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

Lecture 4: September 10

Lecturer: Ryan Tibshirani *Scribes: Lee (Lili) Gao*
Mariya Toneva
Xun Zheng

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

4.1 Last Lecture Leftovers

Relaxing nonaffine equality constaints

Consider an optimization problem of the form:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & x \in C \end{aligned}$$

If we were to take an enlarged constraint set $\tilde{C} \supseteq C$, the optimal value is always smaller or equal to that of the original problem. This technique is called a relaxation.

Relaxation is especially useful when an optimization problem has nonaffine equality constraints of the form:

$$h_j(x) = 0, j = 1, ...r$$

where $h_j, j = 1, ...r$ are convex but nonaffine. Since the convexity of the optimization problem requires affine equality constraints, these convex nonaffine constraints can be relaxed to convex inequalities of the form:

$$h_j(x) \le 0, j = 1, ...r$$

To illustrate the importance of relaxation of nonaffine equalities in optimization, we consider a few examples.

Examples

1. Maximum utility problem

the maximum utility problem models investment/consumption:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & x \in C \end{aligned}$$

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

where b_t is the budget, x_t is the amount consumed at time t, f is an investment return function, u is a utility function. Both f and u are concave and increasing.

Because the equality constraint in the original problem is nonaffine, we can use relaxation to make the criterion concave (as we need to maximize it).

2. Principal Component Analysis (PCA)

Given $X \in \mathbb{R}^{n \times p}$, consider the low rank approximation problem:

$$\begin{aligned} \min_{R \in \mathbb{R}^{n \times p}} \quad & \|X - R\|_F^2 \\ \text{subject to} \quad & rank(R) = k \end{aligned}$$

If $X = UDV^T$, a well-known solution can be found through singular value decomposition. The solution is of the form:

$$R = U_k D_k V_k^T$$

where U_k, V_k are the first k columns of U, V and D_k is the first k diagonal elements of D.

This problem is not convex because rank is not a convex function. However, we can recast this problem into convex form by using relaxation.

To begin, we can rewrite the above as:

$$\begin{aligned} \min_{Z \in \mathbb{S}^p} \quad & \|X - XZ\|_F^2 \\ \text{subject to} \quad & rank(Z) = k \end{aligned}$$

Here the constraint set is the nonconvex set

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

where $\lambda_i(Z), i = 1, ...n$ are the eigenvalues of Z. The solution is

$$Z = V_k V_k^T$$

where V_k gives the first k columns of V.

Next, we can relax the constraint set to $\mathbb{F} = conv(C)$, its convex hull:

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

Note that this is exactly the fantope of order k. Since fantope projections are convex, we can rewrite the above nonconvex problem in a convex form:

$$\begin{aligned} \min_{Z \in \mathbb{S}^p} \quad & \|X - XZ\|_F^2 \\ \text{subject to} \quad & Z \in \mathbb{F} \end{aligned}$$

This reformulation admits the same solution as the nonconvex PCA problem and is thus equivalent, because the convex relaxation is tight at the solution.

4.2 Linear Programs

A linear program (LP) is an optimization program of the following form, which is always a convex optimization problem:

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

LP is the fundamental problem in convex optimization and there are many applications. It is first introduced by Kantorovich in the late 1930s and Dantzig in the late 1940s. Dantzig's simplex algorithm gives a direct solver for LPs.

Examples

1. Diet Problem

The diet problem is to find cheapest combination of foods that satisfies some nutritional requirements. It can be formalized in the following LP problem:

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

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

2. Transportation Problem

The transportation problem is to minimize the cost of shipping commodities from given sources to destinations. It can be formalized in the following LP problem:

$$
\begin{aligned}
\min_{x} \quad & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \\
\text{subject to} \quad & \sum_{j=1}^{n} x_{ij} \le s_i, i = 1, ..., m \\
& \sum_{i=1}^{m} x_{ij} \ge d_j, j = 1, ..., n \\
& x \ge 0
\end{aligned}
$$

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

3. Basis Pursuit

Given $y \in R^n$ and $X \in R^{n \times p}$ with $p > n$. We are seeking th sparsest solution to underdetermined system of equations $X\beta = y$.

