Semidefinite and Second Order Cone Programming Seminar
Fall 2012
Lecture 7

Instructor: Farid Alizadeh
Scribe: Deniz S.Eskandani

10/22/2012

1 Overview

We introduce an algorithm that guarantees the reduction of primal-dual potential function. Then every $O(\sqrt{n})$ iterations gives another bit of accuracy.

2 Potential Function

We suppose that we already have an initial feasible primal dual solution (X_k, y_k, S_k).

Let $q > 0$, and z_0 be a lower bound on the primal solution.

Define the *primal potential function*:

$$\phi(X, z_0) = q \log(C \bullet X - z_0) - \log \det X,$$

And also, the *primal-dual potential function* as:

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

If y is a dual feasible solution then we can set $z_0 = \mathbf{b}^T y$ because z_0 is any arbitrary lower bound on the value of the objective function. Then we have:

$$\psi(X, \mathbf{s}) = \phi(X, z_0) - \log \det(S)$$

We will present an algorithm where in each iteration either X will change, or (y, S). Suppose only X changes, and (y, S) is unchanged from the previous iteration. Then the relation above shows that if ϕ is reduced by a constant amount (say δ), this would imply that ψ is also reduced by the same amount (as $\ln \det S$ is fixed).

3 Useful Lemmas

Before describing the algorithm we state some lemmas. Recall that if Y is a symmetric matrix we have $\|Y\|_F = \sqrt{\sum y_{ij}^2} = \sqrt{\sum \lambda_i^2(Y)}$ is the *Frobenius norm*, and $\|Y\|_2 = \max_i |\lambda_i(Y)|$ is the *spectral norm*.

Lemma 1 *If $0 \prec X \prec I$, that is every eigenvalue of X is strictly between 0 and 1, then*

$$\log\det X \geq \operatorname{trace} X - n - \frac{\|X - I\|_F^2}{2(1 - \|X - I\|_2)}$$

Proof: For $0 < z < 1$ we have

$$\begin{aligned}
\log z &= \log(1 - (1 - z)) \\
&= -(1 - z) - \frac{(1 - z)^2}{2} - \frac{(1 - z)^3}{3} - \cdots \\
&\geq -(1 - z) - \frac{(1 - z)^2}{2} - \frac{(1 - z)^3}{2} - \cdots \\
&= z - 1 + \frac{(1 - z)^2}{2(1 - (1 - z))}
\end{aligned}$$

Now substitute $\lambda_j(X)$ for z and add them up:

$$\sum \log \lambda_i(X) \geq \sum \lambda_i(X) - n - \frac{\sum (1 - \lambda_i(X))^2}{2(1 - \max \lambda_i(X - I))}$$

$$\log\det X \geq \operatorname{trace} X - n - \frac{\|X - I\|_F^2}{2(1 - \|X - I\|_2)}$$

■

Note 2 *The derivative of the primal potential function is:*

$$\nabla_x \phi(X, z_0) = \frac{q}{C \bullet X}.C - X^{-1}$$

Lemma 3 *If $X_0, X_1 \succ 0$ and $\|X_0^{-1/2}(X_1 - X_0)X_0^{-1/2}\|_2 < 1$, then*

$$\phi(X_1, z_0) - \phi(X_0, z_0) \leq \nabla_X \phi(X_0, z_0) \bullet (X_1 - X_0) + \frac{\|X_0^{-1/2}(X_1 - X_0)X_0^{-1/2}\|_F^2}{2(1 - \|X_0^{-1/2}(X_1 - X_0)X_0^{-1/2}\|_2)}$$

