A Nonsmooth Newton-Type Method for Nonlinear Semidefinite Programming

Li Chengjin, Sun Wenyu

(School of Mathematics and Computer Science, Nanjing Normal University, Nanjing 210097, China)

Abstract A nonsmooth Newton s method for nonlinear semidefinite programming was discussed by using 4-tensor analysis. The locally quadratic convergence for this nonsmooth Newton s method was also established

Key words nonlinear semidefinite programming nonsmooth Newton-type method k-tensor convergence

CLC number O221. 2 **Document code** A **Article ID** 1001-4616(2008)02-0001-07

针对非线性半定规划的一类非光滑牛顿型方法

李成进, 孙文瑜

(南京师范大学数学与计算机科学学院, 江苏 南京 210097)

[摘要] 通过 4阶张量分析讨论了一类针对非线性半定规划的非光滑牛顿法. 并给出了这种非光滑牛顿法的局部二次收敛性.

[关键词] 非线性半定规划, 非光滑牛顿型方法, k张量, 收敛性

In this paper, we consider the nonlinear semidefinite programming (nonlinear SDP):

$$\min_{X\in S^n} f(X) \text{ s. t } g(X)=0,\ X\succeq 0, \tag{1}$$

where $f:S^n\to\mathbf{R}$, $g:S^n\to\mathbf{R}^m$ are twice differentiable functions, S^n denotes the subspace of all symmetric matrices in $\mathbf{R}^{n\times n}$, and $X\succeq 0$ a symmetric positive semidefinite matrix. If f and g are all linear (affine) functions, the nonlinear SDP problem (1) reduces to a normal linear SDP problem, which has been extensively studied during the last decade[1,2].

It is well known that the KKT conditions of (1) is

$$\Theta(X,\lambda,S)=\begin{pmatrix} Df(X)+\sum_{r=1}^{m}\lambda_r Dg_r(X)-S \\ g(X) \\ P_{S_+^n}(X-S)-X \end{pmatrix}=0, \tag{2}$$

where S_+^n is the cone of all $n\times n$ symmetric positive semidefinite matrices, $P_{S_+^n}(\cdot)$ is the orthogonal projection on S_+^n and $Df(X)\,(D^2f(X))$ is (twice) F-derivative of f at point X. We use $\langle A,B\rangle:=\mathrm{Tr}(A^{\mathrm T}B)$ (where Tr denote the trace of a matrix) and $\|X\|_F\,(\|b\|_2)$ to denote the inner product of $A,B\in\mathbf{R}^{n\times n}$ and F-norm (2-norm) respectively.

Tensor analysis and computation are a useful tool for optimization[3]. The aims of this paper is to discuss the

Received date 2008-03-21.

Foundation item: Supported by the National Natural Science Foundation of China (10231060, 10501024), the Specialized Research Fund of Doctoral Program of Higher Education of China (20040319003), the Natural Science Fund of Jiangsu Province (BK2006214), and the Key Subject Fund of Nanjing Normal University.

Corresponding author: Li Chengjin, doctoral, majored in numerical mathematics. E-mail: chengjin98298@sina.com

nonsmooth Newton's method for nonlinear semidefinite programming by 4-tensor analysis and to prove the convergence result of this method by an equivalent condition. The equivalent condition used in this paper can be viewed as a generalization of the results in paper [4].

1 Nonsmooth Newton's Method

In this section, we present a nonsmooth Newton's method for solving problem (2) by use of 4-tensor analysis. The definitions of B-subdifferential (C-subdifferential) and k-tensor can be found in [5] and [3] respectively. Thanks to Malick and Sendov[6], where they give the explicit formula of the B-subdifferential of $P_{S^n_+}(\cdot)$, i. e, for all $\boldsymbol{Q}\in S^n$,

$$\partial_B P_{S^n_+}(\boldsymbol{Q}) = O(n)^{\boldsymbol{Q}}\cdot(\mathrm{Diag}^{(12)}\mathscr{L}(\boldsymbol{Q})), \tag{3}$$

where, for more details of $O(n)$, $O(n)^{\boldsymbol{Q}}$, $\mathrm{Diag}^{(12)}$, '$\cdot$' and $\mathscr{L}(\boldsymbol{Q})$, please refer paper [6].

Let $\lambda: S^n\rightarrow \mathbf{R}^n$[6] be a function which maps a symmetric matrix to the vector of its eigenvalues with non-increasing order, and we can define the following three index sets:

$$\alpha: = \{i:\lambda_i(\boldsymbol{Q})>0\},$$
$$\beta: = \{i:\lambda_i(\boldsymbol{Q})=0\},$$
$$\gamma: = \{i:\lambda_i(\boldsymbol{Q})<0\}.$$

Moreover, suppose $\boldsymbol{Q}_{\alpha\beta}$ is the submatrix of matrix $\boldsymbol{Q}$ with the row index $i\in\alpha$ and column index $j\in\beta$, and $|\alpha|$, $|\beta|$, $|\gamma|$ denote the number of elements in the set α, β, γ, respectively.

