Semidefinite and Second Order Cone Programming Seminar
Fall 2012
Lecture 5

Instructor: Farid Alizadeh
Scribe: Ritwik Mitra

10/08/2012

1 Overview

In what follows we first provide a proof for the "General Complementary Slackness Theorem". Then we jump into algorithms. We talk about general barrier methods for convex programming and show its application in the linear programming paradigm which then segues naturally into its application in semidefinite programming problems.

2 Generalized Complementary Slackness Theorem

Theorem 1 *Let $\mathcal{K} \subset \mathbb{R}^n$ be a proper cone and $\mathcal{K}^* \subset \mathbb{R}^n$ be its dual. Consider the set $\mathcal{C}(\mathcal{K}, \mathcal{K}^*) \subset \mathbb{R}^{2n}$ defined as*

$$\mathcal{C}(\mathcal{K}, \mathcal{K}^*) = \left\{ \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} : \mathbf{x} \in \mathcal{K}, \mathbf{s} \in \mathcal{K}^*, \ \langle \mathbf{x}, \mathbf{s} \rangle = 0 \right\} \tag{1}$$

The set $\mathcal{C}(\mathcal{K}, \mathcal{K}^*)$ is an n-dimensional manifold homeomorphic to $\mathbb{R}^n$

Remark 2 *We note that two sets A and B being homeomorphic means that there is continuous mapping $f : A \to B$, such that its inverse $f^{-1} : B \to A$ is also continuous. It means that topologically this two sets are equivalent. f is called a homeomorphism.*

Proof: We make use of the following lemma:

Lemma 3 *Let $C \subset \mathbb{R}^n$ be any convex set and $\mathbf{a} \in \mathbb{R}^n$ is any point. Then there exists a unique point $\mathbf{x} \in C$ which is closest to $\mathbf{a}$. This point $\mathbf{x}$ is called a projection of $\mathbf{a}$ in C and denoted by $\Pi_C(\mathbf{a})$*

Let $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{x} = \Pi_{\mathcal{K}}(\mathbf{a})$. Let us define $\mathbf{s} = \mathbf{x} - \mathbf{a}$. By Lemma 3, this point $\mathbf{x}$ is unique and so is $\mathbf{s}$. Let us define a map f as follows

$$f : \mathbf{a} \to \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} \text{ such that } \mathbf{x} = \Pi_{\mathcal{K}}(\mathbf{a}) \quad \mathbf{s} = \mathbf{x} - \mathbf{a}$$

Clearly for all $\begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix}$ we can also find $\mathbf{a} = \mathbf{x} - \mathbf{s}$. So f^{-1} is defined as

$$f^{-1} : \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} \to \mathbf{a} \text{ such that } \mathbf{a} = \mathbf{x} - \mathbf{s} \text{ where } \mathbf{x} \in \mathcal{K}, \mathbf{s} \in \mathcal{K}^*$$

Now we note that we already have the following

1. Both f and f^{-1} are defined every where in $\mathbb{R}^n$ and $\mathcal{C}(\mathcal{K}, \mathcal{K}^*)$ respectively
2. f and f^{-1}, being defined in terms projection and subtraction, are continuous
3. $\mathbf{x} \in \mathcal{K}$ by definition of $\mathbf{x}$

We claim the following

Lemma 4

$$\mathbf{s} = \mathbf{x} - \mathbf{a} \in \mathcal{K}^* \quad \text{and} \quad \langle \mathbf{s}, \mathbf{x} \rangle = 0 \tag{2}$$

Proof:

$\boxed{\langle \mathbf{s}, \mathbf{x} \rangle = 0}$

Let $\mathbf{u}$ in $\mathcal{K}$ and define $\mathbf{u}_\alpha = \alpha \mathbf{u} + (1-\alpha)\mathbf{x}$ with $0 \leq \alpha \leq 1$ and $\zeta_\alpha = \|\mathbf{a} - \mathbf{u}_\alpha\|^2$. Then we have

$$\begin{aligned} \frac{\partial \zeta_\alpha}{\partial \alpha} &= \frac{\partial}{\partial \alpha} (\mathbf{a} - \alpha \mathbf{u} - (1-\alpha)\mathbf{x})^\top (\mathbf{a} - \alpha \mathbf{u} - (1-\alpha)\mathbf{x}) \\ &= 2(\mathbf{x} - \mathbf{u})^\top (\mathbf{a} - \mathbf{x}) \\ &= 2\mathbf{s}^\top (\mathbf{u} - \mathbf{x}) \quad \text{since } \mathbf{s} = \mathbf{x} - \mathbf{a} \end{aligned} \tag{3}$$

Now $\zeta_\alpha = \|\mathbf{a} - \mathbf{u}_\alpha\|^2$ is non decreasing in α. Thus $\frac{\partial \zeta_\alpha}{\partial \alpha} \geq 0 \ \forall \mathbf{u} \in \mathcal{K}$. First taking $\mathbf{u} = 2\mathbf{x}$, and then $\mathbf{u} = 1/2\mathbf{x}$ we have from (3), $\mathbf{s}^\top \mathbf{x} \geq 0$ and $\mathbf{s}^\top \mathbf{x} \leq 0$ respectively implying $\mathbf{s}^\top \mathbf{x} = 0$ or $\mathbf{s} \perp \mathbf{x}$.

