A new merit function and a descent method for semidefinite complementarity problems[1]

Nobuo Yamashita and Masao Fukushima

Department of Applied Mathematics and Physics,
Graduate School of Engineering, Kyoto University,
Kyoto 606-01, Japan

June 13, 1997

Abstract. Recently, Tseng extended several merit functions for the nonlinear complementarity problem to the semidefinite complementarity problem (SDCP) and investigated various properties of those functions. In this paper, we propose a new merit function for the SDCP based on the squared Fischer-Burmeister function and show that it has some favorable properties. Particularly, we give conditions under which the function provides a global error bound for the SDCP and conditions under which it has bounded level sets. We also present a derivative-free method for solving the SDCP and prove its global convergence under suitable assumptions.

Key Words: Semidefinite complementarity problem, merit function, error bound, bounded level sets, derivative-free descent method.

---

[1] The work of the authors was supported in part by the Scientific Research Grant-in-Aid from the Ministry of Education, Science, Sports and Culture, Japan. The work of the first author was also supported by the Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

1. Introduction

In the last several years, the semidefinite programming problem (SDP) has been studied intensively; for a survey, see [18]. One of the reasons why many researchers pay much attention to it is that SDP can be applied to some important problems such as linear programming problems, control problems and combinatorial optimization problems. Particularly, it is well known that the SDP relaxation for some combinatorial optimization problems can provide a better bound for the original problem than the LP and Lagrangian relaxations. As a natural generalization of SDP, the semidefinite complementarity problem (SDCP) is defined as follows [16]: Find a matrix $x \in \mathcal{S}$ such that

$$[\text{SDCP}] \qquad \langle x, F(x)\rangle = 0,\ x \in \mathcal{K},\ F(x) \in \mathcal{K},$$

where $\mathcal{S}$ denotes the linear space of $n \times n$ real symmetric matrices, $\mathcal{K} \subset \mathcal{S}$ denotes the closed convex cone of positive semidefinite symmetric matrices, F is a mapping from $\mathcal{S}$ into itself, and $\langle \cdot, \cdot \rangle$ is the inner product defined by

$$\langle x, y\rangle := \mathrm{tr}[xy],$$

where $x, y \in \mathcal{S}$ and $\mathrm{tr}[\cdot]$ denotes the matrix trace. When the set $\mathcal{S}$ is restricted to be the space of diagonal matrices, this problem reduces to the nonlinear complementarity problem (NCP) [7, 14]. When $F : \mathcal{S} \to \mathcal{S}$ is affine, then the SDCP is called the semidefinite linear complementarity problem [11].

Since SDCP is considered an extension of NCP, it is natural to expect that the SDCP can be solved by applying the solution methods that have been developed for NCP. In particular, various reformulations of NCP as a minimization problem have recently been proposed and studied vigorously [2, 3, 4, 5, 9, 10, 12, 13]. In this paper, we study a reformulation of SDCP to a minimization problem. We call the objective function of an equivalent minimization problem a merit function for the SDCP. A pioneering work in the study of merit functions for SDCP has recently been done by Tseng [17]. He showed that some well-known merit functions for NCP such as the gap function [1], the regularized gap function [4], the implicit Lagrangian [13], the squared Fischer-Burmeister function [3] and Luo and Tseng's class of functions [12] can be extended to SDCP. The gap function and the regularized gap function constitute an equivalent minimization problem on $\mathcal{K}$, whereas the implicit Lagrangian, the squared Fischer-Burmeister function and Luo and Tseng's class of functions constitute an equivalent minimization problem on $\mathcal{S}$ [17]. Tseng [17] also gave sufficient conditions under which any stationary point of these merit functions is a solution of the SDCP. Specifically, such a sufficient condition is the strong monotonicity of F for the regularized gap function and the implicit Lagrangian, while it is the

monotonicity of F for the squared Fischer-Burmeister function and Luo and Tseng's class of functions. Some of those conditions have recently been refined by Qi and Chen [15]. Moreover, Tseng [17] showed that the regularized gap function and the implicit Lagrangian extended to SDCP provide error bounds for the SDCP if F is strongly monotone and Lipschitz continuous. However, it has yet to be answered whether or not there exist merit functions that provide error bounds for the SDCP under the strong monotonicity of F only, or when a merit function has bounded level sets. Error bounds play an important role in analyzing the convergence rate of iterative methods for solving the SDCP. The boundedness of level sets of a merit function ensures that the sequence generated by a descent method has at least one accumulation point. For NCP, Kanzow et al. [10] proposed a merit function that provides a global error bound under the strong monotonicity of F only and has bounded level sets under the monotonicity of F and the strict feasibility of the NCP. Hence, it is natural to expect that there exists a merit function for the SDCP that enjoys such desirable properties. The purpose of the paper is to propose such a merit function and investigate its properties.