For the definitions of sets $\mathscr{D}_{\{01\}}(m)$, $\mathscr{D}_{[01]}(m)$, $\mathscr{D}_{\{01\}}(\boldsymbol{Q})$ and $\mathscr{D}_{[01]}(\boldsymbol{Q})$, please refer paper [6], and we know the following relation holds from paper [6]: for all $\boldsymbol{Q}\in S^n$,

$$\mathscr{D}_{\{01\}}(\boldsymbol{Q})\subsetneq \mathscr{L}(\boldsymbol{Q})\subseteq\mathscr{D}_{[01]}(\boldsymbol{Q}). \tag{4}$$

The nonsmooth Newton's method for (2) would produce each iteration step by solving the following equations:

$$\begin{cases} D^2 f(\boldsymbol{X}^k)(\Delta\boldsymbol{X}) + \sum_{r=1}^m \lambda_r^k D^2 g_r(\boldsymbol{X}^k)(\Delta\boldsymbol{X}) + \sum_{r=1}^m \Delta\lambda_r Dg_r(\boldsymbol{X}^k) - \Delta\boldsymbol{S} = -\boldsymbol{R}_1^k, \\ \langle Dg_r(\boldsymbol{X}^k), \Delta\boldsymbol{X}\rangle = -g_r(\boldsymbol{X}^k),\ r = 1,2,\cdots,m, \\ V^k(\Delta\boldsymbol{X}, \Delta\boldsymbol{S}) = -\boldsymbol{R}_2^k, \end{cases} \tag{5}$$

where $\boldsymbol{Q}^k: =\boldsymbol{X}^k-\boldsymbol{S}^k$, $V^k\in\partial_B(P_{S^n_+}(\boldsymbol{Q}^k)-\boldsymbol{X}^k)$, $\boldsymbol{R}_1^k = Df(\boldsymbol{X}^k)+\sum_{r=1}^m\lambda_r^k Dg_r(\boldsymbol{X}^k)-\boldsymbol{S}^k$ and $\boldsymbol{R}_2^k = P_{S^n_+}(\boldsymbol{Q}^k)-\boldsymbol{X}^k$. It is easy to know $V^k(\Delta\boldsymbol{X}, \Delta\boldsymbol{S}) = \tilde{\mathcal{T}}^k(\Delta\boldsymbol{X})-\tilde{\mathcal{T}}^k(\Delta\boldsymbol{S})-\Delta\boldsymbol{X}$, where $\tilde{\mathcal{T}}^k\in\partial_B P_{S^n_+}(\boldsymbol{Q}^k)$.

Let $\boldsymbol{I}_{ij}\in S^n$ be a zero matrix with one at positions (i, j) and (j, i) for all $j=1, \cdots, n$, $i=j, \cdots, n$ and

$$\tilde{\boldsymbol{I}}_{ij}: = \begin{cases} \boldsymbol{I}_{ij}, & \text{if } i=j, \\ \dfrac{\sqrt{2}}{2}\boldsymbol{I}_{ij}, & \text{if } i\neq j, \end{cases}$$

$$\boldsymbol{F}_{ij}^k: = D^2 f(\boldsymbol{X}^k)(\tilde{\boldsymbol{I}}_{ij}),$$

$$\boldsymbol{G}_{ij}^{r,k}: = D^2 g_r(\boldsymbol{X}^k)(\tilde{\boldsymbol{I}}_{ij}).$$

It is well known that operator *svec* is defined as follows.

$$svec(\boldsymbol{X}): = (X_{11}, \sqrt{2}X_{21}, \cdots, \sqrt{2}X_{n1}, X_{22}, \sqrt{2}X_{32}, \cdots)^{\mathrm{T}}\in\mathbf{R}^{\frac{n(n+1)}{2}},$$

where X_{ij} denotes the ij-element in X. Let

$$\boldsymbol{\Omega}_{ij}^k = \boldsymbol{F}_{ij}^k + \sum_{r=1}^m\lambda_r^k\boldsymbol{G}_{ij}^{r,k}\in S^n,$$

$$\boldsymbol{\Omega}^k = (svec(\boldsymbol{\Omega}_{11}^k), svec(\boldsymbol{\Omega}_{21}^k), \cdots, svec(\boldsymbol{\Omega}_{n1}^k), svec(\boldsymbol{\Omega}_{22}^k), svec(\boldsymbol{\Omega}_{32}^k), \cdots)$$

and

$$\boldsymbol{A}^k = (svec(Dg_1(\boldsymbol{X}^k)), svec(Dg_2(\boldsymbol{X}^k)), \cdots, svec(Dg_m(\boldsymbol{X}^k))).$$

So, the first and second equations in (5) can be reformulated as

$$\begin{aligned} \boldsymbol{\Omega}^k svec(\Delta \boldsymbol{X}) + \boldsymbol{A}^k \Delta \boldsymbol{\lambda} - svec(\Delta \boldsymbol{S}) &= -svec(\boldsymbol{R}_1^k), \\ (\boldsymbol{A}^k)^{\mathrm{T}} svec(\Delta \boldsymbol{X}) &= -\boldsymbol{g}, \end{aligned} \tag{6}$$

where $\boldsymbol{g} = (g_1(\boldsymbol{X}^k), g_2(\boldsymbol{X}^k), \cdots, g_m(\boldsymbol{X}^k))^{\mathrm{T}}$.