$\boxed{\mathbf{s} \in \mathcal{K}^*}$

From (3) we have $2\mathbf{s}^\top \mathbf{u} \geq 0 \ \forall \mathbf{u} \in \mathcal{K}$ which implies $\mathbf{s} \in \mathcal{K}^*$ ■

Lemma 4 proves that the continuous and everywhere (in $\mathbb{R}^n$) defined function f is in fact a map $f : \mathbb{R}^n \to \mathcal{C}(\mathcal{K}, \mathcal{K}^*)$. Conversely we show the following for f^{-1}

Lemma 5 *Let* $\begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} \in \mathcal{C}(\mathcal{K}, \mathcal{K}^*)$ *and define* $\mathbf{a} = \mathbf{x} - \mathbf{s}$*. Then*

$$\Pi_{\mathcal{K}}(\mathbf{a}) = \mathbf{x}$$

Proof: Suppose Not. So $\exists\ \mathbf{u} \in \mathcal{K}$ such that $\|\mathbf{a} - \mathbf{u}\|_2 < \|\mathbf{a} - \mathbf{x}\|_2$. Squaring and expanding on both sides, we have

$$\mathbf{u}^\top \mathbf{u} - 2\mathbf{a}^\top \mathbf{u} - \mathbf{x}^\top \mathbf{x} + 2\mathbf{a}^\top \mathbf{x} < 0$$

Putting $\mathbf{a} = \mathbf{x} - \mathbf{s}$, we get

$$\begin{aligned}
& \mathbf{u}^\top \mathbf{u} - 2(\mathbf{x} - \mathbf{s})^\top \mathbf{u} - \mathbf{x}^\top \mathbf{x} + 2(\mathbf{x} - \mathbf{s})^\top \mathbf{x} < 0 \\
\Rightarrow\ & \mathbf{u}^\top \mathbf{u} + \mathbf{x}^\top \mathbf{x} - 2\mathbf{x}^\top \mathbf{u} + 2\mathbf{s}^\top \mathbf{u} - 2\mathbf{s}^\top \mathbf{x} < 0 \\
\Rightarrow\ & \mathbf{u}^\top \mathbf{u} + \mathbf{x}^\top \mathbf{x} - 2\mathbf{x}^\top \mathbf{u} + 2\mathbf{s}^\top \mathbf{u} < 0 \quad \text{since } \mathbf{x}^\top \mathbf{s} = 0 \\
\Rightarrow\ & \|\mathbf{u} - \mathbf{x}\|_2 + 2\mathbf{u}^\top \mathbf{s} < 0 \\
\Rightarrow\ & \mathbf{u}^\top \mathbf{s} < 0
\end{aligned}$$

But $\mathbf{u} \in \mathcal{K}$ and $\mathbf{s} \in \mathcal{K}^*$ meaning $\mathbf{u}^\top \mathbf{s} > 0$. A Contradiction! Hence Proved.

■

■

Now we provide a few examples.

3 Examples

3.1 Non Negative Orthant $\mathcal{L}$

$$\mathcal{C}(\mathcal{L}, \mathcal{L}) = \left\{ \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} :\ x_i \geq 0,\ s_i \geq 0,\ x_i s_i = 0\ \forall\ i = 1, \cdots, n \right\}$$

This is obvious since $\sum_i x_i s_i = 0$, and each $x_, s_i \geq 0$. That is, the sum of n nonnegative numbers is zero, thus each and everyone of them is zero. This is where we make a leap from a single equation ($\mathbf{x}^\top \mathbf{s} - 0$ to n equations $x_i s_i = 0$ in the presence of $\nu x \geq 0$ and $\mathbf{s} \geq 0$. This of course is the basis of the complementary slackness theorem in linear programming.

3.2 Semidefinite Cone $\mathcal{P}_{\mathbf{n} \times \mathbf{n}}$

$$\mathcal{C}(\mathcal{P}_{n \times n}, \mathcal{P}_{n \times n}) = \left\{ \begin{pmatrix} X \\ S \end{pmatrix} :\ X \succcurlyeq 0,\ S \succcurlyeq 0,\ XS = 0 \right\}$$

To see that if $X, S \succcurlyeq 0$ and $X \bullet S = 0$ implies $XS = 0$, note that

$$0 = X \bullet X = \text{Trace}(XS) = \text{Trace}\big(X^{1/2} X^{1/2} S\big) = \text{Trace}\big(X^{1/2} S X^{1/2}\big)$$

The last equality is due to the fact Trace(AB) = Trace(BA) for any pair of matrices A, B. But the matrix $X^{1/2}SX^{1/2}$ is positive semidefinite and thus all of its eigenvalues are nonnegative, and their sum, the trace, add up to zero. Thus, all eigenvalues of $X^{1/2}SX^{1/2}$ are zero implying that $X^{1/2}SX^{1/2} = X^{1/2}S^{1/2}S^{1/2}X^{1/2} = 0$. This in turn implies that $X^{1/2}S^{1/2} = S^{1/2}X^{1/2} = 0$ which implies $XS = SX = 0$. Here again the leap from a single equation Trace$(XS) = 0$ to $n(n+1)/2$ equations $XS = 0$ came from that fact that eigenvalues are all nonnegative and add up to zero, so each one of the must be zero.

