Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 11, No. 3, pp. 557–579, July 2015

A DIFFERENTIABLE MERIT FUNCTION FOR SHIFTED PERTURBED KARUSH-KUHN-TUCKER CONDITIONS OF THE NONLINEAR SEMIDEFINITE PROGRAMMING

Yuya Yamakawa and Nobuo Yamashita

Abstract: In this paper we consider a primal-dual interior point method for solving nonlinear semidefinite programming problems, which is based on shifted perturbed Karush-Kuhn-Tucker (KKT) conditions. The main task addressed by the interior point method is to obtain a point that approximately satisfies shifted perturbed KKT conditions. First, we propose a differentiable merit function whose stationary points always satisfy the conditions. This function is an extension of the one proposed by Forsgren and Gill for nonlinear programming problems. Next, we develop a Newton-type method that finds a stationary point of the merit function. We show the global convergence of the proposed Newton-type method under some mild conditions. Finally, we report some numerical results, which show that the performance of the proposed method is comparable to the existing primal-dual interior point method based on perturbed KKT conditions.

Key words: *merit function, Newton-type method, nonlinear semidefinite programming, primal-dual interior point method, shifted perturbed Karush-Kuhn-Tucker conditions*

Mathematics Subject Classification: *90C22, 90C26, 90C51*

1 Introduction

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

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

where $f:\mathbf{R}^n\to\mathbf{R}$, $g:\mathbf{R}^n\to\mathbf{R}^m$ and $X:\mathbf{R}^n\to\mathbf{S}^d$ are twice continuously differentiable functions, and $\mathbf{S}^d$ denotes the set of $d\times d$ real symmetric matrices. Let $\mathbf{S}^d_{++}$ ($\mathbf{S}^d_{+}$) denote the set of $d\times d$ real symmetric positive (semi)definite matrices. For a matrix $M\in\mathbf{S}^d$, $M\succeq 0$ and $M\succ 0$ mean that $M\in\mathbf{S}^d_{+}$ and $M\in\mathbf{S}^d_{++}$, respectively. If the functions f, g and X are linear, the nonlinear SDP (1.1) can be reduced to a linear SDP.

Nonlinear SDP includes a wide class of mathematical programming problems and it has many applications [2, 6, 8, 22, 26]. Linear programming, second order cone programming, linear SDP and nonlinear programming can all be recast as nonlinear SDP. Linear SDP has been studied extensively by many researchers [1, 4, 7, 23–25]. However, there exist important applications that are formulated as nonlinear SDP, but cannot be reduced to linear SDP. For example, the Gaussian channel capacity problem [26], the minimization (or maximization) of the minimal (or maximal) eigenvalue problem [18], the nearest correlation matrix problem

[19] and the static output feedback problem [20] are such applications. Thus, it would be useful to develop methods for solving nonlinear SDP.

Previous studies have proposed several solution methods for nonlinear SDP [6, 10, 11, 14, 21, 28]. Basically, these methods are extensions of existing methods for nonlinear programming, such as sequential quadratic programming methods, successive linearization methods, augmented Lagrangian methods and interior point methods.

Freund, Jarre and Vogelbusch [6] proposed a sequential semidefinite programming method for nonlinear SDP. However, they only considered the case where the objective function is quadratic and the constraint functions are affine. Kanzow, Nagel, Kato and Fukushima [11] extended the successive linearization method using a certain exact penalty function and a trust region-type technique. They showed that the extended method is globally convergent under rather strong assumptions on the generated sequence, which are not verified in advance. Stingl [21] presented an augmented Lagrangian method for nonlinear SDP. He showed its global convergence under rather restrictive conditions such as the second order sufficient optimality condition. Yamashita, Yabe and Harada [28] applied the primal-dual interior point method to nonlinear SDP and they exploited a nondifferentiable L_1 merit function to determine a step length. They showed the global convergence of their algorithm under some unclear assumptions regarding the generated sequences. These assumptions are discussed in Section 4.3.

The aim of the present study is to propose an interior point method for (1.1) that converges globally under milder conditions compared with the methods described above. In particular, we specify the conditions related to the problem data, i.e., f, g and X. We also show that these conditions hold for linear SDP.

Recently, Kato, Yabe and Yamashita [12] proposed a primal-dual interior point method based on shifted perturbed Karush-Kuhn-Tucker (KKT) conditions, which is an extension of the method proposed by Forsgren and Gill [5] for nonlinear programming. This method generates points that satisfy shifted perturbed KKT conditions at each iteration. In order to find such points, Kato, Yabe and Yamashita [12] used a merit function, which is an extension of [27]. However, since the merit function is rather complicated, it might be difficult to implement it appropriately. In this paper, we propose a new merit function F whose stationary points satisfy shifted perturbed KKT conditions. This is an extension of a merit function [5] developed for nonlinear programming. It consists of simple functions of matrices, such as log-determinant and trace, and hence it is easy to implement. We show the following important properties of the merit function F.

(i) The merit function F is differentiable;

(ii) Any stationary point of the merit function F is a shifted perturbed KKT point;

(iii) The level set of the merit function F is bounded under some reasonable assumptions.

Kato, Yabe and Yamashita [12] also showed that their merit function satisfies (i) and (ii), but they did not show the property (iii). These properties mean that we can find a point that satisfies shifted perturbed KKT conditions by minimizing the merit function F. To minimize F, we also propose a Newton-type method based on nonlinear equations in shifted perturbed KKT conditions. We show that the Newton direction is sufficiently descent for the merit function F. As a result, we prove the global convergence of the proposed Newton-type method. These details are provided in Section 4.

The present paper is organized as follows. In Section 2, we introduce some operators and important concepts, which are used in the subsequent sections. In Section 3, we present a primal-dual interior point algorithm based on shifted perturbed KKT conditions. In Section

4, we first propose a merit function F for a shifted perturbed KKT point and present its properties. Secondly, we propose a Newton-type algorithm that minimizes the merit function. Moreover, we prove the global convergence of the Newton-type algorithm. In Section 5, we report some numerical results for the proposed method. Finally, we make some concluding remarks in Section 6.

Throughout this paper, we use the following notations. Let p and q be positive integers. For matrices $A, B \in \mathbf{R}^{p\times q}$, $\langle A, B\rangle$ denotes the inner product of A and B defined by $\langle A, B\rangle \equiv \mathrm{tr}(A^\top B)$, where $\mathrm{tr}(M)$ denotes the trace of a square matrix M, and the superscript $\top$ denotes the transposition of a vector or a matrix. Note that if $q = 1$, then $\langle\cdot,\cdot\rangle$ denotes the inner product of vectors in $\mathbf{R}^p$. For a given vector $w \in \mathbf{R}^p$ and a matrix $W \in \mathbf{R}^{p\times q}$, w_i denotes the i-th element of the vector w, and W_{ij} denotes the (i, j)-th element of the matrix W. Moreover, $\|w\|$ denotes the Euclidean norm of the vector w defined by $\|w\| \equiv \sqrt{\langle w, w\rangle}$, and $\|W\|_F$ denotes the Frobenius norm of the matrix W defined by $\|W\|_F \equiv \sqrt{\langle W, W\rangle}$. Let $\mathcal{V} \equiv \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^d$. For a given $v \in \mathcal{V}$, we use the following notations for simplicity.

$$v = \begin{bmatrix} x \\ y \\ Z \end{bmatrix} \quad \text{or} \quad v = (x, y, Z),$$

where $x \in \mathbf{R}^n$, $y \in \mathbf{R}^m$ and $Z \in \mathbf{S}^d$, respectively. We also define the inner product $\langle\cdot,\cdot\rangle$ and the norm $\|\cdot\|$ on $\mathcal{V}$ as $\langle v_1, v_2\rangle \equiv \langle x_1, x_2\rangle + \langle y_1, y_2\rangle + \langle Z_1, Z_2\rangle$ and $\|v\| \equiv \sqrt{\langle v, v\rangle}$, where $v_1 = (x_1, y_1, Z_1) \in \mathcal{V}$ and $v_2 = (x_2, y_2, Z_2) \in \mathcal{V}$. For a given matrix $U \in \mathbf{S}^d$, $\lambda_1(U), \ldots, \lambda_d(U)$ denote the eigenvalues of the matrix U. In particular, $\lambda_{\min}(U)$ and $\lambda_{\max}(U)$ denote the minimum and maximum eigenvalues of the matrix U, respectively. For a given matrix $V \in \mathbf{S}^d_+$, $V^{\frac{1}{2}} \in \mathbf{S}^d_+$ denotes the matrix such that $V = V^{\frac{1}{2}}V^{\frac{1}{2}}$. Note that $V^{\frac{1}{2}} \equiv Q\Lambda Q^\top$, where

$$\Lambda = \begin{bmatrix} \sqrt{\lambda_1(V)} & & \mathrm{O} \\ & \ddots & \\ \mathrm{O} & & \sqrt{\lambda_d(V)} \end{bmatrix},$$

and Q is a certain orthogonal matrix such that $V = Q\Lambda^2 Q^\top$. Let $\Phi : P_1 \times P_2 \to P_3$, where P_1 and P_2 are open sets. We denote a Fréchet derivative of Φ as $\nabla\Phi$. We further denote a Fréchet derivative of Φ with respect to a variable $Z \in P_1$ as $\nabla_Z\Phi$. Moreover, if Φ is a vector-valued function, then J_Φ denotes a Jacobian of Φ.

2 Preliminaries

In this section, we first introduce some operators. Then we present some useful properties of the log-determinant function on $\mathbf{S}^d$. Moreover, we introduce the (approximate) KKT conditions related to the primal-dual interior point method for nonlinear SDP.

2.1 Some operators and their properties

Let $U, V \in \mathbf{S}^d, P, Q \in \mathbf{R}^{d\times d}$ and $x, w \in \mathbf{R}^n$. We use the following notations.

(i) The product $\circ$ of the matrices U and V is defined by $U \circ V \equiv \frac{UV+VU}{2}$.

(ii) The partial derivative of $X(x)$ with respect to x_i is denoted by $A_i(x) \in \mathbf{S}^d$, that is, $A_i(x) \equiv \frac{\partial}{\partial x_i}X(x)$ for $i = 1, \ldots, n$.

(iii) The operator $\mathcal{A}(x)$ from $\mathbf{R}^n$ to $\mathbf{S}^d$ is defined by $\mathcal{A}(x)w \equiv w_1 A_1(x) + \ldots + w_n A_n(x)$.