Before reformulating the last equation in (5), we need to prove the following Proposition 1. We first introduce an operation in paper[6]: for all 4-tensor T on $\mathbf{R}^n$ and all $n \times n$ matrix $\boldsymbol{M}$, the operation result $T(\boldsymbol{M})$ should be defined as an $n \times n$ matrix with element $t_{ij} = \sum_{t,s=1}^{n} T^{t,i}_{s,j} M_{ts}$. By $\mathcal{Q}(\boldsymbol{A})$, we denote the index set $\{(i,j) \mid A_{ij} = 1\}$ for all $\boldsymbol{A} \in \mathcal{D}_{\{01\}}(|\beta|)$.

Proposition 1 $(\widetilde{T}^k(\boldsymbol{X}))^{\mathrm{T}} = \widetilde{T}^k(\boldsymbol{X})$ for all $\boldsymbol{X} \in S^n$ and $\widetilde{T}^k \in \partial_B P_{S^n_+}(\boldsymbol{Q}^k)$ with $\mathcal{S}(|\beta^k|) = \mathcal{D}_{\{01\}}(|\beta^k|)$.

Proof For simplicity, we omit the superscript k and let $\boldsymbol{B} \in \mathcal{D}_{\{01\}}(|\beta|)$. It is sufficient to prove that $(\widetilde{T}(\boldsymbol{X}))_{t_2 t_4} = (\widetilde{T}(\boldsymbol{X}))_{t_4 t_2}$ holds for all $\boldsymbol{X} \in S^n$ and $\widetilde{T} \in \partial_B P_{S^n_+}(Q)$ with this middle main sub-matrix $\boldsymbol{B}$.

From above definition, we have

$$(\widetilde{T}(\boldsymbol{X}))_{t_2 t_4} = \widetilde{T}^{1,t_2}_{1,t_4} X_{11} + \widetilde{T}^{1,t_2}_{2,t_4} X_{12} + \cdots + \widetilde{T}^{1,t_2}_{n,t_4} X_{1n} + \widetilde{T}^{2,t_2}_{1,t_4} X_{21} + \widetilde{T}^{2,t_2}_{2,t_4} X_{22} + \cdots + \widetilde{T}^{n,t_2}_{n,t_4} X_{nn},$$

and the formula of $(\widetilde{T}(\boldsymbol{X}))_{t_4 t_2}$ is similar. So, it is sufficient to show that each coefficient of different $X_{ij}(i=1, \cdots, n, j=1, \cdots, n)$ in $(\widetilde{T}(\boldsymbol{X}))_{t_2 t_4}$ and $(\widetilde{T}(\boldsymbol{X}))_{t_4 t_2}$ are equivalent, i. e., $\widetilde{T}^{t_1,t_2}_{t_3,t_4} = \widetilde{T}^{t_1,t_4}_{t_3,t_2}$ for all $t_1 = t_3$, and $\widetilde{T}^{t_1,t_2}_{t_3,t_4} + \widetilde{T}^{t_3,t_2}_{t_1,t_4} = \widetilde{T}^{t_1,t_4}_{t_3,t_2} + \widetilde{T}^{t_3,t_4}_{t_1,t_2}$ for all $t_1 \neq t_3$. We only to prove that $\widetilde{T}^{t_1,t_2}_{t_3,t_4} = \widetilde{T}^{t_1,t_4}_{t_3,t_2}$ for all $t_1 = t_3$ since the proof of case $t_1 \neq t_3$ is similarly.

After directly computing, together with the formula (3), (4), we can show the formula of $\widetilde{T}^{t_1,t_2}_{t_3,t_4}$ as follows:

$$\begin{aligned} \widetilde{T}^{t_1,t_2}_{t_3,t_4} = & \sum_{i_1=1}^{|\alpha|} \sum_{i_2=1}^{|\alpha|} U_{t_1 i_1} U_{t_2 i_2} U_{t_3 i_2} U_{t_4 i_1} + \Big(\sum_{i_1=1}^{|\alpha|} \sum_{i_2=|\alpha|+1}^{|\alpha|+|\beta|} U_{t_1 i_1} U_{t_2 i_2} U_{t_3 i_2} U_{t_4 i_1} + \\ & \sum_{i_1=|\alpha|+1}^{|\alpha|+|\beta|} \sum_{i_2=1}^{|\alpha|} U_{t_1 i_1} U_{t_2 i_2} U_{t_3 i_2} U_{t_4 i_1} \Big) + \sum_{(i_1,i_2) \in Q(B)} U_{t_1 i_1} U_{t_2 i_2} U_{t_3 i_2} U_{t_4 i_1} + \\ & \sum_{i_1=1}^{|\alpha|} \sum_{i_2=|\alpha|+|\beta|+1}^{n} \frac{\lambda_{i_1}}{\lambda_{i_1} - \mu_{i_2}} (U_{t_1 i_1} U_{t_2 i_2} U_{t_3 i_2} U_{t_4 i_1} + U_{t_1 i_2} U_{t_2 i_1} U_{t_3 i_1} U_{t_4 i_2}), \end{aligned} \tag{7}$$

