Journal of the Operations Research Society of Japan
Vol. 58, No. 1, March 2015, pp. 24–60

A SURVEY OF NUMERICAL METHODS FOR NONLINEAR SEMIDEFINITE PROGRAMMING

Hiroshi Yamashita
NTT DATA Mathematical Systems Inc.

Hiroshi Yabe
Tokyo University of Science

(Received September 16, 2014; Revised December 22, 2014)

Abstract Nonlinear semidefinite programming (SDP) problems have received a lot of attentions because of large variety of applications. In this paper, we survey numerical methods for solving nonlinear SDP problems. Three kinds of typical numerical methods are described; augmented Lagrangian methods, sequential SDP methods and primal-dual interior point methods. We describe their typical algorithmic forms and discuss their global and local convergence properties which include rate of convergence.

Keywords: Optimization, nonlinear semidefinite programming, augmented Lagrangian method, sequential SDP method, primal-dual interior point method

1. Introduction

This paper is concerned with the nonlinear SDP problem:

$$\begin{array}{lll} \text{minimize} & f(x), & x \in \mathbf{R}^n, \\ \text{subject to} & g(x) = 0, & X(x) \succeq 0 \end{array} \tag{1.1}$$

where the functions $f : \mathbf{R}^n \to \mathbf{R}$, $g : \mathbf{R}^n \to \mathbf{R}^m$ and $X : \mathbf{R}^n \to \mathbf{S}^p$ are sufficiently smooth, and $\mathbf{S}^p$ denotes the set of pth-order real symmetric matrices. We also define $\mathbf{S}^p_+$ to denote the set of pth-order symmetric positive semidefinite matrices. By $X(x) \succeq 0$ and $X(x) \succ 0$, we mean that the matrix $X(x)$ is positive semidefinite and positive definite, respectively.

When f is a convex function, X is a concave function and g is affine, this problem is a convex optimization problem. Here the concavity of $X(x)$ means that

$$X(\lambda u + (1-\lambda)v) - \lambda X(u) - (1-\lambda)X(v) \succeq 0$$

holds for any $u, v \in \mathbf{R}^n$ and any λ satisfying $0 \le \lambda \le 1$. We note that problem (1.1) is an extension of a linear SDP problem. When all the functions f and g are linear and the matrix $X(x)$ is defined by

$$X(x) = \sum_{i=1}^{n} x_i A_i - B$$

with given matrices $A_i \in \mathbf{S}^p, i = 1, \ldots, n$, and $B \in \mathbf{S}^p$, the problem reduces to a linear SDP problem. The linear SDP problems include linear programming problems, convex quadratic programming problems and second-order cone programming problems [1] as special cases. Linear SDP model has been one of the most active research field for several decades. There are many researches on theories and applications, and polynomial-time algorithms based on interior point methods for linear SDP problems. These results can be found in survey papers by Vandenberghe and Boyd [68] and Todd [66], and in the books by Boyd, Ghaoui, Feron

and Balakrishnan [10], Wolkowicz, Saigal and Vandenberghe [70], Ben-Tal and Nemirovski [4], and Anjos and Lasserre [3], for example.

Though the linear SDP model is very useful in practical applications, it is insufficient if one wants to deal with more general problems. The nonlinear SDP problems arise from several application fields, for example, control theory (especially LMI-constrained problems and BMI-constrained problems), structural optimization, material optimization, eigenvalue problems, finance and so forth. See [17, 18, 23, 34, 40, 43, 57, 63, 69] and references therein. Thus, it is desired to develop numerical methods for solving nonlinear SDP problems. Though the interior point methods are main tools for linear SDP problems, nonlinear SDP problems (1.1) can have various algorithms. Typical studies on numerical methods include the three categories; (1) augmented Lagrangian method, (2) the sequential linear programming (SLP) method, or the sequential quadratic programming (SQP) method, and (3) interior point methods.

The nonlinear SDP contains the nonlinear second-order cone programming (SOCP). There are several researches on numerical methods for nonlinear SOCP problems, but these topics are not included in this paper. We list Kanzow, Ferenczi and Fukushima [35], Yamashita and Yabe [74] and Fukuda, Silva and Fukushima [24] for references.

The present paper is organized as follows. In Section 2, we introduce optimality conditions for problem (1.1) and fundamental notions that are used in the subsequent sections. In Section 3, we review the augmented Lagrangian method and its convergence properties. Section 4 describes sequential SDP methods from the view point of the rate of local convergence and the global convergence properties within a framework of a line search strategy, the trust region strategy and the filter method. In Section 5, we focus on primal-dual interior point methods and their local and global convergence properties. Since researches by the current authors in nonlinear SDP area have been on primal-dual interior point methods, the description of this section may be more detailed than the other methods. Finally, we give some concluding remarks in Section 6.

Notations: Throughout this paper, we define the inner product $\langle U, V\rangle$ by $\langle U, V\rangle = \mathrm{tr}(UV)$ for any matrices U and V in $\mathbf{S}^p$, where $\mathrm{tr}(M)$ denotes the trace of the matrix M. The superscript T denotes the transpose of a vector or a matrix. For $U, V \in \mathbf{S}^p$, we define the multiplication $U \circ V$ by

$$U \circ V = \frac{UV + VU}{2}. \tag{1.2}$$

We will implicitly make use of various useful relations described in [2] and Appendix of [67]. For $P \in \mathbf{R}^{p_1 \times p_2}$ and $Q \in \mathbf{R}^{q_1 \times q_2}$, the Kronecker product is defined by

$$P \otimes Q = [P_{ij}Q] \in \mathbf{R}^{p_1 q_1 \times p_2 q_2}.$$

For $U \in \mathbf{R}^{p_2 \times q_2}$, we have

$$(P \otimes Q)\mathrm{vec}(U) = \mathrm{vec}(PUQ^T),$$

where the notation $\mathrm{vec}(U)$ is defined by

$$\mathrm{vec}(U) = (U_{11}, U_{21}, \ldots, U_{p_2 1}, U_{12}, U_{22}, \ldots, U_{p_2 2}, U_{13}, \ldots, U_{p_2 q_2})^T \in \mathbf{R}^{p_2 q_2}.$$

For $U \in \mathbf{S}^p$, $P \in \mathbf{R}^{p \times p}$ and $Q \in \mathbf{R}^{p \times p}$, we define the operator

$$(P \odot Q)U = \frac{1}{2}(PUQ^T + QUP^T)$$

and the symmetrized Kronecker product

$$(P \otimes_S Q)\mathrm{svec}(U) = \mathrm{svec}((P \odot Q)U),$$

where the operator svec is defined by

$$\mathrm{svec}(U) = (U_{11}, \sqrt{2}U_{21}, \ldots, \sqrt{2}U_{p1}, U_{22}, \sqrt{2}U_{32}, \ldots, \sqrt{2}U_{p2}, U_{33}, \ldots, U_{pp})^T \in \mathbf{R}^{p(p+1)/2}.$$

We note that, for any $U, V \in \mathbf{S}^p$,

$$\langle U, V \rangle = \mathrm{tr}(UV) = \mathrm{svec}(U)^T \mathrm{svec}(V) \tag{1.3}$$

holds, and that for any $U, V \in \mathbf{R}^{p_1 \times p_2}$,

$$\mathrm{tr}(U^T V) = \mathrm{vec}(U)^T \mathrm{vec}(V). \tag{1.4}$$

In what follows, $\| \bullet \|$, $\| \bullet \|_1$ and $\| \bullet \|_\infty$ denote the l_2, l_1 and l_∞ norms for vectors, and $\| \bullet \|_F$ denotes the Frobenius norm for matrices. $\lambda_{\min}(M)$ denotes the minimum eigenvalue of a matrix $M \in \mathbf{S}^p$.

2. Optimality Conditions and Preliminaries for Analysis of Local Behavior

This section introduces optimality conditions for problem (1.1) and related quantities. We first define the Lagrangian function of problem (1.1) by

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

where $w = (x, y, Z)$, and $y \in \mathbf{R}^m$ and $Z \in \mathbf{S}^p$ are the Lagrange multiplier vector and matrix for the equality and positive semidefiniteness constraints, respectively. We also define matrices

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

for $i = 1, \ldots, n$. Then the Karush-Kuhn-Tucker (KKT) conditions for optimality of problem (1.1) are given by the following (see [11]):

$$r_0(w) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ X(x)Z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.1}$$

and

$$X(x) \succeq 0, \qquad Z \succeq 0. \tag{2.2}$$

Here $\nabla_x L(w)$ is the gradient vector of the Lagrangian function given by

$$\nabla_x L(w) = \nabla f(x) - \nabla g(x) y - \mathcal{A}^*(x) Z,$$

where $\nabla g(x)$ is defined by

$$\nabla g(x) = (\nabla g_1(x), \ldots, \nabla g_m(x)) \in \mathbf{R}^{n \times m}$$

and $\mathcal{A}^*(x)$ is the operator such that for Z,

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

We will use the norm $\|r_0(w)\|$ defined by

$$\|r_0(w)\| = \sqrt{\left\| \begin{pmatrix} \nabla_x L(w) \\ g(x) \end{pmatrix} \right\|^2 + \|X(x)Z\|_F^2}$$

in this paper.

The complementarity condition $X(x)Z = 0$ will appear in various forms in the following. We will occasionally deal with the multiplication $X(x) \circ Z$ instead of $X(x)Z$. It is known that $X(x) \circ Z = 0$ is equivalent to the relation $X(x)Z = ZX(x) = 0$ for symmetric positive semidefinite matrices $X(x)$ and Z. We also note that for symmetric positive semidefinite matrices $X(x)$ and Z, $X(x)Z = 0 \Longleftrightarrow \langle X(x), Z\rangle = 0$.

In the rest of this section, we briefly present some definitions that are necessary for the analysis of local behavior of methods surveyed below. We also describe the definitions of a stationary point, the Mangasarian-Fromovitz constraint qualification condition, the quadratic growth condition, the strict complementarity condition and the nondegeneracy condition, and the second order necessary / sufficient conditions for optimality. More comprehensive description can be found in [7]–[9, 56, 59, 60]. We recommend the paper by Shapiro and Scheinberg [60] for a good introduction to these subjects. Alternative derivation of the optimality conditions is described in Forsgren [21].

Definition 2.1. *A point $x^* \in \mathbf{R}^n$ is said to be a stationary point of problem (1.1) if there exist Lagrange multipliers (y, Z) such that (x^*, y, Z) satisfies the KKT conditions (2.1) and (2.2).*

Let $\Lambda(x^*)$ denote the set of Lagrange multipliers (y, Z) such that (x^*, y, Z) satisfies the KKT conditions. In this paper, when we refer to a point $w^* = (x^*, y^*, Z^*)$, then it means that w^* is a KKT point.

Definition 2.2. *We say that the Mangasarian-Fromovitz constraint qualification (MFCQ) condition holds at a feasible point x if the matrix $\nabla g(x)$ is of full column rank and there exists a nonzero vector $v \in \mathbf{R}^n$ such that*

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

It can be shown that, if the Mangasarian-Fromovitz condition holds at a stationary point x^*, then the set $\Lambda(x^*)$ is bounded.

The set $C(x^*)$ denote the critical cone of (1.1) at x^* that is defined as follows. Let $T_{\mathbf{S}_+^p}(X(x^*))$ denote the tangent cone of $\mathbf{S}_+^p$ at $X(x^*)$, which is defined by

$$T_{\mathbf{S}_+^p}(X(x^*)) = \{D \mid \operatorname{dist}(X(x^*) + tD, \mathbf{S}_+^p) = o(t),\ t \geq 0\},$$

where $\operatorname{dist}(P, \mathbf{S}_+^p) = \inf\{\|P - Q\|_F \mid Q \in \mathbf{S}_+^p\}$. The set $C(x^*)$, the critical cone at x^*, is defined by

$$C(x^*) = \left\{ h \in \mathbf{R}^n \ \mid \ \nabla g(x^*)^T h = 0,\ \sum_{i=1}^{n} h_i A_i(x^*) \in T_{\mathbf{S}_+^p}(X(x^*)),\ \nabla f(x^*)^T h = 0 \right\}.$$

Definition 2.3. *The second order necessary condition for local optimality of x^* under the MFCQ condition is given by*

$$\sup_{(y,Z)\in\Lambda(x^*)} h^T(\nabla_x^2 L(x^*, y, Z) + \hat{H}(x^*, Z))h \geq 0, \text{ for all } h \in C(x^*). \tag{2.3}$$

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

$$(\hat{H}(x, Z))_{ij} = 2\mathrm{tr}(A_i(x)X(x)^\dagger A_j(x)Z) \tag{2.4}$$

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

The matrix $\hat{H}(x, Z)$ contains a curvature information of the domain defined by the constraint $X(x) \succeq 0$.

Corresponding to the above necessary condition, we have the following sufficient condition.

Definition 2.4. *The second order sufficient condition for local optimality of x^* under the MFCQ condition is given by*

$$\sup_{(y,Z)\in\Lambda(x^*)} h^T(\nabla_x^2 L(x^*, y, Z) + \hat{H}(x^*, Z))h > 0, \text{ for all } h \in C(x^*)\backslash\{0\}. \tag{2.5}$$

It is said that a quadratic growth condition holds at a feasible point x^* of problem (1.1) if there exists $c > 0$ such that the following inequality holds

$$f(x) \geq f(x^*) + c\|x - x^*\|^2 \tag{2.6}$$

for any feasible point x in a neighborhood of x^*. The quadratic growth condition implies that x^* is a strict local optimal solution of problem (1.1). Suppose that the MFCQ condition holds. Then the quadratic growth condition holds if and only if the above second order sufficient conditions (2.5) are satisfied. By comparing conditions (2.3) and (2.5), we note that these conditions give a pair of "no gap" second order optimality conditions. By "no gap", we mean that the weak inequality sign in (2.3) is changed to the strict inequality sign in (2.5).

The matrix $\hat{H}(x^*, Z)$ can be written in the following form

$$\hat{H}(x^*, Z) = 2A(x^*)^T \left(Z \otimes X(x^*)^\dagger\right) A(x^*),$$

where

$$A(x) = [\mathrm{vec}(A_1(x)), \ldots, \mathrm{vec}(A_n(x))] \in \mathbf{R}^{p^2\times n}. \tag{2.7}$$

Since $X(x^*) \succeq 0$ and $Z \succeq 0$, the matrix $Z \otimes X(x^*)^\dagger$ is positive semidefinite (see Appendix of [67]), $\hat{H}(x^*, Z)$ is positive semidefinite. Then the following second order sufficient condition

$$\sup_{(y,Z)\in\Lambda(x^*)} h^T\nabla_x^2 L(x^*, y, Z)h > 0, \text{ for all } h \in C(x^*)\backslash\{0\}, \tag{2.8}$$

which will also be used in the following, is a stronger condition than (2.5). If condition (2.8) holds, condition (2.5) holds. Then the quadratic growth condition is also satisfied under the MFCQ condition.

Since $X(x^*)Z^* = 0$, the matrices $X(x^*)$ and Z^* commute. Therefore they can be simultaneously diagonalized. Thus because of the complementarity condition, we can assume without loss of generality that the matrix $X(x^*)$ and Z^* are represented by

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

respectively, where X_B^* and Z_N^* are diagonal and positive definite matrices.