(iv) The adjoint operator of $\mathcal{A}(x)$ is denoted by $\mathcal{A}^*(x)$, that is, $\mathcal{A}^*(x)U = [\langle A_1(x), U\rangle, \ldots, \langle A_n(x), U\rangle]^\top$ for all $U \in \mathbf{S}^d$.

(v) The operator $P \odot Q$ from $\mathbf{S}^d$ to $\mathbf{S}^d$ is defined by

$$(P \odot Q)U \equiv \frac{1}{2}(PUQ^\top + QUP^\top). \tag{2.1}$$

If $X(x) = x_1 A_1 + \ldots + x_n A_n$ with some constant matrices $A_i \in \mathbf{S}^d, i = 1, \ldots, n$, then $A_i(x) = A_i, i = 1, \ldots, n$. Note that $U \circ V = 0$ is equivalent to $UV = 0$ if U and V are symmetric positive semidefinite.

2.2 Properties of the log-determinant function

Let $\phi : \mathbf{S}^d_{++} \to \mathbf{R}$ be defined by $\phi(M) \equiv -\log\det M$. Let Ω be defined by $\Omega \equiv \{x \in \mathbf{R}^n | X(x) \succ 0\}$, and let $\varphi : \Omega \to \mathbf{R}$ be defined by

$$\varphi(x) \equiv \phi(X(x)). \tag{2.2}$$

We first give the differentiability and convexity of φ.

Proposition 2.1.

(a) *The function φ is differentiable on Ω, and its derivative is given by $\nabla\varphi(x) = -\mathcal{A}^*(x)X(x)^{-1}$.*

(b) *Suppose that*

$$X(\lambda u + (1-\lambda)v) - \lambda X(u) - (1-\lambda)X(v) \succeq 0 \quad \text{for } \lambda \in [0,1] \text{ and } u, v \in \Omega. \tag{2.3}$$

Then φ is convex on Ω. Moreover, if X is injective on Ω, then φ is strictly convex.

(c) *Suppose that (2.3) holds. Suppose also that $A_1(x), \ldots, A_n(x)$ are linearly independent for all $x \in \Omega$. Then φ is strictly convex.*

Proof. (a) From [23, Section 5] and the chain rule, the desired equality holds.
(b) First note that $\det A \leq \det B$ if $0 \preceq A$ and $0 \preceq B - A$ from [9, Corollary 7.7.4]. It then follows from (2.3) that for any $\lambda \in [0,1]$ and $u, v \in \Omega$, $\det[\lambda X(u) + (1-\lambda)X(v)] \leq \det[X(\lambda u + (1-\lambda)v)]$. Since $-\log$ is a decreasing function on $(0,\infty)$ and ϕ is strictly convex from [9, Theorem 7.6.7], we have $\varphi(\lambda u + (1-\lambda)v) \leq \lambda\varphi(u) + (1-\lambda)\varphi(v)$.

Suppose that $u \neq v$. Then, since X is injective on Ω, $X(u) \neq X(v)$. Moreover, since ϕ is strictly convex, $\varphi(\lambda u + (1-\lambda)v) < \lambda\varphi(u) + (1-\lambda)\varphi(v)$ for $\lambda \in (0,1)$. Thus, φ is strictly convex.
(c) Since X is twice differentiable, $X(v + \lambda(u-v)) - X(v) = \lambda\mathcal{A}(v)(u-v) + o(\lambda)$ for $u, v \in \Omega$ and $\lambda \in (0,1)$. Then (2.3) can be written as $\lambda\mathcal{A}(v)(u-v) - \lambda(X(u) - X(v)) + o(\lambda) \succeq 0$. Dividing both sides by λ, we have $\mathcal{A}(v)(u-v) - X(u) + X(v) + \frac{o(\lambda)}{\lambda} \succeq 0$. Letting $\lambda \to 0$ yields

$$\mathcal{A}(v)(u-v) - X(u) + X(v) \succeq 0.$$

Let $M \equiv \mathcal{A}(v)(u-v) - X(u) + X(v)$. Since M and $X(v)^{-1}$ are symmetric positive semidefinite, there exist $M^{\frac{1}{2}}$ and $X(v)^{-\frac{1}{2}}$. Then we have

$$\left\langle X(v)^{-1}, M \right\rangle = \operatorname{tr}(X(v)^{-1}M) = \operatorname{tr}(X(v)^{-\frac{1}{2}} M^{\frac{1}{2}} M^{\frac{1}{2}} X(v)^{-\frac{1}{2}}) = \|M^{\frac{1}{2}} X(v)^{-\frac{1}{2}}\|_F^2.$$

From the convexity of ϕ, (2.2) and $\nabla\phi(M) = -M^{-1}$, we obtain

$$\begin{aligned}
\varphi(u) - \varphi(v) &= \phi(X(u)) - \phi(X(v)) \\
&\geq \left\langle -X(v)^{-1}, X(u) - X(v) \right\rangle \\
&= \left\langle X(v)^{-1}, M \right\rangle + \left\langle X(v)^{-1}, -\mathcal{A}(v)(u-v) \right\rangle \\
&= \|M^{\frac{1}{2}} X(v)^{-\frac{1}{2}}\|_F^2 + \left\langle -\mathcal{A}^*(v) X(v)^{-1}, u - v \right\rangle \\
&\geq \left\langle \nabla\varphi(v), u - v \right\rangle,
\end{aligned} \tag{2.4}$$

where the last inequality follows from (a).

Since φ is convex by (b), it suffices for (c) to show that $u = v$ if and only if $\varphi(u) - \varphi(v) = \langle \nabla\varphi(v), u - v \rangle$. If $u = v$, then it is clear that $\varphi(u) - \varphi(v) = \langle \nabla\varphi(v), u - v \rangle$. Conversely, suppose that $\varphi(u) - \varphi(v) = \langle \nabla\varphi(v), u - v \rangle$, then the equality holds in (2.4). It follows from (2.4) that $\|M^{\frac{1}{2}} X(v)^{-\frac{1}{2}}\|_F = 0$ and $\phi(X(u)) - \phi(X(v)) = \left\langle -X(v)^{-1}, X(u) - X(v) \right\rangle$. Then, we have $\mathcal{A}(v)(u - v) = 0$ from the definition of M. Since $A_1(x), \ldots, A_n(x)$ are linearly independent for all $x \in \Omega$, we have $u = v$. □

Note that Proposition 2.1 (b) does not assume the differentiability of X.

We next give sufficient conditions under which matrices in a level set of ϕ is uniformly positive definite, which is a key property for the level boundedness of the merit function proposed in Section 4.

Proposition 2.2. *For a given $\gamma \in \mathbf{R}$, let $\mathcal{L}_\phi(\gamma) = \{U \in \mathbf{S}^d_{++} | \phi(U) \leq \gamma\}$. Let Γ be a bounded subset of $\mathbf{S}^d$. Then, there exists $\underline{\lambda} > 0$ such that $\lambda_{\min}(U) \geq \underline{\lambda}$ for all $U \in \mathcal{L}_\phi(\gamma) \cap \Gamma$.*

Proof. Suppose the contrary, that is, there exists a sequence $\{U_j\} \subset \mathcal{L}_\phi(\gamma) \cap \Gamma$ such that $\lambda_{\min}(U_j) \to 0$ as $j \to \infty$. Then

$$-\log \lambda_{\min}(U_j) \to \infty. \tag{2.5}$$

Since $U_j \in \mathcal{L}_\phi(\gamma)$, we have $\gamma \geq \phi(U_j) = -\log\det U_j = -\sum_{i=1}^d \log \lambda_i(U_j)$. Then, (2.5) implies that there exist an index k and an infinite subset $\mathcal{J}$ such that $\lim_{j\to\infty, j\in\mathcal{J}} -\log \lambda_k(U_j) = -\infty$, that is, $\lim_{j\to\infty, j\in\mathcal{J}} \lambda_k(U_j) = \infty$. However, this is contrary to the boundedness of $\{U_j\}$. Therefore, there exists $\underline{\lambda} > 0$ such that $\lambda_{\min}(U) \geq \underline{\lambda}$ for all $U \in \mathcal{L}_\phi(\gamma) \cap \Gamma$. □

2.3 Shifted perturbed KKT conditions for nonlinear SDP

We first introduce the optimality conditions for nonlinear SDP (1.1). Let $v = (x, y, Z)$. The Lagrangian function L of (1.1) is given by

$$L(v) \equiv f(x) - g(x)^\top y - \langle X(x), Z \rangle,$$

where $y \in \mathbf{R}^m$ and $Z \in \mathbf{S}^d$ are the Lagrange multiplier vector and matrix for $g(x) = 0$ and $X(x) \succeq 0$, respectively. The gradient of the Lagrangian function L with respect to x is given by

$$\nabla_x L(v) = \nabla f(x) - J_g(x)^\top y - \mathcal{A}^*(x) Z.$$

The Karush-Kuhn-Tucker (KKT) conditions of (1.1) are written as

$$\begin{bmatrix} \nabla_x L(v) \\ g(x) \\ X(x)Z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.6}$$

and

$$X(x) \succeq 0, \qquad Z \succeq 0. \tag{2.7}$$

Most of the solution methods for nonlinear SDP is developed to find a point $v = (x, y, Z)$ that satisfies the KKT conditions. However, it is difficult to get such a point directly due to the complementarity condition $X(x)Z = 0$ with $X(x) \succeq 0$ and $Z \succeq 0$. To overcome this difficulty, the primal-dual interior point method proposed by Yamashita, Yabe and Harada [28] exploits the following perturbed KKT conditions with a parameter $\mu > 0$.

$$\begin{bmatrix} \nabla_x L(v) \\ g(x) \\ X(x)Z - \mu I \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad X(x) \succ 0, \qquad Z \succ 0. \tag{2.8}$$

They [28] proposed a Newton-type algorithm to get a point satisfying the perturbed KKT conditions.

In this paper, we focus on the following shifted perturbed KKT conditions. For $\mu > 0$,

$$\begin{bmatrix} \nabla_x L(v) \\ g(x) + \mu y \\ X(x)Z - \mu I \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.9}$$

and

$$X(x) \succ 0, \qquad Z \succ 0. \tag{2.10}$$

The above shifted perturbed KKT conditions are derived by Forsgren and Gill [5] for nonlinear programming. In what follows, we call a point v satisfying the shifted perturbed KKT conditions a *shifted perturbed KKT point*. Furthermore, we define the set $\mathcal{W} \subset \mathcal{V}$ by

$$\mathcal{W} \equiv \{(x, y, Z) \in \mathcal{V} \mid X(x) \succ 0, Z \succ 0\}.$$

We call a point $v \in \mathcal{W}$ an *interior point*.