where $\boldsymbol{U} \in O(n)^{\boldsymbol{Q}}$, λ_i, μ_i are positive and negative eigenvalue of $\boldsymbol{Q}$ respectively. We let $i_1' = i_2$, $i_2' = i_1$, then, since $t_1 = t_3$, the first term of (7) becomes

$$\sum_{i_2'=1}^{|\alpha|} \sum_{i_1'=1}^{|\alpha|} U_{t_1 i_2'} U_{t_2 i_1'} U_{t_3 i_1'} U_{t_4 i_2'} = \sum_{i_1'=1}^{|\alpha|} \sum_{i_2'=1}^{|\alpha|} U_{t_1 i_1'} U_{t_4 i_2'} U_{t_3 i_2'} U_{t_2 i_1'},$$

and the right hand side of the above equation is exactly the first term of $\widetilde{T}^{t_1,t_4}_{t_3,t_2}$. Similarly, the second term of (7) can be reformulated to the second term of $\widetilde{T}^{t_1,t_4}_{t_3,t_2}$, and from the symmetry of $\boldsymbol{B} \in \mathcal{D}_{\{01\}}(\beta)$ we know the third term of (7) can be reformulated as the similar formula of the last term of (7). It follows that we only need to show the symmetry of the last term.

Now, we show that each item in the sum of the last term of (7) is symmetric relating to t_2 and t_4, i. e.,
$U_{t_1 i_1} U_{t_2 i_2} U_{t_3 i_2} U_{t_4 i_1} + U_{t_1 i_2} U_{t_2 i_1} U_{t_3 i_1} U_{t_4 i_2} = U_{t_1 i_2} U_{t_4 i_1} U_{t_3 i_1} U_{t_2 i_2} + U_{t_1 i_1} U_{t_4 i_2} U_{t_3 i_2} U_{t_2 i_1} = U_{t_1 i_1} U_{t_4 i_2} U_{t_3 i_2} U_{t_2 i_1} + U_{t_1 i_2} U_{t_4 i_1} U_{t_3 i_1} U_{t_2 i_2}$, so we have $\widetilde{T}^{t_1,t_2}_{t_3,t_4} = \widetilde{T}^{t_1,t_4}_{t_3,t_2}$.

From the relations of $\mathcal{D}_{\{01\}}(m)$ and $\mathcal{D}_{[01]}(m)$, we know that the elements in the latter set can be reformulated as linear combination of elements in front set. So, Proposition 1 holds for all $\mathcal{S}(|\beta^k|) = \mathcal{D}_{[01]}(m)$, which, together with (4), concludes that Proposition 1 holds for all $\widetilde{T}^k \in \partial_B P_{S^n_+}(\boldsymbol{Q}^k)$ too.

By $\mathcal{Z}^k$ we denote the set $\{\widetilde{T}_k \in \partial_B P_{S^n_+}(\boldsymbol{Q}^k)$ with its middle main sub-matrix $\boldsymbol{B} \in \mathcal{D}_{\{01\}}(|\beta^k|)\}$. Of course,

$\mathscr{Z}^k := \{D\,P_{S^n_+}(\boldsymbol{Q}^k)\}$ when $|\beta^k| = 0$. Let

$$T_k^{i,j} := \left(\tilde{T}_k^{1,i}{}_{1,j},\ \frac{\sqrt{2}}{2}(\tilde{T}_k^{2,i}{}_{1,j} + \tilde{T}_k^{1,i}{}_{2,j}),\ \frac{\sqrt{2}}{2}(\tilde{T}_k^{3,i}{}_{1,j} + \tilde{T}_k^{1,i}{}_{3,j}),\ \cdots,\ \frac{\sqrt{2}}{2}(\tilde{T}_k^{n,i}{}_{1,j} + \tilde{T}_k^{1,i}{}_{n,j}),\ \cdots\right),$$

where $j = 1,\cdots,n$, $i = j,\cdots,n$ and $\tilde{T}_k \in \mathscr{Z}^k$. So we can define

$$\mathscr{L}^k := \{\boldsymbol{T}^k \mid \boldsymbol{T}^k = (T_k^{1,1},\ \sqrt{2}T_k^{2,1},\ \cdots,\ \sqrt{2}T_k^{n,1},\ T_k^{2,2},\ \sqrt{2}T_k^{3,2},\ \cdots)^{\mathrm{T}} \text{ with } \tilde{T}_k \in \mathscr{Z}^k\}.$$

