A Note on Superlinear Convergence of a Primal-dual Interior Point Method for Nonlinear Semi-definite Programming

Chee-Khian Sim[1]

Department of Applied Mathematics

The Hong Kong Polytechnic University

Hung Hom, Kowloon

Hong Kong

[1]Email address: macksim@inet.polyu.edu.hk

Abstract

We replace one of the assumptions (nondegeneracy assumption) in [9] to show that the main results in [9] still hold. We also provide a simple example to show that the new assumption is satisfied, while the original assumption is not satisfied, with other assumptions being satisfied. This example shows that the new assumption does not implied the original assumption.

Keywords. Nonlinear semi-definite programming; primal-dual interior point method; superlinear convergence; MFCQ condition.

1 Introduction.

The study of superlinear convergence of a linear semi-definite program is quite an active area of research [1]-[7], while superlinear convergence of a linear program is well understood [8]. Not much work has been done so far on the local behavior of interior point method on nonlinear semi-definite programs. In [9], the authors consider a primal-dual interior point method for solving a nonlinear semi-definite program, which is stated as follows:

$$\min \quad f(x)$$

subject to

$$\begin{aligned} g(x) &= 0, \\ X(x) &\succeq 0, \\ x &\in \Re^n, \end{aligned}$$

where the functions $f : \Re^n \to \Re, g : \Re^n \to \Re^m$ and $X : \Re^n \to S^p$ are sufficiently smooth.

The authors in [9] consider interior point algorithms corresponding to the AHO, HKM, and NT search directions in linear semi-definite programs. The key to the superlinear convergence results of various algorithms in [9] is based on Theorem 1 in [9], which is established under four assumptions. We state these assumptions here for the sake of completeness[1]:

Assumptions

(A1) The second derivatives of the functions f, $g_i, i = 1, \ldots, m$, and X are Lipschitz continuous at x^*.

(A2) The second order sufficient condition for optimality of the nonlinear semi-definite problem holds at x^*. That is,

$$\sup_{(y,Z)\in\Lambda(x^*)} h^T(\nabla_x^2 L(x^*, y, Z) + \hat{H}(x^*, Z))h > 0,$$

for all $h \in C(x^*)\backslash\{0\}$.

(A3) The strict complementarity condition holds at x^*. That is, there exists $(y^*, Z^*) \in \Lambda(x^*)$ such that

$$\text{rank}(X(x^*)) + \text{ rank}(Z^*) = p.$$

[1]The various notations used in the below assumptions will be defined in Subsection 1.1

(A4) The nondegeneracy condition is satisfied at x^*. That is, the n dimensional vectors $\nabla g_i(x^*), i = 1, \ldots, m$, and $\begin{pmatrix} (A_{N1})(x^*)_{ij} \\ \vdots \\ (A_{Nn})(x^*)_{ij} \end{pmatrix}$, $i, j = 1, \ldots, |N|$ are linearly independent, where $|N|$ denotes the size of Z_N^*.

Here, x^* is a stationary point of the given nonlinear semi-definite program.

In this note, we replace one of these assumptions (Assumption (A4)), which is only used in the proof of Theorem 1 in [9], by Assumption (A4)′. Using Assumptions (A1)-(A3) and (A4′), we show that Theorem 1 of [9] holds. Hence, by the arguments in [9], which is based on Theorem 1, superlinear convergence of various algorithms in [9] hold as a result.

1.1 Notations

The space of symmetric $p \times p$ matrices is denoted by S^p. Given matrices X and Y in $\Re^{p \times q}$, the standard inner product is defined by $X \bullet Y \equiv \mathrm{Tr}(X^T Y)$, where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. If $X \in S^p$ is positive semi-definite (resp., positive definite), we write $X \succeq 0$ (resp., $X \succ 0$). The cone of positive semi-definite (resp., positive definite) symmetric matrices is denoted by S^p_+ (resp., S^p_{++}).