Definition 2.5. *We say that the strict complementarity condition holds at x^* if there exists $(y^*, Z^*) \in \Lambda(x^*)$ such that*

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

If the strict complementarity condition holds, we have

$$X(x^*) = \begin{pmatrix} X_B^* & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad Z^* = \begin{pmatrix} 0 & 0 \\ 0 & Z_N^* \end{pmatrix}, \tag{2.9}$$

where $\mathrm{rank}(X_B^*) + \mathrm{rank}(Z_N^*) = p$. Corresponding to (2.9), we partition the matrices $X(x)$ and Z as

$$X(x) = \begin{pmatrix} X_B(x) & X_U(x) \\ X_U(x)^T & X_N(x) \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} Z_B(x) & Z_U(x) \\ Z_U(x)^T & Z_N(x) \end{pmatrix}$$

in a neighborhood of $w^* = (x^*, y^*, Z^*)$. Similarly, we partition the matrix $A_i(x)$ as

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

for $i = 1, \ldots, n$. Then the critical cone at x^* can be specifically represented by

$$C(x^*) = \left\{ h \in \mathbf{R}^n \;\middle|\; \nabla g(x^*)^T h = 0,\ \sum_{i=1}^{n} h_i A_{Ni}(x^*) = 0 \right\}.$$

Definition 2.6. *We say that the nondegeneracy condition holds at x^* if the n-dimensional vectors*

$$\nabla g_i(x^*), i = 1, \ldots, m \quad \text{and} \quad \begin{pmatrix} e_i^T A_1(x^*) e_j \\ \vdots \\ e_i^T A_n(x^*) e_j \end{pmatrix},\ i, j = 1, \ldots, p - r$$

are linearly independent, where $r = \mathrm{rank}(X(x^*))$ and the vectors $e_1, \ldots, e_{p-r}$ form a basis of the null space of the matrix $X(x^*)$.

If the strict complementarity condition holds, the nondegeneracy condition means that the vectors

$$\nabla g_i(x^*), i = 1, \ldots, m \quad \text{and} \quad \begin{pmatrix} (A_{N1}(x^*))_{ij} \\ \vdots \\ (A_{Nn}(x^*))_{ij} \end{pmatrix},\ i, j = 1, \ldots, |N|$$

are linearly independent, where $|N|$ denotes the size of Z_N^*. If the strict complementarity condition holds at x^*, then $\Lambda(x^*)$ is a singleton if and only if the nondegeneracy condition is satisfied. It is known that the nondegeneracy condition is stronger than the MFCQ condition, i.e., if the nondegeneracy condition holds at x^*, then the MFCQ condition also holds at x^*.

3. Augmented Lagrangian Method

As in ordinary nonlinear programming, it is possible to device a class of algorithms based on the augmented Lagrangian function. In this section, we survey two kinds of approaches that use the augmented Lagrangian type merit function. The first one is based on a quadratic penalty type augmentation and the second one is based on Polyak's modified barrier function. For other methods which are not reviewed here, refer to [28]–[31].

3.1. Augmented Lagrangian method based on quadratic penalty function

To solve the following nonlinear optimization problem:

$$\begin{array}{lll} \text{minimize} & f(x), & x \in \mathbf{R}^n, \\ \text{subject to} & g(x) = 0, & x \geq 0, \end{array}$$

the augmented Lagrangian by the quadratic penalty is of the form (see, for example, [5])

$$F_\mu(x, y, z) = f(x) + \frac{\mu}{2} \sum_{i=1}^{n} \left\{ \left(\max \left\{ 0, z_i - \frac{x_i}{\mu} \right\} \right)^2 - z_i^2 \right\} - y^T g(x) + \frac{1}{2\mu} \|g(x)\|^2, \quad (3.1)$$

where $\mu > 0$ is a penalty parameter. We repeat alternately an unconstrained minimization of $F_\mu(x, y_k, z_k)$ for computing next primal estimate x_{k+1} and a multiplier update for computing next dual estimates y_{k+1}, z_{k+1} for $k = 0, 1, \ldots$. Because

$$\begin{aligned} 0 &= \nabla_x F_\mu(x_{k+1}, y_k, z_k) \\ &= \nabla f(x_{k+1}) - \sum_{i=1}^{n} \max \left\{ 0, (z_k)_i - \frac{(x_{k+1})_i}{\mu} \right\} - \nabla g(x_{k+1})(y_k - g(x_{k+1})/\mu), \end{aligned}$$

if x_{k+1} is at the minimum of $F_\mu(x, y_k, z_k)$, the usual practice is to update the dual variables by

$$\begin{aligned} y_{k+1} &= y_k - g(x_{k+1})/\mu, \\ (z_{k+1})_i &= \max \left\{ 0, (z_k)_i - \frac{(x_{k+1})_i}{\mu} \right\}, i = 1, \ldots, n, \end{aligned}$$

which gives $\nabla_x L(x_{k+1}, y_{k+1}, z_{k+1}) = 0$. The update of dual variables corresponds to the maximization step of the dual function.

An extension of the above algorithm to problem (1.1) is rather straightforward, and the augmented Lagrangian is defined by

$$F_\mu(x, y, Z) = f(x) + \frac{\mu}{2} \left\{ \mathrm{tr} \left(\left| Z - \frac{1}{\mu} X(x) \right|_+^2 \right) - \mathrm{tr}(Z^2) \right\} - y^T g(x) + \frac{1}{2\mu} \|g(x)\|^2, \quad (3.2)$$

where the matrix $|A|_+$ for $A \in \mathbf{S}^p$ is defined by $|A|_+ = S\mathrm{diag}(|\lambda_1|_+, \ldots, |\lambda_p|_+)S^T$ where $A = S\mathrm{diag}(\lambda_1, \ldots, \lambda_p)S^T$ is an eigenvalue decomposition of A, and $|\lambda_i|_+ = \max\{0, \lambda_i\}, i = 1, \ldots, p$. Therefore the matrix $|A|_+$ is the projection of A onto $\mathbf{S}_+^p$. We will also use the notation $|A|$ for $A \in \mathbf{S}^p$ as the matrix $|A| = S\mathrm{diag}(|\lambda_1|, \ldots, |\lambda_p|)S^T$ later.

Since

$$\nabla_x F_\mu(x, y, Z) = \nabla f(x) - \mathcal{A}^*(x) \left| Z - \frac{1}{\mu} X(x) \right|_+ - \nabla g(x)(y - g(x)/\mu), \quad (3.3)$$

and $\nabla_x F_\mu(x_{k+1}, y_k, Z_k) = 0$ at the minimum x_{k+1} of $F_\mu(x, y_k, Z_k)$ (see, for example, [65]), similarly to the nonlinear optimization case, the next multiplier estimates are computed by

$$\begin{aligned} y_{k+1} &= y_k - g(x_{k+1})/\mu, \\ Z_{k+1} &= \left| Z_k - \frac{1}{\mu} X(x_{k+1}) \right|_+, \end{aligned}$$

yielding $\nabla_x L(x_{k+1}, y_{k+1}, Z_{k+1}) = 0$. We note that Z_{k+1} is always positive semidefinite.

From (3.3), it is easy to verify that $\nabla_x F_\mu(x^*, y^*, Z^*) = \nabla_x L(x^*, y^*, Z^*) = 0$ at a KKT point w^*.

In order to calculate the values of the merit function and its derivatives at a point x with this type of the augmented Lagrangian function, we have to calculate all the eigenvalues of the matrix $Z - \frac{1}{\mu}X(x)$. This may cause some difficulties in practical computation, especially with large scale problems.

We note that Noll, Torki and Apkarian [54] solved nonlinear LMI/BMI problems by handling the equality constraints with an ordinary Lagrangian augmentation as above, and treating the matrix inequalities by interior point methods.

3.1.1. Local properties of quadratic penalty augmented Lagrangian

In this subsection, we describe local properties of the above augmented Lagrangian near the KKT point w^*. The following results are given by Sun, Zhang and Wu [65]. In [65], only problems with inequality constraints are considered, therefore we omit the term containing $g(x)$ and y in this subsection. We note that it is a straightforward matter to include equality constraints in the following analysis. The local convexification and global saddle point condition are proved in the following theorem.

Theorem 3.1. *Assume that the functions f and X are twice continuously differentiable. Assume further that the Mangasarian-Fromovitz constraint qualification, the strict complementarity condition and second order sufficient condition (2.8) hold at w^*. Then the augmented Lagrangian function $F_\mu(x, Z)$ defined by (3.2) satisfies the following properties:*

(i) $\nabla_x F_\mu(x^*, Z^*) = 0$.

(ii) $v^T \nabla_x^2 F_\mu(x^*, Z^*) v > 0, \forall v \in C(x^*) \backslash \{0\}$ *when* $\mu > 0$. *Moreover there is a positive constant $\bar{\mu} > 0$ such that $\nabla_x^2 F_\mu(x^*, Z^*)$ is positive definite when $\mu \in (0, \bar{\mu}]$.*

(iii) *There exists a positive $\mu_0 \leq \bar{\mu}$ such that the global saddle point condition*

$$F_\mu(x^*, Z) \leq F_\mu(x^*, Z^*) \leq F_\mu(x, Z^*) \quad \text{for all } x \in \mathbf{R}^n, Z \in \mathbf{S}^p$$

holds for all positive $\mu \leq \mu_0$.

Based on these properties, various types of augmented Lagrangian methods can be derived. The following simple algorithm is presented for the analysis of local convergence properties.

Algorithm 3.1.

Step 0. Let $x_0 \in \mathbf{R}^n$, $Z_0 \succeq 0$, $\bar{\mu} > \mu > 0$ and $\varepsilon > 0$ be given. Set $k = 0$.

Step 1. (Primal update) Solve the following subproblem, and calculate its minimizer $x_{k+1} \in \mathbf{R}^n$:

$$\text{minimize } F_\mu(x, Z_k), \ x \in \mathbf{R}^n.$$

Step 2. If $Z_k X(x_{k+1}) = 0$, then stop.

Step 3. (Dual update) Calculate the next multiplier estimate by

$$Z_{k+1} = \left| Z_k - \frac{1}{\mu} X(x_{k+1}) \right|_+ .$$

Step 4. Set $k = k + 1$ and go to Step 1. □

The local convergence property of the above algorithm is given by the following theorem.

Theorem 3.2. *Assume that the functions f and X are twice continuously differentiable. Assume further that the Mangasarian-Fromovitz constraint qualification, the strict complementarity condition and second order sufficient condition (2.8) hold at w^*. Then there exist $\delta > 0, \epsilon > 0, \bar{\mu} > 0, \mu_1 > 0$,and $\mu_2 > 0$ with $\mu_1 < \mu_2 \leq \bar{\mu}$, such that for any $\mu \in [\mu_1, \mu_2]$ and for $Z \in \{Z \,|\, \|Z - Z^*\|_F \leq \delta\}$, the following statements hold:*

(i) There exists a unique vector $\hat{x} \in \mathbf{R}^n$ such that $\nabla F_\mu(\hat{x}, Z) = 0$ and

$$\hat{x} = \arg\min \{F_\mu(x, Z) \,|\, \|x - x^*\| \leq \epsilon, x \in \mathbf{R}^n\}.$$

(ii) Denote $U = Z - \frac{1}{\mu}X(\hat{x})$ and $\hat{Z} = |U|_+$. Assume that the inverse of the matrix

$$\begin{pmatrix} \nabla_x^2 F_\mu(\hat{x}, \hat{Z}) & -A(\hat{x})^T \\ -\frac{1}{2}A(\hat{x})^T - \frac{1}{2}V & -\mu p^2 I \end{pmatrix}$$

is bounded for Z in a neighborhood of Z^* and μ sufficiently small, where $A(x)$ is defined in (2.7), and

$$V = [\mathrm{vec}(V_1), \ldots, \mathrm{vec}(V_n)]$$

with

$$V_i = (|U| \odot I)^{-1}(U \odot I)A_i(\hat{x}), i = 1, \ldots, n.$$

Then, the following estimates hold:

$$\begin{aligned} \|\hat{x} - x^*\| &\leq c\mu \|Z - Z^*\|_F, \\ \left\|\hat{Z} - Z^*\right\|_F &\leq c\mu \|Z - Z^*\|_F \end{aligned}$$

where the constant c is independent of μ.

The above theorem shows the rate of convergence is linear, but its speed can be adjusted by decreasing the value of μ.

We note that some local properties of the quadratic penalty type augmented Lagrangian is studied by Shapiro and Sun [61]. Sun, Sun and Zhang [64] proved the similar linear rate of convergence without assuming the strict complementarity condition.

3.1.2. Globally convergent algorithm

It is possible to globalize the quadratic penalty augmented Lagrangian method. The essence of the modifications/assumptions of possible variations is to maintain the primal and dual sequences as bounded somehow, and let the penalty parameter $\mu_k \to 0$. We describe the algorithm proposed by Wu, Luo, Ding and Chen [71] in this subsection. We note that Luo, Wu and Chen [47] proposed other variants of globally convergent algorithms. Let

$$\sigma_k \equiv \max\left\{\mu_k \left\|\left|Z_k - \frac{1}{\mu_k}X(x_{k+1})\right|_+ - Z_k\right\|_F, \|g(x_{k+1})\|_\infty\right\}.$$

It can be shown that, if $\|\nabla F_{\mu_k}(x_{k+1}, y_k, Z_k)\| \leq \epsilon$ and $\sigma_k = 0$, then the point (x_{k+1}, y_k, Z_k) is an ϵ-approximate KKT point which satisfies

$$\begin{aligned} \|\nabla f(x_{k+1}) - \nabla g(x_{k+1})y_k - \mathcal{A}^*(x_{k+1})Z_k\| &\leq \epsilon, \\ g(x_{k+1}) &= 0, \\ X(x_{k+1})Z_k &= 0, \\ X(x_{k+1}) &\succeq 0, Z_k \succeq 0. \end{aligned}$$

Therefore searching for a point that satisfies $\|\nabla F_{\mu_k}(x, y_k, Z_k)\| \leq \epsilon$ and $\sigma_k = 0$ along with the semidefinite conditions may give a desired approximate KKT point.

Algorithm 3.2.

Step 0. Let $x_0 \in \mathbf{R}^n$, $y_0 \in \mathbf{R}^m, Z_0 \succeq 0, \mu_0 > 0, \mu_0' > 0, \epsilon > 0$ and $\theta \in (0,1)$ be given. Set $k = 0$.

Step 1. (Primal update) Find a point $x_{k+1} \in \mathbf{R}^n$ that satisfies

$$\|\nabla F_{\mu_k}(x, y_k, Z_k)\| \leq \mu_k',$$

by solving the problem

$$\text{minimize } F_{\mu_k}(x, y_k, Z_k), \ x \in \mathbf{R}^n.$$

Step 2. If $\sigma_k = 0$ and $\mu_k' \leq \epsilon$, then stop. If $\sigma_k = 0$ and $\mu_k' > \epsilon$, then set $\mu_{k+1}' = \theta\mu_k', k = k+1$, and go to Step 1. Otherwise go to Step 3.

Step 3. (Dual update) Calculate the next multiplier estimate by

$$Z_{k+1} = \left| Z_k - \frac{1}{\mu_k} X(x_{k+1}) \right|_+, \ y_{k+1} = y_k - g(x_{k+1})/\mu_k.$$

Step 4. Set $0 < \mu_{k+1} \leq \mu_k, k = k+1$ and go to Step 1. □

The following theorem shows the convergence of the above algorithm.

Theorem 3.3. *Assume that the functions f, g and X are continuously differentiable.*

(i) If the sequence generated by Algorithm 3.2 stops in Step 2 at the kth iteration, then $\{x_{k+1}, y_k, Z_k\}$ is an ϵ-approximate KKT point.

(ii) If the generated sequence is not finite, and if $\{y_k, Z_k\}$ is bounded and $\mu_k \to 0$ as $k \to \infty$, then any limit point $\bar{x}$ of $\{x_k\}$ is either infeasible, or does not satisfy the MFCQ condition, or a KKT point.

In [71], another more complicated algorithm that converges under more relaxed conditions is also described.

3.2. Augmented Lagrangian function based on modified barrier function

Representatives of this line were first done by Mosheyev and Zibulevsky [49], and later developed by Kočvara and Stingl ([38]–[40, 62]). Kočvara and Stingl developed the software PENNON [38, 41], and applied it to many practical problems. The augmented Lagrangian function (3.1) for ordinary NLP contains the penalty term $\frac{\mu}{2}\{\left(\max\left\{0, z_i - \frac{x_i}{\mu}\right\}\right)^2 - z_i^2\}$ for the constraint $x_i \geq 0$, that goes to infinity as x_i goes to $-\infty$. Polyak proposed the modified barrier function method [55] that uses the penalty/barrier term $-\mu \log(\frac{x_i}{\mu} + 1)z_i$, for example, for the constraint $x_i \geq 0$, that goes to infinity as x_i goes to $-\mu < 0$ from the above.

