10-725/36-725: Convex Optimization **Fall 2016**

Lecture 4: September 12

Lecturer: Ryan Tibshirani *Scribes: Jay Hennig, Yifeng Tao, Sriram Vasudevan*

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

4.1 Previous Lecture

4.1.1 Eliminating Equality Constraints

If the problem is of the form

$$\begin{aligned} \min_x\ & f(x) \\ \text{s.t } & g_i(x) \le 0,\ i = 1, ..., m \\ & Ax = b \end{aligned}$$

then x can be expressed as $My + x_0$ (where $Ax_0 = b$ and $\text{col}(M) = \text{null}(A)$). Doing so allows us to rewrite the above problem as:

$$\begin{aligned} \min_y\ & f(My + x_0) \\ \text{s.t } & g_i(My + x_0) \le 0,\ i = 1, ..., m \end{aligned}$$

4.1.2 Introducing Slack Variables

The concept of slack variables is opposite to that of eliminating equality constraints. Thus the first formulation in the previous section can be written as:

$$\begin{aligned} \min_{x,s}\ & f(x) \\ \text{s.t } & s_i \ge 0,\ i = 1, ..., m \\ & g_i(x) + s_i = 0,\ i = 1, ..., m \\ & Ax = b \end{aligned}$$

This problem however is not convex unless the g_i are all affine.

4.1.3 Relaxing Nonaffine Equality Constraints

Given an optimization problem $\min_x f(x)$ such that $x \in C$, we can consider an enlarged set $\tilde{C} \supseteq C$ and solve $\min_x f(x)$ such that $x \in \tilde{C}$ instead. This is known as relaxation, and its optimal value is always lesser than or equal to that of the original problem.

An important special case is that of replacing convex nonaffine equality constraints $h_j(x) = 0,\ j = 1, ..., r$ with $h_j(x) \leq 0,\ j = 1, ..., r$.

4.1.4 Examples

1. **Maximum Utility Problem:**
   This problem models investment/consumption. It can be formulated as:

$$\begin{aligned} \max_{x,b} \ & \sum_{t=1}^{T} \alpha_t u(x_t) \\ \text{s.t. } & b_{t+1} = b_t + f(b_t) - x_t,\ t = 1, ..., T \\ & 0 \leq x_t \leq b_t,\ t = 1, ..., T \end{aligned}$$

   with b_t being the budget and x_t being the amount consumed at time t. f is the investment return function, u is the utility function, and both are concave and increasing. The equality constraint is nonaffine, but if we relax it to an inequality, the problem doesn't change (relaxation is tight), and the problem is now convex.

2. **Principal Component Analysis:**
   Given $X \in \mathbb{R}^{n \times p}$, consider the low rank approximation problem $\min_R ||X - R||_F^2$ such that rank(R) $= k$. Here $||A||_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} A_{ij}^2$, the entrywise squared l_2 norm. This is equivalent to the PCA problem where $R = U_k D_k V_k^T$ with U_k and V_k being the first k columns of U and V, and D_k being the first k diagonal elements of D ($X = UDV^T$, the SVD decomposition of X).

   This is not a convex problem. To see this, suppose we take a matrix A in the set $C = \{R : \text{Rank}(R) = k\}$, then $-A \in C$, but $0.5A + 0.5(-A) \notin C$.

   This problem can be recast in a convex form by first rewriting the problem as

$$\begin{aligned} & \min_{Z \in \mathbb{S}^P} ||X - XZ||_F^2 \text{ subject to rank}(R) = k \\ \Leftrightarrow & \max_{Z \in \mathbb{S}^P} tr(SZ) \text{ subject to rank}(R) = k \end{aligned}$$

   where Z is a projection and $S = X^T X$. Hence the constraint set is the nonconvex set

$$C = \left\{ Z \in \mathbb{S}^P : \lambda_i(Z) \in \{0, 1\},\ i = 1, ..., p,\ tr(Z) = k \right\}$$

   where $\lambda_i(Z)$ are the n eigenvalues of Z. For this formulation, the solution becomes $Z = V_k V_k^T$ where V_k gives first k columns of V.

   If we relax the constraint set to $\mathcal{F} = \text{conv}(C)$, its convex hull, we have a linear maximixation over the fantope of order k, which is convex: $\max_{Z \in \mathcal{F}} tr(SZ)$. This is equivalent to the nonconvex PCA problem, i.e., it admits the same solution.

