Math. Program., Ser. A 95: 431–474 (2003)

Digital Object Identifier (DOI) 10.1007/s10107-002-0306-1

Xin Chen · Paul Tseng

Non-Interior continuation methods for solving semidefinite complementarity problems*

Received: October 1999 / Accepted: April 2002
Published online: December 19, 2002 –

Abstract. There recently has been much interest in non-interior continuation/smoothing methods for solving linear/nonlinear complementarity problems. We describe extensions of such methods to complementarity problems defined over the cone of block-diagonal symmetric positive semidefinite real matrices. These extensions involve the Chen-Mangasarian class of smoothing functions and the smoothed Fischer-Burmeister function. Issues such as existence of Newton directions, boundedness of iterates, global convergence, and local superlinear convergence will be studied. Preliminary numerical experience on semidefinite linear programs is also reported.

Key words. semidefinite complementarity problem – smoothing function – non-interior continuation – global convergence – local superlinear convergence

1. Introduction

There recently has been much interest in semidefinite linear programs (SDLP) and, more generally, semidefinite linear complementarity problems (SDLCP), which are extensions of LP and LCP, respectively, with the cone of nonnegative real vectors replaced by the cone of symmetric positive semidefinite real matrices. Accordingly, there has been considerable effort to extend solution approaches for LP and LCP to SDLP and SDLCP. The main focus has been on extending the interior-point approach to solve SDLP (see [1, 2, 27, 31, 33, 37, 38, 51] and references therein), monotone SDLCP [31, 48], and semidefinite (nonlinear) complementarity problems (SDCP) [41]. Recently, extensions of the merit function approach have also been considered [49, 54]. In this paper, we consider extensions of a third approach, that of non-interior continuation, which has been extensively studied in the settings of LP and CP.

We describe the semidefinite complementarity problem (SDCP) below, using the notation in [49]. Let $\mathcal{X}$ denote the space of $n \times n$ block-diagonal real matrices with m blocks of sizes $n_1, \ldots, n_m$, respectively (the blocks are fixed). Thus, $\mathcal{X}$ is closed under matrix addition $x + y$, multiplication xy, transposition x^T, and inversion x^{-1}, where $x, y \in \mathcal{X}$. We endow $\mathcal{X}$ with the inner product and norm

$$\langle x, y\rangle := \mathrm{tr}[x^T y], \qquad \|x\| := \sqrt{\langle x, x\rangle},$$

X. Chen, P. Tseng: Department of Mathematics, University of Washington Seattle, Washington 98195, USA, (e-mail: xchen@math.washington.edu, tseng@math.washington.edu)

* This research is supported by National Science Foundation Grant CCR-9731273.

where $x, y \in \mathcal{X}$ and tr[$\cdot$] denotes the matrix trace, i.e., $\mathrm{tr}[x] = \sum_{i=1}^{n} x_{ii}$. [$\|x\|$ is the Frobenius-norm of x and ":=" means "define".] Let $\mathcal{O}$ denote the set of $p \in \mathcal{X}$ that are orthogonal, i.e., $p^T = p^{-1}$. Let $\mathcal{S}$ denote the subspace comprising those $x \in \mathcal{X}$ that are symmetric, i.e., $x^T = x$. Let $\mathcal{S}_+$ (respectively, $\mathcal{S}_{++}$) denote the convex cone comprising those $x \in \mathcal{S}$ that are positive semidefinite (respectively, positive definite). Our problem is to find, for a given mapping $F : \mathcal{S} \mapsto \mathcal{S}$, an $(x, y) \in \mathcal{S} \times \mathcal{S}$ satisfying

$$x \in \mathcal{S}_+, \quad y \in \mathcal{S}_+, \quad \langle x, y\rangle = 0, \quad F(x) - y = 0. \tag{1}$$

We will assume that F is continuously differentiable. We denote by $\mathcal{Z}$ the set of solutions of SDCP, i.e., $\mathcal{Z} := \{(x, y) \in \mathcal{S} \times \mathcal{S} : (x, y) \text{ satisfy (1)}\}$, which we assume is nonempty. This problem contains as special cases the SDLP (for which $n_2 = \cdots = n_m = 1$, and F is affine and skew-symmetric in the sense that $\langle x - y, F(x) - F(y)\rangle = 0$ for all $x, y \in \mathcal{S}$) and CP (for which $n_1 = \cdots = n_m = 1$).