In the following, the outline of the algorithm by Kočvara and Stingl is described. The problem to be solved is

$$\begin{array}{lll} \text{minimize} & f(x), & x \in \mathbf{R}^n, \\ \text{subject to} & X(x) \succeq 0. & \end{array} \tag{3.4}$$

Here the equality constraints are omitted for the sake of simplicity of discussion. Kočvara and Stingl proposed two approaches for handling equality constraints in their implementation. One is that an equality constraint $g_i(x) = 0$ is converted to two inequality constraints $g_i(x) \leq 0$ and $g_i(x) \geq 0$, and the other is that $g_i(x) = 0$ is converted to the inequality constraints $-\varepsilon \leq g_i(x) \leq \varepsilon$, where ε is a small positive constant. They claimed that these two approaches are equally efficient.

In this section, we describe various properties of this type following the contents in [62].

3.2.1. Properties of modified barrier augmented Lagrangian function

In this subsection, we assume that the KKT point x^* satisfies

$$x^* \in \arg\min\{f(x) \,|x \in \Omega\},$$

where $\Omega = \{x \in \mathbf{R}^n \,|X(x) \succeq 0\}$, the strict complementarity condition, the nondegeneracy condition, and the second order sufficient condition (2.5), and that

$$\exists \pi > 0 \text{ and } \tau > 0 \text{ such that } \max\{\|X(x)\|_F \,|x \in \Omega_\pi\} \leq \tau, \tag{3.5}$$

with

$$\Omega_\mu = \{x \in \mathbf{R}^n \,|X(x) \succeq -b\mu I\},$$

where b is a positive constant and μ is a positive number. Condition (3.5) is called a growth condition.

We introduce a penalty/barrier matrix function $\Phi_\mu : \mathbf{S}^p \to \mathbf{S}^p$ that satisfies a number of constraints described below. Let $X(x) = S(x)\mathrm{diag}(\lambda_1(x), \ldots, \lambda_p(x))S(x)^T$ be an eigenvalue decomposition of $X(x)$. Then the primary matrix function Φ_μ is defined by

$$\Phi_\mu(X(x)) = S(x) \begin{pmatrix} \varphi_\mu(\lambda_1(x)) & 0 & \cdots & 0 \\ 0 & \varphi_\mu(\lambda_2(x)) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \varphi_\mu(\lambda_p(x)) \end{pmatrix} S(x)^T,$$

where $\varphi_\mu : \mathbf{R} \to \mathbf{R}$ characterizes the above matrix function and specific choice of φ_μ given below assures

$$X(x) \succeq 0 \Longleftrightarrow \Phi_\mu(X(x)) \succeq 0,$$

in particular, for any value of penalty parameter $\mu > 0$.

The augmented Lagrangian of (3.4) is defined by

$$F_\mu(x, Z) = f(x) - \langle \Phi_\mu(X(x)), Z \rangle \,. \tag{3.6}$$

In order for the augmented Lagrangian (3.6) to have desired properties, we require the function φ to satisfy the following properties, and let the above φ_μ be defined by $\varphi_\mu(t) = -\mu\varphi(-t/\mu)$:

Definition 3.1. *Let $\varphi : (-\infty, b) \to \mathbf{R}$, where $b \in (0, \infty]$ is a given number, be a function with the following properties.*

(φ_0) *φ is strictly convex, strictly monotone increasing and twice continuously differentiable,*
(φ_1) *$\varphi(0) = 0$,*
(φ_2) *$\varphi'(0) = 1$,*
(φ_3) *$\exists c_1$ such that $\varphi'(t) \geq c_1$ for any $t < 0$,*
(φ_4) *$\exists c_2$ such that $\varphi'(\sigma/\mu) \leq c_2\mu$ for any $\sigma < 0$ and $\mu > 0$,*
(φ_5) *$\exists c_3$ such that $\varphi''(\rho\sigma) \leq c_3\mu^2$ for any $\sigma < 0$ and $\mu > 0$,*
(φ_6) *φ' is convex, and $\lim_{t\to b} \varphi' = \infty, \lim_{t\to-\infty} \varphi' = 0$.*

Well known examples of the function φ are:

- The logarithmic penalty function

$$\varphi_{\log}(t) = -\log(1 - t).$$

- The hyperbolic penalty function

$$\varphi_{\rm hyp}(t) = \frac{1}{1-t} - 1.$$

These penalty functions are introduced by Polyak [55], and he called them the modified barrier functions.

Let $t_1, \ldots, t_p$ be real values, and define

$$\begin{aligned}
\Delta\varphi_\mu(t_i, t_j) &= \begin{cases} \frac{\varphi_\mu(t_i)-\varphi_\mu(t_j)}{t_i-t_j}, & \text{for } i \neq j, \\ \varphi'_\mu(t_i), & \text{for } i = j, \end{cases} \\
\Delta^2\varphi_\mu(t_i, t_j, t_k) &= \begin{cases} \frac{\Delta\varphi_\mu(t_i,t_k)-\Delta\varphi_\mu(t_j,t_k)}{t_i-t_j}, & \text{for } i \neq j, \\ \frac{\Delta\varphi_\mu(t_i,t_j)-\Delta\varphi_\mu(t_k,t_j)}{t_i-t_k}, & \text{for } i = j \neq k, \\ \varphi''_\mu(t_i), & \text{for } i = j = k, \end{cases}
\end{aligned}$$

where φ'_μ and φ''_μ are the first and the second derivatives of φ_μ, respectively. Further, let $\lambda'_1(x) < \lambda'_2(x) < \ldots < \lambda'_{p'(x)}(x)$ be distinct eigenvalues of $X(x)$, and let $X(x) = S(x)\mathrm{diag}(\lambda'_1(x), \ldots, \lambda'_1(x), \ldots, \lambda'_{p'(x)}(x), \ldots, \lambda'_{p'(x)}(x))S(x)^T$ be an eigenvalue decomposition of $X(x)$ where each eigenvalue occurs in its multiplicity at x. The Frobenius covariant matrices $P_i(x) \in \mathbf{S}^p, i = 1, \ldots, p'(x)$ of $X(x)$ is defined by

$$P_i(x) = S(x)\mathrm{diag}(0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)S(x)^T, i = 1, \ldots, p'(x),$$

where the nonzeros in the diagonal matrix occur exactly in the positions of $\lambda'_i(x)$ in $\mathrm{diag}(\lambda'_1(x), \ldots, \lambda'_{p'(x)}(x)) \in \mathbf{S}^p$.

With the above definition of the augmented Lagrangian, we can prove the following theorem that shows the desired properties at the KKT point w^*.

Theorem 3.4. *The augmented Lagrangian function $F_\mu(x, Z)$ defined by (3.6) satisfies the following properties:*

(i) $F_\mu(x^*, Z^*) = f(x^*)$.
(ii) $\nabla_x F_\mu(x^*, Z^*) = \nabla f(x^*) - \mathcal{A}^*(x^*)Z^* = 0$.
(iii) $\nabla^2_x F_\mu(x^*, Z^*) = \nabla^2_x L(x^*, Z^*) + H_\mu(x^*, Z^*) + M/\mu$, *where*

$$\begin{aligned}
H_\mu(x^*, Z^*) &= 2\left[\left\langle Z^*, A_i(x^*)\left(\sum_{k=1}^{p'(x^*)-1} \Delta^2\varphi_\mu(t_k, 0, 0)P_k(x^*)\right)A_j(x^*)\right\rangle\right]_{i,j=1}^n \in \mathbf{S}^n, \\
M &= 2\left[\left\langle Z^*, A_i(x^*)\varphi''_\mu(0)P_{p'(x^*)}(x^*)A_j(x^*)\right\rangle\right]_{i,j=1}^n \in \mathbf{S}^n,
\end{aligned}$$

and $\lim_{\mu\to 0} H_\mu(x^*, Z^*) = \hat{H}(x^*, Z^*)$ *(see (2.4)),* $\mathrm{Ker}M = C(x^*)$ *and* $v^T M v > 0, \forall v \notin C(x^*)$.

If f is convex and X is concave, then

(iv) *$F_\mu(x, Z)$ is convex in x for all $x \in \Omega_\mu$.*

From (iii) of the above theorem, we see that the matrix $\nabla^2_x F_\mu(x^*, Z^*)$ is positive definite when $\mu > 0$ is sufficiently small. The following theorem shows properties of the augmented Lagrangian in a neighborhood of x^*.

Theorem 3.5. *The augmented Lagrangian function $F_\mu(x, Z)$ satisfies the following properties:*

(i) *There exists $\bar{\mu} > 0$ such that $F_\mu(x, Z^*)$ is strongly convex for all $\mu \leq \bar{\mu}$ in a neighborhood of x^*.*

(ii) *There exist $\varepsilon > 0$ and $\bar{\mu} > 0$ such that, for any $\mu < \bar{\mu}$,*

$$x^* = \arg\min \{F_\mu(x, Z^*) \,|\, \|x - x^*\| \leq \varepsilon\}.$$

Moreover, if f is convex and X is concave, then for any $\mu > 0$,

$$x^* = \arg\min \{F_\mu(x, Z^*) \,|\, x \in \mathbf{R}^n\}.$$

3.2.2. Algorithm and its local convergence property

As usual, we perform the unconstrained minimization of $F_\mu(x, Z)$ with fixed multiplier estimate Z and current penalty parameter μ. Then we update the multiplier matrix Z and decrease the penalty parameter μ if necessary, and repeat this primal and dual iteration until some termination criterion is satisfied. This procedure will be described more precisely below.

By using the matrix function chain rule, it is possible to write $\left\langle \frac{\partial \Phi_\mu(X(x))}{\partial x_i}, Z \right\rangle$ as a trace of a product matrix of $A_i(x)$ and some matrix. We write this latter matrix as $D\Phi_\mu(X(x))[Z]$, and interpret it as a directional derivative of $\Phi_\mu(X(x))$ with respect to $X(x)$ along the direction Z. More precisely, we have (see (6.6.25) of [27])

$$\frac{\partial \Phi_\mu(X(x))}{\partial x_i} = \sum_{j,k=1}^{p'(x)} \Delta\varphi_\mu(\lambda'_j(x), \lambda'_k(x)) P_j(x) A_i(x) P_k(x),$$

then we have

$$\begin{aligned}
\frac{\partial F_\mu(x, Z)}{\partial x_i} &= \frac{\partial f(x)}{\partial x_i} - \left\langle \frac{\partial \Phi_\mu(X(x))}{\partial x_i}, Z \right\rangle \\
&= \frac{\partial f(x)}{\partial x_i} - \sum_{j,k=1}^{p'(x)} \left\langle \Delta\varphi_\mu(\lambda'_j(x), \lambda'_k(x)) P_j(x) A_i(x) P_k(x), Z \right\rangle \\
&= \frac{\partial f(x)}{\partial x_i} - \sum_{j,k=1}^{p'(x)} \left\langle A_i(x), \Delta\varphi_\mu(\lambda'_j(x), \lambda'_k(x)) P_k(x) Z P_j(x) \right\rangle \\
&= \frac{\partial f(x)}{\partial x_i} - \langle A_i(x), D\Phi_\mu(X(x))[Z] \rangle,
\end{aligned}$$

where

$$D\Phi_\mu(X(x))[Z] = \sum_{j,k=1}^{p'(x)} \Delta\varphi_\mu(\lambda'_j(x), \lambda'_k(x)) P_j(x) Z P_k(x).$$

The formal statement of the algorithm is as follows:

Algorithm 3.3.

Step 0. Let $x_0 \in \mathbf{R}^n$, $Z_0 \succ 0$, $\mu_0 > 0$ and $\varepsilon > 0$ be given. Set $k = 0$.

Step 1. If

$$\|r_0(w_k)\| \leq \varepsilon, X(x_k) \succeq 0, Z_k \succeq 0,$$

then stop.

Step 2. (Primal update) Solve the following subproblem, and calculate the minimizer $x_{k+1} \in \mathbf{R}^n$:

$$\text{minimize } F_{\mu_k}(x, Z_k),\ x \in \mathbf{R}^n.$$

Step 3. (Dual update) Calculate the next multiplier estimate by

$$Z_{k+1} = D\Phi_{\mu_k}(X(x_{k+1}))\,[Z_k]\,.$$

Step 4. Set $\mu_{k+1} \leq \mu_k$, $k = k+1$ and go to Step 1. □

By using the above update formula for Z_k, we have

$$\nabla_x F_\mu(x_{k+1}, Z_k) = \nabla_x L(x_{k+1}, Z_{k+1}) = 0,$$

as desired. It is also possible to show that if $Z_k \succ 0$, then $Z_{k+1} \succ 0$ with the above update formula (see Lemma 7.9 in [62]).

Local convergence properties of the above algorithm is given below. Let $|N|$ be the size of the matrix Z_N^*. For ϵ and Θ $(0 < \epsilon < (Z_N^*)_{ii} < \Theta, i = 1, \ldots, |N|)$, define the set $V = V(Z^*, \bar{\mu}, \delta, \epsilon, \Theta)$ by

$$\begin{aligned} V \;=\; & \{(Z,\mu) \in \mathbf{S}_+^p \times \mathbf{R} \,|\, \|Z - Z^*\|_F \leq \delta/\mu, \mu < \bar{\mu}\} \cap \\ & \{(Z,\mu) \in \mathbf{S}_+^p \times \mathbf{R} \,|\, \|Z\|_F \leq \Theta\} \cap \\ & \{(Z,\mu) \in \mathbf{S}_+^p \times \mathbf{R} \,|\, (Z_N)_{ii} \geq \epsilon, i = 1, \ldots, |N|\}\,. \end{aligned}$$

Theorem 3.6. *There exist a penalty parameter $\bar{\mu} > 0$ and a small enough number $\delta > 0$ such that for any $(Z, \mu) \in V$,*

(i) *There exists a vector*

$$\hat{x} = \arg\min\{F_\mu(x, Z) \,|\, x \in \mathbf{R}^n\}$$

such that $\nabla_x F_\mu(\hat{x}, Z) = 0$.

(ii) *For the pair $\hat{x}$ and $\hat{Z} = D\Phi_\mu(X(\hat{x}))\,[Z]$, the estimate*

$$\max\left\{\|\hat{x} - x^*\|, \left\|\hat{Z} - Z^*\right\|_F\right\} \leq \mu c \|Z - Z^*\|_F$$

holds, where c is a constant independent of μ.

(iii) *$x^* = \arg\min\{F_\mu(x, Z^*) \,|\, x \in \mathbf{R}^n\}$ and $Z^* = D\Phi_\mu(X(x^*))\,[Z^*]$.*

(iv) *The function $F_\mu(x, Z)$ is strongly convex with respect to x in a neighborhood of $\hat{x}$.*

From the above theorem, the following local convergence result is obtained.

Theorem 3.7. *For $\mu_0\,(< \bar{\mu})$ small enough and $(Z_0, \mu_0) \in V$, Algorithm 3.3 converges to w^* with a linear rate. If $\mu_{k+1} < \mu_k$ for all $k \geq 0$, and $\lim_{k\to\infty} \mu_k = 0$, then the rate of convergence is superlinear.*

3.2.3. Globally convergent algorithm

In [62], a globally convergent augmented Lagrangian algorithm was proposed and its convergence property was proved.

Algorithm 3.4.

Step 0. Let $x_0 \in \mathbf{R}^n$, $Z_0 \succ 0$, $\mu_0 > 0, \mu_0' > 0$ and $\varepsilon > 0$ be given. Set $k = 0$.

Step 1. If

$$\|r_0(w_k)\| \leq \varepsilon, X(x_k) \succeq 0, Z_k \succeq 0,$$

then stop.

Step 2. (Primal update) Find a point $x_{k+1} \in \mathbf{R}^n$ that satisfies

$$\|\nabla F_{\mu_k}(x, Z_k)\| \leq \mu'_k,$$

by minimizing $F_{\mu_k}(x, Z_k)$.

