| | |
|---|---|
| **Semidefinite Programming** | **Lecture 26** |
| **OR 6327 Spring 2012** | **April 26, 2012** |
| **Scribe: Zhe Yu** | |

We will finish discussing the Burer, Monteiro and Y. Zhang method.

Burer Monteiro and Y. Zhang's nonlinear transformation method:

$$(P) \qquad \begin{array}{lrcl} \max & C \bullet X & & \\ & \operatorname{diag}(X) & \geq & d \\ & \mathcal{A}X & = & b \\ & X & \succeq & 0, \end{array}$$

$$(D) \qquad \begin{array}{lrcl} \max & d^T z + b^T y & & \\ & \operatorname{Diag}(z) + \mathcal{A}^* y - S & = & C \\ & z & \leq & 0 \\ & S & \succeq & 0. \end{array}$$

Assume that both $(P)$ and $(D)$ have strictly feasible solutions and that the $A_i$'s are linear independent.

For every $y$, for every $z$ large enough, $\operatorname{Diag}(z) + \mathcal{A}^* y - C$ is positive definite. So $S \succ 0$ and it has a Cholesky factorization.

Burer, Monteiro, and Zhang (BMZ) turn this around: for every $y$ and diagonal of the Cholesky factor, they obtain $z$ and the rest of the factor.

**Theorem 1** *a) For every $y \in \mathbb{R}^m$ and $w \in \mathbb{R}^n_{++}$, there is a unique $z = z(w,y) \in \mathbb{R}^n$ and strictly lower triangular $\hat{L} = \hat{L}(w,y) \in \mathbb{R}^{n \times n}$ such that $\operatorname{Diag}(z) + \mathcal{A}^* y - C = (\operatorname{Diag}(w) + \hat{L})(\operatorname{Diag}(w) + \hat{L})^T$.*
*b) The functions $z(w,y)$ and $\hat{L}(w,y)$ are infinite differentiable on $\mathbb{R}^n_{++} \times \mathbb{R}^m$.*
*c) There is a bijection between $(w,y) \in \mathbb{R}^n_{++} \times \mathbb{R}^m$ and $(y,z,S) \in \{\mathbb{R}^m \times \mathbb{R}^n \times M^m_{++} : \operatorname{Diag}(z) + \mathcal{A}^* y - S = C\}$, given by $z = z(w,y)$, $S = LL^T$, where $L := L(w,y) := Diag(w(z,y)) + \hat{L}(w,y)$.*

(Note that the sign conditions on $z$ are not present here.)

Sketch of the proof of (a) (Constructive):
Consider $n = 3$, and let $T := \mathcal{A}^* y - C$. We want

$$\operatorname{Diag}(z) + T = (\operatorname{Diag}(w) + \hat{L})(\operatorname{Diag}(w) + \hat{L})^T = \begin{pmatrix} w_1 & 0 & 0 \\ l_{21} & w_2 & 0 \\ l_{31} & l_{32} & w_3 \end{pmatrix} \begin{pmatrix} w_1 & 0 & 0 \\ l_{21} & w_2 & 0 \\ l_{31} & l_{32} & w_3 \end{pmatrix}^T.$$

So,

$$\begin{gathered} z_1 + t_{11} = w_1^2 \to z_1 \\ t_{21} = (\operatorname{Diag}(z) + T)_{21} = l_{21} w_1 \to l_{21} \\ t_{31} = (\operatorname{Diag}(z) + T)_{31} = l_{31} w_1 \to l_{31} \\ z_2 + t_{22} = l_{21}^2 + w_2^2 \to z_2 \\ t_{32} = l_{31} l_{21} + l_{32} w_2 \to l_{32} \ldots \end{gathered}$$

This proves part (a) of the theorem. $\square$

Problem: $z(w,y)$ and $\hat{L}(w,y)$ do not extend smoothly to the boundary of $\mathbb{R}^n_{++} \times \mathbb{R}^m$, and the bijection similarly cannot be extended. Nevertheless, by the bijection,

$$\begin{array}{lccc} \inf_{y,z,S} & d^T z + b^T y & & \\ & \mathrm{Diag}\,(z) + \mathcal{A}^* y - S & = & C \\ & z & < & 0 \\ & S & \succ & 0, \end{array}$$

is equivalent to

$$\begin{array}{lccc} \inf_{w,y} & d^T z(w,y) + b^T y & & \\ & z(w,y) & < & 0 \\ & w & > & 0. \end{array}$$

Note that, by our assumptions, both these problems have feasible solutions, and the first approximates $(D)$.