We describe below the non-interior smoothing/continuation approach to solve SDCP. This approach was considered in the setting of LP/CP by Smale [42], B. Chen and Harker [11, 12], and Kanzow [28, 29], and was substantially generalized by C. Chen and Mangasarian [15, 16], based on an earlier work of Kreimer and Rubinstein. It has subsequently been extensively studied [5, 6, 9, 10, 14, 17, 18, 26, 50, 52] (also see [20, 22] for further references). In this approach, we construct a continuously differentiable function $\phi_\mu : \mathcal{S} \times \mathcal{S} \mapsto \mathcal{S}$, parameterized by a "smoothing parameter" $\mu > 0$, having the property that

$$\phi_\mu(a, b) \to 0 \text{ and } (a, b, \mu) \to (x, y, 0) \quad \Longrightarrow \quad x \in \mathcal{S}_+,\ y \in \mathcal{S}_+,\ \langle x, y\rangle = 0. \tag{2}$$

Accordingly, (1) is approximated by the smooth equation $H_\mu(x, y) = 0$, where

$$H_\mu(x, y) := (\phi_\mu(x, y), F(x) - y). \tag{3}$$

Then, starting with any $\mu > 0$ and $z \in \mathcal{S} \times \mathcal{S}$, we fix μ and apply a few Newton-type steps for $H_\mu(z) = 0$ to update z, and then we decrease μ and re-iterate. Instead of applying Newton steps to $H_\mu(z) = 0$, one can fix μ and minimize $\|H_\mu(z)\|$, possibly inexactly, using standard methods for unconstrained minimization and then decrease μ [13, 15, 16, 23]. Here, we denote $\|(a, b)\| := \sqrt{\|a\|^2 + \|b\|^2}$ for $(a, b) \in \mathcal{S} \times \mathcal{S}$.

We now consider possible choices of the smoothing function ϕ_μ. Following [50], let $\mathcal{CM}$ denote the class of convex continuously differentiable functions $g : \Re \to \Re$ with the properties that $\lim_{\tau\to-\infty} g(\tau) = 0$ and $\lim_{\tau\to\infty} g(\tau) - \tau = 0$ and $0 < g'(\tau) < 1$ for all $\tau \in \Re$. For any $g \in \mathcal{CM}$, consider the following choice of ϕ_μ based on a proposal of C. Chen and Mangasarian [15, 16] in the LP/CP case:

$$\phi_\mu(x, y) := x - \mu g((x - y)/\mu), \tag{4}$$

where, by convention, for any $a \in \mathcal{S}$ we have $g(a) = p^T \mathrm{diag}[g(\lambda_1), ..., g(\lambda_n)]p$, with $p \in \mathcal{O}$ and $\lambda_1, ..., \lambda_n \in \Re$ satisfying $a = p^T \mathrm{diag}[\lambda_1, ..., \lambda_n]p$ [25, Sec. 6.2]. It is known that $g(a)$ is well defined (independent of the ordering of $\lambda_1, ..., \lambda_n$ and the choice of p) and belongs to $\mathcal{S}$. There are many choices of $g \in \mathcal{CM}$. One is a function proposed independently by B. Chen and Harker [11, 12], Kanzow [28, 29], and Smale [42] (CHKS):

$$g(\tau) := ((\tau^2 + 4)^{1/2} + \tau)/2, \tag{5}$$

The second is the neural network function considered by C. Chen and Mangasarian [15, 16]:

$$g(\tau) := \ln(e^\tau + 1). \tag{6}$$

For these two choices of g, we have, respectively, $g(a) = ((a^2 + 4I)^{1/2} + a)/2$ and $g(a) = \ln(e^a + I)$, where I denotes the $n \times n$ identity matrix, e^a is the matrix exponential of a and $\ln(\cdot)$ denotes its inverse [25]. These two matrix functions were recently considered by Auslender [3] in the context of penalty and barrier methods for SDLP. Notice that $e^a + I$ is positive definite, so $\ln(e^a + I)$ is well defined. As one referee noted, we can associate with each $g \in \mathcal{CM}$ a dual function g°, defined by $g^\circ(\tau) := \tau + g(-\tau)$ for all $\tau \in \Re$, which also belongs to $\mathcal{CM}$ and satisfies

$$x - \mu g((x-y)/\mu) = y - \mu g^\circ((y-x)/\mu) \quad \forall \mu > 0,\ \forall x, y \in \mathcal{S}.$$

Moreover, $(g^\circ)^\circ = g$. Notice that the functions in (5) and (6) are self-dual, i.e., $g^\circ = g$. In fact, for any $g \in \mathcal{CM}$ we can derive a self-dual function $\hat{g}(\tau) := (g(\tau) + g^\circ(\tau))/2$ in $\mathcal{CM}$. We will also consider the following choice of ϕ_μ, based on a proposal of Kanzow [28] in the CP case:

$$\phi_\mu(x, y) = x + y - (x^2 + y^2 + 2\mu^2 I)^{1/2}. \tag{7}$$

This choice, called the smoothed Fischer-Burmeister (FB) function, has properties similar to (4)–(5) but does not belong to the Chen-Mangasarian class (4). In Secs. 2 and 3, we show that the above choices of ϕ_μ are continuously differentiable and satisfy (2); see Cors. 1, 2 and Lemma 5. Notice that ϕ_μ given by (4) or (7) is homogeneous of degree 1 in μ, i.e.,

$$\phi_{\mu\nu}(x, y) = \mu\phi_\nu(x/\mu, y/\mu) \qquad \forall \mu, \nu > 0,\ \forall x, y \in \mathcal{S}. \tag{8}$$