In this paper, we propose the following merit function for the SDCP:

$$f(x) := \psi(\langle x, F(x)\rangle) + \phi(x, F(x)), \tag{1}$$

where $\psi : \Re \to \Re$ and $\phi : \mathcal{S} \times \mathcal{S} \to \Re$ are defined by

$$\psi(t) := \frac{1}{4}\max\{0, t\}^4 \tag{2}$$

and

$$\phi(a, b) := \frac{1}{2}\|(a^2 + b^2)^{\frac{1}{2}} - a - b\|^2, \tag{3}$$

respectively, and the norm $\|\cdot\|$ is defined by

$$\|x\| := \sqrt{\langle x, x\rangle} = \sqrt{\mathrm{tr}[x^2]}.$$

The function ϕ is the squared Fischer-Burmeister function extended to the space $\mathcal{S}$. As shown later, the function f inherits various favorable properties of the squared Fischer-Burmeister function. Moreover, we show that f provides an error bound for the SDCP under strong monotonicity of F only and has bounded level sets under suitable conditions. Note that the function f does not reduce to an existing merit function for NCP when the set $\mathcal{S}$ is restricted to be the space of diagonal matrices. Hence it is considered a new merit function for NCP as well. Note also that it is considered a modification of the function of Luo and Tseng's class of functions. In fact, if the second term $\phi(x, F(x))$ of $f(x)$ is replaced with $\frac{1}{2}\|[(x^2 + F(x)^2)^{\frac{1}{2}} - x - F(x)]_+\|^2$, where $[\cdot]_+$ denotes the orthogonal projection onto $\mathcal{K}$, then f becomes the function

$f_{LT} : \mathcal{S} \to R$ defined by

$$f_{LT}(x) := \psi(\langle x, F(x)\rangle) + \frac{1}{2}\|[(x^2 + F(x)^2)^{\frac{1}{2}} - x - F(x)]_+\|^2,$$

which is a function of Luo and Tseng's class [12]. Tseng [17] showed a condition for a stationary point of f_{LT} to be a solution of the SDCP, but did not investigate properties of f_{LT} concerning error bounds and the boundedness of level sets. As the function f_{LT} is similar to f, f_{LT} in fact enjoys the same properties as those of f which will be shown later in this paper. Nevertheless, we focus on and analyze the function f because of the following two reasons. One is that f is simpler to deal with than f_{LT} since ϕ does not involve the projection $[\cdot]_+$ onto $\mathcal{K}$. The other is that f_{LT} does not seem to enjoy any particular properties f does not have as long as SDCP is concerned. For NCP, f_{LT} is convex under some conditions [12], but f is not convex in general under the same conditions. However, this advantage of f_{LT} over f does not carry over to SDCP. In fact, Qi and Chen [15] gave a counterexample which shows that f_{LT} for SDCP is not convex under the conditions given in [12].

The paper is organized as follows. In Section 2, we mention some basic results that will be used in the subsequent sections. In Section 3, we show that the function f defined by (1) serves as a merit function for SDCP. We also give conditions under which any stationary point of f is a solution of the SDCP. In Section 4, we consider conditions under which f provides an error bound for SDCP and conditions under which the level sets of f are bounded. We propose a derivative-free descent method for solving SDCP and prove global convergence of the method in Section 5. In Section 6, we conclude the paper with some remarks.

2. Basic properties

In this section, we mention some basic results that will play an important role in the subsequent analyses.

First, we define some concepts that will frequently be used in the paper. We basically adopt the notations used in [17].

Definition 2.1. **(a)** *For any matrix $x \in \mathcal{S}$, $[x]_+$ and $[x]_-$ denote the orthogonal projections of x onto $\mathcal{K}$ and $-\mathcal{K}$, respectively.*

(b) *For any $n \times n$ matrix x,* sym$[x]$ *is defined by*

$$\mathrm{sym}[x] := x + x^T,$$

where x^T denotes the transpose of x.

(c) *For any $c \in \mathcal{K}$, the set $\mathcal{S}_c \subset \mathcal{S}$ denotes the subspace comprising those $x \in \mathcal{S}$ whose nullspace contains the nullspace of c.*

(d) *For any $c \in \mathcal{K}$, the linear mapping $L_c : \mathcal{S}_c \to \mathcal{S}_c$ is defined by*

$$L_c[x] := cx + xc.$$

(e) *For any subspace $\mathcal{T} \subseteq \mathcal{S}$, a mapping $G : \mathcal{T} \to \mathcal{T}$ is said to be positive definite on $\mathcal{T}$ if*

$$\langle x, G[x] \rangle > 0 \quad \textit{for any } x \in \mathcal{T} \ \ (x \neq 0).$$

With these definitions, the following properties hold.

Lemma 2.1 *[17]*

(a) *For any $x \in \mathcal{S}$, we have*

$$x = [x]_+ + [x]_-, \ \ [x]_+[x]_- = 0.$$

(b) *For any $a, b \in \mathcal{S}$, let $c := (a^2 + b^2)^{\frac{1}{2}}$. Then $a, b \in \mathcal{S}_c$.*

(c) *For any $c \in \mathcal{K}$, the linear mapping L_c is positive definite and hence invertible on $\mathcal{S}_c$. Specifically, for any $x \in \mathcal{S}_c$, $L_c^{-1}[x]$ is the unique matrix $y \in \mathcal{S}_c$ satisfying*

$$cy + yc = x.$$

(d) *For any $c \in \mathcal{K}$ and $x, y \in \mathcal{S}_c$, we have*

$$\langle y, L_c[x] \rangle = \langle L_c[y], x \rangle, \ \ xL_c^{-1}[c] = L_c^{-1}[c]x = \frac{1}{2}x$$

and

$$xL_c^{-1}[x] = 0 \quad \Rightarrow \quad x = 0.$$

(e) *For any $a \in \mathcal{K}$ and $b \in \mathcal{S}$, if $a^2 - b^2 \in \mathcal{K}$, then $a - b \in \mathcal{K}$.*

(f) *For any $a \in \mathcal{K}$ and $b \in \mathcal{S}$, we have $\langle a, b \rangle \leq \langle a, [b]_+ \rangle$.* □

The proposed function f consists of the function ψ defined by (2) and the squared Fischer-Burmeister function ϕ defined by (3). We show some properties of ψ and ϕ, which will be used in Section 3 where we consider conditions under which any stationary point of f is a solution of the SDCP. The following lemma is concerned with the function ψ. The proof is easy and hence omitted.