Proof:

$$\begin{aligned}
& \phi(X_1, z_0) - \phi(X_0, z_0) \\
= \; & q \log(C \bullet X_1 - z_0) - \log\det X_1 - q\log(C \bullet X_0 - z_0) + \log\det X_0 \\
\leq \; & \nabla_X \phi(X_0, z_0) \bullet (X_1 - X_0) + \nabla_X \log\det X_0 \bullet (X_1 - X_0) - \log\det X_1 + \log\det X_0 \\
= \; & \nabla_X \phi(X_0, z_0) \bullet (X_1 - X_0) + X_0^{-1} \bullet (X_1 - X_0) - \log\det(X_0^{-1/2} X_1 X_0^{-1/2}) \\
\leq \; & \nabla_X \phi(X_0, z_0) \bullet (X_1 - X_0) + X_0^{-1} \bullet (X_1 - X_0) - \mathrm{trace}(X_0^{-1/2} X_1 X_0^{-1/2}) \\
& + n + \frac{\|X_0^{-1/2}(X_1 - X_0)X_0^{-1/2}\|_F^2}{2(1 - \|X_0^{-1/2}(X_1 - X_0)X_0^{-1/2}\|_2)} \\
= \; & \nabla_X \phi(X_0, z_0) \bullet (X_1 - X_0) + \frac{\|X_0^{-1/2}(X_1 - X_0)X_0^{-1/2}\|_F^2}{2(1 - \|X_0^{-1/2}(X_1 - X_0)X_0^{-1/2}\|_2)}
\end{aligned}$$

The first inequality is derived from the fact that if $f(x)$ is a smooth concave function then $\nabla_x f|_{x=x_0}(x - x_0) \geq f(x) - f(x_0)$; here $f(X) = q \ln(C \bullet X - z_0)$. The second inequality is from lemma 1 by replacing $X_0^{-1/2} X_1 X_0^{-1/2}$ in X and the last equality is derived from the fact that $\mathrm{trace}(X_0^{-1/2} X_1 X_0^{-1/2}) = \mathrm{trace}(X_0^{-1} X_1) = X_0^{-1} \bullet X_1$. ■

4 Scaling

In the strategy of interior point methods if the current point is well inside the interior of the feasible set, then in the Newton direction we can go a long way. However if the current point is close to boundary it is possible that in the Newton direction we cannot go a long way and you hit the boundary very quickly. So there is an incentive to be near the center of the feasible region, so that we can take long steps towards improving the objective function. On the other hand the optimal solution is at the boundary. So eventually we wish to be close to boundary. The brilliant trick that Karmarkar came up with is as follows. Consider linear programming. Suppose that current point is x_0 is well centered and the next iteration point is x_1. This point is still an interior point but x_1 may not be well centered. The idea is to make a linear transformation such that the new point is sent back to the center again while the whole optimization problem is preserved, in the sense that the value of the objective function is unchanged. Karamrkar's original idea was to use *projective transformation*. However, it was later recognized that linear transformations can also be used. In summary there is a linear scaling that map the cone (the positive orthant in LP, the semidefinite cone in SDP, and Lorentz cone ins SOCP) back to itself.

The best centered point for semidefinite program is the identity. Thus we seek a linear transformation that maps the positive semidefinite cone back to itself (i.e. it is an *automorphism* of the cone) and brings the current primal feasible point, say X_0 to the identity. Suppose that X is an arbitrary point in the interior of semidefinite cone. Now we want to define a linear transformation that maps in particular X_0, the current solution, to identity. For this purpose we can use this transformation:

$$X \to X_0^{-1/2} X X_0^{-1/2}$$

Under this rescaling the primal problem:

$$\begin{array}{ll} \text{min:} & C \bullet X \\ \text{s.t.} & A_i \bullet X = b_i \\ & X \succeq 0 \\ & X_0 \succ 0 \text{ is feasible} \end{array} \tag{1}$$

will be transformed as follows. Let us define $\tilde{X}$, $\underset{\sim}{C}$ and $\underset{\sim}{A_i}$ as follow:

$$\begin{array}{rcl} \tilde{X} & \stackrel{\text{def}}{=} & X_0^{-1/2} X X_0^{-1/2} \\ \underset{\sim}{C} & \stackrel{\text{def}}{=} & X_0^{1/2} C X_0^{1/2} \\ \underset{\sim}{A_i} & \stackrel{\text{def}}{=} & X_0^{1/2} A_i X_0^{1/2} \end{array}$$

Then we have $\underset{\sim}{C} \bullet \tilde{X} = C \bullet X$. So the rescaling of primal problem is:

$$\begin{array}{ll} \text{min:} & \underset{\sim}{C} \bullet \tilde{X} \\ \text{s.t.} & \underset{\sim}{A_i} \bullet \tilde{X} = b_i \\ & \tilde{X} \succeq 0 \\ & \text{I is feasible} \end{array} \tag{2}$$

