Semidefinite Programming	**Lecture 24**
OR 6327 Spring 2012	**April 19, 2012**
Scribe: Alex Fix	

1 Least Squares SDP (Malick)

Consider the following modified objective function for an SDP:

$$\text{(LSSDP)} \qquad \min_X \left\{ \frac{1}{2}||X + C||_F^2 : \mathcal{A}X = b, X \succeq 0 \right\}.$$

We'll assume throughout this section that there exists some *strictly feasible* solution to this problem.

Note that this objective function is similar to a regularized version of the usual objective, since

$$\frac{1}{2}||X + \lambda C||_F^2 = \frac{1}{2}||X||_F^2 + \lambda C \bullet X + \frac{1}{2}\lambda^2 ||C||_F^2.$$

That is, if λ is large, this is a regularized version of SDP where we want the size of X to be small.

Since this is a least-squares problem, we essentially have to project $-C$ onto the intersection of two sets: the linear constraints $\mathcal{A}X = b$ and the semidefinite cone $\mathbb{M}^n_+$.

Since projection onto $\mathbb{M}^n_+$ is straightforward, we dualize the $\mathcal{A}X = b$ constraints to get

$$\begin{aligned}
\text{(LSSDP)} &\equiv \min_{X \succeq 0} \max_y \left\{ \frac{1}{2}||X + C||_F^2 - y^T(\mathcal{A}X - b) \right\} \\
&= \max_y \min_{X \succeq 0} \left\{ \frac{1}{2}||X - (\mathcal{A}^*y - C)||_F^2 - \frac{1}{2}||\mathcal{A}^*y - C||_F^2 + \frac{1}{2}||C||_F^2 + b^T y \right\}.
\end{aligned}$$

The inner minimization is now exactly a projection: we want to find the closest point in the semidefinite cone to $(\mathcal{A}^*y - C)$.

Proposition 1 *For any matrix $D \in \mathbb{M}^n$, the solution to $\min_{X \succeq 0} \frac{1}{2}||X - D||_F^2$ is $X = D_+$ where if $D = Q\Lambda Q^T$ is the eigenvalue decomposition of D and $\Lambda = \text{Diag}(\lambda)$, then $D_+ = Q\Lambda_+ Q^T$ where $\Lambda_+ = \text{Diag}(\lambda_+)$, $\lambda_+ = (\max\{0, \lambda_j\})$.*

If we write $\lambda = \lambda_+ + \lambda_-$ and set $D_- = Q\Lambda_- Q^T$ (with $\Lambda_- = \text{Diag}(\lambda_-)$), then $D = D_+ + D_-$ and

$$\frac{1}{2}||D||_F^2 = \frac{1}{2}||D_+||_F^2 + \frac{1}{2}||D_-||_F^2.$$

Proof: This follows easily by writing $Y = Q^T X Q$ and noting that $||X - D||_F = ||Y - \Lambda||_F$, so that the minimum is achieved exactly when Y is 0 off the diagonal, equal to Λ on its positive diagonal entries, and equal to 0 for its negative diagonal entries, hence $Y = \Lambda_+$. $\square$

Hence we solve for the inner minimization:

$$\begin{aligned}\theta(y) &:= \min_{X \succeq 0} \left\{ \frac{1}{2}||X - (\mathcal{A}^* y - C)||_F^2 - \frac{1}{2}||\mathcal{A}^* y - C||_F^2 + \frac{1}{2}||C||_F^2 + b^T y \right\} \\ &= -\frac{1}{2}||(A^* y - C)_+||_F^2 + b^T y + \frac{1}{2}||C||_F^2\end{aligned}$$

and the minimum is attained by $X = X(y) = (\mathcal{A}^* y - C)_+$.

Proposition 2 *θ is concave and continuously differentiable with $\nabla\theta(y) = -\mathcal{A}(\mathcal{A}^* y - C)_+ + b$.*

Proof: θ is the minimum of a family of affine functions, so it is concave.

The differentiability and form of the derivative follow from general convex analysis results using the Moreau-Yosida regularization

$$g(W) := \min_{X \in \mathbb{M}^n} \left\{ I_{\mathbb{M}^n_+}(X) + \frac{1}{2}||X - W||_F^2 \right\}$$

of the convex function $I_{\mathbb{M}^n_+}$. $\square$

Using this proposition, we can apply some smooth unconstrained optimization algorithm to maximize $\theta(y)$ and note that if the gradient is 0 (small) then $X(y)$ is (nearly) feasible.

2 A Lagrangian method for standard form SDPs

The following is due to Malick, Povh, Rendl, and Wiegele.

Begin with the usual primal and dual problems:

$$(P) \quad \begin{array}{rrcl} \min & & C \bullet X & \\ & \mathcal{A}X & = & b \\ & X & \succeq & 0, \end{array} \qquad (D) \quad \begin{array}{rrcl} \max & & b^T y & \\ & \mathcal{A}^* y + S & = & C \\ & S & \succeq & 0, \end{array}$$

and assume that both have strictly feasible solutions.

2.1 Regularizing the primal

Assume we have an iterate $W \in \mathbb{M}^n$ and consider

$$(P_\rho(W)) \quad \begin{array}{rrcl} \min_X & C \bullet X + \frac{1}{2\rho}||X - W||_F^2 & & \\ & \mathcal{A}X & = & b \\ & X & \succeq & 0. \end{array}$$

Note that (P) is equivalent to

$$\min_X \min_W \left\{ C \bullet X + \frac{1}{2\rho}||X - W||_F^2 : \mathcal{A}X = b, X \succeq 0 \right\}$$

since the inner minimization is attained by $W = X$, and then the outer minimization is just the original problem (P).