Lemma 2.2 *The function ψ is differentiable and $t\nabla\psi(t) \geq 0$ for all $t \in \Re$. Moreover, if $t\nabla\psi(t) = 0$, then $t \leq 0$.* □

Next we give some important properties of the function ϕ.

Lemma 2.3 **(a)** *$\phi(a,b) \geq 0$ for all $(a,b) \in \mathcal{S} \times \mathcal{S}$. $\phi(a,b) = 0$ and $(a,b) \in \mathcal{S} \times \mathcal{S}$ if and only if $a \in \mathcal{K}$, $b \in \mathcal{K}$ and $\langle a, b\rangle = 0$.*

(b) *ϕ is differentiable on $\mathcal{S} \times \mathcal{S}$ and*

$$\begin{aligned} \nabla_a \phi(a,b) &= \mathrm{sym}[L_c^{-1}[c-a-b](a-c)] \in \mathcal{S}, \\ \nabla_b \phi(a,b) &= \mathrm{sym}[L_c^{-1}[c-a-b](b-c)] \in \mathcal{S}, \end{aligned}$$

where $c := (a^2 + b^2)^{\frac{1}{2}} \in \mathcal{K}$.

(c) *For every $(a,b) \in \mathcal{S} \times \mathcal{S}$, we have*

$$\langle \nabla_a \phi(a,b), \nabla_b \phi(a,b)\rangle \geq \|(c-a-b)g\|^2,$$

where $c := (a^2 + b^2)^{\frac{1}{2}} \in \mathcal{K}$ and $g := L_c^{-1}[c-a-b] \in \mathcal{S}$.

(d) *For every $(a,b) \in \mathcal{S} \times \mathcal{S}$, we have*

$$\langle a, \nabla_a \phi(a,b)\rangle + \langle b, \nabla_b \phi(a,b)\rangle = \|c-a-b\|^2,$$

where $c := (a^2 + b^2)^{\frac{1}{2}} \in \mathcal{K}$.

Proof. Parts (a), (b) and (c) are proven in [17, Lemma 6.3]. So we prove (d) only. By (b) we have

$$\begin{aligned} \langle a, \nabla_a \phi(a,b)\rangle + \langle b, \nabla_b \phi(a,b)\rangle &= \mathrm{tr}\left[L_c^{-1}[c-a-b]((a-c)a + a(a-c) + (b-c)b + b(b-c))\right] \\ &= \mathrm{tr}\left[L_c^{-1}[c-a-b]L_c[c-a-b]\right] \\ &= \|c-a-b\|^2, \end{aligned}$$

where the first equality follows from the fact that $\mathrm{tr}[xyz] = \mathrm{tr}[yzx] = \mathrm{tr}[zxy]$ for any $x, y, z \in \mathcal{S}$, the second equality follows from the definitions of c and L_c, and the last equality follows from Lemma 2.1 (d). □

In the following, we present some lemmas that will be useful in characterizing the asymptotic behavior of the merit function f.

Lemma 2.4 *For any $a \in \mathcal{S}$, we have*

$$\|[a+b]_+\| \geq \|[a]_+\| \quad \textit{for all } b \in \mathcal{K}.$$

Proof. For $a \in \mathcal{S}$, let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be its eigenvalues. For every $b \in \mathcal{K}$, there exists an orthogonal matrix p such that pbp^T is diagonal. Let $\mu_1, \mu_2, \cdots, \mu_n$ be the eigenvalues of b. Then $\mu_i \geq 0$ for all i, and b is written as

$$b = \sum_{i=1}^{n} \mu_i p_i p_i^T,$$

where p_i is the ith column of p. Thus we have

$$a + b = a + \sum_{i=1}^{n} \mu_i p_i p_i^T.$$

Let $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \cdots \leq \hat{\lambda}_n$ be the eigenvalues of $a+b$. Then, by [6, Theorem 8.1.5], the inequality $\hat{\lambda}_i \geq \lambda_i$ holds for every i. Hence we have

$$\|[a+b]_+\|^2 = \mathrm{tr}\left[[a+b]_+^2\right] = \sum_{i:\hat{\lambda}_i > 0} \hat{\lambda}_i^2 \geq \sum_{i:\lambda_i > 0} \lambda_i^2 = \mathrm{tr}\left[[a]_+^2\right] = \|[a]_+\|^2,$$

which shows the desired inequality. □

Lemma 2.5 *For every $(a, b) \in \mathcal{S} \times \mathcal{S}$, we have*

$$2\|(a^2+b^2)^{\frac{1}{2}} - a - b\|^2 \geq 2\|[(a^2+b^2)^{\frac{1}{2}} - a - b]_+\|^2 \geq \|[-a]_+\|^2 + \|[-b]_+\|^2.$$

Proof. The first inequality follows from the fact that

$$\|x\|^2 = \|[x]_+\|^2 + \|[x]_-\|^2$$

holds for any $x \in \mathcal{S}$. Next we prove the second inequality. Since $(a^2+b^2)^{\frac{1}{2}} - a \in \mathcal{K}$ by Lemma 2.1 (e), we have from Lemma 2.4

$$\|[(a^2+b^2)^{\frac{1}{2}} - a - b]_+\|^2 \geq \|[-b]_+\|^2.$$

In a similar way, we obtain the inequality

$$\|[(a^2+b^2)^{\frac{1}{2}} - a - b]_+\|^2 \geq \|[-a]_+\|^2.$$

Adding the above two inequalities yields the desired inequality. □

Now we give a key lemma for the boundedness of level sets of f, which will be used in Section 4. This result is an extension of [9, Lemma 2.6] which states the growth behavior of Fischer-Burmeister function on the space $\Re^2$.