From the above assumptions, we know that the third equation in (5) can be reformulated as

$$\boldsymbol{T}^k svec(\Delta\boldsymbol{X}) - \boldsymbol{T}^k svec(\Delta\boldsymbol{S}) - svec(\Delta\boldsymbol{X}) = -svec(\boldsymbol{R}_2^k), \tag{8}$$

where $\boldsymbol{T}^k \in \mathscr{L}^k$. Together with equations (6) and (8), the equation (5) can be reformulated as

$$\boldsymbol{\Phi}\begin{pmatrix} svec(\Delta\boldsymbol{X}) \\ \Delta\boldsymbol{\lambda} \\ svec(\Delta\boldsymbol{S}) \end{pmatrix} := \begin{pmatrix} \boldsymbol{\Omega}^k & \boldsymbol{A}^k & -\boldsymbol{I} \\ (\boldsymbol{A}^k)^{\mathrm{T}} & 0 & 0 \\ \boldsymbol{T}^k - \boldsymbol{I} & 0 & -\boldsymbol{T}^k \end{pmatrix}\begin{pmatrix} svec(\Delta\boldsymbol{X}) \\ \Delta\boldsymbol{\lambda} \\ svec(\Delta\boldsymbol{S}) \end{pmatrix} = \begin{pmatrix} -svec(\boldsymbol{R}_1^k) \\ -\boldsymbol{g} \\ -svec(\boldsymbol{R}_2^k) \end{pmatrix}, \tag{9}$$

where $\boldsymbol{T}^k \in \mathscr{L}^k$ and $\boldsymbol{\Phi}$ is a matrix, so equation (9) can be solved by several efficient methods for linear equations. Let $Y := (\boldsymbol{X}, \boldsymbol{\lambda}, \boldsymbol{S})$, $\Delta Y := (\Delta\boldsymbol{X}, \Delta\boldsymbol{\lambda}, \Delta\boldsymbol{S})$, we are in a position to state our algorithm in the following.

Algorithm 1 Nonsmooth Newton's Method

(S.0) Choose $Y^0 \in S^n \times R^m \times S^n$, $\varepsilon \geqslant 0$ and set $k := 0$.

(S.1) If $\|\boldsymbol{\Theta}(Y^k)\|_N \leqslant \varepsilon$, STOP. (where, $\|Y\|_N^2 := \|\boldsymbol{X}\|_F^2 + \|\boldsymbol{\lambda}\|_2^2 + \|\boldsymbol{S}\|_F^2$).

(S.2) Choose $\boldsymbol{T}^k \in \mathscr{L}^k$, and find ΔY by solving equation (9).

(S.3) Set $Y^{k+1} = Y^k + \Delta Y$, $k \leftarrow k+1$, go to (S.1).

2 Convergence Analysis

In this section, we are interested in the local convergence of nonsmooth Newton method for nonlinear semidefinite programming. The superscript k defined in Section 1 is replaced by $*$ in this section.

It is clear that the nonsingularity of the matrix

$$\begin{pmatrix} 0 & (\boldsymbol{A}^*)^{\mathrm{T}} \\ -\boldsymbol{T}^*\boldsymbol{A}^* & \boldsymbol{T}^*(\boldsymbol{I} - \boldsymbol{\Omega}^*) - \boldsymbol{I} \end{pmatrix} \tag{10}$$

is equivalent to the nonsingularity of the coefficient matrix of equation (9) for the same $\boldsymbol{T}^*$. $\boldsymbol{T}_0^*$ and $\boldsymbol{T}_1^*$ denote the matrix $\boldsymbol{T}^* \in \mathscr{L}^*$ with the middle main sub-matrix $\boldsymbol{B} \in \mathscr{S}(|\beta^*|)$ being $0_{\beta^*\beta^*}$ and $1_{\beta^*\beta^*}$, respectively, when $|\beta^*| \neq 0$.

Let $\mathscr{A}^*(\cdot) := (\langle D\,g_1(\boldsymbol{X}^*), \cdot\rangle, \cdots, \langle D\,g_m(\boldsymbol{X}^*), \cdot\rangle)^{\mathrm{T}}$ and the operator $(\mathscr{A}^*)^{\dagger}$ be the adjoint of $\mathscr{A}^*$. By $\mathscr{T}_{S^n_+}(\boldsymbol{X}^*)$ and $lin(\mathscr{T}_{S^n_+}(\boldsymbol{X}^*))$, we denote the tangent cone of S^n_+ at $\boldsymbol{X}^*$ and the linear space of $\mathscr{T}_{S^n_+}(\boldsymbol{X}^*)$ respectively[7].

Before showing the convergence theorem, we introduce another formula of $\partial_B\boldsymbol{\Theta}(Y^*)$, which comes from *KN*-nondegeneracy in paper [8] for linear SDP, which indicates that each element in $\partial_B\boldsymbol{\Theta}(Y^*)$ can, in matrix-vector notation, be written as

$$\begin{pmatrix} \boldsymbol{\Omega}^* & \boldsymbol{A}^* & \boldsymbol{I} \\ (\boldsymbol{A}^*)^{\mathrm{T}} & 0 & 0 \\ \frac{1}{2}\boldsymbol{P}^*\sum_-^*(\boldsymbol{P}^*)^{\mathrm{T}} & 0 & \frac{1}{2}\boldsymbol{P}^*\sum_+^*(\boldsymbol{P}^*)^{\mathrm{T}} \end{pmatrix}. \tag{11}$$

