Semidefinite and Second Order Cone Programming Seminar
Fall 2012
Lecture 6

Instructor: Farid Alizadeh
Scribe: Svetlana Soloveva

10/15/2012

1 Overview

We consider barrier function method for SOCP, discuss polynomial time complexity of a SDP problem, introduce path following method, big M method, and potential functions.

2 Barrier function method for second order cone programming (SOCP)

First, let's recall $\circ$ operation for matrices X and S is $X \circ S = \frac{XS+SX}{2}$. Its properties are:

i. if X and S are symmetric matrices then $X \circ S$ is symmetric;

ii. $X \circ (Y \circ Z) \neq (X \circ Y) \circ Z$ (not associative in general);

iii. $X^2 \circ (X \circ Y) = X \circ (X^2 \circ Y)$;

iv. $X \circ (\alpha Y + \beta Z) = \alpha X \circ Y + \beta X \circ Z$ (bilinear operator).

Now let's state the general second order cone programming

$$\begin{cases} \min \mathbf{c}_1^T \mathbf{x}_1 + ... + \mathbf{c}_n^T \mathbf{x}_n \\ A_1 \mathbf{x}_1 + ... + A_n \mathbf{x}_n = \mathbf{b}, \\ \mathbf{x}_i \succeq_Q \mathbf{0},\ \forall i = 1, \ldots, n \end{cases}$$

For simplicity, we will consider only a problem with one block

$$\begin{cases} \min \mathbf{c}^T \mathbf{x} \\ A\mathbf{x} = \mathbf{b}, \\ \mathbf{x} \succeq_Q \mathbf{0} \end{cases}$$

and its dual

$$\begin{cases} \max \mathbf{b}^T \mathbf{y} \\ A^T\mathbf{x} + \mathbf{s} = \mathbf{c}, \\ \mathbf{s} \succeq_Q \mathbf{0} \end{cases}$$

Let's consider cone $Q = \{\mathbf{x} : x_0 \geq \|\overline{\mathbf{x}}\|\}$ where

$$\mathbf{x} = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_0 \\ \overline{\mathbf{x}} \end{pmatrix}$$

We will use the following barrier function: $-\ln(-x_0^2 - \|\overline{\mathbf{x}}\|^2)$. Then we can replace the initial problem with the following

$$\begin{cases} \min \mathbf{c}^T \mathbf{x} - \mu \ln(x_0^2 - \|\overline{\mathbf{x}}\|^2) \\ A\mathbf{x} = \mathbf{b}, \\ \mathbf{x} \succeq_Q \mathbf{0} \end{cases}$$

The Lagrangian is $L(\mathbf{x}, \mathbf{y}) = \mathbf{c}^T\mathbf{x} - \mu \ln(x_0^2 - \|\overline{\mathbf{x}}\|^2) + \mathbf{y}^T(\mathbf{b} - A\mathbf{x})$ and its gradient
$\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) = \mathbf{c}^T - \frac{2\mu}{x_0^2 - \|\overline{\mathbf{x}}\|^2}(x_0, -x_1, ..., -x_n) - \mathbf{y}^T A$,
$\nabla_{\mathbf{y}} L(\mathbf{x}, \mathbf{y}) = \mathbf{b} - A\mathbf{x}$.
Let vector $\mathbf{s}$ be such that $\mathbf{s}^T = (s_0, \dots, s_n) = \frac{2\mu}{x_0^2 - \|\overline{\mathbf{x}}\|^2}(x_0, -x_1, ..., -x_n)$.
Then $\mathbf{s}^T \succ_Q \mathbf{0}$ and

$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) = \mathbf{0}, \\ \nabla_{\mathbf{y}} L(\mathbf{x}, \mathbf{y}) = \mathbf{0} \end{cases}$$

$\Rightarrow$

$$\begin{cases} A\mathbf{x} = \mathbf{b}, \\ A^T\mathbf{y} + \mathbf{s} - \mathbf{c} = \mathbf{0} \\ x_0 s_0 + x_1 s_1 + ... + x_n s_n = 2\mu, \\ x_0 s_i + x_i s_0 = 0 \end{cases}$$

$\Leftrightarrow$

$$\begin{cases} A\mathbf{x} = \mathbf{b}, \\ A^T\mathbf{y} + \mathbf{s} - \mathbf{c} = \mathbf{0} \\ \mathbf{x} \circ \mathbf{s} = 2\mu \boldsymbol{e} \end{cases}$$

We apply Newton's method to get

$$\begin{cases} A\triangle\mathbf{x} = \mathbf{b} - A\mathbf{x}, \\ A^T\triangle\mathbf{y} + \triangle\mathbf{s} = \mathbf{c} - A^T\mathbf{y} - \mathbf{s}, \\ (\mathbf{x} + \triangle\mathbf{x}) \circ (\mathbf{s} + \triangle\mathbf{s}) = 2\mu\boldsymbol{e} \end{cases}$$