If X_0 is a feasible point of (1) in the interior of the feasible region, then by this rescaling I is feasible for the transformed problem. And the values of the objective function at I in the transformed problem equals the value of the objective function at X_0 in the original problem. Now the idea is to try to focus on the (2) problem and try to find new point for this problem and scale it back. We try to replace this problem with easier one. We are looking at primal potential function but this primal potential function is for the original problem. After rescaling we have:

$$\begin{array}{ll} \text{min:} & (X_0^{1/2} \nabla_X \phi(X_0, z_0) X_0^{1/2}) \bullet (X' - I) \\ \text{s.t.} & \underset{\sim}{A_i} \bullet (X' - I) = 0 \\ & \|X' - I\|_F \leq \beta, \quad 0 < \beta < 1 \end{array} \tag{3}$$

The $\|X' - I\|_F \leq \beta$ means that the eigenvalues of X' are between $1 \pm \beta$ and they never can reach zero because β is strictly less than 1. So every feasible point of (3) will be strictly positive definite matrix. The solution of (3) is within the intersection of a ball and an affine space, smaller ball, and for finding the solution, we project $\underset{\sim}{A}$ to smaller ball and then go as far as β.

In semidefinite programming we wrote:

$$A_i \bullet X = b_i \quad i = 1, \cdots, m$$

that it is the same as:

$$(\mathrm{vec}(A_i)^T \mathrm{vec}(X) = \mathbf{b}_i$$
$$A\,\mathrm{vec}(X) = \mathbf{b}$$

So $\mathrm{vec}(\underset{\sim}{A}_i) = (X_0^{1/2} \otimes X_0^{1/2})\,\mathrm{vec}(A_i)$. Then we have $\underset{\sim}{A} = A(X_0^{1/2} \otimes X_0^{1/2})$.

In general if we have a matrix B and a vector $\mathbf{u}$, the formula for the orthogonal projection of $\mathbf{u}$ to the null space of B is:

$$P_B(\mathbf{u}) = (I - B^T(BB^T)^{-1}B)\mathbf{u}.$$

The solution to (3) is given by the orthogonal projection of the $X_0^{1/2}\nabla_X\phi(X_0, z_0)X_0^{1/2}$ to the null space of $\underset{\sim}{A}$ that is:

$$\begin{aligned} P(z_0) &\overset{\text{def}}{=} P_{\underset{\sim}{A}}(X_0^{1/2}\nabla_X\phi(X_0, z_0)X_0^{1/2}) \\ &= [I - (X_0^{1/2} \otimes X_0^{1/2})A^T(AX_0 \otimes X_0A^T)^{-1}A(X_0^{1/2} \otimes X_0^{1/2})] \\ &\quad (X_0^{1/2} \otimes X_0^{1/2})\,\mathrm{vec}(\nabla_X\phi(X_0, z_0)) \end{aligned}$$

So the optimal solution of (3) can be calculated by

$$X' = I - \beta\frac{P(z_0)}{\|P(z_0)\|_F}.$$

This is also a feasible interior point for (2). Therefore, applying the inverse scaling:

$$X_1 \overset{\text{def}}{=} X(z_0) = X_0^{1/2}X'X_0^{1/2} = X_0 - \beta\frac{X_0^{1/2}P(z_0)X_0^{1/2}}{\|P(z_0)\|_F}$$

we get an interior feasible solution for the original SDP problem (1). This point serves as a candidate for our next solution. We will analyze its effect on the primal potential function.