   Note: The fantope of order k is given by:

$$\begin{aligned} \mathcal{F} &= \{Z \in \mathbb{S}^P : \lambda_i(Z) \in [0, 1], i = 1, ..., p, tr(Z) = k\} \\ &= \{Z \in \mathbb{S}^P : 0 \preceq Z \preceq I, tr(Z) = k\} \end{aligned}$$

4.2 Linear Programs

4.2.1 Definition

A linear program (LP) is an optimization problem of the form:

$$\begin{aligned} \min_x \ & c^T x \\ \text{s.t. } & Dx \leq d \\ & Ax = b \end{aligned}$$

Note that this is always convex. A fundamental problem in convex optimization, it has many diverse applications and a rich history. Dantzig's simplex algorithm gives a direct solver.

4.2.2 Examples

Some common LP problems are given below:

1. **Diet Problem:**
   The problem deals with finding the cheapest combination of food items that satisfies some nutritional requirements. It can be formulated as shown below:

   $$\begin{aligned} \min_x \ & c^T x \\ \text{s.t. } & Dx \geq d \\ & x \geq 0 \end{aligned}$$

   where c_j is the per-unit cost of item j, d_i is the minimum intake of nutrient i required, D_{ij} is the amount of nutrient i contained in food j and x_j is the units of food j in the diet.

2. **Transportation Problem:**
   This problem deals with minimizing the costs of shipping the commodities from given sources to destinations. It can be formulated as shown below:

   $$\begin{aligned} \min_x \ & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\ \text{s.t. } & \sum_{j=1}^{n} x_{ij} \leq s_i, \ i = 1, ..., m \\ & \sum_{i=1}^{m} x_{ij} \geq d_{ij}, \ j = 1, ..., n, \ x \geq 0 \end{aligned}$$

   where s_i is the supply at source i, d_j is the demand at destination j, c_{ij} is the per-unit shipping cost from source i to destination j and x_{ij} is number of units shipped from i to j.

3. **Basis Pursuit:**
   Given $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$ (with $p > n$), the aim is to determine the sparsest solution to the underdetermined linear system $X\beta = y$. It can be formulated as below:

   $$\begin{aligned} \min_\beta \ & ||\beta||_0 \\ \text{s.t. } & X\beta = y \end{aligned}$$

where $||\beta||_0 = \sum_{j=1}^{p} 1\{\beta_j \neq 0\}$. This is a nonconvex problem, which can be recast as a linear program through an l_1 approximation known as basis pursuit. This formulation is given below:

$$\begin{aligned} \min_{\beta} \ & ||\beta||_1 \\ \text{s.t. } & X\beta = y \end{aligned}$$

The above problem can be reformulated as:

$$\begin{aligned} \min_{\beta,z} \ & 1^T z \\ \text{s.t. } & z \geq \beta \\ & z \geq -\beta \\ & X\beta = y \end{aligned}$$

4. **Dantzig Selector:**
   The Dantzig selector is a modification of basis pursuit where strict equality is not enforced, i.e., $X\beta \approx y$. Then the formulation becomes:

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

   where $\lambda \geq 0$ is a tuning parameter. This too can be reformulated as a linear program if the constraint is written as:

$$-\lambda \leq X_j^T(y - X\beta) \leq \lambda \ \forall j = 1, ..., p$$

4.2.3 Standard Form

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

$$\begin{aligned} \min_{x} \ & c^T x \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

Any LP can be written in standard form.

4.3 Quadratic Programs

4.3.1 Definition

Convex quadratic program (QP) is a kind of optimization problem of the form:

$$\begin{aligned} \min_{x} \ & c^T x + \frac{1}{2} x^T Q x \\ \text{s.t. } & Dx \leq d \\ & Ax = b \end{aligned}$$

We only discuss the case whose $Q \succeq 0$, since the problem is convex iff $Q \succeq 0$.

4.3.2 Examples

Here are some common QP problems:

1. Portfolio optimization

   We can use the QP:

$$\begin{aligned}\min_x \ & \mu^T x + \frac{\gamma}{2} x^T Q x \\ \text{s.t. } & 1^T x = 1 \\ & x \geq 0\end{aligned}$$

   to trade off performance and risk in a financial portfolio. Here μ is expected assets' returns, Q is covariance matrix of assets' returns, γ is risk aversion, x is portfolio holdings (sum is normalized to be 1).

2. Support vector machine

   Given $y \in \{-1, 1\}^n$, $X \in \mathbb{R}^{n \times p}$ with rows $x_1, x_2, ..., x_n$. SVM problem is:

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

3. Lasso

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

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

   Here $s \geq 0$ is a tuning parameter. This can be rewritten as a quadratic program.

