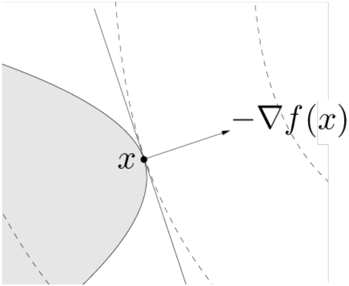
Canonical Problem Forms

Yu-Xiang Wang
CS292F

(Based on Ryan Tibshirani's 10-725)

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)

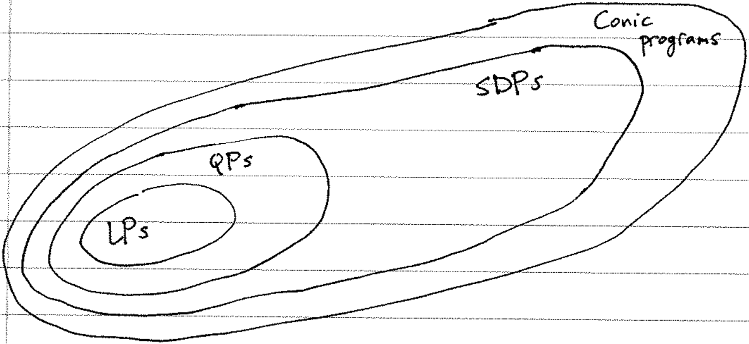
Outline

Today:

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

Hierarchy of Canonical Optimizations

- 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 min cost

$$\begin{array}{ll} \min\limits_{x} & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\ \text{subject to} & \sum_{j=1}^{n} x_{ij} \le s_i,\ i = 1, \ldots, m \\ & \sum_{i=1}^{m} x_{ij} \ge d_j,\ j = 1, \ldots, n, \quad x \ge 0 \end{array}$$

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

Nonconvex formulation:

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

where recall $\|\beta\|_0 = \sum_{j=1}^p 1\{\beta_j \neq 0\}$, the ℓ_0 "norm"

The ℓ_1 approximation, often called basis pursuit:

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

Basis pursuit is a linear program. Reformulation:

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

(Check that this makes sense to you)

Example: Dantzig selector

Modification of previous problem, where we allow for $X\beta \approx y$ (we don't require exact equality), the Dantzig selector:[1]

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

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"

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 $Q \not\succeq 0$

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:

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

Interpretation:

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

Example: support vector machines

Given $y \in \{-1, 1\}^n$, $X \in \mathbb{R}^{n \times p}$ having 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:

$$\begin{array}{ll} \min_{\beta} & \|y - X\beta\|_2^2 \\ \text{subject to} & \|\beta\|_1 \le s \end{array}$$

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

Alternative parametrization (called Lagrange, or penalized form):

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

Now $\lambda \ge 0$ 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.,

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

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

- Basic linear algebra facts, here $\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))$:

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

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

$$X \bullet Y = \operatorname{tr}(XY)$$

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

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

Note: for $x, y \in \mathbb{R}^n$, $\operatorname{diag}(x) \succeq \operatorname{diag}(y) \iff x \geq y$ (recall, the latter is interpreted elementwise)

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 $F_j \in \mathbb{S}^d$, 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!)

Standard form

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

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

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:[2]

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

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

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

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

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

Trace norm approximation:

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

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

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 of appropriate dimensions. Now take $K = Q_1 \times \ldots \times Q_p$

Observe that every LP is an SOCP. Further, 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?

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

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

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

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

LPs $\subseteq$ QPs $\subseteq$ SOCPs $\subseteq$ SDPs $\subseteq$ Conic programs

completing the picture we saw at the start

Approximation Algorithm for MaxCut

- Given a graph with nodes and edges and edge weights. Find a subset S of the nodes such that the sum of the weights w_{ij} of the edges between S and its complement $\bar{S}$ is maximizes.
- Let $x_j = 1$ if $j \in S$ and $x_j = -1$ if $j \in \bar{S}$.

$$
\begin{array}{ll}
\max\limits_{x} & \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(1 - x_i x_j) \\
\text{s.t.} & x_j \in \{-1, 1\}, j = 1, ..., n
\end{array}
$$

- Goemans and Williamson algorithm:
  1. Convex relaxation: solve an SDP instead.
  2. Randomized rounding.
- You get a 0.87856 approximation of an NP-complete problem.

Approximation Algorithm for MaxCut

Reformulation (without changing the problem):

$$\begin{aligned} \max_{Y \in \mathbb{R}^{n \times n}, x \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - Y_{i,j}) \\ \text{s.t.} \quad & Y_{i,i} = 1 \quad \forall j = 1, ..., n \\ & Y = xx^T. \end{aligned}$$

The convex relaxation:

$$\begin{aligned} \max_{Y \in \mathbb{R}^{n \times n}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - Y_{i,j}) \\ \text{s.t.} \quad & Y_{i,i} = 1 \quad \forall j = 1, ..., n \\ & Y \succeq 0. \end{aligned}$$

Goemans and Williamson: Sample v uniformly from the unit sphere in $\mathbb{R}^n$, decompose $Y = UU^T$, output $\mathrm{sign}(Uv)$.

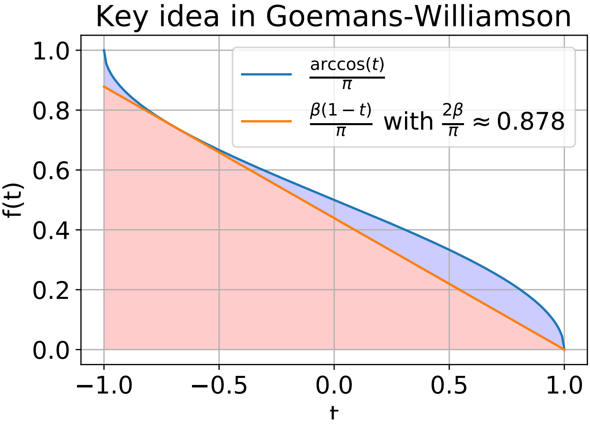
Proof ideas of Goemans and Williamson's algorithm

Observe:

- ith row of matrix U, $u_i \in \mathbb{R}^n$ can be thought of as a nonparametric vector-space embedding of the graph node i.
- $\mathrm{sign}(u_i^T v)$ is the output of a random linear separator.
- The probability that Node i and Node j being classified differently is proportional to θ_{ij}/π where θ_{ij} is the angle between u_i and u_j.
- $\cos(\theta_{ij}) = \langle u_i, u_j \rangle = Y_{i,j}$ (notice that u_i, u_j are both unit vectors (why?)).
- We can lower bound $\theta_{ij} = \arccos(Y_{i,j})$ with $\beta(1 - \langle u_i, u_j \rangle)$ with some constant β. (see next slide for an illustration).
- The objective function of the relaxed problem is larger than that of the original Maxcut problem, but the randomized rounding gives us a "feasible solution" to the original problem, therefore the solution is in expectation a $\beta/\pi \approx 0.878$ constant approximation to Maxcut.

See the more detailed proof in the sketched notes.

Proof ideas of Goemans and Williamson's algorithm



References and further reading

- S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapter 4
- 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
- Goemans, Michel X., and David P. Williamson. "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming" Journal of the ACM (JACM) 42.6 (1995): 1115-1145.