Definition 1 A solution Y^* of the optimality condition (2) is called *KN*-nondegenerate if, for all diagonal matrices $\sum_-$ and $\sum_+$, the following implication holds for any triple $(\Delta\boldsymbol{X}, \Delta\boldsymbol{\lambda}, \Delta\boldsymbol{S}) \in S^n \times R^m \times S^n$:

$$\left.\begin{array}{l} D^2 f(X^*)(\Delta X) + \sum_{r=1}^{m} \lambda_r^* D^2 g_r(X^*)(\Delta X) + (\mathscr{A}^*)^{\dagger}(\Delta \lambda) - \Delta S = 0, \\ \mathscr{A}^*(\Delta X) = 0, \\ \frac{1}{2} P^* \sum_{-}^{*} (P^*)^{\mathrm{T}} svec(\Delta X) + \frac{1}{2} P^* \sum_{+}^{*} (P^*)^{\mathrm{T}} svec(\Delta S) = 0. \end{array}\right\} \Rightarrow \begin{cases} svec(\Delta X) = 0, \\ svec(\Delta S) = 0. \end{cases}$$

Above definition of *KN*-nondegeneracy, which is suit for nonlinear SDP, is a generalization of the relative definition for linear SDP in paper [8]. The following lemma hold obviously.

Lemma 1 Suppose that $A_1^* \in R^{\frac{n(n+1)}{2} \times m}$ is of full column rank, i. e., $rank(A_1^*) = m$. Then the primal constraint nondegeneracy is equivalent to

$$\mathscr{N}^{\perp} \cap (lin(\mathscr{T}_{S_+^n}(X^*)))^{\perp} = \{0\},$$

where $\mathscr{N}$ denotes the null space of $\mathscr{A}^*$.

Now, under above lemma, we can deduce the main theorem in this section.

Theorem 1 Let Y^* be a solution of the optimality conditions (2) satisfying that the matrix (10) is non-singular for $T^* \in \{T_0^*, T_1^*\}$. If for all $\Delta X \in S^n$, $svec(\Delta X)^{\mathrm{T}} \Omega^* svec(\Delta X) \geqslant 0$ and $svec(\Delta X)^{\mathrm{T}} \Omega^* svec(\Delta X) = 0 \Rightarrow V^* svec(\Delta X) = 0$, then Algorithm 1 is locally quadratically convergent.

Proof It is easy to see from the assumptions that

$$0 \leqslant svec(\Delta X)^{\mathrm{T}} \Omega^* svec(\Delta X) = \langle \Delta X, D^2 f(X^*)(\Delta X) + \sum_{r=1}^{m} \lambda_r^* D^2 g_r(X^*)(\Delta X) \rangle \tag{12}$$

for all $\Delta X \in S^n$. The non-singularity of the matirx (10) corresponding to $T^* = T_0^*$ yields the matrix $A_1^* \in R^{\frac{n(n+1)}{2} \times m}$ has full column rank, which together with Lemma 2.6 in paper [4] and Lemma 1 gives

$$\mathscr{N}^{\perp} \cap (lin(\mathscr{T}_{S_+^n}(X^*)))^{\perp} = \{0\}. \tag{13}$$

Let $(\Delta X, \Delta \lambda, \Delta S) \in S^n \times R^m \times S^n$ be any triple satisfying

$$\begin{cases} D^2 f(X^*)(\Delta X) + \sum_{r=1}^{m} \lambda_r^* D^2 g_r(X^*)(\Delta X) + (\mathscr{A}^*)^{\dagger}(\Delta \lambda) - \Delta S = 0, \\ \mathscr{A}^*(\Delta X) = 0, \\ \frac{1}{2} P^* \sum_{-}^{*} (P^*)^{\mathrm{T}} svec(\Delta X) + \frac{1}{2} P^* \sum_{+}^{*} (P^*)^{\mathrm{T}} svec(\Delta S) = 0. \end{cases} \tag{14}$$

Multiplying ΔX to both side of the first equation in (14), we have, from the second equation in (14) and (12), that

$$\langle \Delta X, \Delta S \rangle = \langle \Delta X, D^2 f(X^*)(\Delta X) + \sum_{r=1}^{m} \lambda_r^* D^2 g_r(X^*)(\Delta X) \rangle + \langle \mathscr{A}^*(\Delta X), \Delta \lambda \rangle \geqslant 0.$$

Furthermore, since $P^* = U^* \otimes_S U^*$ is non-singular, where $U^* \in O(n)^{\Omega^*}$, and $\otimes_S$ is the Kronecker product, the last equation in (14) is equivalent to

$$\sum_{-}^{*} (P^*)^{\mathrm{T}} svec(\Delta X) + \sum_{+}^{*} (P^*)^{\mathrm{T}} svec(\Delta S) = 0,$$

which, together with the property of Kronecker product $(P^*)^{\mathrm{T}} svec(\Delta X) = svec((U^*)^{\mathrm{T}} \Delta X U^*)$, yields $\sum_{-}^{*} svec(\widehat{\Delta X}) + \sum_{+}^{*} svec(\widehat{\Delta S}) = 0$, where $\widehat{\Delta X}$ and $\widehat{\Delta S}$ are matrices $(U^*)^{\mathrm{T}} \Delta X U^*$ and $(U^*)^{\mathrm{T}} \Delta S U^*$ respectively. Componentwise, this may be written as