$\Lambda(x^*)$ denote the set of Lagrange multiplier $(y, Z) \in \Re^m \times S^p$ such that (x^*, y, Z) satisfies the KKT condition, that is, (x^*, y, Z) satisfies

$$\begin{pmatrix} \nabla_x L(x^*, y, Z) \\ g(x^*) \\ X(x^*)Z \end{pmatrix} = 0$$

and

$$X(x^*) \succeq 0, \quad Z \succeq 0.$$

$L(x, y, Z)$ is the Lagrangian function of the given nonlinear semi-definite program and is given by

$$L(x, y, Z) = f(x) - y^T g(x) - \mathrm{Tr}(X(x)Z).$$

$\hat{H}(x, Z)$ is a matrix whose (i, j) entry is

$$(\hat{H}(x, Z))_{ij} = 2\mathrm{Tr}(A_i(x)X(x)^\dagger A_j(x)Z),$$

and $\dagger$ denotes the Moore-Penrose generalized inverse.

$C(x^*)$ denotes the critical cone of the given nonlinear semi-definite program, and is defined by

$$C(x^*) = \left\{ h \ ; \ \nabla g(x^*)^T h = 0, \sum_{i=1}^{n} h_i A_i(x^*) \in T_{S_+^p}(X(x^*)), \nabla f(x^*)^T h = 0 \right\},$$

and $T_{S_+^p}(X(x^*))$ denotes the tangent cone of S_+^p at $X(x^*)$.

We have

$$A_i(x) = \frac{\partial X}{\partial x_i},$$

for $i = 1, \ldots, n$.

$\mathcal{A}^*(x)$ is the adjoint operator of $\mathcal{A}(x) : \mathcal{A}(x)v = \sum_{i=1}^{n} v_i A_i(x)$ for $v \in \Re^n$, and hence

$$\mathcal{A}^*(x)Z = \begin{pmatrix} \mathrm{Tr}(A_1(x)Z) \\ \vdots \\ \mathrm{Tr}(A_n(x)Z) \end{pmatrix}.$$

When we partition a symmetric matrix into submatrices, we do it in the following way:

$$Z = \begin{pmatrix} Z_B & Z_U \\ Z_U^T & Z_N \end{pmatrix},$$

where $Z_B \in S^{|B|}$ and $Z_N \in S^{|N|}$. For example,

$$A_i(x) = \begin{pmatrix} A_{Bi}(x) & A_{Ui}(x) \\ \\ A_{Ui}(x)^T & A_{Ni}(x) \end{pmatrix},$$

for $i = 1, \ldots, n$. Hence,

$$\mathcal{A}_N^*(x^*)(U) = \begin{pmatrix} \mathrm{Tr}(A_{N1}(x)U) \\ \\ \vdots \\ \\ \mathrm{Tr}(A_{Nn}(x)U) \end{pmatrix},$$

where $U \in S^{|N|}$.

We have $X(x^*)$ and Z^* commute, and hence they can be simultaneously diagonalized.

If strict complementarity condition holds at x^*, then we can represent $X(x^*)$ and Z^* by

$$X(x^*) = \begin{pmatrix} X_B^* & 0 \\ \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad Z^* = \begin{pmatrix} 0 & 0 \\ \\ 0 & Z_N^* \end{pmatrix},$$

where $|B| + |N| = p$.

2 Main Results.

Here, $w^* = (x^*, y^*, Z^*)$ refers to a KKT point to the nonlinear semi-definite program introduced in the earlier section.

We now introduce the new assumption that we use in this note to replace Assumption (A4) in [9].

Assumption

(A4)′ The Mangasarian-Fromovitz constraint qualification (MFCQ) condition holds at x^*, and $\mathcal{A}_N^*(x^*)(U) = 0 \Rightarrow U = 0$. Also, $\dim(\mathrm{Ker}(\nabla g(x^*)^T)) \geq 2$.

