The 7th International Symposium on Operations Research and Its Applications (ISORA'08)
Lijiang, China, October 31–Novemver 3, 2008

A Primal-Dual Interior-Point Filter Method for Nonlinear Semidefinite Programming

Zhong-Yi Liu[1,*] Wen-Yu Sun[2,†]

[1]College of Science, Hohai University, Nanjing 210098, China
[2]School of Mathematics and Computer Sciences,
Nanjing Normal University, Nanjing 210097, China

Abstract This paper proposes a primal-dual interior-point filter method for nonlinear semidefinite programming, which is an extension of the work of Ulbrich *et al.* (Math. Prog., 100(2):379–410, 2004). We use a mixed norm to tackle with trust region constraints and global convergence to first-order critical points can be proved.

Keywords Nonlinear semidefinite programming; Interior-point method; Primal-dual; Filter; Global convergence

1 Introduction

For the concepts and frame of filter methods, we refer to Fletcher *et al.* [1], Fletcher and Leyffer [2], Fletcher *et al.* [3], Ulbrich *et al.* [5]. We consider the following nonlinear semidefinite programming (SDP) problem:

$$\begin{aligned}&\min\ \ f(x), \quad x\in\mathbb{R}^n,\\ &s.t.\ \ h(x)=0,\ \ X\equiv\mathscr{A}x-B\succeq 0,\end{aligned} \tag{1}$$

where the functions $f:\ \mathbb{R}^n\to\mathbb{R}$ and $h:\ \mathbb{R}^n\to\mathbb{R}^m$ are sufficiently smooth. Here $\mathscr{A}$ is a linear operator defined by $\mathscr{A}x=\sum_{i=1}^{n}x_iA_i$ for $x\in\mathbb{R}^n$, and $A_i\in\mathbb{S}^p,\ i=1,\ldots,n$, and $B\in\mathbb{S}^p$ are given matrices, where $\mathbb{S}^p$ denotes the set of pth order real symmetric matrices. By $X\succeq 0$ and $X\succ 0$, we mean that the matrix X is positive semidefinite and positive definite, respectively.

Throughout this paper, we define the inner product $\langle X,Z\rangle$ by $\langle X,Z\rangle=Tr(XZ)$ for two matrices X and Z in $\mathbb{S}^p$, where $Tr(M)$ denotes the trace of the matrix M. We also define an adjoint operator $\mathscr{A}^*$ of $\mathscr{A}$ such that $\mathscr{A}^*Z$ is an n dimensional vector whose ith element is $Tr(A_iZ)$. Then we have

$$\langle\mathscr{A}x,Z\rangle=x^T(\mathscr{A}^*Z)=(\mathscr{A}^*Z)^Tx,$$

where the superscript T denotes the transpose of a vector or a matrix.

*Corresponding author. Email:zhyi@hhu.edu.cn
†Email:wysun@njnu.edu.cn

2 Interior-point framework

We return to the nonlinear semidefinite programming problem posed in (1).

2.1 Step computation

Let the Lagrangian function of problem (1) be defined by

$$L(x,y,Z) = f(x) - y^T h(x) - \langle X, Z \rangle,$$

where $y \in \mathbb{R}^m$ and $Z \in \mathbb{S}^p$ are the Lagrange multiplier vector and matrix which correspond to the equality and positive semidefiniteness constraints, respectively. Then Karush-Kuhn-Tucker (KKT) conditions for optimality of problem (1) are given by

$$\begin{pmatrix} \nabla_x L(w) \\ h(x) \\ X \circ Z \end{pmatrix} = 0$$

and

$$X \succeq 0, \quad Z \succeq 0.$$

Here $\nabla_x L(x,y,Z)$ is given by

$$\nabla_x L(x,y,Z) = \nabla f(x) - \nabla h(x) y - \mathscr{A}^* Z,$$

and the multiplication $X \circ Z$ is defined by

$$X \circ Z = \frac{XZ + ZX}{2}.$$

It is known that $X \circ Z = 0$ is equivalent to the relation $XZ = ZX = 0$.

