Semidefinite Programming | **Lecture 16**
OR 6327 Spring 2012 | **March 15, 2012**
Scribe: Mark Verheggen

Recall that

$$G(X,y,S;\mu) := \begin{pmatrix} \mathcal{A}^*y + S - C \\ \mathcal{A}X - b \\ -\mu X^{-1} + S \end{pmatrix} = 0$$

defines a path (the *central path*) as long as its derivative (with respect to X, y, and S) is invertible. The derivative has the form

$$\begin{pmatrix} 0 & \mathcal{A}^* & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \mathcal{E} & 0 & \mathcal{F} \end{pmatrix}.$$

Theorem *The linear system*

$$\begin{pmatrix} 0 & \mathcal{A}^* & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \mathcal{E} & 0 & \mathcal{F} \end{pmatrix} \begin{pmatrix} U \\ v \\ W \end{pmatrix} = \begin{pmatrix} P \\ q \\ R \end{pmatrix}$$

has the unique solution

$$\begin{aligned} v &= \left(\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*\right)^{-1}\left(q - \mathcal{A}\mathcal{E}^{-1}\left(R - \mathcal{F}P\right)\right), \\ W &= P - \mathcal{A}^*v, \\ U &= \mathcal{E}^{-1}\left(R - \mathcal{F}W\right) \end{aligned}$$

as long as $\mathcal{E} : \mathbb{M}^n \mapsto \mathbb{M}^n$ is invertible, $\mathcal{E}^{-1}\mathcal{F} : \mathbb{M}^n \mapsto \mathbb{M}^n$ is positive definite (but not necessarily self-adjoint), and the A_is are linearly independent.

This result gives conditions under which the derivative is invertible.

Proof If the linear system has a solution (U, v, W) then

$$W = P - \mathcal{A}^*v.$$

Since we're assuming $\mathcal{E}$ is non-singular.

$$\begin{aligned} U &= \mathcal{E}^{-1}\left(R - \mathcal{F}W\right) \\ &= \mathcal{E}^{-1}(R - \mathcal{F}P) + \mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*v. \end{aligned}$$

Then we must have $\mathcal{A}U = q$, or

$$\left(\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*\right)v = q - \mathcal{A}\mathcal{E}^{-1}\left(R - \mathcal{F}P\right).$$

So we have to show that $\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^* \in \mathbb{R}^{m\times m}$ is non-singular. Suppose a vector y lies in its nullspace:

$$\begin{aligned} \mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^* y &= 0, \text{ so} \\ y^T\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^* y &= 0, \text{ so} \\ (\mathcal{A}^* y) \bullet \mathcal{E}^{-1}\mathcal{F}(\mathcal{A}^* y) &= 0 \end{aligned}$$

Since we are assuming that $\mathcal{E}^{-1}\mathcal{F}$ is positive definite, the last line gives $\mathcal{A}^* y = 0$, and under the assumption that the A_is are independent, this implies $y = 0$. So v is unique, implying W is unique, which implies in turn that U is unique.

Hence the solution, if it exists, is unique, and reversing the argument shows that this proposed solution indeed solves the system. □

Corollary *The set $\{X(\mu), y(\mu), S(\mu) : \mu > 0\}$ is a differentiable path.*

Proof In our case, we have $\mathcal{E}(\cdot) = \mu X^{-1}(\cdot)X^{-1}$ and $\mathcal{F} = \mathcal{I}$, so $\mathcal{E}^{-1}(\cdot) = \mu^{-1}X(\cdot)X = \mathcal{E}^{-1}\mathcal{F}$.

Also

$$\begin{aligned} & U \bullet \mathcal{E}^{-1}\mathcal{F}(U) \\ =\ & U \bullet (\mu^{-1}XUX) \\ =\ & \mu^{-1}\mathrm{tr}(UXUX) \\ =\ & \mu^{-1}\mathrm{tr}(X^{1/2}UX^{1/2}X^{1/2}UX^{1/2}) \\ =\ & \mu^{-1}\|X^{1/2}UX^{1/2}\|_F^2, \end{aligned}$$

which is positive if $U \neq 0$.

Thus we have demonstrated the assumptions that $\mathcal{E}$ is invertible and $\mathcal{E}^{-1}\mathcal{F}$ is positive definite, and the implicit function theorem completes the proof. □

Neighborhoods of the Central Path

The central path equations have two linear parts and one "mildly" nonlinear part. As we approximately follow the path, we maintain equality in the linear part and somehow measure the proximity to the central path by a measure of the residual in the last equation. We define three such measures, and their corresponding neighborhoods about the central path.