Lemma 2.6 *For any sequence* $\{(a^k, b^k)\} \subseteq \mathcal{S} \times \mathcal{S}$, *let* $\lambda_1^k \leq \lambda_2^k \leq \cdots \leq \lambda_n^k$ *and* $\mu_1^k \leq \mu_2^k \leq \cdots \leq \mu_n^k$ *denote the eigenvalues of* a^k *and* b^k, *respectively. Then the following statements hold.*

(a) *If* $\lambda_1^k \to -\infty$, *then* $\phi(a^k, b^k) \to \infty$.

(b) *Suppose that* $\{\lambda_1^k\}$ *and* $\{\mu_1^k\}$ *are bounded below. If* $\lambda_n^k \to \infty$, *then for arbitrary positive definite matrices* $g, h \in \mathcal{S}$, *we have*

$$\langle g, a^k \rangle + \langle h, b^k \rangle \to \infty.$$

Proof. Part (a) follows directly from Lemma 2.5 and the definition (3) of ϕ. To prove (b), let p^k be an orthogonal matrix such that $p^k a^k (p^k)^T = \mathrm{diag}[\lambda_1^k, \cdots, \lambda_n^k]$ and let $g^k := p^k g (p^k)^T$. Then we have

$$\begin{aligned} \langle g, a^k \rangle &= \mathrm{tr}\left[g a^k\right] \\ &= \mathrm{tr}\left[p^k g (p^k)^T p^k a^k (p^k)^T\right] \\ &= \mathrm{tr}\left[g^k \mathrm{diag}[\lambda_1^k, \cdots, \lambda_n^k]\right] \\ &= \sum_{i=1}^{n} \lambda_i^k g_{ii}^k. \end{aligned} \tag{4}$$

Note that g^k is positive definite and hence its diagonal elements g_{ii}^k are all positive. Moreover, we claim that, for each i, the sequence $\{g_{ii}^k\}$ is bounded away from zero. To see this, suppose that there exists an index i and a subsequence $\{g_{ii}^k\}_{k \in \kappa}$ converging to zero. Since the sequence $\{p^k\}$ of orthogonal matrices is bounded, it follows that, by taking a further subsequence if necessary, the sequence $\{p^k\}_{k \in \kappa}$ is convergent to a limit, say p^*, so that $\{g^k\}_{k \in \kappa}$ converges to $g^* = p^* g (p^*)^T$. Since p^* is also an orthogonal matrix, g^* must be positive definite. However this contradicts the hypothesis that g_{ii}^* is zero. Hence $\{g_{ii}^k\}$ is bounded away from zero for each i. Since $\{\lambda_i^k\}$ is bounded below for every i and $\lambda_n^k \to \infty$, it then follows from (4) that

$$\langle g, a^k \rangle \to \infty.$$

In a similar way, we can show that $\{\langle h, b^k \rangle\}$ is bounded below. Hence (b) is proved. □

Finally we recall some concepts about a mapping $F : \mathcal{S} \to \mathcal{S}$.

Definition 2.2. (a) *F is monotone if*

$$\langle x - y, F(x) - F(y) \rangle \geq 0 \quad \textit{for all } x, y \in \mathcal{S}.$$

(b) *F is strongly monotone with modulus $\mu > 0$ if*

$$\langle x - y, F(x) - F(y)\rangle \geq \mu\|x - y\|^2 \quad \textit{for all } x, y \in \mathcal{S}.$$

(c) *$\nabla F(x)$ denotes the Jacobian of F at $x \in S$, viewed as a linear mapping from $\mathcal{S}$ to $\mathcal{S}$.*

(d) *$\nabla F(x)$ is positive semidefinite if*

$$\langle y, \nabla F(x)y\rangle \geq 0 \quad \textit{for all } y \in \mathcal{S}.$$

Note that if F is monotone, then $\nabla F(x)$ is positive semidefinite for all $x \in \mathcal{S}$.

3. Equivalent unconstrained minimization

In this section, we show the equivalence between the SDCP and the minimization problem

$$\min_{x\in\mathcal{S}} \quad f(x), \tag{5}$$

where f is defined by (1). Moreover, we give conditions under which any stationary point of problem (5) is a solution of the SDCP.

Theorem 3.1 *The function f is nonnegative on $\mathcal{S}$. $f(x) = 0$ and $x \in \mathcal{S}$ if and only if x is a solution of the SDCP.*

Proof. The theorem is evident from the definition of f and Lemma 2.3 (a). □

This theorem indicates that we can solve the SDCP by finding a global optimum of the minimization problem (5). However, the merit function f is not convex in general. So it is important to know when any stationary point of f is a global minimum of f. As shown in the next theorem, the function f has this property if F is monotone. Under the same condition, Tseng [17] showed that any stationary point of the squared Fischer-Burmeister function and the function f_{LT} is a solution of the SDCP. The proof given below is similar to that of [17, Proposition 6.1].

Theorem 3.2 *If F is differentiable, then so is f and the gradient of f is given by*

$$\begin{aligned}\nabla f(x) &= \nabla\psi(\langle x, F(x)\rangle)(\nabla F(x)x + F(x)) \\ &\quad +\nabla_a\phi(x, F(x)) + \nabla F(x)\nabla_b\phi(x, F(x)).\end{aligned} \tag{6}$$

Moreover, if F is monotone, then any stationary point of f is a solution of the SDCP.

