Semidefinite Programming	**Lecture 25**
OR 6327 Spring 2012	**April 24, 2012**
Scribe: Marcus Lim	

Methods using Cholesky Factorizations

1 Burer and Monteiro's low-rank method

Consider again the standard problems

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

Recall that if (P) has an optimal solution, it has one of rank at most $\bar{r}$, where $\bar{r}$ is the largest integer r with $r(r+1)/2 \leq m$. So define

$$(P_r) \quad \min C \bullet X, \quad \mathcal{A}X = b,\ X \succeq 0,\ \text{rank}(X) \leq r,$$

for $r > 0$. $(P_r) \equiv (P)$ (same optimal value, subset of same optimal solutions) if $r \geq \bar{r}$. Finally, consider

$$(NLP_r) \quad \min_R C \bullet RR^T, \quad A_i \bullet RR^T = b_i, \text{for } i = 1, \ldots, m,\ R \in \mathbb{R}^{n \times r}.$$

This is a *nonlinear*, *nonconvex* problem in nr variables. But it only has equality constraints, and the positive semidefiniteness constraint has been eliminated. Note: $(P_r) \equiv (NLP_r)$ (same optimal value, corresponding optimal solutions) for any r.

1.1 Optimality conditions for (NLP_r)

Note:

$$\begin{aligned} C \bullet (R+E)(R+E)^T &= C \bullet (RR^T + RE^T + ER^T + EE^T) \\ &= C \bullet RR^T + \text{trace}\,(CRE^T) + \text{trace}\,(CER^T) + \text{trace}\,(CEE^T) \\ &= C \bullet RR^T + 2\,\text{trace}\,(E^T CR) + \text{trace}\,(CEE^T). \end{aligned}$$

So $\nabla(C \bullet RR^T) = 2CR$. Similarly,

$$\nabla(A_i \bullet RR^T) = 2A_i R, \qquad \nabla^2(C \bullet RR^T)[E, E] = 2C \bullet (EE^T).$$

Theorem 1 *Suppose $\bar{R}$ is a local minimizer of (NLP_r), and assume the constraints gradients $2A_i\bar{R}_i$, $i = 1, ..., m$, are linearly independent. Then there exist a unique $\bar{y} \in \mathbb{R}^m$ and $\bar{S} = C - \mathcal{A}^* y \in \mathbb{M}^n$ such that $\bar{S}\bar{R} = 0$.*

Moreover, if $D \in \mathbb{R}^{n \times r}$ satisfies $(A_i R) \bullet D = 0$ for all i, then $\bar{S} \bullet DD^T \geq 0$. □

Corollary 1 *Suppose $\bar{R}$ with $\bar{y}$ and $\bar{S}$ satisfies the conditions of the theorem. Then if $\bar{S} \succeq 0$, $\bar{R}$ is an optimal solution to (NLP_r) and $\bar{X} = \bar{R}\bar{R}^T$ is an optimal solution to (P).*

Proof: $\bar{X}$ is feasible in (P), $(\bar{y}, \bar{S})$ is feasible in (D), and $\bar{S} \bullet \bar{X} = \text{trace}\,(\bar{S}\bar{X}) = \text{trace}\,(\bar{S}\bar{R}\bar{R}^T) = 0$. Hence, there is no duality gap, and $\bar{X}$ is optimal, and thus R is an optimal solution to (NLP_r). □

Corollary 2 *Suppose $\bar{R}$ satisfies the conditions of the theorem. Let $\hat{R} := [\bar{R}, 0] \in \mathbb{R}^{n\times(r+1)}$. If $\hat{R}$ is a local minimizer for (NLP_{r+1}) then $\bar{X} = \bar{R}\bar{R}^T$ is an optimal solution to (P).*

Proof: By the theorem, there is a unique $\bar{y}$ and associated $\bar{S}$ with $\bar{S}\bar{R} = 0$. Also, since the $A_i\hat{R}$'s must also be linearly independent, there is a unique $\hat{y}$ and associated $\hat{S}$ with $\hat{S}\hat{R} = 0$. But $\bar{S}\hat{R} = [\bar{S}\bar{R}, 0]$, so $\hat{y} = \bar{y}, \hat{S} = \bar{S}$.

Moreover, $\bar{S} \bullet DD^T \geq 0$ for all D with $(A_i\hat{R}) \bullet D = 0$ for all i, or equivalently

$$[A_i\bar{R}, 0] \bullet D = 0.$$

In particular, if $D = [\mathbf{0}, d]$, it satisfies these conditions, and since $DD^T = dd^T$ we get $d^T\bar{S}d \geq 0$ for all $d \in \mathbb{R}^n$. Now apply Corollary 1 and we are done. □

1.2 Algorithmic implication

Suppose we (approximately) solve (NLP_r) to get $\bar{R}$. Find the associated $\bar{y}$ and $\bar{S}$. If $\bar{S} \succeq 0$ then $\bar{R}\bar{R}^T$ is an (approximately) optimal solution to (P). If not, find $d \in \mathbb{R}^n$ with $d^T\bar{S}d < 0$, and start the search for an (approximately) optimal solution to (NLP_{r+1}) at $\hat{R} = [\bar{R}, \epsilon d]$ for small ϵ.

- We can restrict R to lower triangular matrices to avoid the redundancy caused by R feasible/optimal iff RQ feasible/optimal for orthogonal $Q \in \mathbb{R}^{r\times r}$.
- We must be caseful to avoid being trapped in a subspace. The iterate is R and all gradients are MR for some $M \in \mathbb{M}^n$. We must avoid zero columns, and more generally, matrices of less than full column rank, since most first-order methods will then get trapped in a subspace.

Burer and Monteiro use an augmented Lagrangian method for "solving" (NLP_r) for each r. This works quite well for random, MAXCUT, and Lovász theta problems and compares well with the spectral bundle method, DSDP, and inexact versions of interior-point methods.

For MAXCUT problems, we can instead solve

$$\min C \bullet \tilde{R}\tilde{R}^T, \quad R \in \mathbb{R}^{n\times r},$$

where if $R = \begin{bmatrix} w_1 \\ \dots \\ w_n \end{bmatrix}$, then $\tilde{R} = \begin{bmatrix} w_1/\|w_1\| \\ \dots \\ w_n/\|w_n\| \end{bmatrix}$, since we want $\text{diag}\,(RR^T) = \begin{bmatrix} \|w_1\|^2 \\ \dots \\ \|w_n\|^2 \end{bmatrix} = e$. This results in an unconstrained minimization.

2 Nonlinear transformation of Burer, Monteiro and Y. Zhang

This applies to:

$$(P) \quad \max C \bullet X, \quad \mathcal{A}X = b,\ \text{diag}\,(X) \geq d,\ X \succeq 0,$$

and its dual

$$(D) \quad \min b^T y + d^T z, \quad \mathcal{A}^* y + \text{Diag}\,(z) - S = C,\ S \succeq 0,\ z \leq 0.$$

Note that the "standard" maximization primal problem and its dual can be written in this way by choosing $d < 0$. We'll assume both (P) and (D) have strictly feasible solutions (note that this follows if they arise from "standard" problems with strictly feasible solutions as above.)