The Nonconvex formulation of this problem can be written as

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

The l_1 approximation to the above problem is called basis pursuit

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

which is an LP and can be reformulated as

$$
\begin{aligned}
\min_{\beta, z} \quad & 1^T z \\
\text{subject to} \quad & z \ge \beta \\
& z \ge -\beta \\
& X\beta = y
\end{aligned}
$$

4. Dantzig Selector

A modification of Basis Pursuit, in which we allow for $X\beta \approx y$ (not enforcing exact equality) is called the Dantzig selector, which can be written as:

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

where λ is a tuning parameter. When $\lambda = 0$, this is equivalent to Basis Pursuit. Again, Dantzig selector can also been refomulated into a LP if we write the constrain as

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

Standard Form

A LP 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 LP can be rewritten in standard form.

4.3 Quadratic Programs

A convex quadratic program (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 $Q \succeq 0$. The QP is convex only if $Q \succeq 0$. The remainder of these notes discuss only QP's in which $Q \succeq 0$.

Examples

1. Portfolio optimization

To trade off performance and risk in a financial portfolio, we can use a QP:

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

where μ represents expected assets' returns, Q the covariance matrix of assets' returns, γ risk aversion, x portfolio holdings (percentages).

2. Support vector machines

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

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

3. Lasso

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

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

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

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

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

Now $\lambda \geq 0$ is a tuning parameter. This can also be rewritten as a quadratic program.

Standard form

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

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

Any quadratic program can be rewritten in standard form.

4.4 Semidefinite Programs

4.4.1 Background and motivation

Recall the linear programming problem:

$$\begin{aligned} \min_x \quad & c^\top x && (4.1)\\ \text{subject to} \quad & Dx \leq d && (4.2)\\ & Ax = b. && (4.3) \end{aligned}$$

The idea of SDP is to generalize element-wise inequality in the constraint to partial orders.

First recall a few definitions of space of symmetric matrices.

- $\mathbb{S}^n$ is the space of all $n \times n$ symmetric matrices. Note that $X \in \mathbb{S}^n \Rightarrow \lambda(X) \in \mathbb{R}^n$, *i.e.*, eigenvalues of a symmetric matrix are real.
- $\mathbb{S}^n_+$ is the space of all $n \times n$ positive semidefinite matrices, *i.e.*,

$$\mathbb{S}^n_+ = \left\{ X \in \mathbb{S}^n \ : \ u^\top X u \geq 0, \ \forall u \in \mathbb{R}^n \right\}. \tag{4.4}$$

  Also $X \in \mathbb{S}^n_+ \Leftrightarrow \lambda(X) \in \mathbb{R}^n_+$, *i.e.*, eigenvalues of a symmetric, positive semidefinite matrix are nonnegative.
- $\mathbb{S}^n_{++}$ is the space of all $n \times n$ positive definite matrices, *i.e.*,

$$\mathbb{S}^n_{++} = \left\{ X \in \mathbb{S}^n \ : \ u^\top X u > 0, \ \forall u \in \mathbb{R}^n \setminus \{0\} \right\}. \tag{4.5}$$

  Similarly $X \in \mathbb{S}^n_{++} \Leftrightarrow \lambda(X) \in \mathbb{R}^n_{++}$, *i.e.*, eigenvalues of a symmetric, positive definite matrix are strictly positive.

The partial ordering over $\mathbb{S}^n$ (the **Loewner ordering**) is defined as follows, given $X, Y \in \mathbb{S}^n$:

$$X \succeq Y \quad \Longleftrightarrow \quad X - Y \in \mathbb{S}^n_+. \tag{4.6}$$

4.4.2 Semidefinite programs

A **semidefinite program (SDP)** is an optimization problem of the form

$$\begin{aligned} \min_x \quad & c^\top x && (4.7)\\ \text{subject to} \quad & x_1 F_1 + \cdots + x_n F_n \preceq F_0 && (4.8)\\ & Ax = b, && (4.9) \end{aligned}$$

where $F_j \in \mathbb{S}^d, j = 0, 1, \ldots, n$, $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$. Notice that the first constraint is the linear matrix inequality (LMI), whose solution set is convex as proven in earlier lectures, therefore it is a convex problem. Observe the similarity between LP and SDP: roughly, we have replaced the columns of matrix D and vector d of LP with a set of matrices $F_1, \ldots, F_n$ and F_0 respectively. Also note that if all F_j are diagonal matrices, then the LMI becomes a set of linear inequalities, thus the problem is reduced to linear programs (LP).