In the case that at least one of X or S is positive semidefinite we can prove that $XS = 0 \Leftrightarrow \frac{XS+SX}{2} = 0$. The $\Rightarrow$ is obvious. To show $\Leftarrow$ note that if $A = XS$, since $A + A^{\top} = 0$ then A is skew symmetric, and so all its eigenvalues are either zero or purely imaginary (i.e. multiple of $i = \sqrt{-1}$.) On the other hand all eigenvalues of XS are real because in general the spectrum (multiset of eigenvalues) of AB and BA are the same. Thus, the spectrum of $A = XS = X^{1/2}X^{1/2}S$ is the same as the spectrum of $X^{1/2}SX^{1/2}$, and the latter has only real eigenvalues since it is symmetric. Thus A's eigenvalues are simultaneously multiples of i and are real, so they are all zero. Thus, $A = 0$. We have shown that

$$\mathcal{C}(\mathcal{P}_{n\times n}, \mathcal{P}_{n\times n}) = \left\{ \begin{pmatrix} X \\ S \end{pmatrix} : \ X \succcurlyeq 0, \ S \succcurlyeq 0, \ \frac{XS+SX}{2} = 0 \right\}$$

3.3 Second Order Cone $\mathcal{Q}$

Let us define

$$\mathbf{x} = \begin{pmatrix} x_0 \\ \bar{x} \end{pmatrix} \quad \mathbf{s} = \begin{pmatrix} s_0 \\ \bar{s} \end{pmatrix} \tag{4}$$

$$\mathcal{C}(\mathcal{Q}, \mathcal{Q}) = \left\{ \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} : \ \mathbf{x} \succcurlyeq_{\mathcal{Q}} 0 \ \mathbf{s} \succcurlyeq_{\mathcal{Q}} 0, \ \langle \bar{\mathbf{x}}, \bar{\mathbf{s}} \rangle = 0, \ x_i s_0 + s_i x_0 = 0 \ \forall \ i = 1, \cdots, n \right\}$$

To see this first note that if either $\mathbf{x} = \mathbf{0}$ or $\mathbf{s} = \mathbf{0}$, then the assertion $x_0 s_i + s_0 x_i = 0$ is trivially true. Thus, we assume that neither $\mathbf{x}$ not $\mathbf{s}$ is zero, which immediately implies $x_0 > 0$ and $s_0 > 0$. From $\mathbf{x}^{\top}\mathbf{s} = 0$, we have

$$\begin{aligned} x_0 s_0 \quad &= -x_1 s_1 - x_2 s_2 - \cdots - x_n s_n \\ &\le |x_1 s_1| + |x_2 s_2| + \cdots + |x_n s_n| \\ &\le \|\bar{\mathbf{x}}\| \|\bar{\mathbf{s}}\| \quad \text{(by Cauchy-Schwatrz inequality)} \\ &\le x_0 s_0 \quad \text{by the fact that } \mathbf{x}, \mathbf{s} \in \mathcal{Q} \end{aligned}$$

Since the first and the last parts are equal, we have equality throughout. In particular, we have $\|\bar{\mathbf{x}}\| = x_0$ and $\|\bar{\mathbf{s}}\| = s_0$. On the other hand, the Cauchy-Schwartz inequality is satisfied with equality only if $\bar{\mathbf{x}}$ and $\bar{\mathbf{s}}$ are proportional. Thus, we have $x_i = \alpha s_i$. Multiplying both sides by x_i and adding up we get $\sum_{i=1}^{n} x_i^2 = \|\bar{\mathbf{x}}\|^2 = \alpha \sum_{i=1}^{n} x_i s_i = -\alpha x_0 s_0$. Similarly multiplying both sides of

$x_i = \alpha s_i$ by s_i and adding up we get $\sum_{i=1}^n s_i x_i = -s_0 s_0 = \alpha \sum_{i=1}^n s_i^2 = \alpha \|\bar{s}\|^2$. Thus,

$$\alpha = -\frac{\|\bar{x}\|}{x_0 s_0} = -\frac{x_0 s_0}{\|\bar{s}\|} = -x_0$$

Plugging into $x_i = \alpha s_i = -x_0 s_i$ which proves our assertion.