In the above assumption, $\dim(\mathrm{Ker}(\nabla g(x^*)^T)) \geq 2$ is a weak assumption and is satisfied if $m \leq n-2$. Also, Assumption (A4) implies MFCQ condition, and $\mathcal{A}_N^*(x^*)(U) = 0 \Rightarrow U = 0$.

Magasarian-Fromovitz constraint qualification (MFCQ) condition holds at a point x^* if the matrix $\nabla g(x^*)$ is of full rank and there exists a nonzero vector $v \in \Re^n$ such that

$$\nabla g(x^*)^T v = 0 \quad \text{and} \quad X(x^*) + \sum_{i=1}^{n} v_i A_i(x^*) \succ 0.$$

MFCQ condition hence implies that there exists an open neighborhood $\mathcal{O}$ of the nonzero vector $v \in \Re^n$ such that every $\bar{v} \in \mathcal{O} \cap \mathrm{Ker}(\nabla g(x^*)^T)$ satisfies $X(x^*) + \sum_{i=1}^{n} \bar{v}_i A_i(x^*) \succ 0$.

The proofs of superlinear convergence results in [9] rely mainly on Theorem 1 in [9] which states that the matrix

$$J_S(w^*) = \begin{pmatrix} \nabla_x^2 L(w^*) & -\nabla g(x^*) & -A(x^*)^T \\ \nabla g(x^*)^T & 0 & 0 \\ (Z^* \otimes_s I)A(x^*) & 0 & X(x^*) \otimes_s I \end{pmatrix},$$

where $\otimes_s$ stands for the symmetric Kronecker product operator, $w^* = (x^*, y^*, Z^*)$ and $A(x^*) = [\mathrm{svec}(A_1(x^*)); \ldots; \mathrm{svec}(A_n(x^*))] \in \Re^{p(p+1)/2 \times n}$, is nonsingular under Assumptions (A1)-(A4). Assumption (A4) is only used in the proof of this theorem in the paper. It is known that Assumption (A4) is stronger than Assumption (A4)$'$ (without the condition on dimensionality). Here, we show nonsingularity of $J_S(w^*)$ under somewhat weaker assumptions by replacing Assumption (A4) in Theorem 1 of [9] by Assumption (A4)$'$ as follows:

Theorem 2.1 *Under Assumptions (A1)-(A3) and (A4)$'$, the matrix $J_s(w^*)$ is nonsingular.*

Proof: In the proof of Theorem 1 of [9], the only place where Assumption (A4) is used is to show that

$$\nabla g(x^*)\Delta y + \mathcal{A}^*(x^*)\begin{pmatrix} 0 & 0 \\ & \\ 0 & \Delta Z_N \end{pmatrix} = 0,$$

that is,

$$\nabla g(x^*)\Delta y + \mathcal{A}_N^*(x^*)(\Delta Z_N) = 0 \tag{1}$$

implies $\Delta y = 0$ and $\Delta Z_N = 0$.

We now show this under Assumption (A4)$'$.

Note that $\Delta Z_N = \Delta Z_N^+ - \Delta Z_N^-$, where $\Delta Z_N^+, \Delta Z_N^- \in S_+^{|N|}$.

We have $\mathcal{A}_N^*(x^*)(\Delta Z_N^+) = z_1 + z_2$, where $z_1 \in \text{Range}(\nabla g(x^*))$ and $z_2 \in \text{Range}(\nabla g(x^*))^\perp =$

$\text{Ker}(\nabla g(x^*)^T)$.

Define the matrix $M := [\nabla g(x^*)\ z_2]$. Then, we have the existence of $y = \begin{pmatrix} \Delta y + \alpha \\ \\ 1 \end{pmatrix}$,

where $\nabla g(x^*)\alpha = z_1$, such that $z := My = \nabla g(x^*)\Delta y + z_1 + z_2$. Also $z = \mathcal{A}_N^*(x^*)(\Delta Z_N^-)$.

Now, $z \in \text{Range}(M) = \text{Ker}(M^T)^\perp$.