Proof. If F is differentiable, then from the definition of f, Lemma 2.2, Lemma 2.3 (b) and the chain rule, f is also differentiable and (6) holds.

Next we prove the last part of the theorem. Define $d(x) \in \mathcal{S}$ by

$$d(x) := -\nabla\psi(\langle x, F(x)\rangle)x - \nabla_b\phi(x, F(x)).$$

Put $\alpha := \nabla\psi(\langle x, F(x)\rangle)$. Note that $\alpha \geq 0$ from the definition of ψ. Then we have

$$\begin{aligned}
\langle d(x), \nabla f(x)\rangle &= -\langle \alpha x + \nabla_b\phi(x, F(x)), \nabla F(x)(\alpha x + \nabla_b\phi(x, F(x))) + \alpha F(x) + \nabla_a\phi(x, F(x))\rangle \\
&\leq -\langle \alpha x + \nabla_b\phi(x, F(x)), \alpha F(x) + \nabla_a\phi(x, F(x))\rangle \\
&= -\alpha^2\langle x, F(x)\rangle - \alpha(\langle x, \nabla_a\phi(x, F(x))\rangle) + \langle F(x), \nabla_b\phi(x, F(x))\rangle \\
&\quad -\langle \nabla_a\phi(x, F(x)), \nabla_b\phi(x, F(x))\rangle \\
&\leq -\alpha^2\langle x, F(x)\rangle - \langle \nabla_a\phi(x, F(x)), \nabla_b\phi(x, F(x))\rangle \\
&\leq -\langle \nabla_a\phi(x, F(x)), \nabla_b\phi(x, F(x))\rangle \qquad (7)
\end{aligned}$$

where the first inequality follows from the monotonicity of F, the second inequality follows from Lemma 2.3 (d), and the third inequality follows from the fact that $\alpha \geq 0$ and $\alpha\langle x, F(x)\rangle \geq 0$ by Lemma 2.2. Moreover, by Lemma 2.3 (c), we have

$$\langle \nabla_a\phi(x, F(x)), \nabla_b\phi(x, F(x))\rangle \geq \|(c - a - b)g\|^2, \qquad (8)$$

where $a := x$, $b := F(x)$, $c := (x^2 + F(x)^2)^{\frac{1}{2}}$ and $g := L_c^{-1}[c - a - b]$. Since $\nabla f(x) = 0$ at a stationary point x, it follows from (7) and (8) that

$$\|(c - a - b)g\|^2 \leq 0,$$

which means $(c - a - b)g = (c - a - b)L_c^{-1}[c - a - b] = 0$. Since $c - a - b \in \mathcal{S}_c$ by Lemma 2.1 (b), we have $c - a - b = 0$ by Lemma 2.1 (d). Hence x is a solution of the SDCP by Lemma 2.3 (a). □

For the NCP, similar results have been proved under weaker conditions. In fact, Facchinei and Soares [2] showed that any stationary point of the squared Fischer-Burmeister function is a solution of the NCP when the function F involved in the NCP is a P_0 function. Tseng raised an interesting question as to how the P_0 property on $\Re^n$ can be extended to the space $\mathcal{S}$. As an answer to this question, Qi and Chen [15] recently proposed the following condition on the mapping $\nabla F(x) : \mathcal{S} \to \mathcal{S}$.

Condition 1. For any $y \in \mathcal{S}$ with $y \neq 0$ and for any nonsingular matrix $p \in \Re^{n\times n}$, there exists an index i such that $(py)_i \neq 0$ and $((py)\nabla F(x)(py)^T)_{ii} \geq 0$.

Qi and Chen [15] called $\nabla F(x)$ with this property a P_0 matrix and showed that any stationary point x of the squared Fischer-Burmeister function is a solution of the SDCP if $\nabla F(x)$ satisfies Condition 1.

We can also prove that, under Condition 1, any stationary point of the function f defined by (1) is a solution of the SDCP. We only state the result without proof since the proof is similar to [15, Theorem 3.3].

Theorem 3.3 *Let $x \in \mathcal{S}$ be a stationary point of f. If $\nabla F(x)$ satisfies Condition 1, then x is a solution of the SDCP.* □

4. Error bound and bounded level sets

In this section we study conditions under which level sets of f are bounded. Moreover, we give conditions under which f provides a global error bound for the SDCP.

First we establish a theorem about the boundedness of level sets of f.

Theorem 4.1 *Suppose that $F : \mathcal{S} \to \mathcal{S}$ is differentiable and monotone and that the SDCP is strictly feasible, i.e., there exists a positive definite matrix $\hat{x} \in \mathcal{S}$ such that $F(\hat{x})$ is positive definite. Then the level set $\mathcal{L}(\gamma) := \{x \in \mathcal{S} \mid f(x) \leq \gamma\}$ is nonempty and bounded for all $\gamma \geq 0$.*

Proof. First we prove the boundedness of level sets. Assume that there exists an unbounded sequence $\{x^k\} \subseteq \mathcal{L}(\gamma)$ for some $\gamma \geq 0$. It can be seen that the sequences of the minimum eigenvalues of $\{x^k\}$ and $\{F(x^k)\}$ are bounded below. In fact, if not, it follows from Lemma 2.6 (a) that $f(x^k) \to \infty$, which contradicts $\{x^k\} \subseteq \mathcal{L}(\gamma)$. Hence, the unboundedness of $\{x^k\}$ implies that the sequence of the maximum eigenvalues of $\{x^k\}$ tends to infinity. Let $\hat{x}$ be a strictly feasible solution of the SDCP. Since F is monotone, we have

$$\langle x^k, F(\hat{x})\rangle + \langle \hat{x}, F(x^k)\rangle \leq \langle x^k, F(x^k)\rangle + \langle \hat{x}, F(\hat{x})\rangle. \tag{9}$$

It follows from Lemma 2.6 (b) and the positive definiteness of $\hat{x}$ and $F(\hat{x})$ that the left-hand side of (9) tends to infinity. Hence, $\langle x^k, F(x^k)\rangle \to \infty$. Then we have $\psi(\langle x^k, F(x^k)\rangle) \to \infty$, which yields $f(x^k) \to \infty$. This contradicts $\{x^k\} \subseteq \mathcal{L}(\gamma)$.