3 Primal-Dual Interior Point Method Based on Shifted Perturbed KKT Conditions

In this section, we introduce a prototype of an interior point algorithm based on the shifted perturbed KKT conditions (2.9) and (2.10). Note that the prototype has already been proposed in [12].

The primal-dual interior point method generates a sequence $\{v_k\} \subset \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^d$ such that the point v_k approximately satisfies the shifted perturbed KKT conditions (2.9) and (2.10) with $\mu = \mu_k > 0$, where $\{\mu_k\}$ is a positive sequence such that $\mu_k \to 0$ $(k \to \infty)$.

To construct a concrete algorithm, it is important to define an approximate shifted perturbed KKT point, and to provide a method for finding an approximate shifted perturbed KKT point.

We first give a concrete definition of an approximate shifted perturbed KKT point. To this end, let

$$r(v;\mu) \equiv \begin{bmatrix} \nabla_x L(v) \\ g(x) + \mu y \\ X(x)Z - \mu I \end{bmatrix} \quad \text{and} \quad \rho(v;\mu) \equiv \sqrt{\left\| \begin{bmatrix} \nabla_x L(v) \\ g(x) + \mu y \end{bmatrix} \right\|^2 + \|X(x)Z - \mu I\|_F^2}.$$

For a given $\varepsilon > 0$, a point $v \in \mathcal{W}$ is called an *approximate shifted perturbed KKT point* if it satisfies $\rho(v;\mu) \leq \varepsilon$. Note that $\rho(v;\mu) = 0$ and $v \in \mathcal{W}$ if and only if v is a shifted perturbed KKT point. Note also that $\rho(v;0) = 0, X(x) \succeq 0$ and $Z \succeq 0$ if and only if v is a KKT point of the nonlinear SDP (1.1).

Now, we give the framework of the primal-dual interior point method.

Algorithm 1.

Step 0. Let $\{\mu_k\}$ be a positive sequence such that $\mu_k \to 0$ as $k \to \infty$. Choose constants $\sigma, \epsilon > 0$. Set $k = 0$.

Step 1. Find an approximate shifted perturbed KKT point v_{k+1} with $\varepsilon = \sigma\mu_k$, that is, $v_{k+1} \in \mathcal{W}$ such that $\rho(v_{k+1};\mu_k) \leq \sigma\mu_k$.

Step 2. If $\rho(v_{k+1};0) \leq \epsilon$, then stop.

Step 3. Set $k = k + 1$ and go to Step 1.

The following theorem gives conditions for the global convergence of Algorithm 1. It can be proved in a way similar to [28, Theorem 1]. Thus, we omit the proof.

Theorem 3.1. *Suppose that an approximate shifted perturbed KKT point v_{k+1} is found in Step 1 at every iteration. Moreover, suppose that the sequence $\{x_k\}$ is bounded and that the Mangasarian-Fromovitz constraint qualification condition holds at any accumulation point of $\{x_k\}$, i.e., for any accumulation point x^* of $\{x_k\}$, the matrix $J_g(x^*)$ is of full rank and there exists a nonzero vector $w \in \mathbf{R}^n$ such that*

$$J_g(x^*)w = 0 \quad \text{and} \quad X(x^*) + \sum_{i=1}^{n} w_i A_i(x^*) \succ 0.$$

Then, the sequences $\{y_k\}$ and $\{Z_k\}$ are bounded, and any accumulation point of $\{v_k\}$ satisfies the KKT conditions (2.6) and (2.7). □

The theorem guarantees the global convergence if an approximate shifted perturbed KKT point v_{k+1} is found at each iteration. Thus it is important to present concrete algorithm that finds the point. In the next section, we will propose a merit function for the shifted perturbed KKT point and a Newton-type algorithm for solving the unconstrained minimization problem of the merit function.

4 Finding a Shifted Perturbed KKT Point

In order to find the approximate shifted perturbed KKT point in Step 1 of Algorithm 1, we may solve the following unconstrained minimization problem:

$$\begin{array}{ll} \text{minimize} & \rho(v;\mu)^2, \\ \text{subject to} & v \in \mathcal{V}, \end{array}$$

where $\mathcal{V} = \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{S}^d$. Unfortunately, a stationary point of the problem is not necessarily a shifted perturbed KKT point unless $\nabla r(v;\mu)$ is invertible. In this section, we first construct a differentiable merit function F whose stationary point is always a shifted perturbed KKT point. Moreover, we show that a Newton direction for the nonlinear equations $r(v;\mu) = 0$ is a descent direction of the merit function F. Next, we propose a Newton-type algorithm for solving the unconstrained minimization of the merit function F. Finally, we show that the proposed algorithm finds a shifted perturbed KKT point under some mild assumptions.

4.1 Merit function and its properties

We propose the following merit function $F : \mathcal{W} \to \mathbf{R}$ for the shifted perturbed KKT point.

$$F(x, y, Z) \equiv F_{BP}(x) + \nu F_{PD}(x, y, Z),$$

where ν is a positive constant, and the functions $F_{BP} : \Omega \to \mathbf{R}$ and $F_{PD} : \mathcal{W} \to \mathbf{R}$ are defined by

$$F_{BP}(x) \equiv f(x) + \frac{1}{2\mu}\|g(x)\|^2 - \mu \log\det X(x),$$

and

$$F_{PD}(x, y, Z) \equiv \frac{1}{2\mu}\|g(x) + \mu y\|^2 + \langle X(x), Z\rangle - \mu \log\det X(x) \det Z,$$

respectively. The functions F_{BP} and F_{PD} are called the primal barrier penalty function and the primal-dual barrier penalty function, respectively. Note that F is convex with respect to x when f is convex and g, X are affine. The merit function F is an extension of the one proposed by Forsgren and Gill [5] for nonlinear programming.

Remark 4.1. For the shifted perturbed KKT conditions, Kato, Yabe and Yamashita [12] also proposed the following merit function $\widetilde{F} : \mathcal{W} \to \mathbf{R}$.

$$\widetilde{F}(x, y, Z) \equiv F_{BP}(x) + \nu \widetilde{F}_{PD}(x, y, Z),$$

where $\widetilde{F}_{PD}(w)$ is defined by

$$\widetilde{F}_{PD}(x, y, Z) \equiv \frac{1}{2}\|g(x) + \mu y\|^2 + \log \frac{\frac{1}{d}\langle X(x), Z\rangle + \|Z^{\frac{1}{2}} X(x) Z^{\frac{1}{2}} - \mu I\|_F^2}{(\det(X(x)Z))^{\frac{1}{d}}}.$$

They showed that $\widetilde{F}$ has nice properties like the merit function F. However, $\widetilde{F}$ is more complicated than F, and hence it might not be easy to implement the Newton-type method based on $\widetilde{F}$ in [12]. Furthermore, even if f is convex and g, X are affine, $\widetilde{F}$ is not necessarily convex with respect to x.

In the rest of this subsection, we present some useful properties of the merit function F such as the differentiability, the equivalence between a stationary point of F and a shifted perturbed KKT point, and the level boundedness.

First of all, we present a concrete formula of the derivatives of the merit function F.

Theorem 4.2. *The merit function F is differentiable on $\mathcal{W}$. Moreover, its derivative is given by*

$$\nabla F(w) = \begin{bmatrix} \nabla F_{BP}(x) + \nu \nabla_x F_{PD}(w) \\ \nu \nabla_y F_{PD}(w) \\ \nu \nabla_Z F_{PD}(w) \end{bmatrix},$$

where $\nabla F_{BP}(x) = \nabla f(x) + \frac{1}{\mu} J_g(x)^\top g(x) - \mu \mathcal{A}^*(x) X(x)^{-1}, \nabla_x F_{PD}(w) = \frac{1}{\mu} J_g(x)^\top (g(x) + \mu y) + \mathcal{A}^*(x)(Z - \mu X(x)^{-1}), \nabla_y F_{PD}(w) = g(x) + \mu y$ *and* $\nabla_Z F_{PD}(w) = X(x) - \mu Z^{-1}$. □

Next, we show the equivalence between a stationary point of the merit function F and a shifted perturbed KKT point.

Theorem 4.3. *A point* $w^* \in \mathcal{W}$ *is a stationary point of the merit function* F *if and only if* w^* *is a shifted perturbed KKT point.*

Proof. First, let $w^* = (x^*, y^*, Z^*) \in \mathcal{W}$ be a stationary point of the merit function F. Theorem 4.2 yields that

$$\nabla f(x^*) + \frac{1}{\mu} J_g(x^*)^\top \{(1+\nu) g(x^*) + \nu \mu y^*\} + \mathcal{A}^*(x^*) \left\{ \nu Z^* - \mu(1+\nu) X(x^*)^{-1} \right\} = 0, \tag{4.1}$$

$$g(x^*) + \mu y^* = 0, \quad X(x^*) - \mu (Z^*)^{-1} = 0. \tag{4.2}$$

Thus we have

$$\begin{aligned} \nabla_x L(w^*) &= \nabla f(x^*) - J_g(x^*)^\top y^* - \mathcal{A}^*(x^*) Z^* \\ &= \nabla f(x^*) + \frac{1}{\mu} J_g(x^*)^\top g(x^*) - \mu \mathcal{A}^*(x^*) X(x^*)^{-1} \\ &= -\frac{\nu}{\mu} J_g(x^*)^\top \{g(x^*) + \mu y^*\} - \nu \mathcal{A}^*(x^*) X(x^*)^{-1} \left\{ X(x^*) - \mu (Z^*)^{-1} \right\} Z^* \\ &= 0, \end{aligned}$$

where the second and third equalities follow from (4.2) and (4.1), respectively. Therefore, w^* is a shifted perturbed KKT point.

