \in \mathbb{S}^d$, $j = 0, 1, \ldots n$ and $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. Observe that this is always a convex optimization problem

Also, any linear program is a semidefinite program (check this!)

Standard form

A semidefinite program is in standard form if it is written as

$$\begin{aligned} \min_{X} \quad & C \bullet X \\ \text{subject to} \quad & A_i \bullet X = b_i,\ i = 1, \ldots m \\ & X \succeq 0 \end{aligned}$$

Any semidefinite program can be written in standard form (check this!)

Example: theta function

Let $G = (N, E)$ be an undirected graph, $N = \{1, \dots, n\}$, and

- $\omega(G)$: clique number of G
- $\chi(G)$: chromatic number of G

The Lovasz theta function:[2]

$$\begin{array}{rcll} \vartheta(G) & = & \max_X & 11^T \bullet X \\ & & \text{subject to} & I \bullet X = 1 \\ & & & X_{ij} = 0, \ (i,j) \notin E \\ & & & X \succeq 0 \end{array}$$

The Lovasz sandwich theorem: $\omega(G) \leq \vartheta(\bar{G}) \leq \chi(G)$, where $\bar{G}$ is the complement graph of G

[2] Lovasz (1979), "On the Shannon capacity of a graph"

Example: trace norm minimization

Let $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear map,

$$\mathcal{A}(X) = \begin{pmatrix} A_1 \bullet X \\ \dots \\ A_p \bullet X \end{pmatrix}$$

for matrices $A_1, \dots A_p \in \mathbb{R}^{m\times n}$ (and where $A_i \bullet X = \mathrm{tr}(A_i^T X)$). Finding the lowest-rank solution to an underdetermined system, nonconvex way:

$$\begin{array}{ll} \min\limits_X & \mathrm{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

Trace norm approximation:

$$\begin{array}{ll} \min\limits_X & \|X\|_{\mathrm{tr}} \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

This is indeed an SDP (but a bit harder to show ...)

Conic program

A conic program is an optimization problem of the form:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & D(x) + d \in K \end{aligned}$$

Here:

- $c, x \in \mathbb{R}^n$, and $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$
- $D : \mathbb{R}^n \to Y$ is a linear map, $d \in Y$, for Euclidean space Y
- $K \subseteq Y$ is a closed convex cone

Both LPs and SDPs are special cases of conic programming. For LPs, $K = \mathbb{R}^n_+$; for SDPs, $K = \mathbb{S}^n_+$

Second-order cone program

A second-order cone program or SOCP is an optimization problem of the form:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & \|D_i x + d_i\|_2 \le e_i^T x + f_i, \; i = 1, \dots p \\ & Ax = b \end{aligned}$$

This is indeed a cone program. Why? Recall the second-order cone

$$Q = \{(x, t) : \|x\|_2 \le t\}$$

So we have

$$\|D_i x + d_i\|_2 \le e_i^T x + f_i \iff (D_i x + d_i, e_i^T x + f_i) \in Q_i$$

for second-order cone Q_i or appropriate dimensions. Now take $K = Q_1 \times \dots \times Q_p$

Observe that every LP is an SOCP. Furthermore, every SOCP is an SDP

Why? Turns out that

$$\|x\|_2 \le t \iff \begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0$$

Hence we can write any SOCP constraint as an SDP constraint

The above is a special case of the Schur complement theorem:

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \iff A - BC^{-1}B^T \succeq 0$$

for A, C symmetric and $C \succ 0$

Hey, what about QPs?

Take a breath (phew!). So far we have established the hierachy

$$\text{LPs} \subseteq \text{SOCPs} \subseteq \text{SDPs} \subseteq \text{Conic programs}$$

What about our old friend QPs? Turns out that QPs "sneak in" nicely into the hierarchy, in between LPs and SOCPs, completing the picture we saw at the start

To see that QPs are SOCPs, start by rewriting a QP as

$$\begin{aligned} \min_{x,t} \quad & c^T x + t \\ \text{subject to} \quad & Dx \leq d, \; \frac{1}{2} x^T Q x \leq t \\ & Ax = b \end{aligned}$$

Now simply introduce a variable $w = Q^{1/2} x$

References and further reading

- D. Bertsimas and J. Tsitsiklis (1997), "Introduction to linear optimization," Chapters 1, 2
- A. Nemirovski and A. Ben-Tal (2001), "Lectures on modern convex optimization," Chapters 1–4
- S. Boyd and L. Vandenberghe (2004), "Convex optimization," Chapter 4