Lemma 4

$$\nabla_X\phi(X_0, z_0) \bullet (X_1 - X_0) = -\beta\|P(z_0)\|_F$$

Proof:

$$\begin{aligned} \nabla_X\phi(X_0, z_0) \bullet (X_1 - X_0) &= \left(X_0^{1/2}\nabla_X\phi(X_0, z_0)X_0^{1/2}\right) \bullet (I - X') \quad \text{using trace definitoin} \\ &= -\beta\left(X_0^{1/2}\nabla_X\phi(X_0, z_0)X_0^{1/2}\right) \bullet \frac{P(z_0)}{\|P(z_0)\|_F} \quad \text{from definition of } X' \\ &= -\beta\left(X_0^{1/2}\nabla_X\phi(X_0, z_0)X_0^{1/2}\right) \bullet \frac{P_{\underset{\sim}{A}}\left(X_0^{1/2}\nabla_X\phi(X_0, z_0)X_0^{1/2}\right)}{\|P(z_0)\|_F} \\ &= -\beta P_{\underset{\sim}{A}}\left(X_0^{1/2}\nabla_X\phi(X_0, z_0)X_0^{1/2}\right) \bullet \frac{P_{\underset{\sim}{A}}\left(X_0^{1/2}\nabla_X\phi(X_0, z_0)X_0^{1/2}\right)}{\|P(z_0)\|_F} \\ &\quad \text{since for projections we have } P_{\underset{\sim}{A}}^2 = P_{\underset{\sim}{A}} \\ &= -\beta\frac{\|P(z_0)\|_F^2}{\|P(z_0)\|_F} = -\beta\|P(z_0)\|_F. \end{aligned}$$

■

Combining Lemma 3 and Lemma 4 we get:

Corollary 5

$$\phi(X_1, z_0) - \phi(X_0, z_0) \leq -\beta \|P(z_0)\|_F + \frac{\beta^2}{2(1-\beta)}$$

5 Algorithm

The goal is to show there is a fixed amount of reduction of primal-dual potential function. According to corollary (5), all we need to do is to show that $\|P(z_0)\|$ is sufficiently large. Now we do a little algebraic here and then give the main lemma. We write $P(z_0)$ a little bit cleaner as follow:

$$P(z_0) = (I - \underset{\sim}{A}^T(\underset{\sim}{A}\underset{\sim}{A}^T)^{-1}\underset{\sim}{A})\,\mathrm{vec}\big(\nabla_X \phi(X_0, z_0)\big)$$

Now we are rewriting $P(z_0)$ as:

$$P(z_0) = \frac{q}{C \bullet X_0 - z_0} X_0^{1/2} S(z_0) X_0^{1/2} - I,$$

where:

$$\begin{aligned} S(z_0) \stackrel{\text{def}}{=} \quad & C - A^T(A(X_0 \otimes X_0)A^T)^{-1} A(X_0^{1/2} \otimes X_0^{1/2}) \\ & \mathrm{vec}\big(X_0^{1/2} C X_0^{1/2} - \tfrac{C \bullet X_0 - z_0}{q} I\big) \end{aligned}$$

We define $y(z_0)$ to be:

$$\mathbf{y}(z_0) \stackrel{\text{def}}{=} (A(X_0 \otimes X_0)A^T)^{-1} A(X_0^{1/2} \otimes X_0^{1/2})\,\mathrm{vec}\big(X_0^{1/2} C X_0^{1/2} - \frac{C \bullet X_0 - z_0}{q} I\big)$$

So $\mathbf{y}(z_0)$ can be written in the form of:

$$\mathbf{y}(z_0) = \mathbf{y}_1(z_0) + \frac{C \bullet X_0 - z_0}{q} \mathbf{y}_2(z_0),$$

where

$$\mathbf{y}_1(z_0) = (A(X_0 \otimes X_0)A^T)^{-1} A(X_0 \otimes X_0)(\vec{C}),$$

$$\mathbf{y}_2(z_0) = (A(X_0 \otimes X_0)A^T)^{-1} A\,\mathrm{vec}(X_0) = (A(X_0 \otimes X_0)A^T)^{-1} b.$$

So far we have just kept z_0 to be any lower bound on the primal. For now we are going to set $z_0 = b^T \mathbf{y}_0$. We already have $(X_0, \mathbf{y}_0, S_0)$ as our current primal and dual feasible solution. Now we are trying to cook up our next iteration. We have defined X_1. Potentially S_1 is equal to $S(z_0)$ and $\mathbf{y}_1$ is the same as $\mathbf{y}(z_0)$, we say potentially because we do not know if $S(z_0)$ is positive semidefinite and therefore it may not be feasible. If $S(z_0)$ is not semidefinite we keep the old $\mathbf{y}$ and S as before and we only change X. But if S is positive semidefinite and has the other conditions outlined below then we change the dual. We are going to show that in either case the primal-dual potential function is going to be reduced.