Now, switch the order of the two minimizations, and we get the equivalent problem $\min_W v(P_\rho(W))$. Let the optimal value of $P\rho(W)$ be $\Phi_\rho(W)$. Expanding terms inside the norms, we get that

$$\Phi_\rho(W) = \min_{\mathcal{A}X=b, X\succeq 0} \frac{1}{2\rho}||X + (\rho C - W)||_F^2 - \frac{\rho}{2}||C||_F^2 + C \bullet W.$$

The last two terms don't depend on X so this is a least squares SDP problem. So, we apply Malick's approach to this inner problem, to get the function

$$\theta_{\rho,W}(\hat{y}) := -\frac{1}{2\rho}||(A^*\hat{y} - (\rho C - W))_+||_F^2 + \frac{1}{\rho}b^T\hat{y} + \frac{1}{2\rho}||\rho C - W||_F^2 - \frac{\rho}{2}||C||_F^2 + C \bullet W$$

and a resulting iteration on $\hat{y}$, and as a byproduct $X(\hat{y})$, where

$$\begin{aligned} X(\hat{y}) &= (\mathcal{A}^*\hat{y} - (\rho C - W))_+ \\ &= \rho\left(\mathcal{A}^*\frac{\hat{y}}{\rho} - C + \frac{1}{\rho}W\right)_+. \end{aligned}$$

We can also set $S(\hat{y}) := (C - \mathcal{A}^*\frac{\hat{y}}{\rho} - \frac{1}{\rho}W)_+$.

Observe that for any $\hat{y}$, we have $X(\hat{y}) \succeq 0, S(\hat{y}) \succeq 0$ and $X(\hat{y}) \bullet S(\hat{y}) = 0$. So, as opposed to interior-point methods, here we have boundary points (which are also infeasible before termination).

2.2 Regularizing the dual

Consider the augmented Lagrangian for the dual:

$$L_\sigma(y, S, W) = b^T y - W \bullet (\mathcal{A}^*y + S - C) - \frac{\sigma}{2}||\mathcal{A}^*y + S - C||_F^2$$

for $\sigma > 0$.

This is the same as the usual Lagrangian for the equivalent problem

$$\max\left\{b^T y - \frac{\sigma}{2}||\mathcal{A}^*y + S - C||_F^2 : \mathcal{A}^*y + S = C, S \succeq 0\right\}.$$

Note that if (y, S) is infeasible, then we can send $L_\sigma(y, S, W)$ to $-\infty$ by choosing an appropriate W. However, if (y, S) is feasible then $L_\sigma(y, S, W)$ is equal to $b^T y$ for all W. Therefore, we have that the value of (D) is equal to

$$\max_{y,S\succeq 0} \min_W L_\sigma(y, S, W).$$

Reversing the minimization and maximization, the value of (D) is also equal to

$$\min_W \max_{y,S\succeq 0} L_\sigma(y, S, W) =: \min_W \Psi_\sigma(W).$$

Here, we have

$$\begin{aligned}\Psi_\sigma(W) &= \max_y \max_{S\succeq 0} \left\{ b^T y - W \bullet (\mathcal{A}^* y + S - C) - \frac{\sigma}{2}\left\|\mathcal{A}^* y + S - C\right\|_F^2 \right\} \\ &= \max_y \left\{ \max_{S\succeq 0} \left\{ -\frac{\sigma}{2}\left\|S + \mathcal{A}^* y - C + \frac{1}{\sigma}W\right\|_F^2 \right\} + \frac{1}{2\sigma}||W||_F^2 + b^T y \right\} \\ &= \max_y \left\{ b^T y - \frac{\sigma}{2}\left\|(\mathcal{A}^* y - C + \frac{1}{\sigma}W)_+\right\|_F^2 + \frac{1}{2\sigma}||W||_F^2 \right\}.\end{aligned}$$

However, when we regularized the primal, we had the problem

$$\begin{aligned}\Phi_\rho(W) &:= \max_{\hat{y}} -\frac{1}{2\rho}\left\|(\mathcal{A}^*\hat{y} - \rho C + W)_+\right\|_F^2 + \frac{1}{2\rho}\left\|\rho C - W\right\|_F^2 + \frac{1}{\rho}b^T\hat{y} - \frac{\rho}{2}||C||_F^2 + C \bullet W \\ &= \max_{\hat{y}} \left\{ -\frac{\rho}{2}\left\|\left(\mathcal{A}^*\frac{\hat{y}}{\rho} - C + \frac{1}{\rho}W\right)_+\right\|_F^2 + \frac{1}{2\rho}||W||_F^2 + b^T\frac{\hat{y}}{\rho} \right\}.\end{aligned}$$

Proposition 3 *If $\rho = \sigma$ then $\Phi_\rho(W) = \Psi_\sigma(W)$.*

2.3 Algorithm schema

Basic idea of the algorithm:

- Outer iteration: update $W \in \mathbb{M}^n$ to minimize $\Phi_\rho(W)$ or $\Psi_\rho(W)$.
- Inner iteration: Given W, find an approximate solution y to evaluate $\Phi_\rho(W)$.

Remarks:

(a) $\Phi_\rho(W)$ is convex and continuously differentiable. Additionally we can find its gradient by solving the inner problem. Also note that taking a gradient step of size ρ goes exactly to $X(y)$ for the optimal y.

(b) In practice, it is much more efficient to take an outer iteration (in W) after *every* inner iteration (in y), as is true for augmented Lagrangian methods in general.