4 Algorithms

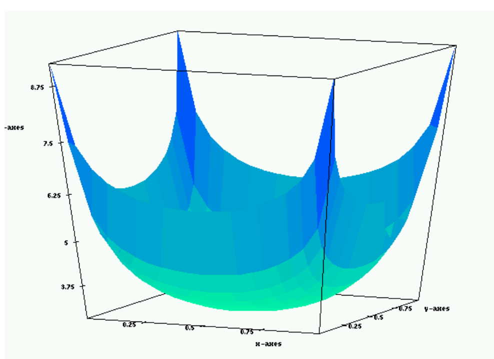
First we talk about a general approach to solving a convex programming problem called the 'Barrier Method'.

4.1 Barrier Method

Let $\mathcal{C}$ be a convex set. Let $b : \mathcal{C} \to \mathbb{R}$ be a function. We require that b has the following properties

1. $b(\mathbf{x}) \geq 0$
2. $b(\mathbf{x})$ is convex
3. Let $\{\mathbf{x}_n\}$ be a sequence in $\mathcal{C}$. Also let $\lim_{n\to\infty} \mathbf{x}_n = \bar{\mathbf{x}}$ where $\bar{\mathbf{x}} \in \partial\mathcal{C}$, the boundary on $\mathcal{C}$. Clearly $\{b(\mathbf{x}_n)\}$ is sequence in $\mathbb{R}$. We assume

$$\lim_{n\to\infty} b(\mathbf{x}_n) = \infty$$



Remark 6 *This definition implies that if we minimize $b(\mathbf{x})$, then b being convex, its local minima is the global minima and it is attained at the interior of the set $\mathcal{C}$*

Consider the following optimization problem

$$\min \ \mathbf{c}^\top \mathbf{x} \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{C} \tag{5}$$

We can rewrite (5) as the following unconstrained problem

$$\min \ \mathbf{c}^\top \mathbf{x} \ + \ \mu b(\mathbf{x}) \quad \text{where } \mu \geq 0 \tag{6}$$

Remark 7 *Note that if $\mu = \infty$, then $\mathbf{c}^\top \mathbf{x}$ does not play any role in the optimization. The minimizer is then at the minimizer of $b(\mathbf{x})$. As $\mu \downarrow 0$, the set of optimizers for each μ traverses a path culminating for $\mu = 0$ at the minimizer of the original problem (5)*

4.2 Barrier Method For LP Problems

Let $\mathbf{x}, \mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$. Consider The following primal problem

$$\begin{array}{ll} \min & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \succcurlyeq 0 \end{array} \tag{7}$$

Let $\mathbf{y} \in \mathbb{R}^m$ and $\mathbf{s} \in \mathbb{R}^n$.The dual of (7) is given by

$$\begin{array}{ll} \max & \mathbf{b}^\top \mathbf{y} \\ \text{s.t.} & A^\top \mathbf{y} + \mathbf{s} = \mathbf{c} \\ & \mathbf{s} \succcurlyeq 0 \end{array} \tag{8}$$

We consider the following logarithmic barrier function

$$b(\mathbf{x}) = -\sum_{i=1}^{n} \log(x_i)$$

We replace (7) as follows

$$\begin{array}{ll} \min & \mathbf{c}^\top \mathbf{x} - \mu \sum_{i=1}^{n} \log(x_i) \\ \text{s.t.} & A\mathbf{x} = \mathbf{b} \end{array} \tag{9}$$

The Lagrangian of the problem in (9) is given by

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathbf{c}^\top \mathbf{x} - \mu \sum_{i=1}^{n} \log(x_i) + \mathbf{y}^\top (\mathbf{b} - A\mathbf{x}) \tag{10}$$

Here $\mathbf{y}$ is the Lagrangian multiplier. KKT conditions yield that a necessary condition for $(\mathbf{x}, \mathbf{y})$ to be optimal to (10) is that

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \mathbf{y}} = 0$$

after differentiation we get

$$\nabla_{\mathbf{x}} \mathcal{L} = 0 \quad \Rightarrow \quad \mathbf{c}^\top - \mu \left(\frac{1}{x_1}, \cdots, \frac{1}{x_n} \right) - \mathbf{y}^\top A = \mathbf{0}^\top \tag{11}$$

$$\nabla_{\mathbf{y}} \mathcal{L} = 0 \quad \Rightarrow \quad (\mathbf{b} - A\mathbf{x})^\top = \mathbf{0}^\top \tag{12}$$

Remark 8 *Note that (11) is the dual feasibility while (12) is the primal feasibility problem.*

Let us define $\mathbf{s}_\mu = \mu\left(\frac{1}{x_1}, \cdots, \frac{1}{x_n}\right)$. Call $(\mathbf{x}_\mu, \mathbf{y}_\mu)$ an optimal solution for a particular choice of μ.

We note that (12) implies $A\mathbf{x}_\mu = \mathbf{b}$ and (11) implies $A^\top \mathbf{y}_\mu + \mathbf{s}_\mu = \mathbf{c}$. Also by definition we have $(\mathbf{x}_\mu)_i(\mathbf{s}_\mu)_i = \mu$ for all $i = 1, \cdots, n$. We note that this last condition looks like the complementary slackness. So, assuming we can solve these equations and letting $\mu \downarrow 0$, we can at least theoretically solve the problem.