A SDP is in the **standard form** if it is written as

$$\begin{aligned} \min_{X \in \mathbb{S}^n} \quad & C \bullet X && (4.10)\\ \text{subject to} \quad & A_i \bullet X = b_i,\ i = 1, \ldots, m && (4.11)\\ & X \succeq 0, && (4.12) \end{aligned}$$

where $C, A_1, \ldots, A_m \in \mathbb{S}^n$ and $X \bullet Y$ is the inner product $\langle X, Y \rangle = \operatorname{tr}\left(X^\top Y\right)$ between two matrices X, Y. Notice the target variable has now became a symmetric matrix instead of a vector. We can also observe similarity between standard form of LP and SDP: instead of nonnegativity constraint, SDP introduced positive semidefinite constraint, which is the matrix version of nonnegativity constraint.

To convert any SDP to standard form, we can again make use of slack variables. In particular, we can first split x into positive and negative parts, *i.e.*, $x = x^+ - x^-$, such that $x^+, x^- \geq 0$. Next, the inequality can be cast into equality by introducing a slack variable $Y \succeq 0$. Then the problem now becomes

$$\begin{aligned} \min_{x^+, x^-, Y} \quad & c^\top x^+ - c^\top x^- && (4.13)\\ \text{subject to} \quad & \left(x_1^+ - x_1^-\right) F_1 + \cdots + \left(x_n^+ - x_n^-\right) F_n + Y = F_0 && (4.14)\\ & Ax^+ - Ax^- = b && (4.15)\\ & x^+ \geq 0,\ x^- \geq 0,\ Y \succeq 0. && (4.16) \end{aligned}$$

The standard form can be realized by constructing block matrices out of x^+, x^-, and Y and rearranging coefficient matrices.

Example: *theta function.* Let $G = (N, E)$ be an undirected graph, $N = \{1, \ldots, n\}$. The **Lovasz theta function** is defined as an SDP:

$$\begin{aligned} \vartheta(G) = \max_X \quad & 11^\top \bullet X && (4.17)\\ \text{subject to} \quad & I \bullet X = 1 && (4.18)\\ & X_{ij} = 0, (i,j) \notin E && (4.19)\\ & X \succeq 0. && (4.20) \end{aligned}$$

Why is this quantity particularly interesting? Denote $\omega(G)$ as the clique number of G, *i.e.* the size of the largest clique in the graph; and $\chi(G)$ as the chromatic number of G, *i.e.* the smallest number of colors needed to color N so that no two adjacent nodes share the same color. Both of these quantities are NP-hard to compute. However, the Lovasz sandwich theorem states that

$$\omega(G) \leq \vartheta(\bar{G}) \leq \chi(G), \tag{4.21}$$

where $\bar{G}$ is the complement graph of G. This is an amazing result since it gives some sense of the two quantities that are NP-hard to compute.

Example: *trace norm minimization.* Let $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ be a linear map,

$$\mathcal{A}(X) = \begin{pmatrix} A_1 \bullet X \\ \cdots \\ A_p \bullet X \end{pmatrix} \tag{4.22}$$

for matrices $A_1, \ldots, A_p \in \mathbb{R}^{m \times n}$. Finding the lowest-rank solution to an underdetermined system can be expressed as

$$\begin{aligned} \min_X \quad & \operatorname{rank}(X) && (4.23)\\ \text{subject to} \quad & \mathcal{A}(X) = b. && (4.24) \end{aligned}$$

If all A_i are diagonal, then we recover the sparse linear system problem. Note that this problem is nonconvex, since rank is a nonconvex function. The trace norm can act as a convex surrogate to the rank function:

$$\min_{X} \quad \|X\|_{\text{tr}} \tag{4.25}$$

$$\text{subject to} \quad \mathcal{A}(X) = b. \tag{4.26}$$

This is an analogy of replacing ℓ_0 norm with ℓ_1 norm in linear systems, but in matrices. The trace norm minimization is an SDP. To show this, recall dual norm of trace norm

$$\|X\|_{\text{tr}} = \max_{\|Y\|_{\text{op}} \leq 1} Y \bullet X. \tag{4.27}$$

Replacing trace norm in the criterion with the variational form, we can rewrite operator norm as SDP constraints.