Step 3. (Dual update) Calculate the next multiplier estimate by

$$Z_{k+1} = D\Phi_{\mu_k}(X(x_{k+1}))\,[Z_k]\,.$$

Step 4. Set $\mu_{k+1} < \mu_k, \mu'_{k+1} < \mu', k = k + 1$ and go to Step 1. □

The global convergence of the above algorithm is shown by the following theorem. As noted in the previous subsection, this theorem also assumes the boundedness of the sequence and $\mu_k \to 0$.

Theorem 3.8. *Assume that there exists $\bar{\mu} > 0$ such that Ω_μ is a compact set for all $\mu \leq \bar{\mu}$, the nondegeneracy condition holds at any limit point of the sequence $\{x_k\}$, and the sequence $\{Z_k\}$ is bounded. Assume further that $\mu_k \to 0$ and $\mu'_k \to 0$. Then the limit point of any convergent subsequence of $\{x_k, Z_k\}$ satisfies the KKT conditions.*

3.2.4. Actual form of penalty function

In order to have computable forms of multiplier update and derivatives, Kočvara and Stingl used the following penalty function:

$$\Phi_\mu(X(x)) = -\mu^2(X(x) + \mu I)^{-1} + \mu I.$$

Let

$$U(x) = (X(x) + \mu I)^{-1}.$$

Since

$$\begin{aligned} 0 &= \frac{\partial}{\partial x_i}\left[U(x)U(x)^{-1}\right] = \frac{\partial}{\partial x_i}\left[U(x)\right]U(x)^{-1} + U(x)\frac{\partial}{\partial x_i}\left[U(x)^{-1}\right] \\ &= \frac{\partial}{\partial x_i}\left[U(x)\right]U(x)^{-1} + U(x)A_i(x), \end{aligned}$$

we have

$$\frac{\partial \Phi_\mu(X(x))}{\partial x_i} = \mu^2 U(x)A_i(x)U(x)$$

for the first derivative of $\Phi_\mu(X(x))$. In order to have

$$\left\langle \frac{\partial \Phi_\mu(X(x_k))}{\partial x_i}, Z_k \right\rangle = \left\langle A_i(x_k), Z_{k+1} \right\rangle,$$

we set

$$Z_{k+1} = \mu_k^2 U(x_{k+1}) Z_k U(x_{k+1})$$

for multiplier update. We note that when Z_k is positive definite, then Z_{k+1} is also positive definite.

Noll [53] also gave a local convergence property similar to the result given in Section 3.2.2 using this particular form of penalty function.

4. Sequential SDP Method

The sequential SDP (S-SDP) method was proposed by using the idea of the SQP (sequential quadratic programming) method [6, 52] for solving nonlinear optimization problems. Such an idea has already been mentioned by Robinson [58] within the more general framework of nonlinear programming problems over convex cones. Given a current point x_k, the S-SDP method generates a search direction Δx_k by solving the quadratic subproblem:

$$\begin{array}{ll} \text{minimize} & \nabla f(x_k)^T \Delta x + \frac{1}{2}\Delta x^T G_k \Delta x, \ \Delta x \in \mathbf{R}^n \\ \text{subject to} & g(x_k) + \nabla g(x_k)^T \Delta x = 0, \\ & \hat{X}_k(\Delta x) \equiv X(x_k) + \sum_{i=1}^n \Delta x_i A_i(x_k) \succeq 0, \end{array} \tag{4.1}$$

where G_k is the Hessian matrix $\nabla_x^2 L(x_k, y_k, Z_k)$ of the Lagrangian function or its approximation. Letting $(\Delta x_k, y_{k+1}, Z_{k+1})$ be a KKT point for minimization problem (4.1), we have the following:

$$\begin{aligned} \nabla f(x_k) + G_k \Delta x_k - \nabla g(x_k) y_{k+1} - \mathcal{A}^*(x_k) Z_{k+1} &= 0, \\ g(x_k) + \nabla g(x_k)^T \Delta x_k &= 0, \\ \hat{X}_k(\Delta x_k) &\succeq 0, \\ Z_{k+1} &\succeq 0, \\ \hat{X}_k(\Delta x_k) Z_{k+1} &= 0, \end{aligned}$$

where y_{k+1} and Z_{k+1} are multipliers corresponding to the equality and positive semidefiniteness constraints, respectively. We note that when $\Delta x_k = 0$, the above conditions become the KKT conditions (2.1) and (2.2) of problem (1.1). As noted in [23], if the matrix G_k is positive definite, then problem (4.1) reduces to a linear SDP, because the convex quadratic term in the objective function can be written as a semidefiniteness constraint or a second-order cone constraint.

Fares, Noll and Apkarian [18] proposed a local S-SDP method. Later Freund, Jarre and Vogelbusch [23] also analyzed the local rate of convergence of the S-SDP method from a different viewpoint. The global convergence of the S-SDP method was shown by Correa and Ramírez [15] under the line search strategy. As another global convergent method, Gómez and Ramírez [25] applied a filter method to the S-SDP method. These methods will be briefly reviewed in this section. Furthermore, other related methods were studied by Jarre [33], Kanzow, Nagel, Kato and Fukushima [36], Diehl, Jarre and Vogelbusch [16], Kanno and Takewaki [34], Li and Sun [46] and Zhu and Zhu [78], for example.

4.1. Local convergence properties of S-SDP methods

Fares, Noll and Apkarian [18] originally dealt with the following minimization problem of a linear objective function subject to linear matrix inequality (LMI) constraints and nonlinear matrix equality constraints, which arises from robust control design:

$$\begin{array}{lll} \text{minimize} & c^T x, & x \in \mathbf{R}^n, \\ \text{subject to} & \mathcal{L}(x) \succeq 0, & \mathcal{B}(x) = 0, \end{array} \tag{4.2}$$

where c is a given vector, $\mathcal{L}(x) : \mathbf{R}^n \to \mathbf{S}^p$ is an affine symmetric matrix function, and $\mathcal{B}(x)$ is a nonlinear matrix valued function. The matrix function $\mathcal{B}(x)$ is bilinear with respect to x in many cases. Fares et. al. [18] solved the problem via the S-SDP method. The following local convergence theorem was shown for solving problem (1.1) (See Theorem 11 of [15]).

Theorem 4.1. *Let (x^*, y^*, Z^*) be a point satisfying the KKT conditions of problem (1.1). Assume that (x^*, y^*, Z^*) satisfies*

$$v^T \nabla_x^2 L(x^*, y^*, Z^*) v > 0 \tag{4.3}$$

for all nonzero vectors $v \in C(x^*)$, where $C(x^*)$ denotes the critical cone of (1.1) at x^*. Suppose that the matrix $((A_1(x^*), \ldots, A_n(x^*))^T, \nabla g(x^*))^T$ has full rank and that G_k approaches $\nabla_x^2 L(x^*, y^*, Z^*)$. Then there exists $\delta > 0$ such that, if $\|x_0 - x^*\| < \delta$, $\sqrt{\|y_0 - y^*\|^2 + \|Z_0 - Z^*\|_F^2} < \delta$ and $\|G_k - \nabla_x^2 L(x^*, y^*, Z^*)\| < \delta$ for all k, the sequence (x_k, y_k, Z_k) generated by the S-SDP method with $x_{k+1} = x_k + \Delta x_k$ is well defined and converges superlinearly to (x^*, y^*, Z^*). Furthermore, the convergence is quadratic if the following holds

$$G_k - \nabla_x^2 L(x^*, y^*, Z^*) = O(\|x_k - x^*\| + \sqrt{\|y_k - y^*\|^2 + \|Z_k - Z^*\|_F^2}).$$

We note that assumption (4.3) is stronger than (2.3).

Freund, Jarre and Vogelbusch [23] chose a motivation that contrasts the S-SDP method with primal-dual interior-point methods and analyzed the local and quadratic convergence based on a sensitivity result for nonlinear SDP problems. They dealt with the following nonlinear SDP problem:

$$\begin{array}{lll} \text{minimize} & c^T x, & x \in \mathbf{R}^n, \\ \text{subject to} & X(x) \succeq 0 & \end{array} \tag{4.4}$$

where $c \in \mathbf{R}^n$ is a given vector. In addition, they extended this result to a more general class of nonlinear SDP problems with equality and inequality constraints.

4.2. Globally convergent algorithms

In order to establish the global convergence of the S-SDP method, Correa and Ramírez [15] introduced the Han penalty function as a merit function of the line search procedure. For a penalty parameter $\sigma > 0$, the merit function is defined by

$$\theta_\sigma(x) = f(x) - \sigma(\min\{0, \lambda_{min}(X(x))\} + \|g(x)\|). \tag{4.5}$$

Then they show that the directional derivative $\theta'_\sigma(x_k; \Delta x_k)$ of the function $\theta_\sigma(x)$ along a search direction Δx_k at x_k satisfies the following estimate.

Lemma 4.1. *Assume that the functions f, g and X are continuously differentiable and that their derivatives are Lipschitz continuous in a neighborhood of x_k. Using the penalty function $\theta_\sigma(x)$ in (4.5), if the point $(\Delta x_k, y_{k+1}, Z_{k+1})$ satisfies the KKT conditions of subproblem (4.1), then the following holds*

$$\begin{aligned} \theta'_\sigma(x_k; \Delta x_k) &\leq \nabla f(x_k)^T \Delta x_k + \sigma(\min\{0, \lambda_{min}(X(x_k))\} - \|g(x_k)\|) \\ &= -\Delta x_k^T G_k \Delta x_k - \mathrm{tr}(Z_{k+1} X(x_k)) - y_{k+1}^T g(x_k) \\ &\quad + \sigma(\min\{0, \lambda_{min}(X(x_k))\} - \|g(x_k)\|). \end{aligned} \tag{4.6}$$

Furthermore, if $\sigma \geq \max\{\mathrm{tr}(Z_{k+1}), \|y_{k+1}\|\}$, we have

$$\theta'_\sigma(x_k; \Delta x_k) \leq -\Delta x_k^T G_k \Delta x_k.$$

This result shows that the search direction becomes a descent direction for $\theta_\sigma(x)$, i.e. $\theta'_\sigma(x_k; \Delta x_k) < 0$, when G_k is positive definite and σ is sufficiently large.

The algorithm proposed by Correa and Ramírez [15] is described for a given current point x_k and a positive definite matrix G_k as follows:

Algorithm 4.1.

Step 0. Give $x_1 \in \mathbf{R}^n$ and $\sigma_0 > 0$. Set $k = 1$.

Step 1. Compute a point $(\Delta x_k, y_{k+1}, Z_{k+1})$ satisfying the KKT conditions of subproblem (4.1).

Step 2. Compute σ_k in a way that the following properties are satisfied:
(a) $\sigma_k \geq \max\{\mathrm{tr}(Z_{k+1}), \|y_{k+1}\|\} + \bar{\sigma}$.
(b) For all $k \geq k_1$, if $\sigma_{k-1} \geq \max\{\mathrm{tr}(Z_{k+1}), \|y_{k+1}\|\} + \bar{\sigma}$, then $\sigma_k = \sigma_{k-1}$.
where a positive integer k_1 and $\bar{\sigma} > 0$ are fixed parameters.
(Remark: The parameter σ_k is updated by $\sigma_k = \max\{1.5\sigma_{k-1},\ \max\{\mathrm{tr}(Z_{k+1}), \|y_{k+1}\|\} + \bar{\sigma}\}$, for example.)

Step 3. The step size α_k is computed by using the Armijo rule, that is, α_k is set to $\alpha_k = \beta^{l_k}$ for the smallest nonnegative integer l_k that satisfies the sufficient decrease condition

$$\theta_{\sigma_k}(x_k + \beta^{l_k}\Delta x_k) \leq \theta_{\sigma_k}(x_k) + \varepsilon_0 \beta^{l_k} \tilde{\Delta}_k,$$

where $0 < \varepsilon_0 < 1$ and $0 < \beta < 1$ are given, and $\tilde{\Delta}_k$ is the upper bound of $\theta'_{\sigma_k}(x_k; \Delta x_k)$ given in (4.6).

Step 4. Set $x_{k+1} = x_k + \alpha_k \Delta x_k$.

Step 5. Set $k := k + 1$ and go to Step 1. □

The global convergence property is shown by the following theorem.

Theorem 4.2. *Assume that the functions f, g and X are continuously differentiable and that their derivatives are Lipschitz continuous. Consider the global algorithm described in Algorithm 4.1. Suppose that G_k is a positive definite matrix for all k. If the sequence $\{G_k\}$ is bounded together with the sequence $\{G_k^{-1}\}$, then one of the following situations occurs for the sequence $\{(x_k, y_{k+1}, Z_{k+1})\}$:*

1. *The sequences $\{\sigma_k\}$ and $\{(y_{k+1}, Z_{k+1})\}$ are unbounded.*

2. *There is an index k_2 such that σ_k is constant for all $k \geq k_2$. In this case, one of the following situations occurs:*
(a) $\theta_{\sigma_k}(x_k) \to -\infty$,
or
(b) $\nabla_x L(x_k, y_{k+1}, Z_{k+1}) \to 0,\ g(x_k) \to 0,\ \min\{0, \lambda_{min}(X(x_k))\} \to 0$,
and $\mathrm{tr}(Z_{k+1}X(x_k)) \to 0$.

As another globally convergent method, Gómez and Ramírez [25] proposed a filter method for solving problem (1.1). The filter methods were first introduced by Fletcher and Leyffer [19] for nonlinear optimization problems. In general, filter methods deal simultaneously with the optimality and feasibility. In the methods, the trial points are accepted when they improve either the objective function or a constraint violation measure. These criteria are compared to previous iterates collected in a filter. Global convergence results of several kinds of filter methods have been studied for nonlinear optimization problems (see [20], for example). Given the current iterate x_k and matrix G_k, Gómez and Ramírez defined the following trust-region local semidefinite approximation of problem (1.1):

$$\begin{array}{ll} \text{minimize} & \nabla f(x_k)^T \Delta x + \frac{1}{2}\Delta x^T G_k \Delta x,\ \Delta x \in \mathbf{R}^n \\ \text{subject to} & g(x_k) + \nabla g(x_k)^T \Delta x = 0, \\ & X(x_k) + \sum_{i=1}^n \Delta x_i A_i(x_k) \succeq 0, \\ & \|\Delta x\|_\infty \leq \Delta_k, \end{array} \tag{4.7}$$

where Δ_k is a trust-region radius. In order to investigate the feasibility, let us define the constraint violation measure by

$$\theta(x) = \|g(x)\| - \min\{0, \lambda_{min}(X(x))\}. \tag{4.8}$$

A filter, denoted by $\mathcal{F}$, is a finite collection of two dimensional vectors. In each vector of the filter, the two components refer to the values of the feasibility function θ and the objective function f, respectively. The new candidate $(\bar{\theta}, \bar{f})$ is called acceptable to the filter $\mathcal{F} = \{(\theta_j, f_j)\}_{j=1}^N$, if, for each $j = 1, \ldots, N$, at least one of the following two conditions is satisfied:

$$\begin{aligned} \bar{\theta} &\leq \beta\theta_j, \\ \bar{f} + \gamma\bar{\theta} &\leq f_j, \end{aligned}$$

where $\beta \in (0,1)$ and $\gamma \in (0, \beta)$ are fixed parameters. If the point $(\bar{\theta}, \bar{f})$ is acceptable for the current filter $\mathcal{F}$, the new filter $Add((\bar{\theta}, \bar{f}), \mathcal{F})$ is defined as

$$Add((\bar{\theta}, \bar{f}), \mathcal{F}) = (\mathcal{F} \cup \{(\bar{\theta}, \bar{f})\}) \backslash \{(\theta_j, f_j) \in \mathcal{F} \mid \bar{\theta} \leq \theta_j, \bar{f} \leq f_j\}.$$

The following is a filter-type algorithm of the S-SDP method.

Algorithm 4.2.