We apply Newton's method to solve the system of equations in (11) and (12)

4.2.1 Newtons Method For Barrier Method In LP

We rewrite (11) and (12) as follows

$$\begin{aligned} &A\mathbf{x} = \mathbf{b} \\ &A^\top \mathbf{y} + \mathbf{s} = \mathbf{c} \\ &\mathbf{x} * \mathbf{s} = \mu\mathbf{1} \end{aligned} \tag{13}$$

Remark 9 *We note that* $(*)$ *denotes element-wise multiplication of two vectors. Also* $\mathbf{1}$ *is a vector of 1's, the identity element in the multiplication operation defined by* $(*)$

Let $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ be the current estimate of the minimizer. Let the new estimate be $(\mathbf{x} + \Delta_x, \mathbf{y} + \Delta_y, \mathbf{s} + \Delta_s)$. Clearly the new solution satisfies (13). So we get

$$\begin{aligned} &A(\mathbf{x} + \Delta_x) = \mathbf{b} \\ &A^\top(\mathbf{y} + \Delta_y) + (\mathbf{s} + \Delta_s) = \mathbf{c} \\ &(\mathbf{x} + \Delta_x) * (\mathbf{s} + \Delta_s) = \mu\mathbf{1} \end{aligned} \tag{14}$$

We simplify (14) by noting that $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ are constants in this setting and the variables are $(\Delta_x, \Delta_y, \Delta_s)$. We put the constant terms to the right of equality and ignore all the terms that are *quadratic* in any of the variables $(\Delta_x, \Delta_y, \Delta_s)$. Thus terms involving $\Delta_x\Delta_y$ or $\Delta_x\Delta_s$ or $\Delta_s\Delta_y$ are all ignored. We have

$$\begin{aligned} &A(\Delta_x) = \mathbf{r}_p \\ &A^\top(\Delta_y) + \Delta_s = \mathbf{r}_d \\ &S\Delta_x + X\Delta_s = \mathbf{r}_c \end{aligned} \tag{15}$$

The identities of the constants are $\mathbf{r}_p = \mathbf{b} - A\mathbf{x}$ and $\mathbf{r}_d = \mathbf{c} - A^\top \mathbf{y} - \mathbf{s}$ and finally $\mathbf{r}_c = \mu\mathbf{1} - \mathbf{x} * \mathbf{s}$. We call these quantities 'Primal Residual', 'Dual residual' and 'Complementarity Residual' respectively. Also S and X are matrices in the equations (15). They are given by

$$S = \text{diag}(\mathbf{s}) \quad X = \text{diag}(\mathbf{x})$$

The systems of equations in $(\Delta_x, \Delta_y, \Delta_s)$, as presented in (15) can be expressed as

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta_x \\ \Delta_y \\ \Delta_s \end{pmatrix} = \begin{pmatrix} \mathbf{r}_p \\ \mathbf{r}_d \\ \mathbf{r}_c \end{pmatrix} \tag{16}$$

We note here that $\Delta_s = \mathbf{r}_d - A^\top \Delta_y$. So we have from the last equation in (15)

$$S\Delta_x + X(\mathbf{r}_d - A^\top \Delta_y) = \mathbf{r}_c$$

$(\Delta_x, \Delta_y, \Delta_s)$ are not really independent of each other. So we drop one of them to make the system independent. We drop Δ_s and use the relation mentioned above. We have

$$\begin{pmatrix} A & 0 \\ S & -XA^\top \end{pmatrix} \begin{pmatrix} \Delta_x \\ \Delta_y \end{pmatrix} = \begin{pmatrix} \mathbf{r}_p \\ \mathbf{r}_c - X\mathbf{r}_d \end{pmatrix} \tag{17}$$

Clearly then we will have a solution (Δ_x, Δ_y) as

$$\begin{pmatrix} \Delta_x \\ \Delta_y \end{pmatrix} = \begin{pmatrix} A & 0 \\ S & -XA^\top \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{r}_p \\ \mathbf{r}_c - X\mathbf{r}_d \end{pmatrix} \tag{18}$$

We find the inverse using block wise Gaussian elimination method. The calculations are given in appendix. The inverse is

$$\begin{pmatrix} A & 0 \\ S & -XA^\top \end{pmatrix}^{-1} = \begin{pmatrix} -(AS^{-1}XA^\top)^{-1}AS^{-1} & (AS^{-1}XA^\top)^{-1} \\ S^{-1} - S^{-1}XA^\top(AS^{-1}XA^\top)^{-1}AS^{-1} & S^{-1}XA^\top(AS^{-1}XA^\top)^{-1} \end{pmatrix}$$