$$\sigma_{ij}^{-} \widehat{\Delta X}_{ij} + \sigma_{ij}^{+} \widehat{\Delta S}_{ij} = 0, \ \forall 1 \leqslant j \leqslant i \leqslant n, \tag{15}$$

which, together with the definitions and properties of diagonal matrices $\sum_{-}$ and $\sum_{+}$, yields

$$\widehat{\Delta S}_{ij} = 0, \ \forall (i,j) \in (\alpha^* \times \alpha^*) \cup (\alpha^* \times \beta^*) \cup (\beta^* \times \alpha^*), \tag{16}$$

$$\widehat{\Delta X}_{ij} = 0, \ \forall (i,j) \in (\beta^* \times \gamma^*) \cup (\gamma^* \times \beta^*) \cup (\gamma^* \times \gamma^*). \tag{17}$$

Furthermore, since $\sigma_{ij}^{-}, \sigma_{ij}^{+} \geqslant 0$, $\sigma_{ij}^{-} + \sigma_{ij}^{+} > 0$ for all $(i,j) \in \beta \times \beta$, it follows from (15) that

$$\widehat{\Delta S}_{ij}\,\widehat{\Delta X}_{ij} \leqslant 0,\ \forall (i,j) \in \beta^* \times \beta^*. \tag{18}$$

Let us partition the matrices $\widehat{\Delta X}$ and $\widehat{\Delta S}$, then we obtain from (16),(17) that

$$\widehat{\Delta X} = \begin{pmatrix} \widehat{\Delta X}_{\alpha^*\alpha^*} & \widehat{\Delta X}_{\alpha^*\beta^*} & \widehat{\Delta X}_{\alpha^*\gamma^*} \\ \widehat{\Delta X}_{\alpha^*\beta^*}{}^{\mathrm{T}} & \widehat{\Delta X}_{\beta^*\beta^*} & 0 \\ \widehat{\Delta X}_{\alpha^*\gamma^*}{}^{\mathrm{T}} & 0 & 0 \end{pmatrix},$$

$$\widehat{\Delta S} = \begin{pmatrix} 0 & 0 & \widehat{\Delta S}_{\alpha^*\gamma^*} \\ 0 & \widehat{\Delta S}_{\beta^*\beta^*} & \widehat{\Delta S}_{\beta^*\gamma^*} \\ \widehat{\Delta S}^{\mathrm{T}}_{\alpha^*\gamma^*} & \widehat{\Delta S}^{\mathrm{T}}_{\beta^*\gamma^*} & \widehat{\Delta S}_{\gamma^*\gamma^*} \end{pmatrix}.$$

Since $\langle \Delta X, \Delta S\rangle \geqslant 0$, we have $\langle \widehat{\Delta X}, \widehat{\Delta S}\rangle \geqslant 0$. Then it follows from the previous representations of $\widehat{\Delta X}$ and $\widehat{\Delta S}$ that

$$0 \leqslant \langle \widehat{\Delta X}, \widehat{\Delta S}\rangle = \mathrm{Tr}(\widehat{\Delta X}_{\alpha^*\gamma^*}\,\widehat{\Delta S}^{\mathrm{T}}_{\alpha^*\gamma^*}) + \mathrm{Tr}(\widehat{\Delta X}_{\beta^*\beta^*}\,\widehat{\Delta S}_{\beta^*\beta^*}) + \mathrm{Tr}(\widehat{\Delta X}^{\mathrm{T}}_{\alpha^*\gamma^*}\,\widehat{\Delta S}_{\alpha^*\gamma^*}). \tag{19}$$

Since

$$\widehat{\Delta X}_{ij} = -\frac{\sigma_{ij}^+}{\sigma_{ij}^-}\widehat{\Delta S}_{ij},\ \forall (i,j) \in \alpha^* \times \gamma^* \tag{20}$$

and $\sigma_{ij}^+, \sigma_{ij}^- > 0$ for all $(i,j) \in \alpha^* \times \gamma^*$, we obtain from Lemma 5.4 in paper [8], (18), and (20) that

$$\mathrm{Tr}(\widehat{\Delta X}_{\alpha^*\gamma^*}\,\widehat{\Delta S}^{\mathrm{T}}_{\alpha^*\gamma^*}) \leqslant 0,\ \mathrm{Tr}(\widehat{\Delta X}_{\beta^*\beta^*}\,\widehat{\Delta S}_{\beta^*\beta^*}) \leqslant 0 \text{ and } \mathrm{Tr}(\widehat{\Delta X}^{\mathrm{T}}_{\alpha^*\gamma^*}\,\widehat{\Delta S}_{\alpha^*\gamma^*}) \leqslant 0.$$

So, we have $\langle \widehat{\Delta X}, \widehat{\Delta S}\rangle = 0$ and $\mathrm{Tr}(\widehat{\Delta X}_{\alpha^*\gamma^*}\,\widehat{\Delta S}^{\mathrm{T}}_{\alpha^*\gamma^*}) = 0$ in view of (19). By (20) and Lemma 5.4 in paper [8], we have $\widehat{\Delta X}_{\alpha^*\gamma^*} = 0$ and $\widehat{\Delta S}_{\alpha^*\gamma^*} = 0$.

