Canonical Problem Forms

Lecturer: 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
Linear program

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

$$\min_x c^T x$$

subject to

$$Dx \leq d$$

$$Ax = b$$

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!)

\[
\begin{align*}
\min_x & \quad c^T x \\
\text{subject to} & \quad Dx \geq d \\
& \quad x \geq 0
\end{align*}
\]

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

\[
\min_x \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}
\]

subject to

\[
\sum_{j=1}^{n} x_{ij} \leq s_i, \quad i = 1, \ldots, m
\]

\[
\sum_{i=1}^{m} x_{ij} \geq d_j, \quad j = 1, \ldots, n, \quad x \geq 0
\]

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 $X \in \mathbb{R}^{n \times p}$, where $p > n$. Suppose that we seek the sparsest solution to underdetermined linear system $X \beta = y$

Nonconvex formulation:

$$\min_{\beta} \|\beta\|_0$$

subject to $X \beta = y$

where recall $\|\beta\|_0 = \sum_{j=1}^{p} 1\{\beta_j \neq 0\}$

$l_1$ approximation, often called basis pursuit:

$$\min_{\beta} \|\beta\|_1$$

subject to $X \beta = y$
Basis pursuit is a linear program. Reformulation:

$$\min_{\beta} \|\beta\|_1 \quad \iff \quad \min_{\beta,z} 1^T z$$

subject to \(X\beta = y\) \quad subject to \(z \geq \beta\)

\(z \geq -\beta\)

\(X\beta = y\)

(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:\(^1\)

$$\min_{\beta} \|\beta\|_1$$

subject to \(\|X^T(y - X\beta)\|_{\infty} \leq \lambda\)

Here $\lambda \geq 0$ 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 $n$”
A linear program is said to be in **standard form** when it is written as

\[ \min_{x} \quad c^{T}x \]

subject to \( Ax = b \)

\( x \geq 0 \)

Any linear program can be rewritten in standard form (check this!)
A convex quadratic program or QP is an optimization problem of the form

$$\min_x c^T x + \frac{1}{2} x^T Q x$$

subject to

$$D x \leq d$$
$$A x = b$$

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

Note that this problem is not convex when $Q \not\succeq 0$

From now on, when we say quadratic program or QP, we implicitly assume that $Q \succeq 0$ (so the problem is convex)
Example: portfolio optimization

Construct a financial portfolio, trading off performance and risk:

\[
\begin{align*}
\max_{x} & \quad \mu^T x - \frac{\gamma}{2} x^T Q x \\
\text{subject to} & \quad 1^T x = 1 \\
& \quad x \geq 0
\end{align*}
\]

Interpretation:
- \(\mu\): expected assets’ returns
- \(Q\): covariance matrix of assets’ returns
- \(\gamma\): risk aversion
- \(x\): portfolio holdings (percentages)
Example: support vector machines

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

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

This is a quadratic program
Example: lasso

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

$$\min_{\beta} \quad \| y - X \beta \|_2^2$$

subject to $\| \beta \|_1 \leq s$

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):

$$\min_{\beta} \quad \frac{1}{2} \| y - X \beta \|_2^2 + \lambda \| \beta \|_1$$

Now $\lambda \geq 0$ is a tuning parameter. And again, this can be rewritten as a quadratic program (check this!)
A quadratic program is in **standard form** if it is written as

\[
\min_x \quad c^T x + \frac{1}{2} x^T Q x
\]

subject to \( Ax = b \)

\( x \geq 0 \)

Any quadratic program can be rewritten in standard form
Motivation for semidefinite programs

Consider linear programming again:

$$\min_x c^T x$$
subject to

$$Dx \leq d$$
$$Ax = b$$

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

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

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

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

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

- Basic linear algebra facts:
  
  \[
  X \in S^n \implies \lambda(X) \in \mathbb{R}^n \\
  X \in S^+_n \iff \lambda(X) \in \mathbb{R}^+_n \\
  X \in S^{++}_n \iff \lambda(X) \in \mathbb{R}^{++}_n 
  \]