Step 0. Set the parameters $\beta \in (0,1), \gamma \in (0, \beta), \Delta^{max} > \Delta^{min} > 0, \sigma \in (0,1), u > 0$ and $k = 1$. Define $\mathcal{F}^0 = \{(u, -\infty)\}$.

Step 1. (Restoration phase) Find a point x_k and a corresponding trust-region radius $\Delta^{max} \geq \Delta_k \geq \Delta^{min}$ such that
(a) $(\theta(x_k), f(x_k))$ is acceptable to $\mathcal{F}^{k-1}$.
(b) Subproblem (4.7) is feasible.

Step 2. (Solving subproblem (4.7))

Step 2.1 Solve subproblem (4.7) to obtain a step Δx.
If subproblem (4.7) is not feasible, then set $\mathcal{F}^k = Add((\theta(x_k), f(x_k)), \mathcal{F}^{k-1})$, $k := k+1$ and go to Step 1.
If $\Delta x = 0$ is obtained, then x_k is a stationary point of problem (1.1) and stop.

Step 2.2 If $(\theta(x_k+\Delta x), f(x_k+\Delta x))$ is not acceptable to the filter $\mathcal{F}^{k-1} \cup \{(\theta(x_k), f(x_k))\}$, then set $\Delta := \Delta/2$ and go to Step 2.1.

Step 2.3 If the following conditions are fulfilled

$$\begin{aligned} &\nabla f(x_k)^T \Delta x + \frac{1}{2}\Delta x^T G_k \Delta x < 0, \\ &f(x_k) + \sigma\left(\nabla f(x_k)^T \Delta x + \frac{1}{2}\Delta x^T G_k \Delta x\right) < f(x_k + \Delta x), \end{aligned}$$

then set $\Delta := \Delta/2$ and go to Step 2.1

Step 2.4 Go to Step 3.

Step 3. If $\nabla f(x_k)^T \Delta x + \frac{1}{2}\Delta x^T G_k \Delta x \geq 0$, then $\mathcal{F}^k = Add((\theta(x_k), f(x_k)), \mathcal{F}^{k-1})$, otherwise set $\mathcal{F}^k = \mathcal{F}^{k-1}$.

Step 4. Define $\Delta_k = \Delta, \Delta x_k = \Delta x$. Set $x_{k+1} = x_k + \Delta x_k, k := k+1$. Finally, reset Δ such that $\Delta^{max} \geq \Delta \geq \Delta^{min}$ is satisfied and go to Step 2. □

In the above algorithm, Step 1 is called restoration phase. The restoration phase is executed at the first iteration and when subproblem (4.7) is not feasible in Step 2.1. The loop between Steps 2 and 4 is a main loop, and the loop between Steps 2.1 and 2.3 solves subproblem (4.7) many times for various trust region radii if necessary.

Now we give the global convergence theorem as follows.

Theorem 4.3. *Suppose that the points sampled by Algorithm 4.2 lie in a nonempty compact set $\Omega \subset \mathbf{R}^n$ and that the MFCQ condition is satisfied at each feasible point of problem (1.1) lying in the set Ω. Assume that there exists a positive constant M such that $\|G_k\|_F \leq M$ for all k. Consider the sequence $\{x_k\}$ generated by Algorithm 4.2. Then one of the following situations occurs:*

1. *The restoration phase (Step 1) fails to find a point x_k satisfying (a) and (b).*
2. *A stationary point of problem (1.1) is found, that is, $\Delta x = 0$ solves subproblem (4.7) for some iteration k.*
3. *There exists an accumulation point of $\{x_k\}$ that is a stationary point of problem (1.1).*

It is known that the trust-region method [14] is a robust strategy to obtain the global convergence property for general nonlinear optimization problems. Kanzow, Nagel, Kato and Fukushima [36] incorporated the idea of the trust-region method into the successive linearization method for nonlinear SDP problems, and showed its global convergence property. Specifically, they considered the following nonlinear semidefinite programming problem:

$$\begin{array}{lll} \text{minimize} & f(X), & X \in \mathbf{S}^p, \\ \text{subject to} & h(X) \leq 0, & X \succeq 0 \end{array} \tag{4.9}$$

where X is a variable, and the functions $f : \mathbf{S}^p \to \mathbf{R}$ and $h : \mathbf{S}^p \to \mathbf{R}^m$ are continuously differentiable functions. By defining the exact l_1 penalty function

$$p_\rho(X) = f(X) + \rho \sum_{i=1}^{m} \max\{0,\ h_i(X)\}$$

with a penalty parameter $\rho > 0$, Kanzow et al. dealt with the penalized problem:

$$\text{minimize } p_\rho(X) \quad \text{subject to } X \succeq 0. \tag{4.10}$$

Letting h_i denote the ith component of h, and letting $Df(X)$ and $Dh_i(X)$ be the Fréchet derivatives of f and h_i, respectively, at X, they defined the first-order approximation of $p_\rho(X + \Delta X)$ by

$$\Phi_\rho(X, \Delta X) = f(X) + \langle Df(X), \Delta X\rangle + \rho \sum_{i=1}^{m} \max\{0,\ h_i(X) + \langle Dh_i(X), \Delta X\rangle\}.$$

Based on the above, they calculated a search direction $\Delta X_k \in \mathbf{S}^p$ by solving the following subproblem:

$$\text{minimize } \frac{1}{2} c_k \langle \Delta X, \Delta X\rangle + \Phi_{\rho_k}(X_k, \Delta X) \quad \text{subject to } X_k + \Delta X \succeq 0, \tag{4.11}$$

where c_k and ρ_k are suitably chosen parameters. Since the quadratic term of the objective function in (4.11) means that the trust-region idea is implicitly used, this algorithm may be

regarded as a successive linearization method for problem (4.9) that employs a trust-region-type globalization technique. We can also consider that the parameter $1/c_k$ plays the role of the trust-region radius. Furthermore, the quadratic term guarantees the strict convexity of the subproblem. In their algorithm, they used criteria based on the ratio

$$r_k = \frac{p_{\rho_k}(X_k) - p_{\rho_k}(X_k + \Delta X_k)}{p_{\rho_k}(X_k) - \Phi_{\rho_k}(X_k, \Delta X_k)},$$

which is the quotient of the actual and the predicted reductions for the function value of the penalty function. Then the point X_k and the parameter c_k are updated by the same method as the trust-region method. In a similar way to this method, Kanno and Takewaki [34] proposed a method for maximum robustness design of structures. Furthermore, Li and Sun [46] applied the filter method to the successive linearization method.

In the algorithms given in this section, the matrix G_k approximates the Hessian matrix $\nabla_x^2 L(x_k, y_k, Z_k)$ of the Lagrangian function by using the quasi-Newton updating formula, if necessary. In this case, G_{k+1} must satisfy the following secant condition

$$G_{k+1} s_k = q_k,$$

where $s_k = x_{k+1} - x_k$ and

$$\begin{aligned} q_k &= \nabla_x L(x_{k+1}, y_{k+1}, Z_{k+1}) - \nabla_x L(x_k, y_{k+1}, Z_{k+1}) \\ &= (\nabla f(x_{k+1}) - \nabla g(x_{k+1}) y_{k+1} - \mathcal{A}^*(x_{k+1}) Z_{k+1}) \\ &\quad -(\nabla f(x_k) - \nabla g(x_k) y_{k+1} - \mathcal{A}^*(x_k) Z_{k+1}) \\ &= \nabla f(x_{k+1}) - \nabla f(x_k) - (\nabla g(x_{k+1}) - \nabla g(x_k)) y_{k+1} - (\mathcal{A}^*(x_{k+1}) - \mathcal{A}^*(x_k)) Z_{k+1}. \end{aligned}$$

In order to preserve the positive definiteness of the matrix G_k, we can use the modified BFGS update proposed by Powell, which is given by the form

$$G_{k+1} = G_k - \frac{G_k s_k s_k^T G_k}{s_k^T G_k s_k} + \frac{\hat{q}_k \hat{q}_k^T}{s_k^T \hat{q}_k},$$

where

$$\begin{aligned} \hat{q}_k &= \psi_k q_k + (1 - \psi_k) G_k s_k, \\ \psi_k &= \begin{cases} 1 & \text{if } s_k^T q_k \geq 0.2 s_k^T G_k s_k \\ \dfrac{0.8 s_k^T G_k s_k}{s_k^T (G_k s_k - q_k)} & \text{otherwise.} \end{cases} \end{aligned}$$

5. Primal-Dual Interior Point Method

As in ordinary NLP problems and linear SDP problems, the interior point methods can be one of most useful numerical methods for solving nonlinear SDP problems. In this section, we introduce the primal-dual interior point methods proposed by Yamashita, Yabe and Harada [76], and Yamashita and Yabe [75]. The proposed algorithm is implemented in the software NUOPT (the current name is Numerical Optimizer), and its numerical performance is described in [76].

We note Jarre [32] gave a primal predictor-corrector type interior point method for solving nonlinear SDP problem, and Leibfritz and Mostafa [45] gave a primal interior point method for solving a special class of nonlinear SDP problem. Burer, Monteiro and Zhang [12,

13] converted problem (1.1) to an equivalent NLP problem, and solved the resultant problem by the interior point method. In [77], Yang and Yu proposed a primal-dual predictor-corrector method based on the homotopy method which is not an interior point type.

A point $w = (x, y, Z)$ satisfying $X(x) \succ 0$ and $Z \succ 0$ is called an interior point, and interior point methods generate such points. To construct an interior point algorithm, a positive barrier parameter μ is introduced as in the interior point methods for ordinary NLP, and the complementarity condition $X(x)Z = 0$ is replaced by $X(x)Z = \mu I$, where I denotes the identity matrix. The primal-dual interior point methods try to find a point that satisfies the barrier KKT (BKKT) conditions:

$$r(w, \mu) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ X(x)Z - \mu I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.1}$$

and the positivity conditions:

$$X(x) \succ 0, \qquad Z \succ 0,$$

for a given μ. By using the symmetrization (1.2), we also define the symmetrized residual $r_S(w, \mu)$ by

$$r_S(w, \mu) = \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ X(x) \circ Z - \mu I \end{pmatrix}. \tag{5.2}$$

This will be used for constructing Newton iteration later, and we denote $r_S(w, 0)$ by $r_{0S}(w)$.

We define the norms $\|r(w, \mu)\|$ and $\|r_S(w, \mu)\|$ by

$$\|r(w, \mu)\| = \sqrt{\left\| \begin{pmatrix} \nabla_x L(w) \\ g(x) \end{pmatrix} \right\|^2 + \|X(x)Z - \mu I\|_F^2}$$

and

$$\|r_S(w, \mu)\| = \sqrt{\left\| \begin{pmatrix} \nabla_x L(w) \\ g(x) \end{pmatrix} \right\|^2 + \|X(x) \circ Z - \mu I\|_F^2},$$

respectively. Note that $\|r_S(w, \mu)\| \leq \|r(w, \mu)\|$ is satisfied, because $\|X(x) \circ Z - \mu I\|_F \leq \|X(x)Z - \mu I\|_F$.

5.1. Algorithm for finding a KKT point

We first describe a procedure for finding a KKT point by using the BKKT conditions. In this subsection, the subscript k denotes an iteration count of the outer iterations.

Algorithm 5.1.

Step 0. (Initialize) Set $\varepsilon > 0$, $M_c > 0$ and $k = 0$. Let a positive sequence $\{\mu_k\}, \mu_k \downarrow 0$ be given.

Step 1. (Approximate BKKT point) Find an interior point w_{k+1} that satisfies

$$\|r(w_{k+1}, \mu_k)\| \leq M_c \mu_k. \tag{5.3}$$

Step 2. (Terminate) If $\|r_0(w_{k+1})\| \leq \varepsilon$, then stop.

Step 3. (Update) Set $k := k + 1$ and go to Step 1. □

The condition (5.3) is called the approximate BKKT condition, and a point that satisfies this condition the approximate BKKT point. The barrier parameter sequence $\{\mu_k\}$ in the above needs not be determined beforehand, and the value of each μ_k may be set adaptively as the iteration proceeds. The concrete procedure in Step 1 of Algorithm 5.1 will be given as Algorithm 5.2 in Section 5.2.

The following theorem shows the convergence property of Algorithm 5.1 under the MFCQ condition.

Theorem 5.1. *Assume that the functions f, g and X are continuously differentiable. Let $\{w_k\}$ be an infinite sequence generated by Algorithm 5.1. Suppose that the sequence $\{x_k\}$ is bounded and that the MFCQ condition is satisfied at any accumulation point of the sequence $\{x_k\}$. Then the sequences $\{y_k\}$ and $\{Z_k\}$ are bounded, and any accumulation point of $\{w_k\}$ satisfies the KKT conditions (2.1) and (2.2).*

5.2. Newton method for finding a barrier KKT point and primal-dual merit function

In [76], a globally convergent algorithm to find a BKKT point for a given fixed barrier parameter $\mu > 0$ was proposed. The algorithm given below is used as an inner iteration of Algorithm 5.1. This subsection describes a Newton-like method to the system of equations (5.4). We denote the Newton directions for the primal and dual variables by $\Delta x \in \mathbf{R}^n, \Delta y \in \mathbf{R}^m$ and $\Delta Z \in \mathbf{S}^p$. We define $\Delta X = \sum_{i=1}^n \Delta x_i A_i(x)$, and note that $\Delta X \in \mathbf{S}^p$. In order to have a Jacobian operator from $\mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ to $\mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ (a square matrix when represented in matrix form), we employ the symmetrized residual representation of the BKKT conditions $r_S(w, \mu) = 0$.

As in the case of linear SDP problems, a scaling of the primal-dual pair $(X(x), Z)$ is used in order to make the resulting Newton equation solvable. In what follows, we denote $X(x)$ simply by X when it is not confusing. Throughout this subsection, we assume that $X \succ 0$ and $Z \succ 0$ hold. We introduce a nonsingular matrix $T \in \mathbf{R}^{p\times p}$ and scale X and Z by

$$\widetilde{X} = TXT^T \quad \text{and} \quad \widetilde{Z} = T^{-T}ZT^{-1},$$

respectively. Using the scaling matrix T, we try to solve the equation $\widetilde{X} \circ \widetilde{Z} = \mu I$ instead of $XZ = \mu I$, and consider the scaled symmetrized residual:

$$\tilde{r}_S(w, \mu) \equiv \begin{pmatrix} \nabla_x L(w) \\ g(x) \\ \widetilde{X} \circ \widetilde{Z} - \mu I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{5.4}$$

to form the Newton directions.

Scalings of ΔX and ΔZ are similarly done by

$$\Delta\widetilde{X} = T\Delta XT^T \quad \text{and} \quad \Delta\widetilde{Z} = T^{-T}\Delta ZT^{-1}.$$

Newton equations for (5.4) are given by

$$G\Delta x - \nabla g(x)\Delta y - \mathcal{A}^*(x)\Delta Z = -\nabla_x L(x, y, Z) \tag{5.5}$$

$$\nabla g(x)^T \Delta x = -g(x) \tag{5.6}$$

$$\frac{1}{2}(\Delta\widetilde{X}\widetilde{Z} + \widetilde{Z}\Delta\widetilde{X} + \widetilde{X}\Delta\widetilde{Z} + \Delta\widetilde{Z}\widetilde{X}) = \mu I - \frac{1}{2}(\widetilde{X}\widetilde{Z} + \widetilde{Z}\widetilde{X}), \tag{5.7}$$

where G denotes the Hessian matrix of the Lagrangian function $L(w)$ or its approximation. By solving these equations, we obtain the Newton direction $\Delta w = (\Delta x, \Delta y, \Delta Z) \in \mathbf{R}^n \times$

$\mathbf{R}^m \times \mathbf{S}^p$. Writing the last part of Newton equations as

$$(\widetilde{Z} \odot I)\Delta\widetilde{X} + (\widetilde{X} \odot I)\Delta\widetilde{Z} = \mu(\widetilde{X} \odot I)\widetilde{X}^{-1} - (\widetilde{X} \odot I)\widetilde{Z},$$

we can formally solve this equation for $\Delta\widetilde{Z}$.

The following theorem [76] gives the desired form of the Newton directions.