Idea: "Solve"

$$(\bar{D}) \quad \begin{array}{lccc} \inf_{w,y} & d^T z + b^T y & & \\ & z(w,y) & \leq & 0 \\ & w & \geq & 0 \end{array}$$

This is a nonconvex NLP in $n+m$ variables.

Suppose we solve $(\bar{D})$ by a log barrier method, using subproblems of the form

$$(B\bar{D}_\mu) \quad \min_{w,y} \quad d^T z(w,y) + b^T y - \mu \sum \ln(-z_j(w,y)) - 2\mu \sum \ln w_j$$

for $\mu > 0$.

Note that, if $S(w,y) = L(w,y)L(w,y)^T$ then this is

$$\min \quad d^T z(w,y) + b^T y - \mu \sum \ln(-z_j(w,y)) - \mu \ln \det S(w,y)$$

since

$$\begin{aligned} \ln \det S(w,y) &= \ln \det(\mathrm{Diag}\,(w) + \hat{L})(\mathrm{Diag}\,(w) + \hat{L})^T \\ &= 2 \ln \det(\mathrm{Diag}\,(w) + \hat{L}) \\ &= 2 \sum \ln w_j. \end{aligned}$$

Moreover, by the bijection, this is equivalent to

$$(BD_\mu) \quad \begin{array}{lccc} \min_{y,z,S} & d^T z + b^T y - \mu \sum(-z_j) - \mu \ln \det S & & \\ & \mathrm{Diag}\,(z) + \mathcal{A}^* y - S & = & C \\ & z & < & 0 \\ & S & \succ & 0. \end{array}$$

BMZ prove

**Theorem 2** *For any $\mu > 0$, there is a unique minimizer, which is also the unique stationary point, for $(B\bar{D}_\mu)$. Moreover, this corresponds by our bijection to the unique minimizer of $(BD_\mu)$.*

Hence, BMZ apply a unconstrained minimization algorithm (they use limited memory BFGS) to $(B\bar{D}\mu)$ to get an approximate stationary point, then update $\mu$ to $\sigma\mu$, $0 < \sigma < 1$ fixed, and continue. They prove global convergence.

In computing derivatives of $z$ and the Lagrangian for $(\bar{D})$, and in the analysis, a key quantity turns out to be a primal estimate $X$.

**Theorem 3** *For any $v \in \mathbb{R}^n$ and any lower triangular $L \in \mathbb{R}^{n\times n}$ with positive diagonal, there is a unique $X \in \mathbb{M}^n$ with* $\operatorname{diag}(X) = v$ *and $XL$ upper triangular.*

Sketch of the proof (for $n = 3$). Compute $X$ from bottom to top and right to left.

$$
\begin{gathered}
x_{33} = v_3 \\
0 = (XL)_{32} = x_{33}l_{32} + x_{32}l_{22} \to x_{32} \\
0 = (XL)_{31} = x_{33}l_{31} + x_{32}l_{21} + x_{31}l_{11} \to x_{31} \\
x_{22} = v_2 \\
0 = (XL)_{21} = x_{21}l_{11} + x_{22}l_{21} + x_{23}l_{31} \to x_{21} \\
\textit{etc.}
\end{gathered}
$$

(Note that, in the last equation, we know $x_{23}$ since $X$ is symmetric.) $\square$

Note: $XL$ upper triangular implies $L^TXL$ is upper triangular and symmetric so diagonal; say it's equal to $\Lambda$.

$$L^TXL = \Lambda \Rightarrow X = L^{-T}\Lambda L^{-1}.$$

If $\Lambda = \mu I$,

$$X = \mu L^{-T}L^{-1} = \mu S^{-1},$$

as on the central path.

For further information, see the original paper.

Semidefinite programming solvers can be found at NEOS at

`http://www.neos-server.org/neos/solvers/index.html`.

Computational results comparing interior-point codes with a code for the spectral bundle method, a code for Burer and Monteiro's low-rank method, and a code for the Burer-Monteiro-Zhang method can be found at

`http://plato.asu.edu/dimacs/paper93.html`.