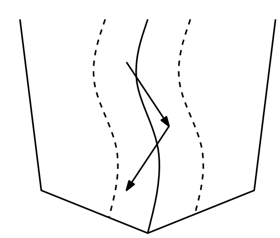

Figure 1: $\mathcal{F}^\circ(P) \times \mathcal{F}^\circ(D)$

Let $\mu = \mu(X, S) = \frac{1}{n} S \bullet X$.

$$\begin{aligned} \|X^{1/2}SX^{1/2} - \mu I\|_F^2 &= \mathrm{tr}(X^{1/2}SXSX^{1/2} - 2\mu X^{1/2}SX^{1/2} + \mu^2 I) \\ &= \mathrm{tr}((S - \mu X^{-1})X(S - \mu X^{-1})X) \\ &= \|S - \mu X^{-1}\|_X^{*\,2} \\ &= \|S + \mu F'(X)\|_X^{*\,2}. \end{aligned}$$

The right-hand side of the first line is equal to $\mathrm{tr}(S^{1/2}XSXS^{1/2} - 2\mu S^{1/2}XS^{1/2} + \mu^2 I) = \|S^{1/2}XS^{1/2} - \mu I\|_F^2$. So by a similar argument $\|S^{1/2}XS^{1/2} - \mu I\|_F = \|X^{1/2}SX^{1/2} - \mu I\|_F = \|X + \mu F'(S)\|_S^* = \|S + \mu F'(X)\|_X^*$.

If we let λ denote the eigenvalues of $X^{1/2}SX^{1/2}$ then the norm can also be expressed as $\|\lambda - \mu e\|_2$. Note that $\lambda(X^{1/2}SX^{1/2}) = \lambda(S^{1/2}XS^{1/2}) = \lambda(XS) = \lambda(SX)$ (since these matrices are all similar).

Similarly, if the operator norm is used instead of the Frobenius norm one can show that $\|X^{1/2}SX^{1/2} - \mu I\| = \|\lambda - \mu e\|_\infty$.

Finally, we will use the semi-norm $\max_j(\mu - \lambda_j)$, which is a one-sided version of the infinity norm.

We define, for $0 \le \beta < 1$, the neighborhoods

$$
\begin{aligned}
\mathcal{N}_F(\beta) &:= \left\{(X,y,S) \in \mathcal{F}^\circ(P) \times \mathcal{F}^\circ(D) : \|X^{1/2}SX^{1/2} - \mu I\|_F \le \beta\mu,\ \mu = \mu(X,S)\right\}, \\
\mathcal{N}_\infty(\beta) &:= \left\{(X,y,S) \in \mathcal{F}^\circ(P) \times \mathcal{F}^\circ(D) : \|X^{1/2}SX^{1/2} - \mu I\| \le \beta\mu,\ \mu = \mu(X,S)\right\}, \\
\mathcal{N}_{-\infty}(\beta) &:= \left\{(X,y,S) \in \mathcal{F}^\circ(P) \times \mathcal{F}^\circ(D) : \lambda(X^{1/2}SX^{1/2}) \ge (1-\beta)\mu e,\ \mu = \mu(X,S)\right\}.
\end{aligned}
$$

Note that $\mathcal{N}_F(\beta) \subseteq \mathcal{N}_\infty(\beta) \subseteq \mathcal{N}_{-\infty}(\beta)$. In fact, any $(X,y,S) \in \mathcal{F}^\circ(P) \times \mathcal{F}^\circ(D)$ lies in $\mathcal{N}_{-\infty}(\beta)$ for β sufficiently close to 1.

Framework for a primal-dual interior point algorithm for SDP

Suppose we have $(X_0, y_0, S_0) \in \mathcal{N}_?(\beta)$ for some $0 \le \beta < 1$ and choice of neighborhood $\mathcal{N}_?$.

- Given $(X_k, y_k, S_k) \in \mathcal{N}_?(\beta)$, choose a direction $(\Delta X, \Delta y, \Delta S)$ and a step size α such that

$$(X_{k+1}, y_{k+1}, S_{k+1}) = (X_k, y_k, S_k) + \alpha(\Delta X, \Delta y, \Delta S) \in \mathcal{N}_?(\beta).$$

- Repeat until $X_k \bullet S_k \le \epsilon X_0 \bullet S_0$.

Usually the directions are chosen as a Newton step for $\mathcal{A}^* y + S = C$, $\mathcal{A}X = b$ and some symmetrization of $XS - \nu I = 0$ where $\nu = \sigma_k \mu_k := \sigma_k \mu(X_k, S_k)$ with $0 \le \sigma_k \le 1$.