Theorem 5.2. *Suppose that the operator $\widetilde{X} \odot I$ is invertible. Then the direction $\Delta\widetilde{Z} \in \mathbf{S}^p$ is given by the form*

$$\Delta\widetilde{Z} = \mu\widetilde{X}^{-1} - \widetilde{Z} - (\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)\Delta\widetilde{X}, \tag{5.8}$$

or equivalently

$$\Delta Z = \mu X^{-1} - Z - (T^T \odot T^T)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)\Delta X. \tag{5.9}$$

Furthermore, the directions $(\Delta x, \Delta y) \in \mathbf{R}^n \times \mathbf{R}^m$ satisfy

$$\begin{pmatrix} G+H & -\nabla g(x) \\ -\nabla g(x)^T & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} \nabla f(x) - \nabla g(x)y - \mu\mathcal{A}^*(x)X^{-1} \\ -g(x) \end{pmatrix}, \tag{5.10}$$

where the elements of the matrix $H \in \mathbf{R}^{n\times n}$ are represented by the form

$$H_{ij} = \left\langle \widetilde{A}_i(x), (\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)\widetilde{A}_j(x) \right\rangle \tag{5.11}$$

with $\widetilde{A}_i(x) = TA_i(x)T^T$.

In addition, if the matrix $G+H$ is positive definite and the matrix $\nabla g(x)$ is of full rank, then the Newton equations (5.5) – (5.7) give a unique search direction $\Delta w = (\Delta x, \Delta y, \Delta Z)$ $\in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$.

In the following, we choose a nonsingular matrix T so that $\widetilde{X}$ and $\widetilde{Z}$ commute, i.e., $\widetilde{X}\widetilde{Z} = \widetilde{Z}\widetilde{X}$. In this case, the matrices $\widetilde{X}$ and $\widetilde{Z}$ share the same eigensystem, and $XZ = \mu I \iff \widetilde{X} \circ \widetilde{Z} = \mu I$. In the algorithm described here, this commutativity is necessary to prove a descent property of the Newton direction with respect to the merit function described later.

Well known examples of the scaling matrix T that satisfy such a condition are given below.

(i) HRVW/KSH/M direction

If we set $T = X^{-1/2}$, then we have $\widetilde{X} = I$ and $\widetilde{Z} = X^{1/2}ZX^{1/2}$, which corresponds to the HRVW/KSH/M direction for linear SDP problems [26, 42, 48]. In this case, the matrices H and ΔZ can be represented by the form:

$$\begin{aligned} H_{ij} &= \mathrm{tr}\left(A_i(x)X^{-1}A_j(x)Z\right), \\ \Delta Z &= \mu X^{-1} - Z - \frac{1}{2}(X^{-1}\Delta X Z + Z\Delta X X^{-1}). \end{aligned}$$

(ii) NT direction

If we set $T = W^{-1/2}$ with $W = X^{1/2}(X^{1/2}ZX^{1/2})^{-1/2}X^{1/2}$, then we have $\widetilde{X} = W^{-1/2}XW^{-1/2}$ $= W^{1/2}ZW^{1/2} = \widetilde{Z}$, which corresponds to the NT direction for linear SDP problems [50, 51]. In this case, the matrices H and ΔZ can be represented by the form:

$$\begin{aligned} H_{ij} &= \mathrm{tr}\left\{A_i(x)W^{-1}A_j(x)W^{-1}\right\}, \\ \Delta Z &= \mu X^{-1} - Z - W^{-1}\Delta X W^{-1}. \end{aligned}$$

Both examples make $(\widetilde{X} \odot I)^{-1}$ in (5.8) tractable by letting $(\widetilde{X} \odot I)^{-1} = I$ (HRVW/KSH /M) or $(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I) = I$ (NT).

Under the above scalings, if the matrices $A_i(x)$ $(i = 1, \ldots, n)$ are linearly independent, the matrix H is symmetric positive definite.

5.3. Global convergence of primal-dual interior point methods

In [76], the line search strategy was used, and the following merit function in the primal-dual space was proposed:

$$F(x, Z) = F_{BP}(x) + \nu F_{PD}(x, Z). \tag{5.12}$$

Here $F_{BP}(x)$ and $F_{PD}(x, Z)$ are the primal barrier penalty function and the primal-dual barrier function, respectively, and they are given by

$$\begin{aligned} F_{BP}(x) &= f(x) - \mu \log(\det X) + \rho \|g(x)\|_1, & (5.13)\\ F_{PD}(x, Z) &= \langle X, Z \rangle - \mu \log(\det X \det Z), & (5.14) \end{aligned}$$

where ν and ρ are positive parameters. Though the functions $F_{BP}(x)$ and $F_{PD}(x, Z)$ depend on the parameters ν, ρ and μ, we use the notation $F(x, Z)$ for simplicity. It follows from the fact $\widetilde{X}\widetilde{Z} = TXZT^{-1}$ that $\left\langle \widetilde{X}, \widetilde{Z} \right\rangle = \langle X, Z \rangle$ and the value of $F_{PD}(x, Z)$ is invariant under the choice of T.

We define the first order approximation F_l of the merit function by

$$F_l(x, Z; \Delta x, \Delta Z) = F(x, Z) + \Delta F_l(x, Z; \Delta x, \Delta Z).$$

Here $\Delta F_l(x, Z; \Delta x, \Delta Z)$ corresponds to the directional derivative and it is defined by the form

$$\Delta F_l(x, Z; \Delta x, \Delta Z) = \Delta F_{BPl}(x; \Delta x) + \nu \Delta F_{PDl}(x, Z; \Delta x, \Delta Z),$$

where

$$\begin{aligned} \Delta F_{BPl}(x; \Delta x) &= \nabla f(x)^T \Delta x - \mu \mathrm{tr}(X^{-1} \Delta X) & (5.15)\\ &\quad + \rho \left(\|g(x) + \nabla g(x)^T \Delta x\|_1 - \|g(x)\|_1 \right), \\ \Delta F_{PDl}(x, Z; \Delta x, \Delta Z) &= \mathrm{tr}(\Delta X Z + X \Delta Z - \mu X^{-1} \Delta X - \mu Z^{-1} \Delta Z). \end{aligned}$$

With these definitions, it is possible to prove that the Newton direction Δw satisfies

$$\Delta F_l(x, Z; \Delta x, \Delta Z) \leq -\Delta x^T (G + H) \Delta x - (\rho - \|y + \Delta y\|_\infty) \|g(x)\|_1.$$

By using this inequality, the following theorem is obtained, which shows that the Newton direction Δw becomes a descent search direction for the proposed primal-dual merit function (5.12).

Theorem 5.3. *Assume that Δw solves (5.5) – (5.7) and that the matrix $G+H$ is symmetric positive definite. Suppose that the penalty parameter ρ satisfies $\rho > \|y + \Delta y\|_\infty$. Then the following hold:*

(i) *The direction Δw becomes a descent search direction for the primal-dual merit function $F(x, Z)$, i.e. $\Delta F_l(x, Z; \Delta x, \Delta Z) \leq 0$.*
(ii) *If $\Delta x \neq 0$, then $\Delta F_l(x, Z; \Delta x, \Delta Z) < 0$.*
(iii) *$\Delta F_l(x, Z; \Delta x, \Delta Z) = 0$ holds if and only if $(x, y + \Delta y, Z)$ is a BKKT point.*

In order to construct a globally convergent algorithm to a BKKT point for a fixed $\mu > 0$, the iterations take the form

$$x_{k+1} = x_k + \alpha_k \Delta x_k, \quad Z_{k+1} = Z_k + \alpha_k \Delta Z_k \quad \text{and} \quad y_{k+1} = y_k + \Delta y_k$$

where α_k is a step size determined by the line search procedure described below. Throughout this subsection, the index k denotes the inner iteration count for a given $\mu > 0$. We also

denote $X(x_k)$ by X_k for simplicity. We note that $X_k \succ 0$ and $Z_k \succ 0$ are maintained for all k in the following.

Since the main iteration is to decrease the value of the merit function (5.12), the step size is determined by the sufficient-decrease rule of the merit function. At the current point (x_k, Z_k), the initial step size is calculated by

$$\bar{\alpha}_{xk} = \begin{cases} -\dfrac{\gamma}{\lambda_{\min}(X_k^{-1}\Delta X_k)} & \text{if } X(x) \text{ is linear} \\ 1 & \text{otherwise} \end{cases} \tag{5.16}$$

and

$$\bar{\alpha}_{zk} = -\frac{\gamma}{\lambda_{\min}(Z_k^{-1}\Delta Z_k)}, \tag{5.17}$$

where $\gamma \in (0,1)$ is a constant. If the minimum eigenvalue in either expression (5.16) or (5.17) is positive, the corresponding step size is set to 1. By trying the values $\ell_k = 0, 1, \ldots$ successively, we find a step to the next iterate given by

$$\alpha_k = \bar{\alpha}_k \beta^{l_k}, \quad \bar{\alpha}_k = \min\left\{\bar{\alpha}_{xk}, \bar{\alpha}_{zk},\ 1\right\},$$

where $\beta \in (0,1)$ is a constant, and l_k is the smallest nonnegative integer such that the sufficient-decrease condition

$$F(x_k + \bar{\alpha}_k\beta^{l_k}\Delta x_k, Z_k + \bar{\alpha}_k\beta^{l_k}\Delta Z_k) \le F(x_k, Z_k) + \varepsilon_0\bar{\alpha}_k\beta^{l_k}\Delta F_l(x_k, Z_k; \Delta x_k, \Delta Z_k) \tag{5.18}$$

and the positive definiteness condition

$$X(x_k + \bar{\alpha}_k\beta^{l_k}\Delta x_k) \succ 0 \tag{5.19}$$

hold, where $\varepsilon_0 \in (0,1)$ is a constant.

Now we summarize the line search algorithm in the following. Since this algorithm should be regarded as the inner iteration of Algorithm 5.1 (Step 1 of Algorithm 5.1), ε' given below corresponds to $M_c\mu$ and an initial point can be set to the approximate BKKT point obtained at the previous outer iteration.

Algorithm 5.2.

Step 0. (Initialize) Let $w_0 \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ ($X_0 \succ 0, Z_0 \succ 0$), $\mu > 0$, $\rho > 0$ and $\nu > 0$ be given. Set $\varepsilon' > 0$, $\gamma \in (0,1)$, $\beta \in (0,1)$ and $\varepsilon_0 \in (0,1)$. Let $k = 0$.

Step 1. (Termination) If $\|r(w_k, \mu)\| \le \varepsilon'$, then stop.

Step 2. (Search direction) Calculate the matrix G_k and the scaling matrix T_k. Determine the direction Δw_k by solving (5.5) – (5.7). If $\Delta x_k = 0$, then set $w_{k+1} = (x_k, y_k + \Delta y_k, Z_k + \Delta Z_k)$, and stop.

Step 3. (Step size) Find the smallest nonnegative integer l_k that satisfies the criteria (5.18) and (5.19), and calculate

$$\alpha_k = \bar{\alpha}_k\beta^{l_k}.$$

Step 4. (Update) Set

$$x_{k+1} = x_k + \alpha_k\Delta x_k, \quad Z_{k+1} = Z_k + \alpha_k\Delta Z_k \quad \text{and} \quad y_{k+1} = y_k + \Delta y_k.$$

Step 5. Set $k := k + 1$ and go to Step 1. □

If the matrix G_k approximates the Hessian matrix $\nabla_x^2 L(w_k)$ by the quasi-Newton formula, the method stated in the last part of Section 4 can be used.

If we want to use the Hessian matrix $\nabla_x^2 L(w_k)$ itself for the matrix G_k, the Levenberg-Marquardt type modification of $\nabla_x^2 L(w_k)$ may be used to obtain a positive semidefinite G_k for global convergence property shown below. Namely, we compute a parameter $\sigma \geq 0$ which gives a positive semidefinite matrix $\nabla_x^2 L(w_k) + \sigma I$. This strategy may be useful for solving large scale problems.

The following global convergence theorem was proved by Yamashita, Yabe and Harada [76].

Theorem 5.4. *If $\Delta x_k = 0$ in Step 2 of Algorithm 5.2, then $(x_k, y_k + \Delta y_k, Z_k)$ is a BKKT point. Otherwise let an infinite sequence $\{w_k\}$ be generated by Algorithm 5.2. Suppose the followings:*

(i) *The functions f, $g_i, i = 1, \ldots, m$, and X are twice continuously differentiable;*
(ii) *The sequence $\{x_k\}$ remains in a compact set Ω of $\mathbf{R}^n$;*
(iii) *The matrix $\nabla g(x_k)$ is of full rank and the matrices $A_1(x_k), \ldots, A_n(x_k)$ are linearly independent for all x_k in Ω;*
(iv) *The scaling matrix T_k is chosen so that $\widetilde{X}_k$ and $\widetilde{Z}_k$ commute;*
(v) *The matrix G_k is uniformly bounded and positive semidefinite;*
(vi) *Both of the sequences $\{T_k\}$ and $\{T_k^{-1}\}$ are bounded;*
(vii) *The penalty parameter ρ is sufficiently large so that $\rho > \|y_k + \Delta y_k\|_\infty$ holds for all k.*

Then there exists at least one accumulation point of $\{w_k\}$, and any accumulation point of the sequence $\{w_k\}$ is a BKKT point.

5.4. The rate of convergence of primal-dual interior point methods

In this section, we survey the local behavior of the primal-dual interior point methods following the study by Yamashita and Yabe [75]. Consider the scaled Newton equations (5.5)-(5.7) and denote the Newton equations by

$$\widetilde{J}_S(w)\Delta w = -\tilde{r}_S(w, \mu), \tag{5.20}$$

where $\widetilde{J}_S(w)$ is the linear operator from $\mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ to $\mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ and $\tilde{r}_S(w, \mu)$ is defined by (5.4). If we choose $T = I$, we call the above equations the unscaled Newton equations and denote the operator $\widetilde{J}_S(w)$ by $J_S(w)$ in this case. This case corresponds to the AHO direction for linear SDP problems [2].

In order to analyze the local behavior of the methods, we make the following assumptions in this subsection: The second derivatives of the functions f, $g_i, i = 1, \ldots, m$, and X are Lipschitz continuous at x^*; The second order sufficient condition (2.5) holds at x^*; The strict complementarity condition holds at x^*; The nondegeneracy condition is satisfied at x^*.

We will discuss a local behavior of the unsymmetric residual $r_0(w)$ in (2.1) or $r(w, \mu)$ in (5.1). For this purpose, we define a linear operator $J : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p \to \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^{p \times p}$ at w by

$$J(w)\Delta w = \begin{pmatrix} \nabla_x^2 L(w)\Delta x - \nabla g(x)\Delta y - \mathcal{A}^*(x)\Delta Z \\ \nabla g(x)^T \Delta x \\ \Delta X Z + X \Delta Z \end{pmatrix}$$

for $\Delta w = (\Delta x, \Delta y, \Delta Z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$, which is an estimate of the first order change of $r_0(w + \Delta w)$ or $r(w + \Delta w, \mu)$.

For the unscaled Newton method, the Newton equations (5.20) is

$$J_S(w)\Delta w = -r_S(w, \mu). \tag{5.21}$$

Under the above assumptions, the regularity of the operators $J_S(w^*)$ and $J(w^*)$ is shown as follows.

Theorem 5.5. *([75]) The operator $J_S(w^*)$ is nonsingular, and the operator $J(w^*)$ is left invertible.*

With the above properties of $J_S(w^*)$ and $J(w^*)$, an algorithm for the analysis of local behavior can be defined as follows.

Algorithm 5.3.

Step 0. (Initialize) Set $\varepsilon > 0$ and $0 < \tau < 1$. Choose $w_0 = (x_0, y_0, Z_0) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ $(X(x_0) \succ 0, Z_0 \succ 0)$. Set $k = 0$.

Step 1. (Termination) If $\|r_0(w_k)\| \leq \varepsilon$, then stop.