Therefore, $z^T w = 0$, for all $w \in \Re^n$ with $M^T w = 0$.

Suppose $z \neq 0$. We are going to show by contradiction that $z = 0$.

Consider

$$\max z^T w$$

subject to

$$\begin{aligned} S + \sum_{i=1}^{n} w_i A_i(x^*) &= X(x^*), \\ S &\succeq 0. \end{aligned}$$

It's dual is

$$\min X(x^*) \bullet X$$

subject to

$$\begin{aligned} A_i(x^*) \bullet X &= z_i, \quad i = 1, \ldots, n, \\ X &\succeq 0. \end{aligned}$$

Note that the optimal value to the above maximization problem is greater than or equal to zero. Also, the feasible set of the dual problem is nonempty because

$$\begin{pmatrix} 0 & 0 \\ 0 & \Delta Z_N^- \end{pmatrix} \in S_+^p$$

is a feasible point. The primal problem has a strictly feasible solution by Assumption (A4)$'$. Therefore, the optimal value to the above two problems exist and are equal to each other. It is easy to see that this common value is zero.

Observe that $z^T v > 0$ for all $v \in \Re^n$ such that $X(x^*) + \sum_{i=1}^n v_i A_i(x^*) \succ 0$. If not, then there exists $\hat{v} \in \Re^n$ such that $X(x^*) + \sum_{i=1}^n \hat{v}_i A_i(x^*) \succ 0$ with $z^T \hat{v} = 0$. Since $z \neq 0$, this implies the existence of a $\hat{\hat{v}} \in \Re^n$ such that $X(x^*) + \sum_{i=1}^n \hat{\hat{v}}_i A_i(x^*) \succ 0$ and $z^T \hat{\hat{v}} < 0$. This is a contradiction to the fact that the optimal value of the above primal-dual problem is zero.

On the other hand, we have $z^T v = 0$ for all $v \in \Re^n$ such that $M^T v = 0$. Hence, for all $v \in \mathrm{Ker}(\nabla g(x^*)^T)$, with $v^T z_2 = 0$, we have $z^T v = 0$. Note that many such v that are nonzero exist since $\dim(\mathrm{Ker}(\nabla g(x^*)^T)) \geq 2$.

Assumption (A4)$'$, in particular, the MFCQ condition, implies that for every nonzero $v \in \mathcal{O} \cap \mathrm{Ker}(\nabla g(x)^T)$, $X(x^*) + \sum_{i=1}^n v_i A_i(x^*) \succ 0$. Choose a nonzero $\bar{v} \in \mathcal{O} \cap \mathrm{Ker}(\nabla g(x)^T)$ such that $\bar{v}^T z_2 = 0$, where $z_2 \in \mathrm{Ker}(\nabla g(x)^T)$. Such $\bar{v}$ exists since by Assumption (A4)$'$, $\dim(\mathrm{Ker}(\nabla g(x^*)^T)) \geq 2$. Then, by above arguments, we have $0 = z^T \bar{v} > 0$, a contradiction.

Hence, $z = 0$, which implies that $\Delta Z_N^- = 0$, again using Assumption (A4)$'$ on $z = \mathcal{A}_N^*(x^*)(\Delta Z_N^-)$. Therefore, we have

$$\nabla g(x^*)\Delta y + \mathcal{A}_N^*(x^*)(\Delta Z_N^+) = 0,$$

where $\Delta Z_N^+ \in S_+^n$. By similar arguments as above, we then show that $\Delta Z_N^+ = 0$, which implies that $\nabla g(x^*)\Delta y = 0$. Hence, $\Delta y = 0$, since $\nabla g(x^*)$ is of full rank. **QED**

We have established in Theorem 2.1 that $J_s(w^*)$ is nonsingular under Assumptions (A1)-(A3) and (A4)$'$. Since Assumption (A4) in [9], which we have replaced by Assumption (A4)$'$, is only used in the proof of Theorem 1 in [9], establishing the theorem here under these assumptions, Assumptions (A1)-(A3), (A4)$'$, means that the superlinear convergence results in [9] also holds.