Let's define some notation:

$$\begin{aligned} \Delta_0 &= \frac{X_0 \bullet S_0}{n}, \quad \text{(current gap)} \\ \Delta &= \frac{X_0 \bullet S(z_0)}{n}, \\ q &= n + \sqrt{n}. \end{aligned}$$

The key lemma is as follows where basically everything comes together.

Lemma 6 *Let* $0 < \alpha < 1$

i) If $S(z_0) \nsucceq 0$ *then*

$$||P(z_0)||_F > 1$$

ii) Else if $||X_0^{1/2} S(z_0) X_0^{1/2}||_2 \geq \alpha\Delta$ *then*

$$||P(z_0)||_F \geq \alpha\sqrt{\frac{n}{n+\alpha^2}}$$

iii) Else if $\Delta \geq (1 - \frac{\alpha}{2\sqrt{n}})\Delta_0$ *then*

$$||P(z_0)||_F \geq 1 - \alpha$$

iv) Else

$$\psi(X_0, S(z_0)) - \psi(X_0, S_0) \leq -\frac{\alpha}{2} + \frac{\alpha^2}{2(1-\alpha)}$$

The first case basically says that if your S is not positive semidefinite then stick with the old S and just update the new X and your new X already gives a good new bound for $||P(z_0)||_F$. In the first three cases we bound $||P(z_0)||$ and therefore we guarantee a reduction in the primal potential function. So we can throw away duals, just keep the primal and the new S is equal to the old S. Therefore the primal dual potential function is the same as the primal potential function minus a constant. So the reduction in the ϕ transfer in the reduction of ψ. In the fourth case we are sticking to the old X but now we change the dual. This also guarantees a fixed amount of reduction to the primal dual potential function.

Proof:

i) If $S(z_0) \nsucceq 0$ then $X_0^{1/2} S(z_)X_0^{1/2} \nsucceq 0$. As $P(z_0) = \frac{q}{C \bullet X - z_0} X_0^{1/2} S(z_0) X_0^{1/2} - I$ and $X_0^{1/2} S(z_)X_0^{1/2} \nsucceq 0$ then $P(z)$ must have an eigenvalue less than -1. So $||P(z_0)||_F = \sqrt{\sum \lambda_i^2} > 1$.

ii) $\|P(z_0)\|_F = \|\frac{q}{n\Delta_0}X_0^{1/2}S(z_0)X_0^{1/2} - \frac{q\Delta}{n\Delta_0}I + \frac{q\Delta}{n\Delta_0}I - I\|_F$.

Trace of $\frac{q}{n\Delta_0}X_0^{1/2}S(z_0)X_0^{1/2} - \frac{q\Delta}{n\Delta_0}I$ is zero, because by definition of Δ we have $n\Delta = X_0 \bullet S(z_0)$. Therefore $\frac{q}{n\Delta_0}X_0^{1/2}S(z_0)X_0^{1/2} - \frac{q\Delta}{n\Delta_0}I$ is orthogonal to $\frac{q\Delta}{n\Delta_0}I - I$. So

$$
\begin{aligned}
(\star) \quad \|P(z_0)\|_F^2 &= \|\frac{q}{n\Delta_0}X_0^{1/2}S(z_0)X_0^{1/2} - \frac{q\Delta}{n\Delta_0}I\|^2 + (\frac{q\Delta}{n\Delta_0} - 1)^2 n \\
&\geq (\frac{q}{n\Delta_0})^2\alpha^2\Delta^2 + (\frac{q\Delta}{n\Delta_0} - 1)^2 n
\end{aligned}
$$

If we set $t = \frac{q\Delta}{n\Delta_0}$, then we have $t^2\alpha^2 + (t-1)^2 n \geq \alpha^2 \frac{n}{n+\alpha^2}$.

