10-725/36-725: Convex Optimization	**Spring 2015**

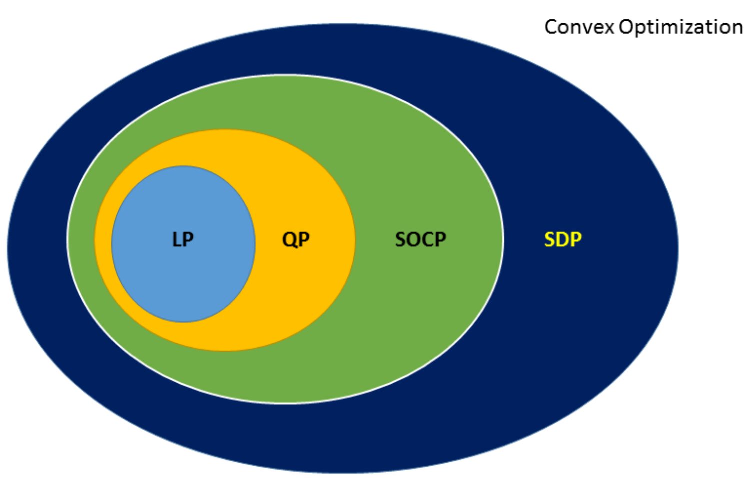
Lecture 4: January 26

Lecturer: Javier Pena *Scribes: Vipul Singh, Shinjini Kundu, Chia-Yin Tsai*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

4.1 Introduction

Optimization is a huge class of problems. There is a hierarchy of convex optimization problems. We'll talk about linear programming, quadratic programming, second-order cone programming, and semidefinite programming today. See below.



4.2 Linear Program

A linear program is a problem of the form:

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

The simplest type of convex optimization problem. Recall that a convex optimization problem is a problem of the form:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to } & g(x) \leq d \\ & Ax = b \end{aligned}$$

where f and g's are convex functions. In LP, objective and all inequality constraints are linear, and linear functions are convex. Aside: Linear programming has an interesting history. Attributed to Dantzig in in 1940s, and has a vast range of applications, especially in game theory.

Examples: Diet problem Find the cheapest combination of foods that satisfies some nutritional requirements.

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

Here, c_j is per-unit cost of food j, d_i is minimum required intake of nutrient i, $D_i j$ is content of nutrient i per unit of food j, and x_j is the units of food j in the diet.

Another example: Transportation problem (see slides)

Another example: L1 minmization is a heuristic to find a sparse solution to an under-determined system of equations,

$$\begin{aligned} \min_x \quad & ||x||_1 \\ \text{subject to } & Ax = b \end{aligned}$$

Where A is a fat matrix (m¡n), having fewer constraints than variables. Solving the combinatorial problem is NP-hard, but the LP problem recovers a sparse solution with high probability.

Another example: Dantzig selector Tweaks the L1 minimization problem, where assuming that b is not just Ax, but is $Ax + \epsilon$, and allows some error in the solution.

Any linear programming problem can be written in standard form.

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

For example, if we have

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

we can add slack variables and write as

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

Can replace $x = y - s$ Optimality conditions: We have that -c is in the normal cone at an optimal solution x*, which equivalent to stating that $c = A^T y * + s*$. Easier to characterize the optimality conditions when the problem is in standard form.

4.3 Quadratic Program

4.3.1 Convex quadratic programming

This is an optimization problem of the form

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

where Q is symmetric and positive semidefinite. The problem is convex iff the matrix Q is positive semidefinite.

Examples: portfolio optimization Model to construct financial portfolio with optimal performance/risk trade-off:

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

Here, μ is the expected assets' returns, Q is the covariance matrix of assets' returns, γ is risk aversion, and x is the portfolio holding (percentages).

Another example: support vector machines The objective is quadratic

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

Why is it a quadratic minimization problem when the C_is don't appear square? The Q matrix can have many zero entries.

4.3.2 Standard form

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

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

$\bar{x}$ is an optimal solution $\Leftrightarrow -c - Q\bar{x} \in N_C(\bar{x}) \Leftrightarrow Q\bar{x} + c = A^T\bar{y} + \bar{s}$ for some $\bar{y}$ and $\bar{s} \geq 0$ such that $\bar{s}^T\bar{x} = 0$. We see that for linear and quadratic convex programs, all we need to formalize the optimal solution are first order conditions. Comparing with the optimality condition for linear programs, we see that if we have a solver for linear programs, we might be able to tweak it a bit to solve convex quadratic programs.