From (14) and (12), we have that

$$svec(\Delta X)^{\mathrm{T}}\boldsymbol{\Omega}^*\, svec(\Delta X) = \langle \Delta X, (D^2 f(X^*) + \sum_{r=1}^{m}\lambda_r^* D^2 g_r(X^*))(\Delta X)\rangle =$$

$$\langle \widehat{\Delta X}, \widehat{\Delta S}\rangle - \langle \mathscr{A}^*(\Delta X), \Delta\boldsymbol{\lambda}\rangle = 0.$$

It follows that $\boldsymbol{\Omega}^* svec(\Delta X) = 0$, which together with $\widehat{\Delta X}_{\alpha^*\gamma^*} = 0$, (14), (17) and Lemma 2.7 in paper [4] concludes $\Delta X = 0$.

Because of (14) and $\Delta X = 0$ we have $\Delta S \in \mathscr{N}^{\perp}$. On the other hand from $\widehat{\Delta S}_{\alpha^*\gamma^*} = 0$ and (14), we get $\Delta S \in (\mathrm{lin}(\mathscr{T}_{S^n_+}(X^*)))^{\perp}$. Then by (13), we conclude $\Delta S = 0$.

At last, since $\Delta X = 0$ and $\Delta S = 0$, by using first equation of (14) and full column rank of A_1^*, we obtain $\Delta\boldsymbol{\lambda} = 0$, it follows that all elements in $\partial_B\Theta(Y^*)$ are non-singular, which together with the result in paper [5, 9], complete the proof.

Corollary 1 Let Y^* be a solution of optimality conditions (2). If $D_X^2 L(Y^*)$ ($L(\cdot)$ is the Lagrange function of (1)) is self-adjoint positive semidefinite on S^n, then the following are equivalent:

(1) The matrix (10) is non-singular for $T^* \in \{T_0^*, T_1^*\}$.

(2) *KN*-nondegeneracy holds at Y^* and A_1^* is of full column rank.

(3) All elements in $\partial_B\Theta(Y^*)$ are non-singular.

(4) The matrix (10) is non-singular for all $T^* \in \mathscr{L}^*$.

(5) All elements in $\partial_C\Theta(Y^*)$ are nonsingular.

(6) The constraint nondegeneracy condition and the strong second order sufficient condition hold at X^*.

Proof From proof of Theorem 1, we have (1) ⇔ (2) ⇔ (3) ⇔ (4) hold since the positive semidefinity of * is equivalent to the self-adjoint positive semidefinity of $D_X^2 L(Y^*)$. Furthermore, the relations (1) (5) (6) come from [4].

From the conversional local quadratic convergence analysis of nonsmooth Newton s method, we have the following conclusion holds.

Corollary 2 Let Y^* be a solution of optimality conditions (2). If $D_X^2 L(Y^*)$ is self-adjoint positive semi-definite on S^n and one condition in Corollary 1 holds, then Algorithm 1 is locally quadratically convergent.

3 Conclusion

In this paper, we discuss a non-smooth Newton s method in [4] for nonlinear semidefinite programming by use of 4-tensor analysis. The convergence result of Algorithm 1 is also established.

[References]

[1] Sun W, Yuan Y. Optimization Theory and Methods: Nonlinear Programming[M]. New York: Springer, 2006.

[2] Wolkowicz H, Saigal R, Vandenberghe L, et al. Handbook of Semidefinite Programming[M]. Boston-Dordrecht-London: Kluwer Academic Publishers, 2000.

[3] Qi L, Sun W, Wang Y. Numerical multilinear algebra and its applications[J]. Frontiers of Mathematics in China, 2007 (2): 501-526.

[4] Li C, Sun W. An equivalent condition in convex semidefinite program[J/OL]. http://www.paper.edu.cn/downloadpaper.php?serial_number=200802-167.

[5] Qi L. Convergence analysis of some algorithms for solving nonsmooth equations[J]. Mathematics of Operations Research, 1993, 18: 227-244.

[6] Malick J, Sendov H S. Clarke generalized jacobian of the projection onto the cone of positive semidefinite matrices[J]. Set-Valued Analysis, 2006, 14: 273-293.

[7] Chan Z, Sun D. Constraint nondegeneracy, strong regularity and nonsingularity in semidefinite programming[J]. SIAM J Opt, 2008, 19: 370-396.

[8] Flegel M L, Kanzow C. A comparison of three nondegeracy conditions in semidefinite programs[J/OL]. http://www.mathematik.uni-wuerzburg.de/~kanzow/paper/DegenPropP.pdf

[9] Cai X, Sun W. A nonmonotone line search algorithm for nonsmoothy discrete minimax problem[J]. Journal of Nanjing Normal University: Natural Science, 2003, 26(4): 16-21.

[责任编辑: 丁 蓉]