Next we prove that $\mathcal{L}(\gamma)$ is nonempty for all $\gamma \geq 0$. Choose an arbitrary $\bar{x} \in \mathcal{S}$ and let $\bar{\gamma} := f(\bar{x})$. Then $\mathcal{L}(\bar{\gamma})$ is nonempty and, as shown above, bounded. Thus $\mathcal{L}(\bar{\gamma})$ contains a local minimizer $\tilde{x}$ of f, which satisfies the stationarity condition $\nabla f(\tilde{x}) = 0$. It then follows from Theorems 3.2 and 3.1 that $f(\tilde{x}) = 0$. Hence, $\mathcal{L}(0)$ is nonempty. Since $\mathcal{L}(0) \subseteq \mathcal{L}(\gamma)$ for all $\gamma \geq 0$, $\mathcal{L}(\gamma)$ is nonempty for all $\gamma \geq 0$. □

In a way similar to the above proof, we can prove the boundedness of level sets of a function of Luo and Tseng's class [12], provided it satisfies the properties stated in Lemma 2.6. Specifically, the function f_{LT} has bounded level sets when F is monotone and the SDCP is strictly feasible. On the other hand, the implicit Lagrangian and the squared Fischer-Burmeister function do not necessarily have bounded level sets under the same conditions, because those functions may be bounded above even if $\langle x, F(x)\rangle \to \infty$. Thus, the term $\psi(\langle x, F(x)\rangle)$ involved in f and f_{LT} plays an important role in ensuring the boundedness of level sets.

Combining Theorem 4.1 and Theorem 3.2, we can easily show the existence of a solution and the boundedness of the solution set of the SDCP.

Corollary 4.1 *Suppose that the assumptions in Theorem 4.1 are satisfied. Then the SDCP has a nonempty bounded solution set.* □

We note that Shida and Shindo [16] show the result stated in Corollary 4.1 under slightly weaker conditions. But the above proof is quite different from that of [16].

Next we show that f provides a global error bound for strongly monotone SDCPs. As in the case of the boundedness of level sets, the term $\psi(\langle x, F(x)\rangle)$ is the key to the error bound result. Note that not only the strong monotonicity of F but also the Lipschitz continuity of F is required for the implicit Lagrangian to provide a global error bound for the SDCP [17].

Theorem 4.2 *Let F be strongly monotone with modules μ. Then there exists a constant $\tau > 0$ such that*

$$\tau\|x - x^*\|^2 \le \max\{0, \langle x, F(x)\rangle\} + \|[-x]_+\| + \|[-F(x)]_+\| \quad \textit{for all } x \in \mathcal{S}, \tag{10}$$

where x^* is the unique solution of the SDCP. Moreover, there exists a constant $\tau' > 0$ such that

$$\tau'\|x - x^*\|^2 \le f(x)^{\frac{1}{4}} + f(x)^{\frac{1}{2}} \quad \textit{for all } x \in \mathcal{S}. \tag{11}$$

Proof. Since F is strongly monotone, we have

$$\begin{aligned}
\mu\|x - x^*\|^2 &\le \langle x - x^*, F(x) - F(x^*)\rangle \\
&= \langle x, F(x)\rangle + \langle x^*, -F(x)\rangle + \langle F(x^*), -x\rangle \\
&\le \max\{0, \langle x, F(x)\rangle\} + \langle x^*, [-F(x)]_+\rangle + \langle F(x^*), [-x]_+\rangle \\
&\le \max\{0, \langle x, F(x)\rangle\} + \|x^*\|\|[-F(x)]_+\| + \|F(x^*)\|\|[-x]_+\| \\
&\le \max\{1, \|x^*\|, \|F(x^*)\|\}\left(\max\{0, \langle x, F(x)\rangle\} + \|[-F(x)]_+\| + \|[-x]_+\|\right),
\end{aligned}$$

where the second inequality follows from Lemma 2.1 (f). Setting $\tau := \mu/\max\{1, \|x^*\|, \|F(x^*)\|\}$ then yields the inequality (10).

By the definition of the function ψ and Lemma 2.5, we have

$$\max\{0, \langle x, F(x)\rangle\} = \frac{1}{\sqrt{2}}\psi(\langle x, F(x)\rangle)^{\frac{1}{4}} \leq \frac{1}{\sqrt{2}} f(x)^{\frac{1}{4}}$$

and

$$\|[-F(x)]_+\| + \|[-x]_+\| \leq \sqrt{2}\phi(x, F(x))^{\frac{1}{2}} \leq \sqrt{2} f(x)^{\frac{1}{2}},$$

which yield the desired inequality (11). □

From this theorem, we obtain another result about the boundedness of level sets of f.