Conversely, let $w^* = (x^*, y^*, Z^*)$ be a shifted perturbed KKT point. Then, we obtain that

$$\nabla_x L(w^*) = 0, \quad g(x^*) + \mu y^* = 0, \quad X(x^*) Z^* - \mu I = 0.$$

From Theorem 4.2, it is clear that $\nabla_y F(w^*) = \nu\{g(x^*) + \mu y^*\} = 0$ and $\nabla_Z F(w^*) = \nu\{X(x^*) - \mu (Z^*)^{-1}\} = \nu\{X(x^*) Z^* - \mu I\}(Z^*)^{-1} = 0$. Moreover,

$$\begin{aligned} \nabla_x F(w^*) &= \nabla f(x^*) + \frac{1}{\mu} J_g(x^*)^\top \{(1+\nu) g(x^*) + \nu \mu y^*\} + \mathcal{A}^*(x^*) \left\{ \nu Z^* - \mu(1+\nu) X(x^*)^{-1} \right\} \\ &= \nabla f(x^*) + \frac{1}{\mu} J_g(x^*)^\top g(x^*) - \mu \mathcal{A}^*(x^*) X(x^*)^{-1} \\ &\qquad + \frac{\nu}{\mu} J_g(x^*)^\top \{g(x^*) + \mu y^*\} + \nu \mathcal{A}^*(x^*) \left\{ Z^* - \mu X(x^*)^{-1} \right\} \\ &= \nabla_x L(x^*) + \frac{\nu}{\mu} J_g(x^*)^\top \{g(x^*) + \mu y^*\} + \nu \mathcal{A}^*(x^*) X(x^*)^{-1} \{X(x^*) Z^* - \mu I\} \\ &= 0. \end{aligned}$$

Therefore, we have $\nabla F(w^*) = 0$, that is, w^* is a stationary point of F. □

This theorem is an extension of [5, Lemma 3.1] for nonlinear programming.

From this theorem, we can find an approximate shifted perturbed KKT point by solving the following unconstrained minimization problem.

$$\begin{array}{ll} \text{minimize} & F(w), \\ \text{subject to } w \in \mathcal{W}. \end{array} \tag{4.3}$$

One of the sufficient conditions under which descent methods find a stationary point is that a level set of the objective function is bounded. Thus, it is worth providing sufficient conditions for the level boundedness of the merit function F. For a given $\alpha \in \mathbf{R}$, we define the level set $\mathcal{L}(\alpha)$ of F by

$$\mathcal{L}(\alpha) = \{w \in \mathcal{W} \mid F(w) \leq \alpha\}.$$

We first give two lemmas. The following lemma follows directly from [28, Lemma 1].

Lemma 4.4. *Let $w = (x, y, Z) \in \mathcal{W}$ and $\mu > 0$. Then the following properties hold.*

(a) $\langle X(x), Z\rangle - \mu \log\det X(x)Z \geq d\mu(1 - \log\mu)$.

(b) $F_{PD}(w) \geq d\mu(1 - \log\mu)$. *The equality holds if and only if* $g(x) + \mu y = 0$ *and* $X(x)Z - \mu I = 0$.

(c) $\lim_{\langle X(x),Z\rangle \downarrow 0} F_{PD}(w) = \infty$ and $\lim_{\langle X(x),Z\rangle \uparrow \infty} F_{PD}(w) = \infty$.

Lemma 4.5. *Suppose that an infinite sequence $\{w_j = (x_j, y_j, Z_j)\}$ is included in $\mathcal{L}(\alpha)$. Suppose also that the sequence $\{x_j\}$ is bounded. Then, the sequences $\{y_j\}$ and $\{Z_j\}$ are also bounded. In addition, the sequences $\{X(x_j)\}$ and $\{Z_j\}$ are uniformly positive definite.*

Proof. Since $\{x_j\}$ is bounded, the sequence $\{-\log\det X(x_j)\}$ is bounded below. Thus, there exists a real number M_1 such that $M_1 \leq F_{BP}(x_j)$ for all j. Then, the definition of F and $w_j \in \mathcal{L}(\alpha)$ imply that $F_{PD}(w_j) \leq \frac{1}{\nu}(\alpha - M_1)$ for all j, which can be rewritten as

$$\frac{1}{2\mu}\|g(x_j) + \mu y_j\|^2 \leq \frac{\alpha - M_1}{\nu} - \langle X(x_j), Z_j\rangle + \mu \log\det X(x_j)Z_j \leq \frac{\alpha - M_1}{\nu} - d\mu(1 - \log\mu),$$

where the last inequality follows from Lemma 4.4 (a). Hence, the sequence $\{y_j\}$ is bounded.

Next, we show that $\{X(x_j)\}$ is uniformly positive definite. From Lemma 4.4 (b), we have

$$M_1 \leq F_{BP}(x_j) = F(w_j) - \nu F_{PD}(w_j) \leq \alpha - \nu F_{PD}(w_j) \leq \alpha - \nu d\mu(1 - \log\mu) \quad \text{for all } j,$$

and hence, the sequence $\{F_{BP}(x_j)\}$ is bounded. It then follows from the boundedness of $\{x_j\}$ and $F_{BP}(x_j) = f(x_j) + \frac{1}{2\mu}\|g(x_j)\|^2 - \mu\log\det X(x_j)$ that $\{-\log\det X(x_j)\}$ is also bounded. From Proposition 2.2, the boundedness of $\{-\log\det X(x_j)\}$ and $\{X(x_j)\}$ implies that $\{X(x_j)\}$ is uniformly positive definite, that is, there exists $\underline{\lambda}$ such that $\lambda_{\min}(X(x_j)) \geq \underline{\lambda} > 0$ for all j.

Next we show that $\{Z_j\}$ is bounded. From Lemma 4.4 (b), we have

$$d\mu(1 - \log\mu) \leq F_{PD}(w_j) \leq \frac{1}{\nu}(\alpha - M_1) \quad \text{for all } j,$$

and hence the sequence $\{F_{PD}(w_j)\}$ is bounded. Then, Lemma 4.4 (c) yields that $\{\langle X(x_j), Z_j\rangle\}$ is bounded. Thus, there exists a real number M_2 such that for all j,

$$M_2 \geq \operatorname{tr}(X(x_j)Z_j) \geq \lambda_{\min}(X(x_j))\operatorname{tr}(Z_j) \geq \underline{\lambda}\operatorname{tr}(Z_j) = \underline{\lambda}\sum_{k=1}^{d}\lambda_k(Z_j) \tag{4.4}$$

where the second inequality follows from [3, Proposition 8.4.13]. Since $\{Z_j\}$ is positive definite, $\lambda_k(Z_j) > 0$ for $k = 1, \dots, d$. Then, (4.4) implies that $\{\lambda_k(Z_j)\}$ is bounded for $k = 1, \dots, d$, and hence $\{Z_j\}$ is bounded.

Finally, we show that $\{Z_j\}$ is uniformly positive definite. Recall that

$$F_{PD}(w_j) = \frac{1}{2\mu}\|g(x_j) + \mu y_j\|^2 + \langle X(x_j), Z_j\rangle - \mu \log\det X(x_j) - \mu \log\det Z_j,$$

and that the sequences $\{x_j\}, \{y_j\}, \{\langle X(x_j), Z_j\rangle\}, \{-\log\det X(x_j)\}$ and $\{F_{PD}(w_j)\}$ are bounded. Therefore, $\{-\log\det Z_j\}$ is also bounded. It then follows from Proposition 2.2 and the boundedness of $\{Z_j\}$ that $\{Z_j\}$ is uniformly positive definite. □

We now give sufficient conditions under which any level set of the merit function F is bounded.

Theorem 4.6. *Suppose that the following five assumptions hold.*

(i) *The function f is convex;*

(ii) *The functions $g_1, \ldots, g_m$ are affine;*

(iii) *The function X satisfies $X(\lambda u + (1-\lambda)v) - \lambda X(u) - (1-\lambda)X(v) \succeq 0$ for $\lambda \in [0,1]$ and $u, v \in \Omega$;*

(iv) *The matrices $A_1(x), \ldots, A_n(x)$ are linearly independent for all $x \in \Omega$;*

(v) *There exists a shifted perturbed KKT point w^*.*

Then, the level set $\mathcal{L}(\alpha)$ of F is bounded for all $\alpha \in \mathbf{R}$.

Proof. Let $\{(x_k, y_k, Z_k)\}$ be an infinite sequence in $\mathcal{L}(\alpha)$. We first show that the sequence $\{x_k\}$ is bounded. In order to prove this by contradiction, we suppose that there exists a subset $\mathcal{I} \subset \{0, 1, \ldots\}$ such that $\lim_{k\to\infty, k\in\mathcal{I}} \|x_k\| = \infty$. Since $F(w_k) \le \alpha$ and $F_{PD}(w_k) \ge d\mu(1 - \log\mu)$ from Lemma 4.4 (b), $F_{BP}(x_k) = F(w_k) - \nu F_{PD}(w_k) \le \alpha - \nu d\mu(1 - \log\mu)$.

On the other hand, since w^* is a shifted perturbed KKT point, Theorem 4.2 implies that

$$0 = \nabla_x L(w^*) = \nabla f(x^*) + \frac{1}{\mu} J_g(x^*)^\top g(x^*) - \mu \mathcal{A}^*(x^*) X(x^*)^{-1} = \nabla F_{BP}(x^*). \tag{4.5}$$

Note that F_{BP} is strictly convex from Proposition 2.1 (c) and the assumptions (i)–(iv). Thus, (4.5) implies that x^* is the unique global minimizer of F_{BP}. Note that $x^* \in \Omega$ and $X(x^*) \succ 0$. Then, there exists $\varepsilon > 0$ such that $\{x^* + \varepsilon u \mid \|u\| = 1\} \subset \Omega$ and $\min\{F_{BP}(x^* + \varepsilon u) \mid \|u\| = 1\} > F_{BP}(x^*)$. Let $d_k \equiv \frac{1}{\varepsilon}(x_k - x^*)$ $(k \in \mathcal{I})$ and $F^\varepsilon_{BP} \equiv \min\{F_{BP}(x^* + \varepsilon u) \mid \|u\| = 1\}$. Note that $\|d_k\| \to \infty$ $(k \to \infty, k \in \mathcal{I})$. Without loss of generality, we suppose that $\|d_k\| > 1$ for all $k \in \mathcal{I}$. From the convexity of F_{BP}, we have

$$\frac{\|d_k\| - 1}{\|d_k\|} F_{BP}(x^*) + \frac{1}{\|d_k\|} F_{BP}(x^* + \varepsilon d_k) \ge F_{BP}\left(x^* + \varepsilon \frac{d_k}{\|d_k\|}\right) \ge F^\varepsilon_{BP},$$

which implies that $F_{BP}(x_k) = F_{BP}(x^* + \varepsilon d_k) \ge \|d_k\|(F^\varepsilon_{BP} - F_{BP}(x^*)) + F_{BP}(x^*)$. Thus, since $F^\varepsilon_{BP} - F_{BP}(x^*) > 0$, we have $F_{BP}(x_k) \to \infty$ $(k \to \infty, k \in \mathcal{I})$. However, this result contradicts $F_{BP}(x_k) \le \alpha - d\mu(1 - \log\mu)$. Hence, for any sequence $\{x_k, y_k, Z_k\} \subset \mathcal{L}(\alpha)$, the sequence $\{x_k\}$ is bounded. Since $\{x_k\}$ is bounded and $\{F(w_j)\}$ is bounded above, it follows from Lemma 4.5 that the sequences $\{y_k\}$ and $\{Z_k\}$ are also bounded. □

Remark 4.7. The level boundedness of the merit function for nonlinear programming is not given in [5]. Applying Theorem 4.6, it is easy to show that the merit function M in [5] is level bounded if the objective function f is convex, the constraint functions c_i $(i \in \mathcal{E})$ are affine, and $\operatorname{rank}(J_c) = n$.