So we'll need the solution of

$$
\begin{pmatrix} 0 & \mathcal{A}^* & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \mathcal{E} & 0 & \mathcal{F} \end{pmatrix}
\begin{pmatrix} \Delta X \\ \Delta y \\ \Delta S \end{pmatrix}
=
\begin{pmatrix} 0 \\ 0 \\ R_{EF} \end{pmatrix}.
$$

[The 0's are because we have a feasible solution; R_{EF} is the residual for the last equation.]

Possible choices of direction

From now on we'll use (X, y, S) for the current iterate (i.e. (X_k, y_k, S_k)) and similarly μ for μ_k. $(\tilde{X}, \tilde{y}, \tilde{S})$ will indicate generic values in problems and systems of equations.

Primal Direction: Using $\tilde{S} - \mu\tilde{X}^{-1} = 0$ as the third equation, we get

$$\begin{aligned}\mathcal{E} &= \nu X^{-1} \odot X^{-1}\\ \mathcal{F} &= \mathcal{I}\\ R_{EF} &= \nu X^{-1} - S,\end{aligned}$$

where $P \odot Q$ is the operator from $\mathbb{R}^{n\times n}$ to $\mathbb{M}^n$ defined by $(P \odot Q)H := \frac{1}{2}(PHQ^T + QH^TP^T)$. ($\odot$ is meant to resemble the Kronecker product $\otimes$; cf. $(P \otimes Q)(\mathrm{vec}(H)) = \mathrm{vec}(QHP^T)$.) This is the same as using the Newton method on the primal barrier problem, and so is not really a primal-dual method.

Dual Direction: Similarly the dual direction corresponds to $\mathcal{E} = \mathcal{I}$, $\mathcal{F} = \nu S^{-1} \odot S^{-1}$, $R_{EF} = \nu S^{-1} - X$.

We want true primal-dual directions depending on both the primal and the dual iterates, so we will look further.

AHO Direction: Here is the simplest symmetrization: $\frac{1}{2}\left(\tilde{X}\tilde{S} + \tilde{S}\tilde{X}\right) = \nu I$. This defines the same path as $\tilde{X}\tilde{S} = \nu I$. Indeed, if $\tilde{X}\tilde{S} = \nu I$, then clearly the symmetrized equation above holds. Conversely, if the symmetrized equation holds, then $\tilde{X}\tilde{S} - \nu I = \frac{1}{2}\left(\tilde{X}\tilde{S} - \tilde{S}\tilde{X}\right)$. Therefore its eigenvalues are both real (eigenvalues of $\tilde{X}^{1/2}\tilde{S}\tilde{X}^{1/2} - \nu I$) and purely imaginary (eigenvalues of a skew symmetric matrix) and hence all zero, implying that $\tilde{X}\tilde{S} - \nu I = 0$. So these equations define the same central path. Then we get $\mathcal{E} = S \odot I$, $\mathcal{F} = X \odot I$, $R_{EF} = \nu I - \frac{1}{2}(XS + SX)$. $\mathcal{E}$ is invertible (at a cost) but $\mathcal{E}^{-1}\mathcal{F}$ may be singular even with $X, S \succ 0$. However, it is invertible as long as $(X, y, S) \in \mathcal{N}_F(\frac{1}{\sqrt{2}})$. This is the Alizadeh-Haeberly-Overton (AHO) direction.

HRVW-KSH Direction: The next choice comes from another motivation. We can view the equations as from $\mathbb{R}^{n\times n} \times \mathbb{R}^m \times \mathbb{M}^n$ to $\mathbb{M}^n \times \mathbb{R}^m \times \mathbb{R}^{n\times n}$ and leave $\tilde{X}\tilde{S} = \nu I$ alone [ignoring symmetry for X]. This gives

$$\begin{aligned}\widehat{\Delta X}S + X\Delta S &= \nu I - XS, \text{ or}\\ \widehat{\Delta X} + X\Delta S S^{-1} &= \nu S^{-1} - X.\end{aligned}$$

Note that ΔS is automatically symmetric from the first equation. Now symmetrize $\widehat{\Delta X}$ to get ΔX satisfying

$$\Delta X + \frac{1}{2}(X\Delta S S^{-1} + S^{-1}\Delta S X) = \nu S^{-1} - X,$$

i.e. we have $\mathcal{E} = \mathcal{I}$, $\mathcal{F} = X \odot S^{-1}$, $R_{EF} = \nu S^{-1} - X$. (Note that, since $\mathcal{A}\widehat{\Delta X} = 0$ and the A_is are symmetric, we also have $\mathcal{A}\Delta X = 0$.) This derivation is due to Helmberg, Rendl, Vanderbei, and Wolkowicz, and also, independently, Kojima, Shindoh, and Hara and so is called the HRVW-KSH direction.