Corollary 4.2 *If F is differentiable and strongly monotone, then the level set $\mathcal{L}(\gamma) := \{x \in \mathcal{S} \mid f(x) \leq \gamma\}$ is nonempty and bounded for all $\gamma \geq 0$.* □

5. Descent method

In this section, we propose a descent method for solving the minimization problem (5) and prove its global convergence. The proposed method uses the following matrix $d(x) \in \mathcal{S}$ as a search direction:

$$d(x) := -\nabla\psi(\langle x, F(x)\rangle)x - \nabla_b \phi(x, F(x)). \tag{12}$$

Since $d(x)$ does not involve the Jacobian $\nabla F(x)$, the method is considered an extension of the derivative-free methods developed for NCP [8, 19]. The proposed method will turn out to be particularly useful for problems in which the Jacobian of F is costly to evaluate.

The next lemma shows that, when $\nabla F(x)$ is positive semidefinite, the matrix $d(x)$ is a descent direction of f whenever x is not a solution of the problem.

Lemma 5.1 *Suppose that x is not a solution of the SDCP. If $\nabla F(x)$ is positive semidefinite, then the matrix $d(x) \in \mathcal{S}$ defined by (12) satisfies the descent condition $\langle \nabla f(x), d(x)\rangle < 0$.*

Proof. As shown in the proof of Theorem 3.2, we have

$$\langle \nabla f(x), d(x)\rangle \leq -\|(c - a - b)g\|^2, \tag{13}$$

where $a := x, b := F(x), c := (a^2 + b^2)^{\frac{1}{2}}$ and $g := L_c^{-1}[c - a - b]$. Hence, we have

$$\langle \nabla f(x), d(x)\rangle \leq 0.$$

Moreover, if $\langle \nabla f(x), d(x)\rangle = 0$, then (13) implies $(c - a - b)g = (c - a - b)L_c^{-1}[c - a - b] = 0$. It then follows from Lemma 2.1 (d) that $c - a - b = 0$. Hence x is a solution of the SDCP by Lemma 2.3 (a). However this contradicts the assumption. Consequently, we have $\langle \nabla f(x), d(x)\rangle < 0$. □

Now we state the descent method that uses the search direction $d(x)$.

Algorithm 5.1

(S0) Choose $x^0 \in \mathcal{S}, \varepsilon \geq 0, \sigma \in (0, 1/2), \beta \in (0, 1)$ and set $k := 0$.

(S1) If $f(x^k) \leq \varepsilon$, then stop.

(S2) Let $d(x^k) := -\nabla\psi(\langle x^k, F(x^k)\rangle)x^k - \nabla_b\phi(x^k, F(x^k))$.

(S3) Compute a steplength $t_k := \beta^{m_k}$, where m_k is the smallest nonnegative integer m satisfying the Armijo-type condition

$$f(x^k + \beta^m d(x^k)) \leq (1 - \sigma\beta^{2m})f(x^k).$$

(S4) Set $x^{k+1} := x^k + t_k d(x^k), k := k + 1$, and go to S1.

Note that the generated sequence $\{x^k\}$ lies in $\mathcal{S}$ because $d(x^k) \in \mathcal{S}$ for every k.

Next we show the global convergence of Algorithm 5.1. Here we suppose that $\varepsilon = 0$ so that the algorithm generally generates an infinite sequence, i.e., x^k is not a solution of the SDCP for every k.

Theorem 5.1 *Suppose that F is continuously differentiable. Suppose either that F is monotone and the SDCP is strictly feasible or that F is strongly monotone. Then the sequence generated by Algorithm 5.1 has at least one accumulation point and any accumulation point is a solution of the SDCP.*

Proof. First, we claim that, whenever x^k is not a solution, there exists a nonnegative integer m_k in S3 of the algorithm. Assume the contrary. Then for any positive integer m, we have

$$f(x^k + \beta^m d(x^k)) - f(x^k) > -\sigma\beta^{2m} f(x^k).$$

Dividing the both sides by β^m and letting $m \to \infty$ yield

$$\langle \nabla f(x^k), d(x^k)\rangle \geq 0.$$

Since $\nabla F(x)$ is positive semidefinite under the given hypotheses, this contradicts Lemma 5.1. Hence we can find an integer m_k in S3.

Next we show that the sequence generated by the algorithm has at least one accumulation point. By the descent property of Algorithm 5.1, the sequence $\{f(x^k)\}$ is monotonically decreasing. Hence, by Theorem 4.1 or Corollary 4.2, the generated sequence is bounded, and hence it has at least one accumulation point.

The last part of the theorem can be proven in a way similar to that of [8, Proposition 11]. For completeness, we give a proof. Let x^* be an arbitrary accumulation point of the generated sequence $\{x^k\}$ and $\{x^k\}_{k\in\kappa}$ be a subsequence converging to x^*. Since $d(\cdot)$ is continuous, $\{d(x^k)\}_{k\in\kappa}$ converges to the limit $d(x^*)$. First consider the case where there exists a constant $\bar{\beta}$ such that $\beta^{m_k} \geq \bar{\beta} > 0$ for all $k \in \kappa$. Then since

$$f(x^{k+1}) \leq (1 - \sigma\bar{\beta}^2) f(x^k)$$

for all $k \in \kappa$ and the entire sequence $\{f(x^k)\}$ is decreasing, we have $f(x^*) = 0$. This implies that x^* is a solution of the SDCP. Next consider the case where there exists a further subsequence such that $\beta^{m_k} \to 0$. Note that by Armijo's rule in S3, we have

$$f(x^k + \beta^{m_k - 1} d(x^k)) - f(x^k) > -\sigma\beta^{2(m_k-1)} f(x^k).$$

Dividing both side by β^{m_k-1} and passing to the limit on the subsequence, we get

$$\langle \nabla f(x^*), d(x^*) \rangle \geq 0.$$

In view of Lemma 5.1, we deduce that x^* is a solution of the SDCP. □

6. Concluding remarks

In this paper, we have proposed a new merit function for the SDCP. We have shown that the function enjoys a number of favorable properties. In particular, it provides a global error bound under strong monotonicity of F only and it has bounded level sets when F is monotone and the SDCP is strictly feasible.