- We can define an inner product over $S^n$: given $X, Y \in S^n$,
  
  \[
  X \bullet Y = \text{tr}(XY) 
  \]

- We can define a partial ordering over $S^n$: given $X, Y \in S^n$,
  
  \[
  X \succeq Y \iff X - Y \in S^+_n 
  \]

  Note: for $x, y \in \mathbb{R}^n$, $\text{diag}(x) \succeq \text{diag}(y) \iff x \succeq y$ (recall, the latter is interpreted elementwise)
A **semidefinite program** or SDP is an optimization problem of the form

\[
\min_x \quad c^T x \\
\text{subject to} \quad x_1 F_1 + \ldots + x_n F_n \preceq F_0 \\
\quad Ax = b
\]

Here \( F_j \in \mathbb{S}^d \), for \( j = 0, 1, \ldots, n \), 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!)
A semidefinite program is in **standard form** if it is written as

\[
\min_{X} \quad C \cdot X \\
\text{subject to} \quad A_i \cdot X = b_i, \quad i = 1, \ldots, m \\
X \succeq 0
\]

Any semidefinite program can be written in standard form (for a challenge, check this!)
Example: theta function

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

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

The Lovasz theta function:\footnote{Lovasz (1979), “On the Shannon capacity of a graph”}

\[
\vartheta(G) = \max_X 11^T \bullet X \\
\text{subject to } I \bullet X = 1 \\
X_{ij} = 0, (i, j) \notin E \\
X \succeq 0
\]

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

Let \( A : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p \) be a linear map,

\[
A(X) = \begin{pmatrix}
A_1 \cdot X \\
\vdots \\
A_p \cdot X
\end{pmatrix}
\]

for matrices \( A_1, \ldots, A_p \in \mathbb{R}^{m \times n} \) (and where \( A_i \cdot X = \text{tr}(A_i^T X) \)).

Finding the lowest-rank solution to an underdetermined system, nonconvex way:

\[
\min_X \quad \text{rank}(X) \\
\text{subject to} \quad A(X) = b
\]

Trace norm approximation:

\[
\min_X \quad \|X\|_{\text{tr}} \\
\text{subject to} \quad A(X) = b
\]

This is indeed an SDP (but harder to show, requires duality ...)

22
Conic program

A conic program is an optimization problem of the form:

$$\min_x \ c^T x$$
subject to $Ax = b$
$$D(x) + d \in K$$

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, $K = \mathbb{S}^n_+$
A second-order cone program or SOCP is an optimization problem of the form:

$$\min_x \quad c^T x$$

subject to

$$\|D_i x + d_i\|_2 \leq e_i^T x + f_i, \quad i = 1, \ldots, p$$

$$Ax = b$$

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

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

So we have

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

for second-order cone $Q_i$ or appropriate dimensions. Now take

$$K = Q_1 \times \ldots \times Q_p$$
Observe that every LP is an SOCP. Furthermore, every SOCP is an SDP.

Why? Turns out that

\[ \|x\|_2 \leq t \iff \begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0 \]

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?

Finally, our old friend QPs “sneak” into the hierarchy. Turns out QPs are SOCPs, which we can see by rewriting a QP as

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

Now write \( \frac{1}{2} x^T Q x \leq t \iff \| (\frac{1}{\sqrt{2}} Q^{1/2} x, \frac{1}{2} (1 - t)) \|_2 \leq \frac{1}{2} (1 + t) \)

Take a breath (phew!). Thus we have established the hierarchy

\[
\text{LPs} \subseteq \text{QPs} \subseteq \text{SOCPs} \subseteq \text{SDPs} \subseteq \text{Conic programs}
\]

completing the picture we saw at the start
References and further reading

- D. Bertsimas and J. Tsitsiklis (1997), “Introduction to linear optimization,” Chapters 1, 2