Remark 4.8. Kato, Yabe and Yamashita [12] showed that their merit function $\widetilde{F}$ is differentiable and its stationary point is a shifted perturbed KKT point. However, they did not discuss the level boundedness of their merit function.

Remark 4.9. Theorem 4.6 assumes that f is convex and g is affine. These assumptions are rather restrictive for some applications. We can replace these assumptions with the following coerciveness condition.

$$\lim_{\|x\|\to\infty, x\in\Omega} \frac{1}{\|x\|}\left(f(x) + \frac{1}{2\mu}\|g(x)\|^2\right) = \infty.$$

Due to Theorems 4.2–4.6, we can solve the unconstrained minimization problem (4.3) by any descent method, such as the quasi-Newton method and the steepest descent method, and hence we can get an approximate shifted perturbed KKT point v_{k+1} in Step 1 of Algorithm 1.

4.2 Newton algorithm for minimization of the merit function

In this subsection, we propose a Newton-type method for the unconstrained minimization problem (4.3) of the merit function F.

We exploit the scaling of $X(x)$ and Z. Let $T \in \mathbf{R}^{d\times d}$ be a nonsingular matrix such that

$$TX(x)T^\top T^{-\top}ZT^{-1} = T^{-\top}ZT^{-1}TX(x)T^\top. \tag{4.6}$$

Let $\widetilde{X}(x)$ and $\widetilde{Z}$ be defined by

$$\widetilde{X}(x) = TX(x)T^\top = (T\odot T)X(x) \quad \text{and} \quad \widetilde{Z} = T^{-\top}ZT^{-1} = (T^{-\top}\odot T^{-\top})Z,$$

respectively. Note that $\widetilde{X}(x)$ and $\widetilde{Z}$ commute, that is, $\widetilde{X}(x)\widetilde{Z} = \widetilde{Z}\widetilde{X}(x)$ from (4.6). As seen later, the scaling enables us to analyze and calculate a Newton direction easily. In the subsequent discussions, for simplicity, we denote $X(x)$ and $\widetilde{X}(x)$ by X and $\widetilde{X}$, respectively.

Next, we give a Newton direction, and show that it is a descent direction for the merit function F. The Newton direction is derived from the nonlinear equations $r(w;\mu) = 0$ in the shifted perturbed KKT conditions (2.9). However, the matrix ΔZ of a pure Newton direction $(\Delta x, \Delta y, \Delta Z)$ for $r(w;\mu) = 0$ is not necessarily symmetric due to $XZ - \mu I = 0$. Thus, we consider the following symmetrized shifted perturbed KKT conditions with scaling.

$$r_S(w;\mu) \equiv \begin{bmatrix} \nabla_x L(w) \\ g(x) + \mu y \\ \widetilde{X}\circ\widetilde{Z} - \mu I \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.7}$$

and

$$\widetilde{X} \succ 0, \qquad \widetilde{Z} \succ 0.$$

Note that $\widetilde{X}\circ\widetilde{Z} - \mu I = 0$ is equivalent to $XZ - \mu I = 0$ if X and Z are symmetric positive semidefinite [28]. Moreover, $\widetilde{X}(x) \succ 0$ and $\widetilde{Z} \succ 0$ if and only if $X(x) \succ 0$ and $Z \succ 0$. Therefore, the symmetrized shifted perturbed KKT conditions (4.7) are essentially the same as the original shifted perturbed KKT conditions (2.9).

We apply the Newton method to the equation (4.7). Then, it follows from [12] that

$$G\Delta x - J_g(x)^\top \Delta y - \mathcal{A}^*(x)\Delta Z = -\nabla_x L(w), \tag{4.8}$$

$$J_g(x)\Delta x + \mu\Delta y = -g(x) - \mu y, \tag{4.9}$$

$$\widetilde{Z}\Delta\widetilde{X} + \Delta\widetilde{X}\widetilde{Z} + \widetilde{X}\Delta\widetilde{Z} + \Delta\widetilde{Z}\widetilde{X} = 2\mu I - \widetilde{X}\widetilde{Z} - \widetilde{Z}\widetilde{X}, \tag{4.10}$$

where G denotes the Hessian matrix of the Lagrangian function L with respect to x or its approximation. In what follows, we call the solution $\Delta w \equiv (\Delta x, \Delta y, \Delta Z)$ of the Newton equations (4.8)–(4.10) the *Newton direction*.

Next, we give the explicit form of the Newton direction Δw. It follows from [12] that

$$\left(G + H + \frac{1}{\mu} J_g(x)^\top J_g(x)\right) \Delta x = -\left(\nabla f(x) + \frac{1}{\mu} J_g(x)^\top g(x) - \mu \mathcal{A}^*(x) X^{-1}\right), \tag{4.11}$$

$$\Delta y = -\frac{1}{\mu}(g(x) + \mu y + J_g(x)\Delta x), \tag{4.12}$$

$$\Delta Z = \mu X^{-1} - Z - (T^\top \odot T^\top)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)\mathcal{A}(x)\Delta x, \tag{4.13}$$

where the elements of $H \in \mathbf{R}^{n\times n}$ are written as

$$H_{ij} = \left\langle A_i(x), (T^\top \odot T^\top)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)A_j(x)\right\rangle. \tag{4.14}$$

Since $J_g(x)^\top J_g(x)$ is positive semidefinite, we can solve (4.11) with respect to Δx if $G + H$ is positive definite. Fortunately, H is positive semidefinite. The following lemma can be proved by using the similar arguments as in [28, Theorem 3].

Lemma 4.10. *Suppose that X and Z are symmetric positive definite. Then, H is symmetric positive semidefinite. Furthermore, if $A_1(x), \ldots, A_n(x)$ are linearly independent for all $x \in \mathbf{R}^n$, then H is symmetric positive definite.* □

Remark 4.11. In the case of linear SDP, $A_1(x), \ldots, A_n(x)$ are usually supposed to be linearly independent for all $x \in \mathbf{R}^n$. Then, H is positive definite from Lemma 4.10.

To summarize the discussion above, we give the concrete formulae of the Newton direction Δw in the following theorem.

Theorem 4.12. *Let $\mu > 0$ and $w = (x, y, Z) \in \mathcal{W}$. Suppose that $G + H$ is positive definite. Then, the Newton equations (4.8)–(4.10) have the unique solution $\Delta w = (\Delta x, \Delta y, \Delta Z)$ such that*

$$\Delta x = -\left(G + H + \frac{1}{\mu} J_g(x)^\top J_g(x)\right)^{-1} \left(\nabla f(x) + \frac{1}{\mu} J_g(x)^\top g(x) - \mu \mathcal{A}^*(x) X^{-1}\right), \tag{4.15}$$

$$\Delta y = -\frac{1}{\mu}(g(x) + \mu y + J_g(x)\Delta x),$$

$$\Delta Z = \mu X^{-1} - Z - (T^\top \odot T^\top)(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)(T \odot T)\mathcal{A}(x)\Delta x.$$

Proof. It is clear that $\frac{1}{\mu} J_g(x)^\top J_g(x)$ is positive semidefinite. Thus, the positive definiteness of $G + H$ and (4.11) yield that

$$\Delta x = -\left(G + H + \frac{1}{\mu} J_g(x)^\top J_g(x)\right)^{-1} \left(\nabla f(x) + \frac{1}{\mu} J_g(x)^\top g(x) - \mu \mathcal{A}^*(x) X^{-1}\right).$$

Furthermore, Δy and ΔZ directly follow from (4.12) and (4.13), respectively. □

One of the main burdens during the computation of the Newton direction Δw is the calculation of the operator $(\widetilde{X} \odot I)^{-1}$ in (4.13) and (4.14). Note that $(\widetilde{X} \odot I)^{-1}$ in (4.13) and (4.14) appears as $(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)$. Hence, when $\widetilde{X} = I$, it is clear that $(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I) = \widetilde{Z} \odot I$. On the other hand, when $\widetilde{X} = \widetilde{Z}$, $(\widetilde{X} \odot I)^{-1}(\widetilde{Z} \odot I)$ is the identity mapping. Thus, if we choose the scaling matrix T such that $\widetilde{X} = I$ or $\widetilde{X} = \widetilde{Z}$, we do not have to handle the operator $(\widetilde{X} \odot I)^{-1}$ explicitly. This is one of the reasons why we exploit the scaling. Note that the choices of T such that $\widetilde{X} = I$ or $\widetilde{X} = \widetilde{Z}$ are well known as the HRVW/KSH/M choice or the NT choice.

(i) HRVW/KSH/M choice

Let $T = X^{-\frac{1}{2}}$. Then we have $\widetilde{X} = I$ and $\widetilde{Z} = X^{\frac{1}{2}} Z X^{\frac{1}{2}}$. This choice corresponds to the dual HRVW/KSH/M choice for the linear SDP [7,13,15].

(ii) NT choice

Let $T = W^{-\frac{1}{2}}$, where $W = X^{\frac{1}{2}}(X^{\frac{1}{2}} Z X^{\frac{1}{2}})^{-\frac{1}{2}} X^{\frac{1}{2}}$. Then we have $\widetilde{X} = W^{-\frac{1}{2}} X W^{-\frac{1}{2}} = W^{\frac{1}{2}} Z W^{\frac{1}{2}} = \widetilde{Z}$. This choice corresponds to the NT choice for linear SDP [16,17].

Next, we show that the Newton direction is a descent direction for the merit function F. For this purpose, we first show the following two lemmas.

Lemma 4.13. *Let $\mu > 0$ and $w = (x, y, Z) \in \mathcal{W}$. Suppose that $G + H$ is positive definite. Let Δx be given by (4.15). Then we have*

$$\nabla F_{BP}(x)^\top \Delta x = -\Delta x^\top \left(G + H + \frac{1}{\mu} J_g(x)^\top J_g(x) \right) \Delta x \leq 0.$$

Furthermore, $\nabla F_{BP}(x)^\top \Delta x = 0$ if and only if $\Delta x = 0$.