We call (x,y,Z) satisfying $X \succ 0$ and $Z \succ 0$ the interior point. To construct an interior-point algorithm, we introduce a positive parameter $\hat{\mu}$, and we replace the complementarity condition $X \circ Z = 0$ by $X \circ Z = \hat{\mu} I$, where I denotes the identity matrix and $\hat{\mu} > 0$. Then we try to find a point that satisfies the following system of equations:

$$\begin{pmatrix} \nabla_x L(x,y,Z) \\ h(x) \\ X \circ Z - \hat{\mu} I \end{pmatrix} = 0 \tag{2}$$

and

$$X \succ 0, \quad Z \succ 0.$$

Throughout we will work with $\hat{\mu} = \sigma\mu$, where $\sigma \in (0,1)$ is a centering parameter and

$$\mu = \frac{Tr(XZ)}{n}.$$

To abbreviate notation we set

$$w = (x,y,Z) \quad \text{and} \quad \Delta w = (\Delta x, \Delta y, \Delta Z).$$

We apply a Newton-like method to the system of equations (2). Let the Newton directions for the primal-dual variables by $\Delta x \in \mathbb{R}^n$ and $\Delta Z \in \mathbb{S}^p$, respectively. We define $\Delta X = \sum_{i=1}^{n} \Delta x_i A_i$ and note that $\Delta X \in \mathbb{S}^p$.

Since $(X+\Delta X)\circ(Z+\Delta Z) = \sigma\mu I$ can be written as

$$(X+\Delta X)(Z+\Delta Z)+(Z+\Delta Z)(X+\Delta X) = 2\sigma\mu I,$$

neglecting the nonlinear parts $\Delta X\Delta Z$ and $\Delta Z\Delta X$ implies the Newton equation for (2)

$$\begin{aligned} G\Delta x - \nabla h(x)\Delta y - \mathscr{A}^*\Delta Z &= -\nabla_x L(x,y,Z) \\ \nabla h(x)^T \Delta x &= -h(x) \\ \Delta X Z + Z\Delta X + X\Delta Z + \Delta Z X &= 2\sigma\mu I - XZ - ZX, \end{aligned}$$

where G denotes the Hessian matrix of the Lagrangian function $L(w)$ or its approximation.

To motivate our choice of the components in the filter and the step decomposition, we rewrite the perturbed KKT-conditions in the form

$$\begin{pmatrix} 0 \\ h(x) \\ X\circ Z - \mu I \end{pmatrix} + \begin{pmatrix} \nabla_x L(x,y,Z) \\ 0 \\ (1-\sigma)\mu I \end{pmatrix} = 0. \tag{3}$$

The first term in the right-hand side of equality measures the proximity bo the quasi-central path, which is defined by

$$P_\mu^q = \{(x,Z):\ h(x)=0,\ X\circ Z = \mu I\}.$$

Therefore it is natural to choose the measure of quasi-centrality

$$\theta(x) = \|h(x)\| + \|X\circ Z - \frac{Tr(XZ)}{n} I\|_{\mathscr{F}}$$

as the first component in the filter. The second term on the left-hand side of (3) measures complementarity and criticality. For the second filter component we choose therefore the optimality measure

$$Tr(XZ)/n + \|\nabla_x L(w)\|^2.$$

With this choice of the filter components it remains to define corresponding tangential and normal components of the trial step. We use the decomposition associated with the splitting (3). For the normal step $s^n = (\Delta x^n, \Delta y^n, \Delta Z^n)$, we thus choose

$$\begin{pmatrix} G\Delta x^n - \nabla h(x)\Delta y^n - \mathscr{A}^*\Delta Z^n \\ \nabla h(x)^T \Delta x^n \\ X\circ \Delta Z^n + Z\circ \Delta X^n \end{pmatrix} = -\begin{pmatrix} 0 \\ h(x) \\ X\circ Z - \mu I \end{pmatrix}, \tag{4}$$

whereas our tangential step $s^t = (\Delta x^t, \Delta y^t, \Delta Z^t)$ is given by

$$\begin{pmatrix} G\Delta x^t - \nabla h(x)\Delta y^t - \mathscr{A}^*\Delta Z^t \\ \nabla h(x)^T \Delta x^t \\ X\circ \Delta Z^t + Z\circ \Delta X^t \end{pmatrix} = -\begin{pmatrix} \nabla_x L(w) \\ 0 \\ (1-\sigma)\mu I \end{pmatrix}. \tag{5}$$

Note that $\Delta w = s^n + s^t$, and the last expression in (4) and (5) are both for matrices, which are used to unify the form. However, it will be crucial that we exploit the flexibility of the step splitting to introduce different stepsizes for s^n and s^t in our trial step computation.