Step 2. (Newton step) Choose a barrier parameter μ_k such that

$$\mu_k = \xi_k \|r_0(w_k)\|^{1+\tau} \tag{5.22}$$

with $\xi_k = \Theta(1)$. Compute the direction Δw_k by solving the Newton equations (5.21). Set $w_{k+1} = w_k + \Delta w_k$.

Step 3. (Update) Set $k := k + 1$ and go to Step 1. □

We note that in the above algorithm, the barrier parameter is decreased faster than $\|r_0(w_k)\|$, and the step sizes are set to 1. By Theorem 5.5, if the iterate w_k is sufficiently close to w^*, the Jacobian operator $J_S(w_k)$ is nonsingular and its inverse is uniformly bounded. Thus the Newton equations have a unique solution and the following relations hold

$$\Delta w_k = \Theta(\|r_S(w_k, \mu_k)\|) = \mathrm{O}(\|r_{0S}(w_k)\|) + \mathrm{O}(\mu_k) = \mathrm{O}(\|r_0(w_k)\|). \tag{5.23}$$

Here the last equality can be obtained by equation (5.22) which shows that the parameter μ_k is decreased sufficiently fast. Then it is possible to show that the unit step iteration gives an interior point and an approximate BKKT point successively. Thus we can prove the superlinear convergence of Algorithm 5.3 in the following theorem.

Theorem 5.6. *Let M_c and τ be given constants satisfying $0 < M_c < 1$ and $0 < \tau < 1$. Let μ_{-1} be a sufficiently small positive number. Assume that an initial interior point w_0 is sufficiently close to w^* and satisfies the approximate BKKT condition $\|r(w_0, \mu_{-1})\| \leq M_c\mu_{-1}$. Then the sequence $\{w_k\}$ generated by Algorithm 5.3 satisfies*

$$\|r(w_k, \mu_{k-1})\| \leq M_c\mu_{k-1}, \quad X(x_k) \succ 0 \quad \text{and} \quad Z_k \succ 0 \tag{5.24}$$

for all $k \geq 0$ and converges locally and superlinearly to w^*.

We note that although no globally convergent algorithm for unscaled Newton method is proposed yet, this method has favorable local properties as seen in the above theorem. We also note that, to solve the unscaled Newton equation, we have to represent it by the following matrix-vector form:

$$\begin{pmatrix} \nabla_x^2 L(w) & -\nabla g(x) & -A_S(x)^T \\ \nabla g(x)^T & 0 & 0 \\ (Z \otimes_S I)A_S(x) & 0 & (X \otimes_S I) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \mathrm{svec}(\Delta Z) \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x, y, Z) \\ -g(x) \\ \mathrm{svec}(\mu I - X \circ Z) \end{pmatrix},$$

where

$$A_S(x) = [\mathrm{svec}(A_1(x)), \ldots, \mathrm{svec}(A_n(x))] \in \mathbf{R}^{p(p+1)/2 \times n}.$$

Next, we proceed to results on the local behavior of the scaled Newton equations. Specifically we show that local and two-step superlinear convergence properties are proved for the primal-dual interior point methods which use the HRVW/KSH/M and the NT directions.

In the following, we describe an algorithm which calculates a KKT point by using the scaled Newton method. Roughly speaking, if the current point w_k is sufficiently close to the BKKT point with sufficiently small μ_k, then the Newton equation (5.20) can be written as

$$J'_S(w_k)\Delta w_k = -r_S(w_k, \mu_k)$$

with an operator $J'_S(w_k)$ which is close to nonsingular $J_S(w_k)$, and Δw_k satisfies

$$\Delta w_k = \mathrm{O}(\|r(w_k, \mu_k)\|).$$

With these properties, the scaled Newton step gives a local and two-step superlinear convergence as the following algorithm and theorem show.

Algorithm 5.4.

Step 0. (Initialize) Set $\varepsilon > 0$ and $0 < \tau < 1$. Choose $w_0 = (x_0, y_0, Z_0) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ $(X(x_0) \succ 0, Z_0 \succ 0)$. Set $k = 0$.

Step 1. (Termination) If $\|r_0(w_k)\| \leq \varepsilon$, then stop.

Step 2. (Scaled Newton steps)

Step 2.1 Choose $\mu_k = \xi_k \|r_0(w_k)\|^{1+\tau}$ with $\xi_k = \Theta(1)$.

Step 2.2 Calculate the direction Δw_k by solving the scaled Newton equations $\widetilde{J}_S(w_k)\Delta w_k = -\tilde{r}_S(w_k, \mu_k)$ at w_k. Set $w_{k+\frac{1}{2}} = w_k + \Delta w_k$.

Step 2.3 Calculate the direction $\Delta w_{k+\frac{1}{2}}$ by solving the scaled Newton equations

$$\widetilde{J}_S(w_{k+\frac{1}{2}})\Delta w_{k+\frac{1}{2}} = -\tilde{r}_S(w_{k+\frac{1}{2}}, \mu_k)$$

at $w_{k+\frac{1}{2}}$. Set $w_{k+1} = w_{k+\frac{1}{2}} + \Delta w_{k+\frac{1}{2}}$.

Step 3. (Update) Set $k := k + 1$ and go to Step 1. □

In the above algorithm, the decreasing speed of the barrier parameter is set to the same as the unscaled algorithm. But the scaled Newton steps with unit step sizes are repeated twice with the same barrier parameter value. The following theorem gives a detailed behavior of the scaled Newton steps near the KKT point.

Theorem 5.7. *Let M_c be a positive constant, and let τ and τ' be positive constants that satisfy*

$$1 > \tau' > \tau \quad \text{and} \quad \tau' > \frac{2\tau}{1-\tau}.$$

Let μ_{-1} be a sufficiently small positive number and satisfy

$$\left(\frac{1}{2M_c}\right)^{1/\tau'} \geq \mu_{-1}.$$

Assume that an initial interior point w_0 is sufficiently close to w^* and satisfies the approximate BKKT condition $\|r(w_0, \mu_{-1})\| \leq M_c \mu_{-1}^{1+\tau'}$. Then the sequence $\{w_k\}$ generated by Algorithm 5.4 with $T_k = X_k^{-1/2}$ or $T_k = W_k^{-1/2}$ satisfies

$$\|r(w_k, \mu_{k-1})\| \leq M_c \mu_{k-1}^{1+\tau'}, \quad X(x_k) \succ 0 \quad \text{and} \quad Z_k \succ 0$$

for all $k \geq 0$ and converges two-step superlinearly to w^* in the sense that

$$\|w_k + \Delta w_k + \Delta w_{k+\frac{1}{2}} - w^*\| = \mathrm{O}(\|w_k - w^*\|^{1+\tau'}) \quad \text{for all } k.$$

5.5. Primal-dual interior point methods with shifted KKT conditions

Interior point methods usually replace the complementarity condition $X(x)Z = 0$ by $X(x)Z = \mu I$ and manage the resulting barrier KKT conditions instead of conditions (2.1) and (2.2). These conditions are intimately related to the necessary condition for optimality of the following problem

$$\min_{X(x)\succ 0} \quad F_{BP}(x,\mu) = f(x) + \rho\|g(x)\|_1 - \mu\log(\det X(x)), \quad x \in \mathbf{R}^n. \tag{5.25}$$

In connection with (1.1), we consider the following problem instead of (5.25):

$$\min_{X(x)\succ 0} \quad F_{BP2}(x,\mu) = f(x) + \frac{1}{2\mu}\|g(x)\|^2 - \mu\log(\det X(x)), \quad x \in \mathbf{R}^n. \tag{5.26}$$

The necessary conditions for optimality of this problem are given by

$$\nabla F_{BP2}(x,\mu) = \nabla f(x) + \frac{1}{\mu}\nabla g(x)g(x) - \mu\mathcal{A}^*(x)X^{-1}(x) = 0,$$

and $X(x) \succ 0$. If we let the variables y and Z satisfy the relations $y = -g(x)/\mu$ and $Z = \mu X(x)^{-1}$, respectively, then the above conditions are written as

$$r(w,\mu) = \begin{pmatrix} \nabla_x L(w) \\ g(x) + \mu y \\ X(x)Z - \mu I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.27}$$

and

$$X(x) \succ 0, \qquad Z \succ 0.$$

These conditions are called the shifted barrier KKT (SBKKT) conditions, and the point which satisfies SBKKT conditions are called the SBKKT point. Motivated by the prior work by Forsgren and Gill [22], Yamashita and Yabe [73] proposed a method that uses SBKKT conditions for ordinary NLP. An extension of the method for solving nonlinear SDP was done by Kato, Yabe and Yamashita [37]. We note that similar approach for nonlinear SDP was also done by Yamakawa and Yamashita [72].

In order to obtain the KKT point, the same prototype algorithm as Algorithm 5.1 is used, where condition (5.3) in Step 1 of Algorithm 5.1 should use the above residual (5.27). As in the previous section, the Newton method for equations (5.27) is considered to search the SBKKT point for fixed μ, and we obtain the Newton step Δw by solving the following linear system:

$$\begin{aligned} G\Delta x - \nabla g(x)\Delta y - \mathcal{A}^*(x)\Delta Z &= -\nabla_x L(x,y,Z) && (5.28)\\ \nabla g(x)^T\Delta x + \mu\Delta y &= -g(x) - \mu y && (5.29)\\ \Delta\widetilde{X}\widetilde{Z} + \widetilde{Z}\Delta\widetilde{X} + \widetilde{X}\Delta\widetilde{Z} + \Delta\widetilde{Z}\widetilde{X} &= 2\mu I - \widetilde{X}\widetilde{Z} - \widetilde{Z}\widetilde{X}. && (5.30) \end{aligned}$$

Then, the equations (5.28) – (5.30) are rewritten by

$$\{G + H + \frac{1}{\mu}\nabla g(x)\nabla g(x)^T\}\Delta x = -\nabla F_{BP2}(x,\mu), \tag{5.31}$$

$$\Delta y = -\frac{1}{\mu}\{g(x) + \mu y + \nabla g(x)^T\Delta x\} \tag{5.32}$$

and

$$\Delta Z = \mu X^{-1} - Z - (T^T \odot T^T)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)\Delta X, \tag{5.33}$$

where the elements of the matrix $H \in \mathbf{R}^{n\times n}$ are given by (5.11).

If the matrix $G + H + \frac{1}{\mu}\nabla g(x)\nabla g(x)^T$ is nonsingular, then the Newton equations (5.28) – (5.30) have a unique search direction $\Delta w = (\Delta x, \Delta y, \Delta Z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ obtained by (5.31) – (5.33). To obtain a unique direction, the interior point method in [76] needs the assumption that $\nabla g(x)$ is of full rank, while this method does not need such an assumption.

In contrast to the merit function $F_{BP}(x, \mu)$ which contains the non-differentiable term and which does not depend on y, the merit function in [37] has the following form:

$$F_{\ell 2}(w, \mu) = F_{BP2}(x, \mu) + \nu F_{PD2}(w, \mu),$$

where $F_{BP2}(x, \mu)$ is defined by (5.26) and $F_{PD2}(w, \mu)$ is a primal-dual barrier function given by

$$F_{PD2}(w, \mu) = \frac{1}{2}\|g(x) + \mu y\|^2 + \log \frac{\langle X, Z\rangle/p + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{\{\det(XZ)\}^{1/p}}.$$

It is possible to prove that the Newton direction is a descent direction with respect to this merit function. The Newton direction Δw satisfies

$$\begin{aligned} D(F_{\ell 2}(w, \mu); \Delta w) \quad \leq \quad & -\Delta x^T \{G + H + \frac{1}{\mu}\nabla g(x)\nabla g(x)^T\}\Delta x \\ & -\nu \frac{\|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2}{\langle X, Z\rangle/p + \|Z^{\frac{1}{2}} X Z^{\frac{1}{2}} - \mu I\|_F^2} - \nu\|g(x) + \mu y\|^2, \end{aligned}$$

where $D(F_{\ell 2}(w, \mu); \Delta w)$ is the directional derivative of the function $F_{\ell 2}(w, \mu)$ along the direction Δw. If the matrix $G + H + \frac{1}{\mu}\nabla g(x)\nabla g(x)^T$ is positive definite, then the Newton direction Δw is a descent direction for the merit function $F_{\ell 2}(w, \mu)$.

The algorithm that finds an approximate SBKKT point given in [37] is similar to Algorithm 5.2. The only difference is that $\Delta F_l(x_k, Z_k; \Delta x_k, \Delta Z_k)$ in Algorithm 5.2 is replaced by the directional derivative $D(F_{\ell 2}(w_k, \mu); \Delta w_k)$, because the merit function $F_{\ell 2}$ is differentiable.

Algorithm 5.5.

Step 0. (Initialize) Set an initial interior point $w_0 = (x_0, y_0, Z_0) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^p$ ($X(x_0) \succ 0, Z_0 \succ 0$), the fixed barrier parameter $\mu > 0$ and a parameter $M_c > 0$. Choose the parameters $\nu > 0$, $\gamma \in (0, 1)$, $\beta \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$. Set $k = 0$.

Step 1. (Termination) If $\|r(w_k, \mu)\| \leq M_c\mu$, then stop.

Step 2. (Search direction) Calculate the matrix G_k and the scaling matrix T_k. Determine a search direction Δw_k by (5.31) – (5.33).

Step 3. (Step size) Calculate an initial step size $\bar{\alpha}_k$ by

$$\bar{\alpha}_k = \min\{\bar{\alpha}_{xk}, \bar{\alpha}_{zk}, 1\},$$

where $\bar{\alpha}_{xk}$ and $\bar{\alpha}_{zk}$ are calculated by (5.16) and (5.17). Find the smallest nonnegative integer ℓ_k that satisfies the conditions

$$F_{\ell 2}(w_k + \bar{\alpha}_k \beta^{\ell^k} \Delta w_k, \mu) \leq F_{\ell 2}(w_k, \mu) + \varepsilon_0 \bar{\alpha}_k \beta^{\ell^k} D(F_{\ell 2}(w_k, \mu); \Delta w_k)$$

and

$$X(x_k + \bar{\alpha}_k \beta^{\ell_k} \Delta x_k) \succ 0.$$

Set $\alpha_k = \bar{\alpha}_k \beta^{\ell_k}$.

Step 4. (Update) Set

$$w_{k+1} = w_k + \alpha_k \Delta w_k.$$

Step 5. Set $k := k + 1$ and go to Step 1. □

The global convergence of the proposed algorithm is proved in [37]:

Theorem 5.8. *Let $\{w_k\}$ be the sequence generated by Algorithm 5.5. Suppose the followings:*

(i) *The functions f, $g_i, i = 1, \ldots, m$, and X are twice continuously differentiable;*
(ii) *The sequence $\{x_k\}$ remains in a compact set Ω of $\mathbf{R}^n$;*
(iii) *The matrix $G_k + H_k + \frac{1}{\mu}\nabla g(x_k)\nabla g(x_k)^T$ is uniformly positive definite and the matrix $G_k + H_k$ is uniformly bounded;*
(iv) *The scaling matrix T_k is chosen so that $\widetilde{X}_k$ and $\widetilde{Z}_k$ commute and both of the sequences $\{T_k\}$ and $\{T_k^{-1}\}$ are bounded.*

Then there exists at least one accumulation point of $\{w_k\}$, and any accumulation point of the sequence $\{w_k\}$ is an SBKKT point.

6. Concluding Remarks

In this paper, current status of numerical methods for solving the nonlinear SDP problems has been surveyed. It is seen that typical methods for ordinary NLP problems, such as the augmented Lagrangian method, the SQP method and the primal-dual interior point method, have their counterparts in this area.

The augmented Lagrangian method, especially the one with the modified barrier type augmentation, has been applied to the wide area of problems using PENNON [38] by Kočvara and Stingl. The prototype algorithm is rather simple, and the subproblems to be solved are ordinary nonlinear optimization problems. The globally convergent algorithm that assures bounded sequences could be somewhat complex though. The primal-dual interior point method proposed by Yamashita, Yabe and Harada [76] has nice global and local convergence properties, and implemented in the software Numerical Optimizer (its old name is NUOPT). The sequential SDP methods seem to have rooms for more research in theory and practice. Also it is not so obvious that the methods that have linear/quadratic SDP subproblems can be practical or not, because these subproblems should be solved by an interior point method anyway. Notable practical implementation of the sequential SDP method is not yet known to the current authors.