$\Rightarrow$

$$\begin{cases} A\triangle \mathbf{x} = \mathbf{b} - A\mathbf{x}, \\ A^T \triangle \mathbf{y} + \triangle \mathbf{s} = \mathbf{c} - A^T \mathbf{y} - \mathbf{s}, \\ \mathbf{s} \circ \triangle \mathbf{x} + \mathbf{x} \circ \triangle \mathbf{s} = 2\mu \mathbf{e} - \mathbf{x} \circ \mathbf{s} \end{cases}$$

$\Leftrightarrow$ Matrix form

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I \\ Arw(S) & 0 & Arw(X) \end{pmatrix} \begin{pmatrix} \triangle \mathbf{x} \\ \triangle \mathbf{y} \\ \triangle \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_p \\ \mathbf{r}_d \\ \mathbf{r}_c \end{pmatrix}$$

where $Arw(X)$ is such that

$$x \circ s = \begin{pmatrix} x_0 s_0 + x_1 s_1 + \cdots + x_n s_n \\ \vdots \\ x_i y_0 + y_i x_0 \\ \vdots \end{pmatrix} = \underbrace{\begin{pmatrix} x_0 & x_1 & \dots & & x_n \\ x_1 & x_0 & 0 & \dots & 0 \\ x_2 & 0 & x_0 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ x_n & 0 & \dots & 0 & x_0 \end{pmatrix}}_{Arw(X)} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

3 The Path Following Method

Consider a primal-dual feasible region. We start from the point called *logarithmic center*. At this point penalty coefficient $\mu \to \infty$, and the optimal solution is as far away from every barrier as possible. Then construct a path following which μ decreases from ∞ to 0 until we reach an optimal solution. Usual way of proving polynomiality of a path following algorithm is to bound number of reductions of μ.
Unless alternative representations of solutions of SDP's are found, we cannot in general speak of polynomial time complexity of a SDP problems. To see this, let's consider the following examples:

$$\begin{cases} \min x_n \\ x_1 = 2, \\ x_{i+1} \geq x_i^2, \forall i = 1, \dots, n \end{cases} \qquad \begin{cases} \min x_n \\ x_1 = \frac{1}{2}, \\ x_{i+1} \geq x_i^2, \forall i = 1, \dots, n \end{cases}$$

As it can be trivially observed, the optimal solution satisfies $x_n^* = 2^{2^n}$, for the first problem and $x_n^* = 2^{-2^n}$ for the second. Therefore, in both cases we would need exponentially many bits in problem size to express the solution of this problem. However, as Khachyian and Porkolab proved, the exponential number of bits is the worst case scenario.

scribe: Svetlana Soloveva
Lecture 6 **Date:** 10/15/2012

4 Big M method

Replace primal problem with the following problem:

$$\begin{cases} \min \mathbf{c}^T\mathbf{x} - Mx' \\ A\mathbf{x} + (\mathbf{b} - A\mathbf{x}_0)x' = \mathbf{b}, \\ \langle \mathbf{c} - A^T\mathbf{y}_0 - \mathbf{s}_0, \mathbf{x}\rangle \le N, \\ \mathbf{x} \succeq_K 0, x' \ge 0 \end{cases}$$

Then its dual is

$$\begin{cases} \max \mathbf{b}^T\mathbf{y} - Ny' \\ A^T\mathbf{y} + \mathbf{s} + (\mathbf{c} - A^T\mathbf{y}_0 - \mathbf{s}_0)y' = \mathbf{c}, \\ (\mathbf{b} - A\mathbf{x}_0)^T\mathbf{y} \ge M, \\ \mathbf{s} \succeq_{K^*} 0, y' \ge 0 \end{cases}$$

If optimal solutions $(\mathbf{x}^*, x^{*\prime})$, $(\mathbf{y}^*, \mathbf{s}^*, y^{*\prime})$ are such that $x^{*\prime} = 0, y^{*\prime} = 0$, then $\mathbf{x}^*$ and $(\mathbf{y}^*, \mathbf{s}^*)$ are optimal for the initial primal and dual problems. However, we need to choose M and N such that x' and y' can achieve zero at optimal solution. If $x^{*\prime} \neq 0$ or $y^{*\prime} \neq 0$, then either the initial primal or the initial dual problem is infeasible respectively, or we have not selected suitable values for M and N. If we know an upper bound on the size of the optimal solution of the original primal and dual problems, then M and N can be set equal to those numbers, respectively.

5 Potential functions

We now develop an approach, based on the notion of *potential functions* which allows us to precisely bound the number of operations required to reduce the duality gap in a pair of primal and dual semidefinite programs to below a prescribed value of ϵ.