The tangential step is the sum of a tangential component s_1^t, which is a solution of the following system of equations

$$\begin{pmatrix} G\Delta x^t - \nabla h(x)\Delta y^t - \mathscr{A}^*\Delta Z^t \\ \nabla h(x)^T \Delta x^t \\ X \circ \Delta Z^t + Z \circ \Delta X^t \end{pmatrix} = - \begin{pmatrix} \nabla_x L(w) \\ 0 \\ 0 \end{pmatrix}$$

that attempts to reduce $\|\nabla_x L(w)\|$, with a predictor component s_2^t, which is a solution of the following system of equations

$$\begin{pmatrix} G\Delta x^t - \nabla h(x)\Delta y^t - \mathscr{A}^*\Delta Z^t \\ \nabla h(x)^T \Delta x^t \\ X \circ \Delta Z^t + Z \circ \Delta X^t \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ (1-\sigma)\mu I \end{pmatrix}$$

that seeks the minimization of $\mu = Tr(XZ)/n$. Therefore, the tangential step aims to reduce the optimality measure $\theta_g(w) = Tr(XZ)/n + \|\nabla_x L(w)\|^2$.

We introduce Δ as the positive scalar that primarily controls the length of the step taken along Δw, forcing the damped components $\alpha^n(\Delta)s^n$ and $\alpha^t(\Delta)s^t$, to satisfy

$$\|\alpha^n(\Delta)s^n\|_* \le \Delta, \quad \|\alpha^t(\Delta)s^t\|_* \le \Delta.$$

Having these bounds in mind, and requiring explicitly $\alpha^t(\Delta) \le \alpha^n(\Delta)$, we set

$$\alpha^n(\Delta) = \min\left\{1, \frac{\Delta}{\|s^n\|_*}\right\}, \tag{6}$$

$$\alpha^t(\Delta) = \min\left\{\alpha^n(\Delta), \frac{\Delta}{\|s^n\|_*}\right\} = \min\left\{1, \frac{\Delta}{\|s^n\|_*}, \frac{\Delta}{\|s^t\|_*}\right\}. \tag{7}$$

Hereby, we use for $\Delta > 0$ the natural definition $\alpha^n(\Delta) = 1$ for $\|s^n\|_* = 0$ and $\alpha^t(\Delta) = \alpha^n(\Delta)$ for $\|s^t\|_* = 0$ by using the convention $\min\{1, \infty\} = 1$.

Let also

$$\begin{aligned} w(\Delta) &= (x(\Delta), y(\Delta), Z(\Delta)) = w + \alpha^n(\Delta)s^n + \alpha^t(\Delta)s^t, \\ s(\Delta) &= (s_x(\Delta), s_y(\Delta), s_Z(\Delta)) = w(\Delta) - w = \alpha^n(\Delta) + \alpha^t(\Delta)s^t. \end{aligned}$$

Thus,

$$\|s(\Delta)\|_* \le 2\Delta,$$

and here Δ plays a role comparable to the trust-region radius.

We introduce the notation

$$\theta_h(w) = \|h(x)\|, \quad \theta_c(w) = \left\| X \circ Z - \frac{Tr(XZ)}{n} I \right\|, \quad \theta_L(w) = \|\nabla_x L(w)\|,$$

which allows to rewrite the filter components as

$$\theta_c(w),\ \theta_h(w),\ \text{and } \theta_g(w) = \frac{Tr(XZ)}{n} + \|\nabla_x L(w)\|^2.$$

Since $X \circ Z$ might not be zero matrix, a point w that satisfies $\theta_c(w) = \theta_h(w) = \theta_L(w) = 0$ and $X \succeq 0, Z \succeq 0$, might not be a KKT point. The definition of $\theta_g(w)$, however, guarantees that a point w verifying $\theta_c(w) = \theta_h(w) = \theta_g(w) = 0$ and $X \succeq 0, Z \succeq 0$, indeed a KKT point.

With the purpose of achieving a reduction on the function θ_g, we introduce, at a given point w, the quadratic model

$$\begin{aligned}
& m(w(\Delta)) \\
&= \frac{Tr(XZ)}{n} + \frac{Tr((X(\Delta)-X)Z) + Tr((Z(\Delta)-Z)X)}{n} \\
&\quad + \|\nabla_x L(w) + \nabla^2_{xw} L(w)(w(\Delta)-w)\|^2 \\
&= \frac{Tr(X(\Delta)Z(\Delta)) - Tr((X(\Delta)-X)(Z(\Delta)-Z))}{n} \\
&\quad + \|\nabla_x L(w) + \nabla^2_{xw} L(w)(w(\Delta)-w)\|^2,
\end{aligned}$$

by adding to the linearization of $Tr(XZ)/n$ the squared norm of the linearization of $\nabla_x L(w)$. To shorten notation we also set

$$\mu(\Delta) = \frac{Tr(X(\Delta)Z(\Delta))}{n}.$$

In order to prevent $(X(\Delta), Z(\Delta))$ from approaching the boundary of the positive cone too rapidly we will keep the iteration in the neighborhood

$$\mathcal{N}(\gamma, M) = \left\{ w:\ X \succ 0,\ Z \succ 0,\ X \circ Z \succeq \gamma \frac{Tr(XZ)}{n},\ \theta_h(w) + \theta_L(w) \le M \frac{Tr(XZ)}{n} \right\}$$

with fixed $\gamma \in (0,1)$ and $M > 0$. We will set in the next subsection that $w \in \mathcal{N}(\gamma, M)$ implies $w(\Delta) \in \mathcal{N}(\gamma, M)$ whenever $\Delta \in (0, \Delta_{\min}]$ for a given constant $\Delta_{\min} > 0$.

