Convergence Conditions for Combining Multiple Higher Order Corrections in SDP

Zhifeng Li
FannieMae

Joint with Sanjay Mehrotra
Dept. of IE/MS, Northwestern University

INFORMS 2006
Pittsburgh, PA

Outline

- Problem formulation.
- Interior-point method for SDP.
- Symmetrization and computational issue.
- Convergence conditions for combining multiple higher order corrections in SDP.
- Future Research.

Problem Formulation

We consider the following SDP:

$$
\begin{array}{lll}
\text{(Primal)} & \min & C \bullet X \\
& s.t. & A_i \bullet X = b_i, \; i = 1, \dots, m, \\
& & X \succeq 0,
\end{array}
$$

where $C \in \mathcal{S}^n$, $A_i \in \mathcal{S}^n$ and $\mathcal{S}^n$ denotes the set of $n \times n$ symmetric matrices; $b_i \in \mathbb{R}$, $i = 1, \dots, m$; $C \bullet X = tr(CX)$ and $X \succeq (\succ) 0$ means that X is a positive semidefinite (definite) matrix.

Optimality Conditions

The dual problem is

$$
\begin{aligned}
\text{(Dual)}\quad \max \quad & b^T y \\
s.t. \quad & A^T y + S = C, \\
& S \succeq 0,
\end{aligned}
$$

where $A^T = [\text{vec}A_1, ..., \text{vec}A_m]$. Under some conditions, the optimality conditions for the SDP problem are

$$
\begin{aligned}
A^T y + \text{vec}S - \text{vec}C &= 0, \\
A(\text{vec}X) - b &= 0, \\
XS &= 0, \\
X \succeq 0, S &\succeq 0.
\end{aligned}
$$

Problem: ΔX is not symmetric if linearized directly.

Symmetrization Schemes

The symmetrization scheme of Zhang-Monteiro bases on the operator $\mathcal{H}_P$: $\mathbb{R}^{n\times n} \mapsto \mathcal{S}^n$ defined as

$$\mathcal{H}_P(A) = \frac{1}{2}[PAP^{-1} + (PAP^{-1})^T],$$

for a given nonsingular matrix $P \in \mathbb{R}^{n\times n}$. For instance the AHO direction for $P = I$, the HKM direction for $S^{1/2}$, the dual HKM direction for $X^{-1/2}$, and the NT direction for $W^{-1/2}$.

Alternative symmetrization schemes have been proposed. Different schemes generated different search directions. There are more than 20 search directions [Todd98].

Interior-Point Method for SDP

SDP can be solved using interior-point method due to Nesterov and Nemirovski [94]. The perturbed optimality conditions are given as follows:

$$
\begin{aligned}
A^T y + \text{vec}S - \text{vec}C = 0, \\
A(\text{vec}X) - b = 0, \\
\mathcal{H}(XS) = \sigma\mu I, \\
X \succeq 0, S \succeq 0,
\end{aligned}
$$

where σ is a parameter, $\mu = \frac{X \bullet S}{n}$.

Linearized Optimality Conditions

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ E & 0 & F \end{bmatrix} \begin{bmatrix} \text{vec}\Delta X \\ \Delta y \\ \text{vec}\Delta S \end{bmatrix} = \begin{bmatrix} \text{vec}R_d \\ r_p \\ \text{vec}R_c \end{bmatrix}$$

where

$$\begin{aligned} \text{vec}R_d &= -(A^T y + \text{vec}S - \text{vec}C), \\ r_p &= -(A\text{vec}X - b), \\ E &= 2S \otimes S, \\ F &= SX \otimes I + I \otimes SX, \end{aligned}$$

where $\otimes$ is the Kronecker-product.

Computational Challenge

$$[A(E^{-1}F)A^T]\Delta y = r_p + A[E^{-1}(F\text{vec}R_d - \text{vec}R_c)],$$

$$\Delta S = R_d - \sum_{i=1}^{m} \Delta y_i A_i,$$

$$\Delta X = \sigma\mu S^{-1} - X - [X(\Delta X)S^{-1} + (X(\Delta S)S^{-1})^T]/2.$$

Note that $E^{-1}F = \frac{X \otimes S^{-1} + S^{-1} \otimes X}{2}$ is s.p.d. Therefore, $A(E^{-1}F)A^T$ is s.p.d. whenever A has full rank. Direct methods that using the Cholesky factorization of $A(E^{-1}F)A^T$ can be numerical intensive even for sparse A. Iterative methods and nonlinear programming approach have been used for large SDP.

Interior-Point Methods with Higher Order Corrections

- Mehrotra’s [92] predictor-corrector methods for LP.
- Gondzio’s [96] centrality corrections.
- Jarre-Wech [99] corrections.
- Alizadeh et al. [99] extended Mehrotra and Gondzio’s corrections to SDP.
- Mehrotra-Li’s [05] algorithm combines direct method and iterative method for solving LPs.

Main Ideas

- Reuse Cholesky factors as much as possible.
- Combine direct method and iterative method to gain efficiency and stability.
- Combine different higher order correction directions in SDP.

$$\Delta X(\rho^p) = \rho^p_{aff} \Delta X^{aff} + \rho^p_1 \Delta X^1 + \ldots + \rho^p_j \Delta X^j + \rho^p_c \Delta X^c,$$

$$\Delta y(\rho^d) = \rho^d_{aff} \Delta y^{aff} + \rho^d_1 \Delta y^1 + \ldots + \rho^d_j \Delta y^j + \rho^d_c \Delta y^c,$$

$$\Delta S(\rho^d) = \rho^d_{aff} \Delta S^{aff} + \rho^d_1 \Delta S^1 + \ldots + \rho^d_j \Delta S^j + \rho^d_c \Delta S^c,$$

where $\rho^p = (\rho^p_{aff}, \rho^p_1, \ldots, \rho^p_j, \rho^p_c)$ and $\rho^d = (\rho^d_{aff}, \rho^d_1, \ldots, \rho^d_j, \rho^d_c)$.