   An alternative way to parametrize the lasso problem is in the penalized / Lagrange form:

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

   Here λ is the tuning parameter. The can also be rewritten in a quadratic form.

4.3.3 Standard Form

Any QP can be rewritten in the standard form:

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

4.4 Semidefinite programs (SDPs)

4.4.1 Motivation

Recall that Linear programs (LPs) have the following form:

$$\begin{aligned} \min_x \; & c^T x \\ \text{subject to } & Dx \leq d \\ & Ax = b \end{aligned} \tag{4.1}$$

Here, x is a vector. But we can generalize this problem to optimize over matrices, X, by changing $\leq$ to $\preceq$. This defines a partial ordering over matrices. (More on this later.)

4.4.2 Background

Recall:

- $\mathbb{S}^n$ is the space of $n \times n$ symmetric matrices.
- $\mathbb{S}^n_+$ is the space of positive semidefinite matrices:
$$\mathbb{S}^n_+ = \{X \in \mathbb{S}^n \mid u^T X u \geq 0 \text{ for all } u \in \mathbb{R}^n\}$$
- $\mathbb{S}^n_{++}$ is the space of positive definite matrices.
$$\mathbb{S}^n_{++} = \{X \in \mathbb{S}^n \mid u^T X u > 0 \text{ for all } u \in \mathbb{R}^n \backslash \{0\}\}$$

If X is a matrix in one of the above sets, this constrains its eigenvalues. Letting $\lambda(X)$ denote the eigenvalues of a matrix X:

- $X \in \mathbb{S}^n \Rightarrow \lambda(X) \in \mathbb{R}^n$
- $X \in \mathbb{S}^n_+ \Leftrightarrow \lambda(X) \in \mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$
- $X \in \mathbb{S}^n_{++} \Leftrightarrow \lambda(X) \in \mathbb{R}^n_{++} = \{x \in \mathbb{R}^n \mid x > 0\}$

We can define an inner product between two symmetric matrices $X, Y \in \mathbb{S}^n$ using the trace operator:

$$X \bullet Y = tr(XY) = \sum_{i,j} X_{i,j} Y_{i,j}$$

We can also partially order $\mathbb{S}^n$ by defining $\succeq$ as follows:

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

If we consider diagonal matrices, then this ordering for matrices becomes the same as our ordering for vectors. Below, $diag(x)$ denotes a matrix $X \in \mathbb{S}^n$ which has the vector $x \in \mathbb{R}^n$ as its diagonal elements, and 0 elsewhere. Now let $x, y \in \mathbb{R}^n$. Then:

$$diag(x) \succeq diag(y) \Leftrightarrow x \geq y$$

4.4.3 Semidefinite programs (SDPs)

An SDP is an optimization problem of the form:

$$\begin{aligned} \min_x \, & c^T x \\ \text{subject to } & x_1 F_1 + \cdots + x_n F_n \preceq F_0 \\ & Ax = b \end{aligned} \tag{4.2}$$

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

Recall that in an LP we have the constraint $Dx \leq d$. If we let D_i be the i^{th} column of D, then this is the same as $\sum_i x_i D_i \leq d$. So here, the SDP simply generalizes the vectors D_i and d to be symmetric matrices, F_i.

This problem is always convex because linear matrix inequalities are convex.

An SDP is said to be in standard form if it is written as:

$$\begin{aligned} \min_{X \in \mathbb{S}^n} \, & C \bullet X \\ \text{subject to } & A_i \bullet X = b_i, \; i = 1, \cdots, m \\ & X \succeq 0 \end{aligned} \tag{4.3}$$

With $C \in \mathbb{S}^n$, $A_i \in \mathbb{S}^n$, and $b_i \in \mathbb{R}$.

Any SDP can be written in this form, though a proof of this fact will require some effort!

Finally, any linear program is also a semidefinite program. To see this, consider the SDP where $X = diag(x)$.

Example: trace norm minimization

Let $A_1, \cdots, A_p \in \mathbb{R}^{m \times n}$. Then the following is a linear mapping from $\mathbb{R}^{m \times n} \to \mathbb{R}^p$:

$$A(X) = \begin{bmatrix} A_1 \bullet X \\ \cdots \\ A_p \bullet X \end{bmatrix} \tag{4.4}$$

(Note that because A_i is not necessarily symmetric, we use the standard defintion of trace: $A_i \bullet X = tr(A_i^T X)$.)