2.2 Step estimates

The following lemma measures the decrease on complementarity obtained by the new iterate $w(\Delta)$.

Lemma 1. *Let $A, B \in \mathbb{S}^{n\times n}$, $\lambda_i(A)$ be its eigenvalue, $\Lambda = diag(\lambda_i(A))$, and $\rho(A)$ be its spectral radius, i.e., $\rho(A) = \max\{|\lambda_i(A)|,\ i = 1, \ldots, n\}$. Then*

$$A \circ B \preceq \rho(A)\rho(B) I$$

Proof. Since

$$\begin{aligned}
& 2\rho(A)\rho(B)I - (Q^TAQQ^TBQ + Q^TBQQ^TAQ) \\
= \ & \rho(A)\rho(B)I - \Lambda Q^TBQ + \rho(A)\rho(B)I - Q^TBQ\Lambda \\
\succeq \ & \Lambda\rho(B)I - \Lambda Q^TBQ + \rho(B)\Lambda - Q^TBQ\Lambda \\
= \ & \Lambda(\rho(B)I - Q^TBQ) + (\rho(B)I - Q^TBQ)\Lambda \\
\succeq \ & 0
\end{aligned}$$

holds for some orthogonal matrix Q, then the lemma follows easily. □

Using the above lemma and Lemma 1 in Ulbrich *et al.* [5], we can show if the current point $w = (x, y, Z)$ satisfies the centrality requirement $X \circ Z \succeq \gamma I$, so does the next point $w(\Delta) = (x(\Delta), y(\Delta), Z(\Delta))$, provided Δ is sufficiently small.

3 The interior-point filter method

We choose θ and θ_g to form a filter entry, where θ measures feasibility and θ_g measures optimality. For the definitions of 'dominance', 'filter', 'acceptable' and 'add', see Ulbrich *et al.* [5].

Algorithm 1. *Primal-dual interior-point filter method*
Step 0. *Choose* $\sigma \in (0,1)$, $\nu \in (0,1)$, $\gamma_1, \gamma_2 > 0$, $0 < \beta, \eta, \kappa < 1$, *and* $\gamma_{\mathscr{F}} \in (0, 1/3)$. *Set* $\mathscr{F} := \varnothing$. *Choose* $(X_0, Z_0) \succ 0$ *and* y_0, *and determined* $\gamma \in (0,1)$ *such that* $X_0 \circ Z_0 \geq \gamma \mu_0$ *with* $\mu_0 = Tr(X_0 Z_0)/n$. *Further, choose* $M > 0$ *such that* $\theta_h(w_0) + \theta_L(w_0) \leq M\mu_0$. *Choose* $\Delta_0^{in} > 0$ *and set* $k := 0$.
Step 1. *Set* $\mu_k := Tr(X_k S_k)/n$ *and compute* s_k^n *and* s_k^t *by solving the linear systems (4) and (5), respectively, with* $(w, \mu) = (w_k, \mu_k)$.
Step 2. *Compute* $\Delta_k' \in [0, \Delta_k^{in}]$ *such that*

$$X_k(\Delta) \succ 0, \;\; Z_k(\Delta) \succ 0, \;\; X_k(\Delta) \circ Z_k(\Delta) \succeq \gamma \mu_k I \;\; for \;\; all \; \Delta \in [0, \Delta_k'].$$

Step 3. *Compute the largest* $\Delta_k'' \in (0, \Delta_k']$ *such that*

$$\theta_h(w_k(\Delta_k'')) + \theta_L(w_k(\Delta_k'')) \leq M\mu_k(\Delta_k'').$$

Set $\Delta_k := \Delta_k''$.
Step 4. *If* $\min\{\theta(w_k), \theta(w_k(\Delta))\} \leq \Delta_k \min\{\gamma_1, \gamma_2 \Delta_k^{\beta}\}$, *then go to Step 5. Otherwise,* **add** w_k *to the filter and call a restoration algorithm that produces a point* w_{k+1} *such that:*

$w_{k+1} \in \mathscr{N}(\gamma, M)$ *with* $\mu_{k+1} = Tr(X_{k+1} Z_{k+1})/n$;
w_{k+1} *is acceptable to the filter;*
$\min\{\theta(w_{k+1}), \theta(w_{k+1}(\Delta))\} \leq \Delta_{k+1} \min\{\gamma_1, \gamma_2 \Delta_k^{\beta}\}$ *with* $\Delta_{k+1} = \Delta_k$.