A Small SDP

$$\max t$$

$$X + \Delta X(\rho^d) \succeq 0,\ S + \Delta S(\rho^d) \succeq 0, \quad (1)$$

$$\|D^{-T}\mathrm{vec}\Delta X(\rho^p)\| \le \delta c_1,\ \|D\mathrm{vec}\Delta S(\rho^d)\| \le \delta c_1, \quad (2)$$

$$\|r_p(\rho)\| \le \delta\|r_p\|,\ \|r_p - r_p(\rho)\| \le (1-\delta)\|r_p\|, \quad (3)$$

$$\|\mathrm{vec}R_d(\rho)\| \le \delta\|\mathrm{vec}R_d\|,\ \|vecR_d - \mathrm{vec}R_d(\rho)\| \le (1-\delta)\|\mathrm{vec}R_d\|, \quad (4)$$

$$X \bullet \Delta S(\rho^d) + \Delta X(\rho^p) \bullet S \le \delta(\beta_2 - 1)X \bullet S), \quad (5)$$

$$\mathcal{H}(X\Delta S(\rho^d) + \Delta X(\rho^p)S) \succeq \delta(\beta_1\mu I - \mathcal{H}(XS)), \quad (6)$$

$$C \bullet (X + \Delta X(\rho^p)) - b^T(y + \Delta y(\rho^d)) \le (1-t)\bar{\mu} \quad (7)$$

$$\rho_i^p, \rho_i^d, 0 \le t \le \delta \le 1,$$

IPHOC Algorithm

Given an initial point (X^0, y^0, S^0)

For $k = 0, 1, 2, \ldots$

Step 1) Set $(X, y, S) = (X^k, y^k, S^k), \tilde{\mu} = \tilde{\mu}_k$ and let $\mu = \frac{X \bullet S}{n}$, $\bar{\mu} = \max\{C \bullet X - b^T y, X \bullet S, \tilde{\mu}\}, \tilde{\mu} = \tilde{\mu}_k$. Check termination criteria.

Step 2) Compute $(\Delta X^{aff}, \Delta y^{aff}, \Delta S^{aff})$, $(\Delta X^c, \Delta y^c, \Delta S^c)$, and generate j correction directions $(\Delta X^0, \Delta y^0, \Delta S^0), \ldots, (\Delta X^j, \Delta y^j, \Delta S^j)$. Solve the small SDP to obtain the search direction $(\Delta X(\rho^x), \Delta y(\rho^s), \Delta S(\rho^s))$;

Step 3) Set $\Delta X = \Delta X(\rho^x)$, $\Delta y = \Delta y(\rho^s)$, $\Delta S = \Delta S(\rho^s)$, and compute the step length α along this direction.

Step 4) Update $k \leftarrow k+1$ and the new iterate

$$(X^{k+1}, y^{k+1}, S^{k+1}) = (X^k, y^k, S^k) + \alpha(\Delta X, \Delta y, \Delta S),$$

End (For)

Convergence Conditions

We use the following wide central neighborhood.

$$\begin{aligned} \Lambda(\mathcal{H}(XS)) &\geq \gamma \frac{X \bullet S}{n}, \\ X \bullet S \geq \gamma_p \|r_p\| &\quad \text{or} \quad \|r_p\| \leq \epsilon_p, \\ X \bullet S \geq \gamma_d \|\text{vec} R_d\| &\quad \text{or} \quad \|\text{vec} R_d\| \leq \epsilon_d, \end{aligned}$$

where Λ is the spectrum of a matrix, given $\widehat{\gamma} \in (0,1)$, γ_1, γ_p, and γ_d are some parameters. We also ensure duality gap reduction.

$$X_{k+1} \bullet S_{k+1} \leq (1 - \alpha\delta(1 - \beta_2)) X_k \bullet S_k.$$

Lower Bound for the Step Length

The step length in Step 3 of the IPHOC algorithm has the following lower bound.

$$\alpha > \bar{\alpha} = \min\left\{1, \frac{(\beta_1 - \gamma\beta_2)\gamma}{(1+\gamma)c_0 n^2 \sqrt{\kappa}\delta}, \frac{\beta_1}{\sqrt{\kappa} c_0 n^2 \delta}, \frac{\beta_2 - \beta_1}{\sqrt{\kappa} c_0 n^2 \delta}\right\},$$

where $\kappa = \frac{\lambda_1}{\lambda_n}$ is the condition number of $S^{1/2} X S^{1/2}$, and

$$1 \leq \kappa \leq \frac{n}{\gamma}.$$

Convergence Theorem

The IPHOC algorithm terminates after $O(n^{2.5}\ln(\frac{1}{\epsilon}))$ iterations. It either generates (X^k, S^k) such that $\|X\| > w^*$ or $\|S\| > w^*$ or identifies an approximate solution to the primal and dual SDP problems such that $X \bullet S \leq \epsilon X_0 \bullet S_0$.

Proof: Use the analysis in Zhang [98] and Mehrotra-Li [05].

Summary & Future Research

- We gave convergence conditions for interior-point method with higher order corrections for SDP.
- Practical implementation considerations:
  - The small SDP does not need to be solved exactly.
  - Special techniques can be utilized to solve it fast.