We present next an example which satisfies Assumptions (A1)-(A3) and (A4)$'$, but not Assumption (A4).

Example 2.1 *Let* $n = 3$, $m = 1$, *and* $p = 2$. *Also, we have* $|N| = 2$.

Let $f(x) = \sum_{i=1}^3 x_i^2 + 2x_2 + x_1$, $g(x) = b^T x$, *and* $X(x) = \sum_{i=1}^3 x_i A_i - B$.

Also, let

$$B = 0, \quad A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$b = \begin{pmatrix} 1 \\ -1/2 \\ 1 \end{pmatrix}.$$

We now verify that

$$x^* = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad y^* = 0, \quad Z^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

satisfied KKT conditions, Assumptions (A1)-(A3) and (A4)$'$, but do not satisfied Assumption (A4).

It is easy to verify that KKT conditions are satisfied at (x^*, y^*, Z^*).

Assumption (A1) is automatically satisfied.

Assumption (A2) is satisfied, since it is easy to check that the quadratic growth condition holds at x^**, and suppose MFCQ condition holds, Assumption (A2) holds if and only if the quadratic growth condition holds (see [9] for example).*

Assumption (A3) is also satisfied, as strict complementarity condition holds at x^*.

Assumption (A4)′ is satisfied, since dim($Ker(\nabla g(x^*)^T)) = 2$. Also,

$$\nabla g(x^*) = b = \begin{pmatrix} 1 \\ -1/2 \\ 1 \end{pmatrix},$$

is of full rank, and there exists

$$v = \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix}$$

such that $b^T v = 0$ and

$$X(x^*) + \sum_{i=1}^{3} v_i A_i(x^*) = \begin{pmatrix} 3/2 & 0 \\ 0 & 1 \end{pmatrix} \succ 0.$$

Also, it can be checked easily that

$$Tr(A_1 U) = Tr(A_2 U) = Tr(A_3 U) = 0 \quad \textit{for } U \in S^p$$

implies $U = 0$.

However, we see that

$$\begin{pmatrix} 1 \\ -1/2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

are not linearly independent. Hence, Assumption (A4) does not hold.

To ensure that the interior point algorithms in [9] can be used on this nonlinear semi-definite program, we note that there exists strictly feasible points for the problem. Take, for example,

$$x = \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix}, \quad y = 0, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

References

[1] F. Alizadeh, J.A. Haeberly and M. Overton, *Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results,* SIAM Journal on Optimization, Vol. 8(1998), pp. 746-768.

[2] M. Kojima, M. Shida and S. Shindoh, *Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs,* Mathematical Programming, Series A, Vol. 80(1998), pp. 129-160.

[3] M. Kojima, M. Shida and S. Shindoh, *A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction,* SIAM Journal on Optimization, Vol. 9(1999), pp. 444-465.

[4] F.A. Potra and R. Sheng, *Superlinear convergence of interior-point algorithms for semidefinite programming,* Journal of Optimization Theory and Applications, Vol. 99(1998), pp. 103-119.

[5] F.A. Potra and R. Sheng, *A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming,* SIAM Journal on Optimization, Vol. 8(1998), pp. 1007-1028.

[6] M. Preiß and J. Stoer, *Analysis of infeasible-interior-point paths and high-order methods for solving SDLCPs,* Parametric Optimization and Related Topics, VII. Aportaciones Matemáticas: Investigación, 18(2004), pp. 235-251.

[7] C.-K. Sim, *Superlinear convergence of an infeasible predictor-corrector path-following interior point algorithm for a SDLCP using the HKM direction,* preprint, 2010.

[8] S. J. Wright, *Primal-Dual Interior-Point Methods,* Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[9] H. Yamashita & H. Yabe, *Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming,* Mathematical Programming, Series A. Published Online: 15 May 2010.