4.4 Semi-definite programming

Semi-definite programs are a much bigger subset of convex optimization problems than convex quadratic programs. We can extend from linear programs to semi-definite programs by changing the order ($\leq$) involved in the inequality constraint $Dx \leq d$ to a different kind of order in some vector space. We work with the vector-space $\mathbb{S}^n$ now.

4.4.1 Notation and Definitions

- $\mathbb{S}^n$ is the vector space of symmetric $n \times n$ real matrices.
- Inside this vector space resides the cone of positive semi-definite matrices:

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

- We will be using a couple of facts from linear algebra:
  - The eigenvalues of a symmetric matrix are always real.
  - The eigenvalues of a positive semi-definite matrix are always non-negative.
- The canonical inner product in $\mathbb{S}^n$ is:

  $$\langle X, Y \rangle = X \bullet Y := \text{trace}(XY) = \sum_{i=1}^n \sum_{j=1}^n X_{ij} Y_{ij}$$

  Trace satisfies the property that if the product ABC is well-defined and the result is a square matrix, then $\text{tr}(ABC) = \text{tr}(BCA)$.
- $\mathbb{S}^n_+$ is a closed convex cone.
- The interior of $\mathbb{S}^n_+$ is the cone of positive definite matrices defined as:

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

- $X \in \mathbb{S}^n_{++} \Leftrightarrow \lambda(X) \in \mathbb{R}^n_{++}$, where $\lambda(X)$ is the map defining the eigenvalues of X.
- Loewner ordering: Given $X, Y \in \mathbb{S}^n$,

$$X \succcurlyeq Y \Leftrightarrow X - Y \in \mathbb{S}^n_+$$

4.4.2 The Optimization Problem

A semi-definite program (SDP) is of the form:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to } & \Sigma_{j=1}^n F_j x_j \preccurlyeq F_0 \\ & Ax = b \end{aligned}$$

Here, $F_j \in \mathbb{S}^d,\ j = 0, 1, ..., n$ and $A \in \mathbb{R}^{m \times n}, c \in \mathbb{R}^n, b \in \mathbb{R}^m$.

In the LP formulation, we had constraints $Dx \leq d$, i.e., $\Sigma_{j=1}^n d_j x_j \leq d$, where d_j's are the columns of D. This is a system of linear inequalities. For the SDP, we have replaced d_j with F_j, d with F_0, and the order $\leq$ with $\preccurlyeq$. By analogy, this is called a system of Linear Matrix Inequalities.

A semidefinite program is a convex optimization problem.

4.4.3 Standard form

$$\begin{aligned} \min_X \quad & C \bullet X \\ \text{subject to } & A_i \bullet X = b_i,\ i = 1, ..., m \\ & X \succcurlyeq 0 \end{aligned}$$

where $A_1, ..., A_m$ and C are given symmetric $n \times n$ matrices, and $X \in \mathbb{S}^n$ is the matrix variable.

Every linear program is an SDP. We see this using the following:

- $x \in \mathbb{R}^n_+ \Leftrightarrow \text{Diag}(x) \succcurlyeq 0$
- $c^T x = \text{Diag}(c) \bullet \text{Diag}(x)$
- The objective as well as constraints can be written in matrix form. The constraints for off-diagonal elements being 0 can be enforced by equalities.

4.4.4 History of semidefinite programming

- Eigenvalue optimization, LMI problems (1960s - 1970s) linear matrix inequality in control
- Lovasz theta function (1979) in information theory [lovasz]
- Interior-point algorithm for SDP (1980s, 1990s)
- Advancements in theory, algorithms, application (1990s)
- New algorithm and applications in data and imaging science (2000s-)

4.5 Application of semidefinite programming

4.5.1 theta function

Assume $G = (N, E)$, N is the node and E is the edge.

$\omega(G)$ is the clique number of G. The largest set of nodes that are completely connected.

$\chi(G)$ is the chromatic number of G. The minimum number of colors that suffice to color the nodes of the graph.

The theta function:

$$\begin{aligned} \theta(G) := \max_X \ & \mathbf{11^T \bullet X} \\ \text{subject to } & I \bullet X = 1 \\ & X_{ij} = 0, (i,j) \notin E \\ & X \succcurlyeq 0 \end{aligned}$$

Lovasz sandwich theorem: $\omega(\overline{G}) \leq \theta(G) \leq \chi(\overline{G})$

4.5.2 Nuclear norm minimization

Similar to l^1-norm for vectors.

$$\begin{aligned} \min_X \ & \|X\|_{tr} \\ \text{subject to } & A(X) = b \end{aligned}$$

Here $A : R^{mxn} \longrightarrow p$ linear map, $b \in R^p$. Nuclear norm is $\|X\|_{tr} = \|\sigma(X)\|_1$, the sum of the singular values of X. The dual of the nuclear norm is operator norm: $\|X\|_{op} = \|\sigma(X)\|_\infty = \max \|Xu\|_2 : \|u\|_2 \leq 1$ (Note: the duality is like the p and q norm for vector.)