Thus we have the final solution as

$$\begin{aligned} \Delta_y &= (AS^{-1}XA^\top)^{-1} \left(\mathbf{r}_p - AS^{-1}(\mathbf{r}_c - X\mathbf{r}_d)\right) \\ \Delta_s &= \mathbf{r}_d - A^\top \Delta_y \\ \Delta_x &= S^{-1}(\mathbf{r}_c - X\Delta_s) \end{aligned}$$

Remark 10 *a) The main cost of calculation in the above equations come from computing* $(AS^{-1}XA^\top)^{-1}$. *It usually involves a Cholesky decomposition of the* $m \times m$*-dimensional matrix* $AS^{-1}XA^\top$, *which is* $O(m^3)$. *However this is just for a single* μ. *We are supposed to do the calculations for a grid of* μ *values which increases the computational complexity*

b) Note that $x_i s_i = \mu$ *which means* $x_i/s_i = \mu/s_i^2$. *Thus as* $\mu \downarrow 0$ *and we approach the true optimizer, some of the eigenvalues of* $AS^{-1}XA^\top$ *get closer to 0 and the matrix* $AS^{-1}XA^\top$ *becomes ill conditioned*

c) If D *is a diagonal matrix, then* $ADA^\top$ *has the same sparsity pattern as that of* $AA^\top$

d) The key issue in very large scale linear program is when A, *and* $AS^{-1}XA^\top$ *are very large and very sparse. The goal is to find an ordering of elimination in the Cholesky factorization so that that amount of "fill-in" (that is turning zeros to nonzeros in the process of elimination) as small as possible. This is a combinatorial problem which is NP-hard to solve optimally. However, one can use strong heuristic to find a very good order at the outset. Then, since in the subsequent iterations the zero/nonzero pattern of matrices* $(AS_k^{-1}X_kA^\top)$ *are all the same for all* k, *the same order of elimination may be used. This* *is a key reason why interior point methods for linear programming are considered competitive, if not sometimes even more efficient than the simplex method.*

Now, before we dive into barrier methods for Semi-definite problems, we first digress a little and define two matrix operators and some useful properties.

5 A Digression Into Kronecker Product & vec Operator

Understanding the algebraic properties of Kronecker products makes some presentation of interior point methods more streamlined.

5.1 The vec Operator

Let the matrix $X = (\mathbf{x}_{\cdot 1}\ \mathbf{x}_{\cdot 2} \cdots \mathbf{x}_{\cdot n}) \in \mathbb{R}^{m \times n}$ with $\mathbf{x}_{\cdot i} \in \mathbb{R}^m\ \forall\ i = 1, \cdots, n$, the i^{th} column of X. We define the vec operator on a matrix X as

$$\mathrm{vec}(X) = \begin{pmatrix} \mathbf{x}_{\cdot 1} \\ \mathbf{x}_{\cdot 2} \\ \vdots \\ \mathbf{x}_{\cdot n} \end{pmatrix} \in \mathbb{R}^{mn}$$

5.2 Kronecker Product

Given two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$, we define the Kronecker product of the two matrices as

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} \in \mathbb{R}^{mp \times nq}$$

5.3 A Few Properties

Property 1 $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$

Property 2 $\mathrm{vec}(ABC) = (C^\top \otimes A)\mathrm{vec}(B)$

From these two properties almost all useful properties of Kronecker products can be deduced. For instance we can find out eigenvalues and eigenvectors of Kronecker products.

Property 3 *Let* $A\mathbf{u} = \lambda\mathbf{u}$ *and* $B\mathbf{v} = \omega\mathbf{v}$. *Then*

$$(A \otimes B)(\mathbf{u} \otimes \mathbf{v}) = \lambda\omega(\mathbf{u} \otimes \mathbf{v})$$

Property 3 implies that if $(\lambda, \mathbf{u})$ is an eigen pair for A and $(\omega, \mathbf{v})$ one for B, then $(\lambda\omega, \mathbf{u} \otimes \mathbf{v})$ is an eigen pair for $A \otimes B$

Property 4 *Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$, and let $(\lambda_i(A), \mathbf{u}_i)$ for $i = 1, \ldots, n$, denote the eigenvalue-eigenvectors of a matrix A, and $(\omega_j, \mathbf{v}_j)$ for $j = 1, \ldots, m$. Consider the quantity $A \otimes I_m + I_n \otimes B$. Then eigenvalues of this quantity are given by the pair wise sums of all eigenvalues of A and B. That is*

$$(A \otimes I + I \otimes B)(\mathbf{u}_i \otimes \mathbf{v}_j) = (\lambda_i + \omega_j)(\mathbf{u}_i \otimes \mathbf{v}_j) \quad \forall\ i, j = 1, \cdots, m, n$$

The operation $A \otimes I + I \otimes B$ is sometimes called the *Kronecker sum* of A and B. Not that the Kronecker sum and the Kronecker product hare the same set of eigenvectors, therefore they commute.