Finally, we point out that the function f is a new merit function, even if we regard it as a merit function for NCP. For NCP, f is written as

$$f(x) := \psi(\langle x, F(x)\rangle) + \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{x_i^2 + F_i(x)^2} - x_i - F_i(x)\right)^2,$$

where F is a function from $\Re^n$ into $\Re^n$, and the inner product is simply defined as $\langle x, y\rangle := x^T y$ for $x, y \in \Re^n$. As shown in this paper, f constitutes an unconstrained differentiable minimization problem equivalent to the given NCP. Any stationary point of f is a solution of the NCP if F is monotone. Moreover, f provides a global error bound for the NCP under the monotonicity of F only, and f has bounded level sets if F is monotone and the NCP is strictly feasible. Note that, for NCP, when the first term $\psi(\langle x, F(x)\rangle)$ of f is replaced by $\sum_{i=1}^{n}\psi(x_i F_i(x))$, we can show that the above mentioned properties of f are preserved under the P property of F instead of

the monotonicity of F. Hence, f retains all desirable properties of the following merit function f_{KYF} for the NCP proposed by Kanzow et al. [10].

$$f_{KYF}(x) := \sum_{i=1}^{n} \psi(x_i F_i(x)) + \frac{1}{2} \sum_{i=1}^{n} \left[\sqrt{x_i^2 + F_i(x)^2} - x_i - F_i(x) \right]_+^2 .$$

Moreover, there are at least two advantages of f over the function f_{KYF}. One is that f is twice differentiable at every nondegenerate solution of the NCP, while f_{KYF} is not twice differentiable at any solution of the NCP. The other is that f behaves like the squared Fischer-Burmeister function near each solution of the NCP. Thanks to this property, we may construct a hybrid method for NCP by combining a steepest descent method for minimizing f and Newton's method for the system of equations derived from the Fischer-Burmeister function. In this method, the steepest descent method guarantees the global convergence and Newton's method ensures a rapid local convergence property.

References

[1] A. Auslender, Optimisation: Méthodes Numériques, Masson: Paris, 1976.

[2] F. Facchinei and J. Soares, "A new merit function for nonlinear complementarity problems and a related algorithm," SIAM Journal on Optimization, to appear.

[3] A. Fischer, "An NCP-function and its use for the solution of complementarity problems," in D.-Z. Du, L. Qi and R.S. Womersley (Eds.), Recent Advances in Nonsmooth Optimization, World Scientific Publishers: Singapore, pp. 88–105, 1995.

[4] M. Fukushima, "Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems," Mathematical Programming, Vol. 53, pp. 99–110, 1992.

[5] M. Fukushima, "Merit functions for variational inequality and complementarity problems," in G. Di Pillo and F. Giannessi (Eds.), Nonlinear Optimization and Applications, Plenum Press: New York, pp. 155–170, 1996.

[6] G.H. Golub and C.F. Van Loan, Matrix Computation, The Johns Hopkins University Press: Baltimore, 1989.

[7] P.T. Harker and J.-S. Pang, "Finite-dimensional variational inequality and nonlinear complementarity problem: A survey of theory, algorithms and applications," Mathematical Programming, Vol. 48, pp. 161–220, 1990.

[8] H. Jiang, "Unconstrained minimization approaches to nonlinear complementarity problems, " Journal of Global Optimization, to appear.

[9] C. Kanzow, "Global convergence properties of some iterative methods for linear complementarity problems," SIAM Journal on Optimization, to appear.

[10] C. Kanzow, N. Yamashita and M. Fukushima, "New NCP-functions and their properties," Journal of Optimization Theory and Application, to appear.

[11] M. Kojima, S. Shindoh and S. Hara, "Interior-point methods for the monotone semidefinite linear complementarity problems," SIAM Journal on Optimization, to appear.

[12] Z.-Q. Luo and P. Tseng, "A new class of merit functions for the nonlinear complementarity problem," in M.C. Ferris and J.-S. Pang (Eds.), Complementarity and Variational Problems: State of the Art, SIAM: Philadelphia, pp. 204–225, 1997.

[13] O.L. Mangasarian and M.V. Solodov, "Nonlinear complementarity as unconstrained and constrained minimization," Mathematical Programming, Vol. 62, pp. 277–297, 1993.

[14] J.-S. Pang, "Complementarity problems," in R. Horst and P. Pardalos (Eds.), Handbook of Global Optimization, Kluwer Academic Publishers: Boston, Massachusetts, pp. 271–338, 1994.

[15] H.D. Qi and X. Chen, "On stationary points of merit functions for semi-definite complementarity problems," working paper, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, Beijing, China, 1997.

[16] M. Shida and S. Shindoh, "Monotone semidefinite complementarity problems", Research Report 312, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, Japan, 1996.

[17] P. Tseng, "Merit functions for semi-definite complementarity problems", working paper, Department of Mathematics, University of Washington, Seattle, WA, 1996.

[18] L. Vandenberghe and S. Boyd, "Semidefinite Programming," SIAM Review, Vol. 38, pp. 49–95, 1996.

[19] N. Yamashita and M. Fukushima, "On stationary points of the implicit Lagrangian for nonlinear complementarity problems," Journal of Optimization Theory and Applications, Vol. 84, pp. 653–663, 1995.