Example: Netflix challenge. We would want to find a matrix that is low rank.

Key for proof that this is a semidefinite programming.

Observation For $y \in R^{mxn}$

$$\begin{aligned} \|y\|_{op} \leq 1 & \Longleftrightarrow yy^T \preccurlyeq I_m \\ & \Longleftrightarrow \begin{bmatrix} I_m & y \\ y^T & I_n \end{bmatrix} \succcurlyeq 0 \text{(special case of Schur Complement)} \end{aligned}$$

$$\begin{aligned}
\|x\|_{tr} &= \max_y \ \{\text{trace}(x^T y) : \|y\|_{op} \le 1\} \\
&= \max_y && \text{trace}(x^T y) \\
&\text{subject to} && \begin{bmatrix} I_m & y \\ y^T & I_n \end{bmatrix} \succcurlyeq 0 \\
&= \min_{w_1, w_2} && \frac{1}{2}(I_m w_1 + I_n w_2) \\
&\text{subject to} && \begin{bmatrix} w_1 & x \\ x^T & w_2 \end{bmatrix} \succcurlyeq 0 \\
&\text{by SDP duality}
\end{aligned}$$

The problem becomes a SDP.

$$\begin{aligned}
\min_{X, w_1, w_2} \ & \frac{1}{2}(I_m w_1 + I_n w_2) \\
\text{subject to } & A(X) = b \\
& \begin{bmatrix} w_1 & X \\ X^T & w_2 \end{bmatrix} \succcurlyeq 0
\end{aligned}$$

*Schur Complement theorem: For a matrix $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ and A, C are symmetric, $C \succcurlyeq 0$

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

4.6 Conic programming

LP and SDP are special cases of conic programming.

Conic program

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

where K is a closed convex cone.

4.6.1 second-order conic programming (SOCP)

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

where $Q = Q_{n_1} \times ... \times Q_{n_r}$. and Q_n is defined as $Q_n := \{x = \begin{bmatrix} x_0 \\ \overline{x} \end{bmatrix} \in R^n : x_0 \le \|\overline{x}\|\}$

Standard form:

$$\begin{aligned} \min_x \ & c^T x \\ \text{subject to } & x \succcurlyeq_Q 0 \\ & Ax = b \end{aligned}$$

for $Q = Q_{n_1} \times ... \times Q_{n_r}$.

LP $\subsetneq$ SOCP $\subsetneq$ SDP.

4.6.1.1 Convex QCQP

For a convex inequality $x^T Qx + q^T x + l \le 0$, where $Q = LL^T \in S^n$, $q \in R^n$, and $l \in R$. It can be recast as

$$\left\| \begin{bmatrix} L^T x \\ \frac{1+q^T x + l}{2} \end{bmatrix} \right\| \le \frac{1 - q^T x - l}{2}$$

.

That is,

$$\begin{aligned} & \left\| \begin{bmatrix} Ax + b \\ c^T x + d \end{bmatrix} \right\| \le e^T x + f \\ \Longleftrightarrow & \begin{bmatrix} e^T x + f Ax + b \\ c^T x + d \end{bmatrix} \in Q \end{aligned}$$

Therefore, it can be rewrite to a SOCP if $Q_i \succcurlyeq 0, i = 1, ..., r$

$$\begin{aligned} \min \ & x^T Q_0 x + q_0^T x \\ \text{subject to } & x^T Q_i x + q_i^T x, i = 1, ..., r \end{aligned}$$

since all the inequality can be transformed to a second order cone constraint.

4.6.1.2 Rewrite a second order cone in terms of SDP

We can rewrite $\|x\|_2 \le x_0$. By Shur Complement theorem,

$$\begin{bmatrix} I & x \\ x^T & 1 \end{bmatrix} \succcurlyeq 0 \Longleftrightarrow \|x\|^2 \le 1$$

So,

$$\|x\|^2 \le x_0 \Longleftrightarrow \begin{bmatrix} x_0 I & x \\ x^T & x_0 \end{bmatrix} \succcurlyeq 0$$

Any second order cone constraints can be rewrite to a SD constraint.

References

[lovasz] LOVASZ L, "On the Shannon capacity of a graph." *IEEE Transactions on Information Theory Vol. 25, No. 1*, (Jan. 25, 1979), pp. 1-7