iii) Since $\frac{\Delta}{\Delta_0} \geq (1 - \frac{\alpha}{2\sqrt{n}})$, then

$$
\frac{q\Delta}{n\Delta_0} = (1 + \frac{1}{\sqrt{n}})\frac{\Delta}{\Delta_0} \geq (1 + \frac{1}{\sqrt{n}})(1 - \frac{\alpha}{2\sqrt{n}}) \geq 1.
$$

$$
\begin{aligned}
\|P(z_0)\|_F^2 &\geq (\frac{q\Delta}{n\Delta_0} - 1)^2 n \qquad (\text{From}(\star)) \\
&\geq ((1 + \frac{1}{\sqrt{n}})(1 - \frac{\alpha}{2\sqrt{n}}) - 1)^2 n \\
&= (1 - \frac{\alpha}{2\sqrt{n}} + \frac{1}{\sqrt{n}} - \frac{\alpha}{2n} - 1)^2 n \\
&= (-\frac{\alpha}{2} + 1 - \alpha 2\sqrt{n})^2 \geq (1 - \alpha)^2
\end{aligned}
$$

iv) $\Delta < (1 - \frac{\alpha}{2\sqrt{n}})\Delta_0$

$$
\|X_0^{1/2S(z_0)X_0^{1/2}}\|_F < \alpha\Delta
$$

therefore

$$
\sqrt{n}\log\frac{X_0 \bullet S(z_0)}{X_0 \bullet S_0} = \sqrt{n}\log\frac{\Delta}{\Delta_0} \leq -\frac{\alpha}{2} \quad (\star\star)
$$

The last inequality is derived from the fact that $\log(1 - z) \leq -z$.

$$
\begin{aligned}
& n\log X_0 \bullet S(z_0) - \log\det(X_0^{1/2}S(z_0)X_0^{1/2}) \\
&= n\log\frac{X_0 \bullet S(z_0)}{\Delta} - \log\det\frac{X_0^{1/2}S(z_0)X_0^{1/2}}{\Delta} \\
&= n\log n - \log\det\frac{X_0^{1/2}S(z_0)X_0^{1/2}}{\Delta}(\text{add and subtract}\log\Delta) \\
&\leq n\log n - \frac{X_0 \bullet S(z_0)}{\Delta} + n + \frac{\|\frac{X_0^1/2S(z_0)X_0^{1/2}}{\Delta} - I\|_F^2}{2(1 - \|\frac{X_0^{1/2}S(z_0)X_0^{1/2}}{\Delta} - I\|_2)} \\
&\leq n\log n + \frac{\alpha^2}{2(1-\alpha)}
\end{aligned}
$$

The first inequality is derived from lemma 1 by replacing $\frac{X_0^{1/2}S(z_0)X_0^{1/2}}{\Delta}$ into X. From the above equations, we have

$$n \log X_0 \bullet S(z_0) - \log\det(X_0^{1/2}S(z_0)X_0^{1/2}) \leq n\log n + \frac{\alpha^2}{2(1-\alpha)}(\star\star\star)$$

Now we are applying the arithmetic geometric inequality to the eigenvalues of $X_0 \bullet S_0$, so we have:

$$\frac{\text{trace}\, X_0 S_0}{n} \geq (\det X_0.S_0)^{1/n}$$

$$\log(X_0 \bullet S_0) - \log n \geq \frac{1}{n}\log\det(X_0 S_0)$$

$$n\log n \leq n\log(X_0 \bullet S_0) - \log\det(X_0 S_0)$$

Now from the last inequality and $(\star\star\star)$ we have

$$n \log X_0 \bullet S(z_0) - \log\det(X_0^{1/2}S(z_0)X_0^{1/2}) \leq n\log(X_0 S_0) - \log\det(X_0 S_0) + \frac{\alpha^2}{2(1-\alpha)}.$$

Now if we add up two sides of $(\star\star)$ and the last inequality we get the result and this prove forth case.

■

Basically $\delta = \min\{-\frac{\alpha}{2} + \frac{\alpha^2}{2(1-\alpha)}, \beta\min(\alpha\sqrt{\frac{n}{n+\alpha^2}}, 1-\alpha), \frac{\beta^2}{2(1-\beta)}\}$. For example, if we set $\alpha = 0.43$, $\beta = 0.3$ then it is easy to show by inspection that $\delta > 0.005$. So we are guaranteed to be reduced by 0.005 in each iteration.