Finding a matrix X that satisfies $A(X) = b$ for some b such that X has the lowest rank is a nonconvex problem, because calculating rank is not a convex function. But we can use the trace norm as a surrogate objective, resulting in the following trace norm approximation:

$$\begin{aligned} \min_X \, & \|X\|_{tr} \\ \text{subject to } & A(X) = b \end{aligned} \tag{4.5}$$

This is an SDP, though this is not a trivial fact.

4.5 Conic Programs

4.5.1 Definition

A conic program is an optimization problem of the form:

$$\begin{aligned}\min_x\ & c^T x\\ \text{s.t. } & Ax = b\\ & D(x) + b \in K\end{aligned}$$

Here, $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. $D : \mathbb{R}^n \to Y$ is a linear mapping, $d \in Y$ for Euclidean space Y, $K \subseteq Y$ is a closed convex cone. This is very similar to LP, the only distinction is the set of linear inequalities are replaced with conic inequalities, i.e., $D(x) + d \preceq_K 0$. Notice that if $K = \mathbb{S}^n_+$, we recover SDP. Thus, this is a very broad class of problems.

4.5.2 Examples

Second-order cone program. A second-order cone program (SOCP) is an optimization problem of the form:

$$\begin{aligned}\min_x\ & c^T x\\ \text{s.t. } & \|D_i x + d\|_2 \leq e_i^T x + f_i, i = 1, 2, ..., n\\ & Ax = b\end{aligned}$$

This is a conic program with specific choice of K. In particular, it is a combination of second-order cones that are defined as:

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

From the definition, it is easy to see

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

for appropriate dimensions, then taking $K = Q_1 \times Q_2 \times ... \times Q_p$ will lead to the conic program form.

It is easy to see every LP is SOCP. In addition, to see every SOCP is and SDP, recall the Schur complement theorem:

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

For A, C sysmmetric and $C \succ 0$.

Apply this the theorem to the following matrix,

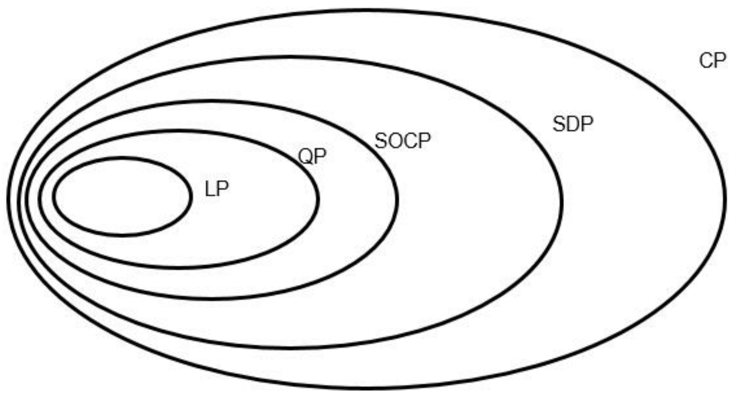
$$\begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0 \Leftrightarrow tI - \frac{xx^T}{t} \preceq 0 \Leftrightarrow \|x\|_2 \leq t.$$

Thus, we can convert the second-order cone constraint to PSD constraint.

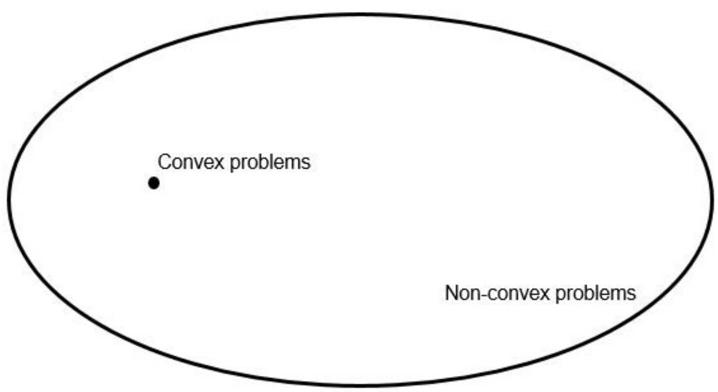
4.6 Relationship between Programs

The relationship between Linear Program (LP), Quadratic Program (QP), Second-Order Cone Program (SOCP), Semidefinite Program (SDP) and Conic Program (CP) is shown in the following figure.

While the relationship between convex problems and non-convex problems is shown in the following figure.



Convex problems just contain the amount of a bubble compared with non-convex ones in this figure.



Acknowledgements

The slides and scribe notes in the former years were referred to while making this scribe note.