6 Barrier Method For SDP

Let $C,\ X \in \mathbb{R}^{n \times n}$ and $A_i, i = 1, \cdots, m$ are matrices. Let $\mathbf{c} = \text{vec}(C)$ and $\mathbf{x} = \text{vec}(X)$. Also let us define

$$A = \begin{pmatrix} \text{vec}(A_1) \\ \text{vec}(A_2) \\ \vdots \\ \text{vec}(A_m) \end{pmatrix}$$

Consider the following primal semidefinite programming problem and its equivalent form

$$\begin{array}{ll} \min & C \cdot X \\ \text{s.t.} & A_i \cdot X = b_i \ i = 1, \cdots, m \\ & X \succcurlyeq_{\mathcal{P}_{n \times n}} 0 \end{array} \quad \equiv \quad \begin{array}{ll} \min & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} & A\mathbf{x} = \mathbf{b} \\ & X \succcurlyeq_{\mathcal{P}_{n \times n}} 0 \end{array} \tag{19}$$

Clearly the right hand side of the equivalence in (19) is similar to that in (7) except that the positivity is defined with respect to the SDP cone in this case. We now need to find a suitable barrier function.

Let $\lambda_i(X)$ denote the i^{th} eigenvalue of X. Define $b(X)$ as

$$\begin{aligned} b(X) &= -\sum_{i=1}^{n} \log \lambda_i(X) \\ &= -\log \prod_{i=1}^{n} \lambda_i(X) \\ &= -\log \det(X) \end{aligned}$$

We recall the following well-known lemma:

Proposition 11 $b(X) = -\log\det(X)$ *satisfies the properties required for it to be a barrier function on the set* $\mathcal{P}_{n\times n}$

(See below for proof based on derivatives.) Now we state the problem in (19) in terms of the barrier function as

$$\begin{array}{ll} \min & \mathbf{c}^\top \mathbf{x} - \mu \log\det(X) \\ \text{s.t.} & A\mathbf{x} = \mathbf{b} \end{array} \tag{20}$$

The Lagrangian of the problem is given by

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathbf{c}^\top \mathbf{x} - \mu \log\det(X) + \mathbf{y}^\top(\mathbf{b} - A\mathbf{x}) \tag{21}$$

We rewrite the Lagrangian in terms of the matrix X, instead of its vectorized version.

$$\mathcal{L}(X, \mathbf{y}) = C \cdot X - \mu \log\det(X) + \sum_{i=1}^{m} y_i(b_i - A_i \cdot X) \tag{22}$$

We note that (21) and (22) are equivalent forms. To find $\nabla_{\mathbf{y}}$ we will use (21). It is the same as in the case of an LP. It is given by

$$\nabla_{\mathbf{y}} = 0 \quad \Rightarrow \quad (\mathbf{b} - A\mathbf{x})^\top = \mathbf{0}^\top \tag{23}$$

On the other hand for ∇_X we will use the form (22). First we note the following lemma

Lemma 12 *Let* $X \in \mathcal{P}_{n\times n} \subset \mathbb{R}^{n\times n}$. *Then*

$$\nabla_X \log\det(X) = X^{-1}$$

Proof: Let X_{-ij} denote the submatrix of $X = (x_{ij})$ obtained by deleting the i^{th} row and j^{th} column. Clearly

$$\det(X) = x_{11}\det(X_{-11}) - x_{12}\det(X_{-12}) + \cdots + (-1)^{n+1}x_{1n}\det(X_{-1n})$$

Then we have

$$(\nabla_X \det(X))_{ij} = \frac{\partial \det(X)}{\partial x_{ij}} = (-1)^{i+j}\det(X_{-ij})$$

Hence we have

$$\nabla_X \log\det(X) = \frac{\big((-1)^{i+j}\det(X_{-ij})\big)}{\det(X)} = X^{-1}$$

■

Proposition 13 *Let* $\mathcal{H} = \nabla_X^2 \log\det X$ *be the Hessian matrix, that is the matrix whose* i, j *entry is the partial derivatives of* $\log\det X$ *with respect to variables* x_i *and* x_j. *Then* $H = X^{-1} \otimes X^{-1}$. *In particular, for* $X \succ 0$, $H \succ 0$, *proving that* $\log\det X$ *is convex for positive definite* X.

Now using Lemma 12 in differentiating (22) w.r.t. X, we have

$$\nabla_X = 0 \quad \Rightarrow \quad C - \mu X^{-1} - \mathbf{y}^\top A = 0^\top \tag{24}$$

Define $S = \mu X^{-1}$. Then from the KKT conditions above we have thy following system of equations

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ A^\top \mathbf{y} + S &= C \\ XS &= \mu I \end{aligned} \tag{25}$$

Remark 14 *We note here that even though both X and S are symmetric, XS may not be symmetric. Thus the final $\Delta.X$ that we will get will not be symmetric. We thus take an equivalent formulation as defined below*

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ A^\top \mathbf{y} + S &= C \\ \frac{XS + SX}{2} &= \mu I \end{aligned} \tag{26}$$