Proof. We easily see that $G + H + \frac{1}{\mu} J_g(x)^\top J_g(x)$ is positive definite from the positive definiteness of $G+H$. Since $\nabla F_{BP}(x) = \nabla f(x) + \frac{1}{\mu} J_g(x)^\top g(x) - \mu \mathcal{A}^*(x) X^{-1}$ from Theorem 4.2, (4.11) yields that

$$\begin{aligned} \nabla F_{BP}(x)^\top \Delta x &= \Delta x^\top \left(\nabla f(x) + \frac{1}{\mu} J_g(x)^\top g(x) - \mu \mathcal{A}^*(x) X^{-1} \right) \\ &= -\Delta x^\top \left(G + H + \frac{1}{\mu} J_g(x)^\top J_g(x) \right) \Delta x \\ &\leq 0. \end{aligned}$$

Furthermore, since $G + H + \frac{1}{\mu} J_g(x)^\top J_g(x)$ is positive definite, $\nabla F_{BP}(x)^\top \Delta x = 0$ if and only if $\Delta x = 0$. □

Lemma 4.14. *Let $\mu > 0$ and $w = (x, y, Z) \in \mathcal{W}$. Suppose that $G + H$ is positive definite. Let $\Delta w = (\Delta x, \Delta y, \Delta Z)$ be given in Theorem 4.12. Then we have*

$$\langle \nabla F_{PD}(w), \Delta w \rangle = -\frac{1}{\mu} \|g(x) + \mu y\|^2 - \|(\widetilde{X}\widetilde{Z})^{-\frac{1}{2}} (\mu I - \widetilde{X}\widetilde{Z})\|_F^2 \leq 0.$$

Furthermore, $\langle \nabla F_{PD}(w), \Delta w \rangle = 0$ if and only if $g(x) + \mu y = 0$ and $XZ - \mu I = 0$.

Proof. Let $\Psi_1 : \mathcal{W} \to \mathbf{R}$ and $\Psi_2 : \mathcal{W} \to \mathbf{R}$ be defined by $\Psi_1(w) \equiv \frac{1}{2\mu} \|g(x) + \mu y\|^2$ and $\Psi_2(w) \equiv \langle X, Z \rangle - \mu \log \det XZ$, respectively. Note that $F_{PD}(w) = \Psi_1(w) + \Psi_2(w)$. Then, we have

$$\langle \nabla \Psi_1(w), \Delta w \rangle = \frac{1}{\mu} (g(x) + \mu y)^\top (J_g(x) \Delta x + \mu \Delta y) = -\frac{1}{\mu} \|g(x) + \mu y\|^2 \leq 0,$$

where the second equation follows from (4.9). The equality holds if and only if $g(x) + \mu y = 0$.

On the other hand, $\langle \nabla \Psi_2(w), \Delta w\rangle = -\|(\widetilde{X}\widetilde{Z})^{-\frac{1}{2}}(\mu I - \widetilde{X}\widetilde{Z})\|_F^2 \leq 0$ holds from [28, Lemma 3]. Moreover the equality holds if and only if $XZ - \mu I = 0$. □

We show that the Newton direction Δw is the descent direction for the merit function F.

Theorem 4.15. *Let $\mu > 0$ and $w = (x, y, Z) \in \mathcal{W}$. Assume that $G + H$ is positive definite. Then, $\Delta w = (\Delta x, \Delta y, \Delta Z)$ given in Theorem 4.12 is a descent direction for the merit function F, i.e.,*

$$\begin{aligned}\langle \nabla F(w), \Delta w\rangle &= -\Delta x^\top \left(G + H + \frac{1}{\mu} J_g(x)^\top J_g(x)\right)\Delta x \\ &\qquad - \frac{\nu}{\mu}\|g(x) + \mu y\|^2 - \nu\|(\widetilde{X}\widetilde{Z})^{-\frac{1}{2}}(\mu I - \widetilde{X}\widetilde{Z})\|_F^2 \\ &\leq 0.\end{aligned}$$

Furthermore, $\langle \nabla F(w), \Delta w\rangle = 0$ if and only if w is a shifted perturbed KKT point.

Proof. It is clear that the first part of this statement holds from Lemmas 4.13 and 4.14. Now, we show the second part of this theorem. Suppose that w is a shifted perturbed KKT point, i.e., $\nabla f(x) - J_g(x)^\top y - \mathcal{A}^*(x)Z = 0,\ g(x) + \mu y = 0$ and $XZ - \mu I = 0$. Then, we obtain $\langle \nabla F_{PD}(w), \Delta w\rangle = 0$ from Lemma 4.14. Moreover, we have

$$\nabla f(x) + \frac{1}{\mu} J_g(x)^\top g(x) - \mu \mathcal{A}^*(x)X^{-1} = \nabla f(x) - J_g(x)^\top y - \mathcal{A}^*(x)Z = 0,$$

and hence $\Delta x = 0$ from (4.15). Then, $\nabla F_{BP}(x)^\top \Delta x = 0$, and hence $\langle \nabla F(w), \Delta w\rangle = 0$.

Conversely, suppose that $\langle \nabla F(w), \Delta w\rangle = 0$. It then follows from Lemmas 4.13 and 4.14 that $\Delta x = 0, g(x) + \mu y = 0$ and $XZ - \mu I = 0$. Then we have from (4.15) that

$$\nabla_x L(w) = \nabla f(x) - J_g(x)^\top y - \mathcal{A}^*(x)Z = \nabla f(x) + \frac{1}{\mu} J_g(x)^\top g(x) - \mu \mathcal{A}^*(x)X^{-1} = 0.$$

Thus, w is a shifted perturbed KKT point. □

Theorem 4.15 guarantees that $F(w + \alpha\Delta w) < F(w)$ for sufficiently small $\alpha > 0$ if w is not a shifted perturbed KKT point.

Now, we discuss how to choose an appropriate step size α such that $F(w+\alpha\Delta w) < F(w)$. The merit function F and the Newton equations (4.8)–(4.10) are well-defined only on $\mathcal{W}$. Therefore, the new point $w+\alpha\Delta w$ is required to be an interior point. Thus, we must choose a step size $\alpha \in (0, 1]$ such that $X(x + \alpha\Delta x) \succ 0$ and $Z + \alpha\Delta Z \succ 0$. To this end, we first calculate

$$\bar{\alpha}_x = \begin{cases} -\frac{\tau}{\lambda_{\min}(X^{-\frac{1}{2}}\Delta X X^{-\frac{1}{2}})} & \text{if } \lambda_{\min}(X^{-\frac{1}{2}}\Delta X X^{-\frac{1}{2}}) < 0 \text{ and } X \text{ is affine} \\ 1 & \text{otherwise} \end{cases}$$

and

$$\bar{\alpha}_z = \begin{cases} -\frac{\tau}{\lambda_{\min}(Z^{-\frac{1}{2}}\Delta Z Z^{-\frac{1}{2}})} & \text{if } \lambda_{\min}(Z^{-\frac{1}{2}}\Delta Z Z^{-\frac{1}{2}}) < 0 \\ 1 & \text{otherwise,} \end{cases}$$

where $\tau \in (0, 1)$ is a given constant. Set

$$\bar{\alpha} = \min\{1, \bar{\alpha}_x, \bar{\alpha}_z\}. \tag{4.16}$$

Then $Z + \alpha\Delta Z \succ 0$ for any $\alpha \in (0, \bar{\alpha}]$. Moreover, $X(x + \alpha\Delta x) \succ 0$ for any $\alpha \in (0, \bar{\alpha}]$ if X is affine. Note that if X is nonlinear, $X(x + \alpha\Delta x)$ is not necessarily positive definite for any $\alpha \in (0, \bar{\alpha}]$.

Next, we choose a step size $\alpha \in (0, \bar{\alpha}]$ such that $F(w + \alpha\Delta w) < F(w)$ and $X(x + \alpha\Delta x) \succ 0$. For this purpose, we adopt the following Armijo's line search rule. Find the smallest nonnegative integer l such that

$$F(w + \bar{\alpha}\beta^l \Delta w) \leq F(w) + \varepsilon_0 \bar{\alpha}\beta^l \langle \nabla F(w), \Delta w \rangle, \quad X(x + \bar{\alpha}\beta^l \Delta x) \succ 0$$

and set $\alpha = \bar{\alpha}\beta^l$, where $\beta, \varepsilon_0 \in (0, 1)$. Note that the second condition is not necessary when X is affine.

Now, we describe a concrete Newton-type method for Step 1 of Algorithm 1. Recall that the script k denotes the k-th iteration of Algorithm 1.

Algorithm 2. (for Step 1 of Algorithm 1)

Step 0. Choose $\beta, \varepsilon_0, \tau \in (0, 1)$ and set $j = 0$ and $w_0 = v_k$.

Step 1. If $\rho(w_j; \mu_k) \leq \sigma\mu_k$, then set $v_{k+1} = w_j$ and return.

Step 2. Obtain the Newton direction $\Delta w_j = (\Delta x_j, \Delta y_j, \Delta Z_j)$ by solving the Newton equations (4.8)–(4.10).

Step 3. Set $\alpha_j = \bar{\alpha}_j \beta^{l_j}$, where $\bar{\alpha}_j$ is given by (4.16) and l_j is the smallest nonnegative integer such that

$$F(w_j + \bar{\alpha}_j \beta^{l_j} \Delta w_j) \leq F(w_j) + \varepsilon_0 \bar{\alpha}_j \beta^{l_j} \langle \nabla F(w_j), \Delta w_j \rangle, \quad X(x_j + \bar{\alpha}_j \beta^{l_j} \Delta x_j) \succ 0.$$

Step 4. Set $w_{j+1} = w_j + \alpha_j \Delta w_j$ and $j = j + 1$, and go to Step 1.

4.3 Global convergence of Algorithm 2

In this subsection, we prove the global convergence of Algorithm 2. For this purpose, we make the following assumptions.

(A1) The functions $f, g_1, \ldots, g_m$ and X are twice continuously differentiable.

(A2) The sequence $\{x_j\}$ generated by Algorithm 2 remains in some compact set Ω of $\mathbf{R}^n$.

(A3) The sequence $\{G_j + H_j + \frac{1}{\mu} J_g(x_j)^\top J_g(x_j)\}$ is uniformly positive definite and the sequence $\{G_j\}$ is bounded.