4.4.3 Conic programs

A **conic program** is an optimization problem of the form:

$$\min_{x} \quad c^\top x \tag{4.28}$$

$$\text{subject to} \quad Ax = b \tag{4.29}$$

$$D(x) + d \in K, \tag{4.30}$$

where $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$; $D : \mathbb{R}^n \to Y$ is a linear map, $d \in Y$ for Euclidean space Y; and $K \subseteq Y$ is a closed convex cone. This is again very similar to LP, the only distinction is the set of linear inequalities are replaced with conic inequalities, *i.e.* $D(x) + d \preceq_K 0$ (see B&V for definition of $\preceq_K$). Notice that if $K = \mathbb{R}^n_+$ the nonnegative orthant, *i.e.* the inequality has the form $D(x) + d \leq 0$, we recover the LP; similarly if $K = \mathbb{S}^n_+$, we recover SDP. Therefore we can see this is a very broad class of problems.

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

$$\min_{x} \quad c^\top x \tag{4.31}$$

$$\text{subject to} \quad \|D_i x + d\|_2 \leq e_i^\top x + f_i, \;\; i = 1, \ldots, p \tag{4.32}$$

$$Ax = b. \tag{4.33}$$

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

$$Q = \{(x, t) : \|x\|_2 \leq t\}. \tag{4.34}$$

From this definition it is easy to see

$$\|D_i x + d\|_2 \leq e_i^\top x + f_i \iff \left(D_i x + d, e_i^\top x + f_i\right) \in Q_i, \tag{4.35}$$

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

It is easy to see every LP is SOCP. Furthermore, to see every SOCP is an SDP, first recall the **Schur complement theorem**:

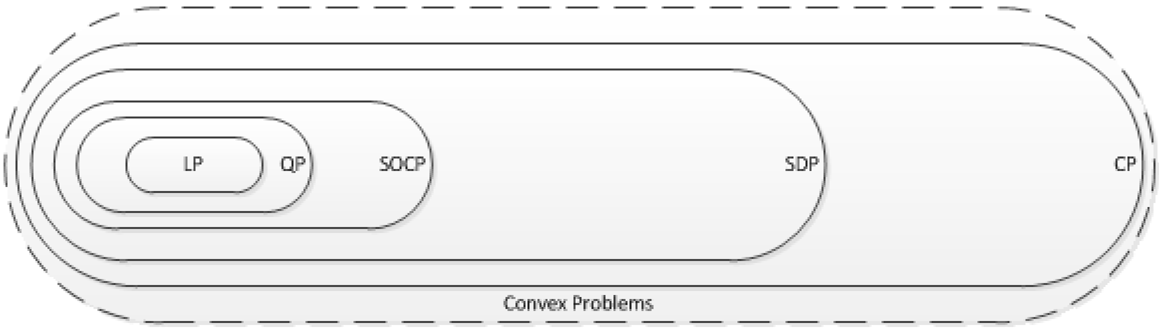
$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \succeq 0 \iff A - BC^{-1}B^\top \succeq 0, \tag{4.36}$$

for A, C symmetric and $C \succ 0$. Apply this theorem to the following matrix,

$$\begin{bmatrix} tI & x \\ x^\top & t \end{bmatrix} \succeq 0 \iff tI - \frac{xx^\top}{t} \succeq 0 \iff \|x\|_2 \leq t. \tag{4.37}$$

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

The relationship between Linear Program (LP), convex quadratic program (QP), second-order cone program (SOCP), semidefinite program (SDP) and conic program (CP) in shown in the following figure



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