Just as we did before, we again apply Newton's Method to the problem in (26). Thus we assume the current solution as $(\mathbf{x}, \mathbf{y}, S)$ and assume the increments needed to get the new solutions as $(\Delta_x(\text{or } \Delta_X), \Delta_y, \Delta_S)$. We have the new system

$$\begin{aligned} A(\mathbf{x} + \Delta_x) &= \mathbf{b} \\ A^\top(\mathbf{y} + \Delta_y) + (S + \Delta_S) &= C \\ \frac{(X + \Delta_X)(S + \Delta_S) + (S + \Delta_S)(X + \Delta_X)}{2} &= \mu I \end{aligned} \tag{27}$$

As before taking the constant parts on the right and ignoring the quadratic parts in $(\Delta_x(\text{or } \Delta_X), \Delta_y, \Delta_S)$ we get

$$\begin{aligned} A(\Delta_x) &= \mathbf{b} - A\mathbf{x} \\ A^\top(\Delta_y) + \Delta_S &= C - A^\top \mathbf{y} - \mathbf{s} \\ \frac{(X\Delta_S) + (\Delta_S X) + (S\Delta_X) + (\Delta_X S)}{2} &= \mu I - \frac{XS + SX}{2} \end{aligned} \tag{28}$$

Let us take a closer look at the last equation in (28). We have

$$\begin{aligned} \text{vec}(X\Delta_S + \Delta_S X) &= \text{vec}(X\Delta_S I + I\Delta_S X) \\ &= (I \otimes X + X \otimes I)\text{vec}(\Delta_S) \quad \text{(by Property 2)} \end{aligned}$$

A similar calculation would yield that

$$\text{vec}(S\Delta_X + \Delta_X S) = (I \otimes S + S \otimes I)\text{vec}(\Delta_X)$$

Note that as in LP formulations, both the quantities $(I \otimes X + X \otimes I)$ and $(I \otimes S + S \otimes I)$ are linear in X and S respectively. A similar procedure as in LP would yield the estimates for $(\Delta_X, \Delta_y, \Delta_S)$. Note that the matrices X and S are different here.

Remark 15 *Note that unlike in LP, even if A and consequently $AA^\top$ (possibly, to a lesser degree) is sparse, the matrix $AS^{-1}XA^\top$ may not be sparse at all. In fact this matrix could be quite dense.*

Appendix

We prove the following

Proposition 16

$$\begin{pmatrix} A & 0 \\ S & -XA^\top \end{pmatrix}^{-1} = \begin{pmatrix} -(AS^{-1}XA^\top)^{-1}AS^{-1} & (AS^{-1}XA^\top)^{-1} \\ S^{-1} - S^{-1}XA^\top(AS^{-1}XA^\top)^{-1}AS^{-1} & S^{-1}XA^\top(AS^{-1}XA^\top)^{-1} \end{pmatrix}$$

Start with

$$\begin{pmatrix} A & 0 & | & I & 0 \\ S & -XA^\top & | & 0 & I \end{pmatrix}$$

pre-multiply $-AS^{-1}$ to 2^{nd} row and add to 1^{st} row to get

$$\begin{pmatrix} 0 & AS^{-1}XA^\top & | & I & -AS^{-1} \\ S & -XA^\top & | & 0 & I \end{pmatrix}$$

pre-multiply S^{-1} to 2^{nd} row to get

$$\begin{pmatrix} 0 & AS^{-1}XA^\top & | & I & -AS^{-1} \\ I & -S^{-1}XA^\top & | & 0 & S^{-1} \end{pmatrix}$$

Pre-multiply $S^{-1}XA^\top(AS^{-1}XA^\top)^{-1}$ to 1^{st} row and add to 2^{nd} row to get

$$\begin{pmatrix} 0 & AS^{-1}XA^\top & | & I & -AS^{-1} \\ I & 0 & | & S^{-1}XA^\top(AS^{-1}XA^\top)^{-1} & S^{-1} - S^{-1}XA^\top(AS^{-1}XA^\top)^{-1}AS^{-1} \end{pmatrix}$$

Pre-multiply $(AS^{-1}XA^\top)^{-1}$ to 1^{st} row to get

$$\begin{pmatrix} 0 & I & | & (AS^{-1}XA^\top)^{-1} & -(AS^{-1}XA^\top)^{-1}AS^{-1} \\ I & 0 & | & S^{-1}XA^\top(AS^{-1}XA^\top)^{-1} & S^{-1} - S^{-1}XA^\top(AS^{-1}XA^\top)^{-1}AS^{-1} \end{pmatrix}$$

Interchange 1^{st} column and 2^{nd} column to get

$$\begin{pmatrix} I & 0 & | & -(AS^{-1}XA^\top)^{-1}AS^{-1} & (AS^{-1}XA^\top)^{-1} \\ 0 & I & | & S^{-1} - S^{-1}XA^\top(AS^{-1}XA^\top)^{-1}AS^{-1} & S^{-1}XA^\top(AS^{-1}XA^\top)^{-1} \end{pmatrix}$$