(A4) The sequences $\{T_j\}$ and $\{T_j^{-1}\}$ are bounded.

Note that Assumption (A2) holds under the assumptions of Theorem 4.6. Assumption (A3) guarantees that the Newton equations (4.8)–(4.10) have a unique solution.

Remark 4.16. Assumptions (A1)–(A3) hold for linear SDP such that $A_1(x_j), \ldots, A_n(x_j)$ are linearly independent. In fact, it is clear that Assumption (A1) holds. Theorem 4.6 guarantees that Assumption (A2) holds. Moreover, H_j is positive definite from Remark 4.11 and $G_j = 0$. Thus, Assumption (A3) holds.

Remark 4.17. Yamashita, Yabe and Harada [28] showed the global convergence of their Newton-type algorithm under the boundedness of the sequence $\{y_j\}$, in addition to Assumptions (A1)–(A4). However, they did not give sufficient conditions for the boundedness of $\{y_j\}$.

Remark 4.18. Kato, Yabe and Yamashita [12] also showed that a Newton-type algorithm with the merit function $\widetilde{F}$ can find a shifted perturbed KKT point under the same assumptions. However, concrete sufficient conditions were not given for Assumption (A2).

First of all, we show that the sequence $\{w_j\}$ generated by Algorithm 2 is bounded.

Lemma 4.19. *Suppose that Assumptions (A2) holds. Then, the sequence* $\{w_j = (x_j, y_j, Z_j)\}$ *generated by Algorithm 2 is bounded. Furthermore, the matrices* $\{X_j\}$ *and* $\{Z_j\}$ *are uniformly positive definite.*

Proof. Since the sequence $\{F(w_j)\}$ is monotonically decreasing, we have $F(w_j) \leq F(w_0)$ for all j. From Assumption (A2) and Lemma 4.5, we have the desired results. □

Note that the lemma above guarantees that Assumption (A4) holds if the scaling matrix T is given by HRVW/KSH/M choice or NT choice.

Lemma 4.20. *Suppose that Assumptions (A2)–(A4) hold. Then, the sequence* $\{\Delta w_j\}$ *generated by Algorithm 2 is bounded.*

Proof. Assumptions (A2)–(A4), Lemma 4.19 and Theorem 4.12 yield that the sequence $\{\Delta w_j\}$ generated by Algorithm 2 is bounded. □

We now show the global convergence of Algorithm 2.

Theorem 4.21. *Suppose that Assumptions (A1)–(A4) hold. Then the sequence* $\{w_j = (x_j, y_j, Z_j)\}$ *generated by Algorithm 2 is bounded. Moreover, any accumulation point* $w^* = (x^*, y^*, Z^*)$ *of* $\{w_j\}$ *is a shifted perturbed KKT point.*

Proof. Since the sequence $\{w_j\}$ is bounded from Lemma 4.19, it has at least one accumulation point w^*.

Next, we prove that w^* is a shifted perturbed KKT point. To this end, we first show that the sequence $\{\bar{\alpha}_j\}$ given in Step 3 of Algorithm 2 is bounded away from zero, that is, there exists a real number $\bar{\alpha}$ such that $0 < \bar{\alpha} \leq \bar{\alpha}_j$ for all j. Note that from Lemmas 4.19 and 4.20, the sequences $\{X_j\}$, $\{Z_j\}$, $\{\Delta X_j\}$ and $\{\Delta Z_j\}$ are bounded. Moreover, the matrices $\{X_j\}$ and $\{Z_j\}$ are uniformly positive definite. Thus, the sequences $\{\lambda_{\min}(X_j^{-\frac{1}{2}} \Delta X_j X_j^{-\frac{1}{2}})\}$ and $\{\lambda_{\min}(Z_j^{-\frac{1}{2}} \Delta Z_j Z_j^{-\frac{1}{2}})\}$ are also bounded. Then, the definition of $\bar{\alpha}_j$ yields that there exists a real number $\bar{\alpha}$ such that $0 < \bar{\alpha} \leq \bar{\alpha}_j$ for all j.

Next, we show $\langle \nabla F(w_j), \Delta w_j \rangle \to 0$ as $j \to \infty$. From the Armijo's line search strategy in Step 3, we have $F(w_{j+1}) - F(w_j) \leq \varepsilon_0 \bar{\alpha}_j \beta^{l_j} \langle \nabla F(w_j), \Delta w_j \rangle$ and $X(x_j + \bar{\alpha}_j \beta^{l_j} \Delta x_j) \succ 0$. Summing up the above inequalities from $j = 1$ to $j = \widetilde{j}$, we have

$$F(w_{\widetilde{j}+1}) - F(w_1) \leq \varepsilon_0 \sum_{j=1}^{\widetilde{j}} \bar{\alpha}_j \beta_j \langle \nabla F(w_j), \Delta w_j \rangle .$$

It then follows from $\langle \nabla F(w_j), \Delta w_j \rangle \leq 0$ by Theorem 4.15 and $\bar{\alpha} \leq \bar{\alpha}_j$ that

$$F(w_{\widetilde{j}+1}) - F(w_1) \leq \varepsilon_0 \bar{\alpha} \sum_{j=1}^{\widetilde{j}} \beta^{l_j} \langle \nabla F(w_j), \Delta w_j \rangle .$$

Since the sequence $\{w_j\}$ is bounded, the sequence $\{F(w_j)\}$ is also bounded, and hence

$$-\infty < \sum_{j=1}^{\infty} \beta^{l_j} \langle \nabla F(w_j), \Delta w_j \rangle \leq 0.$$

Therefore, we have

$$\lim_{j\to\infty} \beta^{l_j} \langle \nabla F(w_j), \Delta w_j \rangle = 0.$$

Now, we consider two cases: $\liminf_{j\to\infty} \beta^{l_j} > 0$ and $\liminf_{j\to\infty} \beta^{l_j} = 0$.

Case 1: $\liminf_{j\to\infty} \beta^{l_j} > 0$. Then, we have $\lim_{j\to\infty} \langle \nabla F(w_j), \Delta w_j \rangle = 0$.

Case 2: $\liminf_{j\to\infty} \beta^{l_j} = 0$. In this case, there exists a subset $\mathcal{J} \subset \{0, 1, \dots\}$ such that $\lim_{j\to\infty, j\in\mathcal{J}} l_j = \infty$. Since $\{X(x_j)\}$ is uniformly positive definite and $\{\Delta x_j\}$ is bounded, there exists $\bar{l}$ such that $X(x_j + \bar{\alpha}_j \beta^{l_j} \Delta x_j) \succ 0$ for all $l_j > \bar{l}$. Therefore, without loss of generality, we suppose that $X(x_j + \bar{\alpha}_j \beta^{l_j - 1} \Delta x_j) \succ 0$ for all $j \in \mathcal{J}$. Furthermore, since $l_j - 1$ does not satisfy the Armijo rule in Step 3, we have

$$\varepsilon_0 t_j \langle \nabla F(w_j), \Delta w_j \rangle < F(w_j + t_j \Delta w_j) - F(w_j),$$

where $t_j \equiv \bar{\alpha}_j \beta^{l_j - 1}$. Let $h(t) \equiv F(w_j + t\Delta w_j)$. By the mean value theorem for h, there exists $\theta_j \in (0, 1)$ such that

$$\begin{aligned} \varepsilon_0 t_j \langle \nabla F(w_j), \Delta w_j \rangle &< F(w_j + t_j \Delta w_j) - F(w_j) \\ &= h(t_j) - h(0) \\ &= t_j h'(\theta_j t_j) \\ &= t_j \langle \nabla F(w_j + \theta_j t_j \Delta w_j), \Delta w_j \rangle, \end{aligned}$$

which yields that

$$\begin{aligned} 0 < (\varepsilon_0 - 1) \langle \nabla F(w_j), \Delta w_j \rangle &< \langle \nabla F(w_j + \theta_j t_j \Delta w_j) - \nabla F(w_j), \Delta w_j \rangle \\ &\leq \|\nabla F(w_j + \theta_j t_j \Delta w_j) - \nabla F(w_j)\| \|\Delta w_j\|, \quad (4.17) \end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality. Since $\{w_j\}$ and $\{\Delta w_j\}$ are bounded and $\lim_{j\to\infty, j\in\mathcal{J}} t_j = 0$, we have from Assumption (A1)

$$\lim_{j\to\infty, j\in\mathcal{J}} \|\nabla F(w_j + \theta_j t_j \Delta w_j) - \nabla F(w_j)\| = 0.$$

Then, $\lim_{j\to\infty, j\in\mathcal{J}} \langle \nabla F(w_j), \Delta w_j \rangle = 0$ from (4.17).

From both cases, we can conclude that

$$\lim_{j\to\infty} \langle \nabla F(w_j), \Delta w_j \rangle = 0. \quad (4.18)$$

From the boundedness of $\{w_j\}$ and Assumptions (A3) and (A4), there exists a subset $\mathcal{K} \subset \{0, 1, \dots\}$ such that $\{w_j\}_{\mathcal{K}}, \{G_j\}_{\mathcal{K}}$ and $\{T_j\}_{\mathcal{K}}$ converge to w^*, G^* and T^*, respectively. Moreover from (2.1), the sequences $\{T_j \odot T_j\}_{\mathcal{K}}$ and $\{T_j^\top \odot T_j^\top\}_{\mathcal{K}}$ converge to $T^* \odot T^*$ and $(T^*)^\top \odot (T^*)^\top$, respectively. Then we have from (4.14) that $\{H_j\}_{\mathcal{K}}$ converges to H^*.