In order to have a brief comparison of the algorithms described in this paper, we cite the comparative study of numerical results of a class of problems solved by PENNON and Numerical Optimizer given in [76]. For detailed explanation about computational aspects of the comparison, we refer to the original paper. The following example problem area is the so called static output feedback (SOF) problems from $COMPl_eib$ library [44]. The SOF-$\mathcal{H}_2$ type problem is defined as follows:

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(X), \\ \text{subject to} & Q \succeq 0, \\ & A(F)Q + QA(F)^T + B_1 B_1^T \preceq 0, \\ & \begin{pmatrix} X & C(F)Q \\ QC(F)^T & Q \end{pmatrix} \succeq 0, \end{array}$$

where $X \in \mathbf{S}^{n_z \times n_z}, F \in \mathbf{R}^{n_u \times n_y}$ and $Q \in \mathbf{S}^{n_x \times n_x}$ are variable matrices to be determined. The matrices $A \in \mathbf{R}^{n_x \times n_x}, B \in \mathbf{R}^{n_x \times n_u}, B_1 \in \mathbf{R}^{n_x \times n_w}, C \in \mathbf{R}^{n_y \times n_x}, C_1 \in \mathbf{R}^{n_z \times n_x}, D_{11} \in$

$\mathbf{R}^{n_z \times n_w}$, $D_{12} \in \mathbf{R}^{n_z \times n_u}$ and $D_{21} \in \mathbf{R}^{n_y \times n_w}$ are given constant matrices, and form the matrices $A(F), B(F), C(F), D(F)$ which appear in the problem definition as follows:

$$\begin{aligned} A(F) &= A + BFC, \\ B(F) &= B_1 + BFD_{21}, \\ C(F) &= C_1 + D_{12}FC, \\ D(F) &= D_{11} + D_{12}FD_{21}. \end{aligned}$$

Table 8 in [76] shows the required CPU(sec) data for solutions of the SOF-$\mathcal{H}_2$ problems by Numerical Optimizer and PENBMI which is a specialized BMI-version of PENNON. The number of variables of the test problem ranges from 7 to 160. The required CPU by Numerical Optimizer varies from 0.03 sec to 12.19 sec, and that by PENBMI varies from 0.18 sec to 223 sec. As the table shows, the performances of these two codes are comparable, and it is difficult to draw any firm conclusion about relative merits of these two algorithms from the table. However we should point out that interior point methods need an interior point for starting the calculation, and it is not so easy to find an interior point for this type of problems.

In summary, the area of nonlinear SDP problems seems to be still in the early stage. We believe there appear new proposals of algorithms and their variations in the future as in ordinary NLP world.

Acknowledgement. The authors are grateful to the referees for their valuable comments that greatly helped to improve the paper. The second author was supported in part by the Grant-in-Aid for Scientific Research (C) 25330030 of Japan Society for the Promotion of Science.

References

[1] F. Alizadeh and D. Goldfarb: Second-order cone programming. *Mathematical Programming*, **95** (2003), 3–51.

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

[3] M.F. Anjos and J.B. Lasserre (eds.): *Handbook on Semidefinite, Conic and Polynomial Optimization* (Springer, New York, 2012).

[4] A. Ben-Tal and A. Nemirovski: *Lectures on Modern Convex Optimization*, MPS-SIAM Series on Optimization (SIAM, Philadelphia, 2001).

[5] D.P. Bertsekas: *Nonlinear Programming – Second Edition* (Athena Scientific, Belmont, MA, 1999).

[6] P.T. Boggs and J.W. Tolle: Sequential quadratic programming. *Acta Numerica*, **4** (1995), 1–51.

[7] J.F. Bonnans, R. Cominetti, and A. Shapiro: Sensitivity analysis of optimization problems under second order regular constraints. *Mathematics of Operations Research*, **23** (1998), 806–831.

[8] J.F. Bonnans, R. Cominetti, and A. Shapiro: Second order optimality conditions based on parabolic second order tangent sets. *SIAM Journal on Optimization*, **9** (1999), 466–492.

[9] J.F. Bonnans and A. Shapiro: *Perturbation Analysis of Optimization Problems* (Springer Verlag, New York, 2000).

[10] S. Boyd, L.E. Ghaoui, E. Feron, and V. Balakrishnan: *Linear Matrix Inequalities in System and Control Theory*, SIAM Studies in Applied Mathematics, **15** (SIAM, Philadelphia, 1994).

[11] S. Boyd and L. Vandenberghe: *Convex Optimization* (Cambridge University Press, 2004).

[12] S. Burer, R.D.C. Monteiro, and Y. Zhang: Solving a class of semidefinite programs via nonlinear programming. *Mathematical Programming*, **93** (2002), 97–122.

[13] S. Burer, R.D.C. Monteiro, and Y. Zhang: Interior-point algorithms for semidefinite programming based on a nonlinear formulation. *Computational Optimization and Applications*, **22** (2002), 49–79.

[14] A.R. Conn, N.I.M. Gould, and P.L. Toint: *Trust-Region Methods*, MPS-SIAM Series on Optimization (SIAM, Philadelphia, 2000).

[15] R. Correa and C.H. Ramírez: A global algorithm for nonlinear semidefinite programming. *SIAM Journal on Optimization*, **15** (2004), 303–318.

[16] M. Diehl, F. Jarre, and C.H. Vogelbusch: Loss of superlinear convergence for an SQP-type method with conic constraints. *SIAM Journal on Optimization*, **16** (2006), 1201–1210.

[17] B. Fares, P. Apkarian, and D. Noll: An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. *International Journal of Control*, **74** (2001), 348–360.

[18] B. Fares, D. Noll, and P. Apkarian: Robust control via sequential semidefinite programming. *SIAM Journal on Control and Optimization*, **40** (2002), 1791–1820.

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

[20] R. Fletcher, S. Leyffer, and P. Toint: *A brief history of filter methods*. Preprint ANL/MCS-P1372-0906, September 26, 2006 (revised October 9, 2006), Argonne National Laboratory, Illinois, USA.

[21] A. Forsgren: Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming*, **88** (2000), 105–128.

[22] A. Forsgren and Ph.E. Gill: Primal-dual interior methods for nonconvex nonlinear programming. *SIAM Journal on Optimization*, **8** (1998), 1132–1152.

[23] R.W. Freund, F. Jarre, and C.H. Vogelbusch: Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. *Mathematical Programming*, **109** (2007), 581–611.

[24] E.H. Fukuda, P.J.S. Silva, and M. Fukushima: Differentiable exact penalty functions for nonlinear second-order cone programs. *SIAM Journal on Optimization*, **22** (2012), 1607–1633.

[25] W. Gómez and H. Ramírez: A filter algorithm for nonlinear semidefinite programming. *Computational and Applied Mathematics*, **29** (2010), 297–328.

[26] C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz: An interior-point method for semidefinite programming. *SIAM Journal on Optimization*, **6** (1996), 342–361.

[27] R.A. Horn and C.R. Johnson: *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, 1991).

[28] X.X. Huang, K.L. Teo, and X.Q. Yang: Approximate augmented Lagrangian functions and nonlinear semidefinite programs. *Acta Mathematica Sinica, English Series*, **22** (2006), 1283–1296.

[29] X.X. Huang, X.Q. Yang, and K.L. Teo: A sequential quadratic penalty method for nonlinear semidefinite programming. *Optimization*, **52** (2003), 715–738.

[30] X.X. Huang, X.Q. Yang, and K.L. Teo: Augmented Lagrangian and nonlinear semidefinite programs. *Variational Analysis and Applications. Nonconvex Optimization and Its Applications*, **79** (2005), 513–529.

[31] X.X. Huang, X.Q. Yang, and K.L. Teo: Lower-order penalization approach to nonlinear semidefinite programming. *Journal of Optimization Theory and Applications*, **132** (2007), 1–20.

[32] F. Jarre: An interior method for nonconvex semidefinite programs. *Optimization and Engineering*, **1** (2000), 347–372.

[33] F. Jarre: Some aspects of nonlinear semidefinite programming. In E.W. Sachs and R. Tichatschke (Eds.): *System Modeling and Optimization XX* (Kluwer Academic Publishers, 2003), 55–69.

[34] Y. Kanno and I. Takewaki: Sequential semidefinite program for maximum robustness design of structures under load uncertainty. *Journal of Optimization Theory and Applications*, **130** (2006), 265–287.

[35] C. Kanzow, I. Ferenczi, and M. Fukushima: On the local convergence of semismooth Newton methods for linear and nonlinear second-order cone programs without strict complementarity. *SIAM Journal on Optimization*, **20** (2009), 297–320.

[36] C. Kanzow, C. Nagel, H. Kato, and M. Fukushima: Successive linearization methods for nonlinear semidefinite programs. *Computational Optimization and Applications*, **31** (2005), 251–273.

[37] A. Kato, H. Yabe, and H. Yamashita: An interior point method with a primal-dual quadratic barrier penalty function for nonlinear semidefinite programming. *Journal of Computational and Applied Mathematics*, **275** (2015), 148–161.

[38] M. Kočvara and M. Stingl: PENNON – A code for convex nonlinear and semidefinite programming. *Optimization Methods and Software*, **18** (2003), 317–333.

[39] M. Kočvara and M. Stingl: PENNON – A generalized augmented Lagrangian method for semidefinite programming. In G. Di Pillo and A. Murli (eds.): *High Performance Algorithms and Software for Nonlinear Optimization* (Kluwer Academic Publishers, Boston, 2003), 303–321.

[40] M. Kočvara and M. Stingl: Solving nonconvex SDP problems of structured optimization with stability control. *Optimization Methods and Software*, **19** (2004), 595–609.

[41] M. Kočvara and M. Stingl: On the solution of large-scale SDP problems by the modified barrier method using iterative solvers. *Mathematical Programming*, **109** (2007), 413–444.

[42] M. Kojima, S. Shindoh, and S. Hara: Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, **7** (1997), 86–125.

[43] H. Konno, N. Kawadai, and D. Wu: Estimation of failure probability using semi-definite Logit model. *Computational Management Science*, **1** (2003), 59–73.

[44] F. Leibfritz and W. Lipinski: Description of the benchmark examples in $COMPl_eib$ 1.0. Technical Report, Department of Mathematics, University of Trier, D-54286, Trier, Germany, 2003.

[45] F. Leibfritz and E.M.E. Mostafa: An interior point constrained trust region method

for a special class of nonlinear semidefinite programming problems. *SIAM Journal on Optimization*, **12** (2002), 1048–1074.

[46] C.J. Li and W.Y. Sun: On filter-successive linearization methods for nonlinear semidefinite programming. *Science in China Series A Mathematics*, **52** (2009), 2341–2361.

[47] H.Z. Luo, H.X. Wu, and G.T. Chen: On the convergence of augmented Lagrangian methods for nonlinear semidefinite programming. *Journal of Global Optimization*, **54** (2012), 599–618.

[48] R.D.C. Monteiro: Primal-dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, **7** (1997), 663–678.

[49] L. Mosheyev and M. Zibulevsky: Penalty/barrier multiplier algorithm for semidefinite programming. *Optimization Methods and Software*, **13** (2000), 235–261.

[50] Y.E. Nesterov and M.J. Todd: Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations Research*, **22** (1997), 1–42.

[51] Y.E. Nesterov and M.J. Todd: Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimization*, **8** (1998), 324–364.

[52] J. Nocedal and S.J. Wright: *Numerical Optimization* (Springer, New York, 1999).

[53] D. Noll: Local convergence of an augmented Lagrangian method for matrix inequality constrained programming. *Optimization Methods and Software*, **22** (2007), 777–802.

[54] D. Noll, M. Torki, and P. Apkarian: Partially augmented Lagrangian method for matrix inequality constraints. *SIAM Journal on Optimization*, **15** (2004), 161–184.

[55] R. Polyak: Modified barrier functions (theory and methods). *Mathematical Programming*, **54** (1992), 177–222.

[56] H.D. Qi: Local duality of nonlinear semidefinite programming. *Mathematics of Operations Research*, **34** (2009), 124–141.

[57] H. Qi and D. Sun: A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications*, **28** (2006), 360–385.

[58] S.M. Robinson: Generalized equations and their solutions, part II: Applications to nonlinear programming. *Mathematical Programming Study*, **19** (1982), 200–221.

[59] A. Shapiro: First and second order analysis of nonlinear semidefinite programs. *Mathematical Programming*, **77** (1997), 301–320.

[60] A. Shapiro and K. Scheinberg: Duality and optimality conditions. In H. Wolkowicz, R. Saigal and L. Vandenberghe (eds.): *Handbook of Semidefinite Programming, Theory, Algorithms and Applications* (International Series in Operations Research and Management Science, Kluwer, Boston, 2000), 67–110.

[61] A. Shapiro and J. Sun: Some properties of the augmented Lagrangian in cone constrained optimization. *Mathematics of Operations Research*, **29** (2004), 479–491.

[62] M. Stingl: *On the Solution of Nonlinear Semidefinite Programs by Augmented Lagrangian Methods*. Doctoral Thesis, University of Erlangen, 2005.

[63] M. Stingl, M. Kočvara, and G. Leugering: A sequential convex semidefinite programming algorithm with an application to multiple-load free material optimization. *SIAM Journal on Optimization*, **20** (2009), 130–155.

[64] D. Sun, J. Sun, and L. Zhang: The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Mathematical Programming*, **114** (2008), 349–391.

[65] J. Sun, L.W. Zhang, and Y. Wu: Properties of the augmented Lagrangian in nonlinear semidefinite optimization. *Journal of Optimization Theory and Applications*, **129** (2006), 437–456.

[66] M.J. Todd: Semidefinite optimization. *Acta Numerica*, **10** (2001), 515–560.

[67] M.J. Todd, K.C. Toh, and R.H. Tütüncü: On the Nesterov-Todd direction in semidefinite programming. *SIAM Journal on Optimization*, **8** (1998), 769–796.

[68] L. Vandenberghe and S. Boyd: Semidefinite programming. *SIAM Review*, **38** (1996), 49–95.

[69] L. Vandenberghe, S. Boyd, and S.P. Wu: Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, **19** (1998), 499–533.

[70] H. Wolkowicz, R. Saigal, and L. Vandenberghe (eds.): *Handbook of Semidefinite Programming – Theory, Algorithms and Applications* (International Series in Operations Research and Management Science, Kluwer, Boston, 2000).

[71] H. Wu, H. Luo, X. Ding, and G. Chen: Global convergence of modified augmented Lagrangian methods for nonlinear semidefinite programming. *Computational Optimization and Applications*, **56** (2013), 531–558.

[72] Y. Yamakawa and N. Yamashita: Differentiable merit function for shifted perturbed Karush-Kuhn-Tucker conditions of nonlinear semidefinite programming. to appear in *Pacific Journal of Optimization*.

[73] H. Yamashita and H. Yabe: An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization. *SIAM Journal on Optimization*, **14** (2003), 479–499.

[74] H. Yamashita and H. Yabe: A primal-dual interior point method for nonlinear optimization over second-order cones. *Optimization Methods and Software*, **24** (2009), 407–426.

[75] H. Yamashita and H. Yabe: Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming*, **132** (2012), 1–30.

[76] H. Yamashita, H. Yabe, and K. Harada: A primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming*, **135** (2012), 89–121.

[77] L. Yang and B. Yu: A homotopy method for nonlinear semidefinite programming. *Computational Optimization and Applications*, **56** (2013), 81–96.

[78] Z.B. Zhu and H.L. Zhu: A filter method for nonlinear semidefinite programming with global convergence. *Acta Mathematica Sinica, English Series*, **30** (2014), 1810–1826.

Hiroshi Yabe
Department of Mathematical Information Science
Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku
Tokyo 162-8601, Japan
E-mail: `yabe@rs.kagu.tus.ac.jp`