First, let's define primal potential function as

$$\phi(X, \underset{\sim}{z}) = q\ln(C \bullet X - \underset{\sim}{z}) - \ln\det X,$$

where $\underset{\sim}{z} < C \bullet X$ $\forall X$ primal feasible. Define the primal-dual potential function as

$$\psi(X, S) = q\ln(X \bullet S) - \ln\det X - \ln\det S.$$

Note that $\psi(X, S) = q\ln(X \bullet S) - \ln\det XS = q\ln(X \bullet S) - \ln\det X^{\frac{1}{2}}SX^{\frac{1}{2}}$

Theorem 1 *Let* $(X_0, \mathbf{y}_0, S_0)$ *be primal-dual feasible with* $X_0 > 0$, $S_0 > 0$. *Set* $q = n + \sqrt{n}$. *Suppose that the sequence* $(X_0, \mathbf{y}_0, S_0)$, $(X_1, \mathbf{y}_1, S_1), \ldots$ *are feasible solutions with* $X_i > 0$, $S_i > 0$ $\forall i$, *and* $\psi(X_k, S_k) - \psi(X_{k+1}, S_{k+1}) > \delta$ *for a fixed constant* $\delta > 0$. *If* $\psi(X_0, S_0) \le \sqrt{n}E$ *for some constant* $E > 0$, *then after* $k = O(\sqrt{n}(E- \mid \ln(\varepsilon) \mid))$ *iterations* $C \bullet X_k - \mathbf{b}^T\mathbf{y}_k = X_k \bullet S_k < \varepsilon$.

Proof: We know $\psi(X_0, S_0) \leq \sqrt{n}E$ or equivalently $c\psi(X_0, S_0) \leq c\sqrt{n}E$ for any $c \geq 0$ and some E, and $\psi(X_k, S_k) - \psi(X_{k+1}, S_{k+1}) > \delta$ for some $\delta > 0$. Then

$$\begin{aligned}
&\psi(X_0, S_0) - \psi(X_1, S_1) > \delta, \\
&\psi(X_1, S_1) - \psi(X_2, S_2) > \delta, ..., \\
&\psi(X_{k-1}, S_{k-1}) - \psi(X_k, S_k) > \delta, \quad \text{Therefore,} \\
&\psi(X_0, S_0) - \psi(X_k, S_k) > k\delta \quad \text{or} \\
&c\psi(X_0, S_0) - c\psi(X_k, S_k) > ck\delta \quad \text{for } c \geq 0 \quad \text{and} \\
&c\psi(X_k, S_k) < c\psi(X_0, S_0) - ck\delta = c\sqrt{n}E - ck\delta.
\end{aligned}$$

Since $k = O(\sqrt{n}(E- \mid \ln(\varepsilon) \mid))$, we have

$$c\psi(X_k, S_k) < c\sqrt{n}E - c(\sqrt{n}(E- \mid \ln(\varepsilon) \mid)\delta \leq c\sqrt{n} \mid \ln(\varepsilon) \mid)$$

(take the positive multiplier c inside of big O notation with this step). By definition,

$$\psi(X_k, S_k) = (n + \sqrt{n}) \ln(X_k \bullet S_k) - \ln \det(X_k S_k),$$

and

$$c(n + \sqrt{n}) \ln(X_k \bullet S_k) - c \ln \det(X_k S_k) < c\sqrt{n} \mid \ln(\varepsilon) \mid)$$

therefore,

$$c\sqrt{n} \ln(X_k \bullet S_k) < -cn \ln(X_k \bullet S_k) + c \ln \det(X_k S_k) + c\sqrt{n} \mid \ln(\varepsilon) \mid)$$

Now, by the arithmetic and geometric means inequality we have:

$$\frac{\sum_i \lambda_i(X_k S_k)}{n} \geq (\prod_i \lambda_i(X_k S_k))^{\frac{1}{n}}$$

Thus,

$$\begin{aligned}
&\ln(X_k \bullet S_k) - \ln n \geq \frac{1}{n} \ln \det(X_k S_k) \\
&n \ln(X_k \bullet S_k) - n \ln n \geq \ln \det(X_k S_k) \\
&- n \ln n \geq -n \ln(X_k \bullet S_k) + \ln \det(X_k S_k)
\end{aligned}$$

Therefore, $c\sqrt{n} \ln(X_k \bullet S_k) < -cn \ln n + c\sqrt{n} \mid \ln(\varepsilon) \mid) < c\sqrt{n} \mid \ln(\varepsilon) \mid) \Rightarrow$ $X_k \bullet S_k < \varepsilon$. ■

The next topic of our discussion will be the algorithm which generates the sequence $(\mathbf{x}_0, \mathbf{y}_0, S_0)$, $(\mathbf{x}_1, \mathbf{y}_1, S_1)$, ... of feasible solutions.