Note that the matrix $G^* + H^* + \frac{1}{\mu} J_g(x^*)^\top J_g(x^*)$ is positive definite from Assumption (A3). Thus, (4.15) implies that the subsequence $\{\Delta x_j\}_\mathcal{K}$ converges to Δx^*, where

$$\Delta x^* = -\left(G^* + H^* + \frac{1}{\mu} J_g(x^*)^\top J_g(x^*) \right)^{-1} \left(\nabla f(x^*) + \frac{1}{\mu} J_g(x^*)^\top g(x^*) - \mu \mathcal{A}^*(x^*) X(x^*)^{-1} \right).$$

Similarly, $\{\Delta y_j\}_\mathcal{K}$ and $\{\Delta Z_j\}_\mathcal{K}$ converge to Δy^* and ΔZ^*, where

$$\begin{aligned} \Delta y^* &= -\frac{1}{\mu}(g(x^*) + \mu y^* + J_g(x^*)\Delta x^*), \\ \Delta Z^* &= \mu X(x^*)^{-1} - Z^* - ((T^*)^\top \odot (T^*)^\top)(\widetilde{X}(x^*) \odot I)^{-1}(\widetilde{Z}^* \odot I)(T^* \odot T^*)\mathcal{A}(x^*)\Delta x^*, \end{aligned}$$

and $\widetilde{Z}^* = ((T^*)^{-\top} \odot (T^*)^{-\top})Z^*$. Then, $\langle \nabla F(w^*), \Delta w^* \rangle = 0$ by (4.18). Therefore, Theorem 4.15 yields that

$$\nabla_x L(w^*) = 0, \quad g(x^*) + \mu y^* = 0 \quad \text{and} \quad X(x^*)Z^* - \mu I = 0,$$

i.e., w^* is a shifted perturbed KKT point. □

5 Numerical Experiments

In this section, we report some numerical experiments for the proposed algorithm (Algorithm 1 with Algorithm 2). We compare the proposed algorithm with the interior point method [28] based on the perturbed KKT conditions (2.8). We present the number of iterations and the CPU time of both algorithms. The program is written in MATLAB R2010a and run on a machine with an Intel Core i7 920 2.67GHz CPU and 3.00GB RAM. The parameter μ_k used by both algorithms is updated by $\mu_{k+1} = \mu_k/10$ with $\mu_0 = 0.1$. Moreover, we use the approximate Hessian G_k updated by the Levenberg-Marquardt type algorithm [28, Remark 3]. We employ the scaling matrix $T = X^{-\frac{1}{2}}$ and the following parameters.

$$\epsilon = 10^{-4}, \quad \sigma = 3.5, \quad \nu = 1.0, \quad \tau = 0.95, \quad \beta = 0.95, \quad \varepsilon_0 = 0.50.$$

We solve the following three test problems described in [28] by using the initial points indicated in [28].

Gaussian channel capacity problem:

$$\begin{aligned} &\text{maximize} \quad \frac{1}{2}\sum_{i=1}^{n} \log(1 + t_i), \\ &\text{subject to} \quad \frac{1}{n}\sum_{i=1}^{n} X_{ii} \le P,\ X_{ii} \ge 0,\ t_i \ge 0,\ \begin{bmatrix} 1 - a_i t_i & \sqrt{r_i} \\ \sqrt{r_i} & a_i X_{ii} + r_i \end{bmatrix} \succeq 0,\ (i = 1, \ldots, n), \end{aligned}$$

where the decision variables are X_{ii} and t_i for $i = 1, \ldots, n$. In the experiment, the constants r_i and a_i for $i = 1, \ldots, n$ are selected randomly from the interval $[0, 1]$, and P is set to 1. Note that the objective function of the problem is concave and the constraint functions are affine.

Minimization of the minimal eigenvalue problem:

$$\begin{aligned} &\text{minimize} \quad \mathrm{tr}(\Pi M(q)), \\ &\text{subject to} \quad \mathrm{tr}(\Pi) = 1,\ \Pi \succeq 0,\ q \in Q, \end{aligned}$$

where $Q \subset \mathbf{R}^p$, and M is a function from $\mathbf{R}^p$ to $\mathbf{S}^n$, and decision variables are $q \in \mathbf{R}^p$ and $\Pi \in \mathbf{S}^n$. In the experiment, p is set to 2 and the function M is given by $M(q) \equiv q_1q_2M_1 + q_1M_2 + q_2M_3$, where $M_1, M_2, M_3 \in \mathbf{S}^n$ are constant matrices whose elements are selected randomly from the interval $[-1, 1]$. The constraint region Q is set to $[-1, 1]\times[-1, 1]$. Note that the objective function is nonconvex and the constraint functions are affine.

Nearest correlation matrix problem:

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n}{\text{minimize}} & \frac{1}{2}\|X - A\|_F^2, \\ \text{subject to} & X \succeq \eta I,\ X_{ii} = 1,\ (i = 1, \dots, n), \end{array}$$

where $A \in \mathbf{S}^n$ is a constant matrix and $\eta \in \mathbf{R}$ is a positive constant. Note that $X \succeq \eta I$ is equivalent to $X - \eta I \succeq 0$. In the experiment, the elements of the matrix A are selected randomly from the interval $[-1, 1]$ with $A_{ii} = 1$ for $i = 1, \dots, n$. Moreover, we set $\eta = 10^{-3}$. Note that the objective function is quadratic and convex, and the constraint functions are affine. Therefore, the problem is convex.

The numerical results are presented in Tables 1–3. In these tables, SDPIP denotes the interior point algorithm of [28]. From Tables 1–3, we see that the results obtained by using Algorithm 1 are comparable to those produced with SDPIP.

6 Concluding Remarks

In this paper, we proposed a new merit function F for shifted perturbed KKT conditions. We also showed the properties of the merit function F. In particular, we gave the level boundedness of the merit function F, which is not given in other related papers for nonlinear SDP. Moreover, we proposed a Newton-type method (Algorithm 2) to find an approximate shifted perturbed KKT point. We further proved the global convergence under weaker assumptions than those in [28]. In the numerical experiments, we showed that the performance of Algorithm 1 was comparable to that of the interior point method [28] based on the perturbed KKT conditions.

As future research, it is worth to show the superlinear convergence of Algorithm 1 under appropriate conditions.

Acknowledgements

The authors would like to thank Professor Fukushima and two anonymous referees for their valuable comments. In particular, Professor Fukushima gave us a lot of constructive suggestions that make the paper more readable.

Table 1: Gaussian channel capacity problem

	Algorithm 1		SDPIP	
n	iteration	time(s)	iteration	time(s)
5	126	5.17	128	5.25
10	142	19.52	153	20.36
15	190	87.58	175	84.04
20	148	184.57	126	169.09
25	169	500.51	145	432.62
30	70	472.58	63	429.41
35	106	1421.61	79	1119.49
40	110	3033.11	80	2020.00

Table 2: Minimization of the minimal eigenvalue problem

	Algorithm 1		SDPIP	
n	iteration	time(s)	iteration	time(s)
5	6	0.23	9	0.28
10	7	1.16	10	1.60
15	7	7.19	10	10.09
20	8	39.03	10	46.88
25	8	108.23	11	162.18
30	8	241.76	14	443.60
35	8	560.41	16	1161.47
40	10	1289.72	16	2092.33

Table 3: Nearest correlation matrix problem

	Algorithm 1		SDPIP	
n	iteration	time(s)	iteration	time(s)
5	8	0.13	9	0.15
10	8	1.52	10	1.79
15	10	10.33	11	11.02
20	11	37.47	12	40.68
25	10	151.93	11	180.84
30	9	307.40	10	328.88
35	11	875.31	11	872.60
40	11	1503.82	11	1461.04

References

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

[2] A. Ben-Tal, F. Jarre, M. Kocvara, A. Nemirovski and J. Zowe, Optimal design of trusses under a nonconveex global buckling constraint, *Optim. Eng.* 1 (2000) 189–213.

[3] D.S. Bernstein, *Matrix Mathematics*, Princeton University Press, New Jersey, 2009.

[4] B. Fares, D. Noll and P. Apkarian, Robust control via sequential semidefinite programming, *SIAM J. Control Optim.* 40 (2002) 1791–1820.

[5] A. Forsgren and P.E. Gill, Primal-dual interior methods for nonconvex nonlinear programming, *SIAM J. Optim.* 8 (1998) 1132–1152.

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

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

[8] C.W. J. Hol, C.W. Scherer, E.G. Van der Meche and O.H. Bosgra, A nonlinear SDP approach to fixed-order controller synthesis and comparison with two other methods applied to an active suspension system, *Eur. J. Control* 9 (2003) 13–28.

[9] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[10] F. Jarre, An interior method for nonconvex semidefinite programs, *Optim. Eng.* 1 (2000) 347–372.

[11] C. Kanzow, C. Nagel, H. Kato and M. Fukushima, Successive linearization methods for nonlinear semidefinite programs, *Comput. Optim. Appl.* 31 (2005) 251–273.

[12] A. Kato, H. Yabe and H. Yamashita, An interior point method with a primal-dual quadratic barrier penalty function for nonlinear semidefinite programming, Technical Report, Department of Mathematical Information Science, Tokyo University of Science, 2012.

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

[14] F. Leibfritz and E.M.E. Mostafa, An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems, *SIAM J. Optim.* 12 (2002) 1048–1074.

[15] R.D.C. Monteiro, Primal-dual path-following algorithms for semidefinite programming, *SIAM J. Optim.* 7 (1997) 663–678.

[16] Y.E. Nesterov and M.J. Todd, Self-scaled barriers and interior-point methods for convex programming, *Math. Oper. Res.* 22 (1997) 1–42 .

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

[18] M.L. Overton, Large-scale optimization of eigenvalues, *SIAM J. Optim.* 2 (1992) 88–120.

[19] H. Qi and D. Sun, A quadratically convergent Newton method for computing the nearest correlation matrix, *SIAM J. Matrix Anal. Appl.* 28 (2006) 360–385.

[20] T.D. Quoc, S. Gumussoy, W. Michiels and M. Diehl, Combining convex-concave decompositions and linearization approaches for solving BMIs, with application to static output feedback, *IEEE Trans. Automat. Control* 57 (2012) 1377–1390.

[21] M. Stingl, On the solution of nonlinear semidefinite programs by augmented Lagrangian methods, Doctoral Thesis, University of Erlangen, 2005.

[22] M. Stingl, M. Kocvara and G. Leugering, A new non-linear semidefinite programming algorithm with an application to multidisciplinary free material optimization, *Internat. Ser. Numer. Math.* 158 (2009) 275–295.

[23] M.J. Todd, Semidefinite optimization, *Acta Numer.* 10 (2001) 515–560.

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

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

[26] L. Vandenberghe, S. Boyd and S.P. Wu, Determinant maximization with linear matrix inequality constraints, *SIAM J. Matrix Anal. Appl.* 19 (1998) 499–533.

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

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

Manuscript received 20 May 2014
revised 27 June 2014
accepted for publication 2 August 2014

Yuya Yamakawa
Department of Applied Mathematics and Physics, Graduate School of Informatics
Kyoto University, Kyoto 606-8501, Japan
E-mail address: yamakawa@amp.i.kyoto-u.ac.jp

Nobuo Yamashita
Department of Applied Mathematics and Physics, Graduate School of Informatics
Kyoto University, Kyoto 606-8501, Japan
E-mail address: nobuo@i.kyoto-u.ac.jp