Set $\Delta_{k+1}^{in} := \Delta_k$, $k := k+1$, *and go to Step 1.*
Step 5. *If* $w_k(\Delta_k)$ *is not acceptable to the filter (and* $m_k(w_k)$-$m_k(w_k(\Delta_k))<\kappa \min\{\theta(w_k)^2, \theta(w_k(\Delta))^2\}$*), then go to Step 11.*
Step 6. *If* $m_k(w_k) - m_k(w_k(\Delta_k)) = 0$, *then set* $\rho_k := 0$. *Otherwise,*

$$\rho_k := \frac{\theta_g(w_k) - \theta_g(w_k(\Delta_k))}{m_k(w_k) - m_k(w_k(\Delta_k))}.$$

Step 7. *If* $\rho_k < \eta$ *and* $m_k(w_k) - m_k(w_k(\Delta_k)) \geq \kappa \min\{\theta(w_k)^2, \theta(w_k(\Delta))^2\}$, *then go to Step 11.*
Step 8. *If* $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa \min\{\theta(w_k)^2, \theta(w_k(\Delta))^2\}$, *then* **add** w_k *to the filter.*
Step 9. *Choose* $\Delta_k^{in} \geq \Delta_k$.
Step 10. *Set* $w_{k+1} := w_k(\Delta_k)$, $k := k+1$, *and go to Step 1.*
Step 11. *Set* $w_{k+1} := w_k, s_{k+1}^n := s_k^n, s_{k+1}^t := s_k^t, \Delta_{k+1}' := \Delta_k/2$. *Set* $k := k+1$ *and go to Step 3.*

4 Global convergence to first-order critical points

The algorithm have two cases, the first is that infinitely many iterates are added to the filter and the second is that the algorithm runs infinitely but the filter is left with a finite number of entries. We summarize both situations in the next theorems. For the details of proof, see Liu and Sun [4].

Theorem 2. *Suppose that infinitely many iterates are added to the filter. Then there exists a subsequence* $\{k_j\}$ *such that*

$$\lim_{j\to\infty} \theta(w_{k_j}) = 0, \quad \lim_{j\to\infty} \theta_g(w_{k_j}) = 0.$$

It remains to consider the case where the algorithm runs infinitely but the filter is left with a finite number of entries.

Theorem 3. *Suppose that the algorithm runs infinitely and only finitely many iterates added to the filter. Then*

$$\lim_{k\to\infty} \theta(w_k) = 0, \quad \liminf_{k\to\infty} \theta_g(w_k) = 0.$$

The main result is obtained by combining Theorem 2 and Theorem 3.

Corollary 4. *The sequence of iterates* $\{w_k\}$ *generated by the primal-dual interior-point method (Algorithm 1) satisfies*

$$\liminf_{k\to\infty} \theta(w_k) + \theta_g(w_k) = 0.$$

5 Concluding remarks

In this paper we extend the interior-point filter method to nonlinear semidefinite programming. This work is trivial and little difficulty happened. From the paper, we can see that it is natural that three-dimensional filter methods can be with thinking, where the first is for feasibility, the second for centrality, and the third for optimization. In Liu and Sun [4], we presented a three-dimensional filter interior-point method for nonlinear semidefinite programming.

References

[1] R. Fletcher, S. Leyffer, and P.L. Toint. On the Global Convergence of and SLP-Filter Algorithm, Report 98/13, Dept of Mathematics, FUNDP, Namur (B), 1998.

[2] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91(2):239–269, 2002.

[3] R. Fletcher, N.I.M. Gould, S. Leyffer, P. L. Toint, and A. Wächter. Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming. *SIAM Journal on Optimization*, 13(3):635–659, 2002.

[4] Z. Liu and W. Sun. A globally convergent primal-dual interior-point three-dimensional filter method for nonlinear semidefinite programming. Submitted to SIAM Journal on Optimization.

[5] M. Ulbrich, S. Ulbrich, and L.N. Vicente. A globally convergent primal-dual interior-point filter method for nonlinear programming. *Mathematical Programming*, 100(2):379–410, 2004.