As with interior-point methods, a convergence analysis requires the iterates (z, μ) to lie in a neighborhood of the "path" defined by $H_\mu(z) = 0$ [5, 6, 9, 14, 50, 52]. We will use the following choice of neighborhood, based on one used for the CP case [6, 9, 50]:

$$\mathcal{N}_\beta := \{\, (z, \mu) \in \mathcal{S} \times \mathcal{S} \times \Re_{++} : \|H_\mu(z)\| \le \beta\mu \,\}, \tag{9}$$

where $\beta \in \Re_{++}$ is a constant. Our method iteratively moves (z, μ) along the Newton direction

$$-(\nabla H_\mu(z)^{-1} H_\mu(z), \sigma\mu)$$

($\sigma \in (0, 1)$) while maintaining it to remain in $\mathcal{N}_\beta$. In Secs. 4 and 5, we derive sufficient conditions for the Newton direction to exist and for (z, μ) to be bounded; see Lemmas 6, 7, 8. In Sec. 6, we describe the method and, in Sec. 7, we analyze the global (linear) convergence of (z, μ), extending the results in [50] to the SDCP setting; see Prop. 1 and Cor. 3. To accelerate local convergence, we also consider moving z along the "pure" Newton direction

$$-\nabla H_\mu(z)^{-1} H_0(z)$$

and decreasing μ superlinearly, where $H_0(z) := \lim_{\mu \downarrow 0} H_\mu(z)$; see Prop. 2. This direction has been used by [9, 10, 14, 17] in the LP/CP case. In Sec. 8, we prove local superlinear convergence under the assumptions of strict complementarity and nondegeneracy–the same as for interior-point path-following methods using non-shrinking neighborhood [2, 27, 32, 33]. Our proof uses the Lipschitz continuity of $\phi_\mu(z)$ in μ and the Lipschitz continuity of $\nabla\phi_\mu(z)$ in z (see Lemma 9). Extension of our results to SDLP and generalized SDCP is discussed in Sec. 9. Preliminary computational experience on SDLP is reported in Sec. 10.

In what follows, we say that F is *monotone* if

$$\langle F(x) - F(y), x - y\rangle \geq 0 \quad \forall x, y \in \mathcal{S}$$

and F is *strongly monotone* if there exists a $\rho \in \Re_{++}$ (the “modulus”) such that

$$\langle F(x) - F(y), x - y\rangle \geq \rho\|x - y\|^2 \quad \forall x, y \in \mathcal{S}.$$

We write $x \succeq y$ (respectively, $x \succ y$) to mean $x - y$ is positive semidefinite (respectively, positive definite). We denote by $\nabla F(x)$ the Jacobian of F at each $x \in \mathcal{S}$, viewed as a linear mapping from $\mathcal{S}$ to $\mathcal{S}$. For a linear mapping $M : \mathcal{S} \mapsto \mathcal{S}$, we denote its operator norm $|||M||| := \max_{\|x\|=1} \|Mx\|$ and we denote the adjoint of M by M^*, i.e., $\langle y, Mx\rangle = \langle M^*y, x\rangle$ for all $x, y \in \mathcal{S}$. For any $x \in \mathcal{S}$, we denote by x_{ij} the (i, j)th entry of x. We use $\circ$ to denote the Hadamard product, i.e., $x \circ y = [x_{ij}y_{ij}]_{i,j=1}^n$. For any $\lambda_1, ..., \lambda_n \in \Re$, we denote by $\mathrm{diag}[\lambda_1, ..., \lambda_n]$ the $n \times n$ diagonal matrix with diagonal entries $\lambda_1, ..., \lambda_n$. We will freely use the following facts about trace [24]: For any $x, y \in \mathcal{X}$ and any $p \in \mathcal{O}$, $\mathrm{tr}[x] = \mathrm{tr}[x^T] = \mathrm{tr}[pxp^T]$, $\mathrm{tr}[xy] = \mathrm{tr}[yx]$, and $\mathrm{tr}[x + y] = \mathrm{tr}[x] + \mathrm{tr}[y]$. Also, $\|\cdot\|$ is a norm on $\mathcal{X}$ and, in particular, the triangle inequality and the Cauchy-Schwartz inequality hold for $\|\cdot\|$. For $x \in \mathcal{S}$, we denote by $[x]_+$ the orthogonal projection of x onto $\mathcal{S}_+$, i.e., $[x]_+ := \arg\min_{y\in\mathcal{S}_+} \|x - y\|$. Also, $\Re_+$, $\Re_{++}$ denote the nonnegative and positive reals.

2. Lipschitzian properties of ϕ_μ

The following lemma shows that ϕ_μ given by (4) or (7) is Lipschitz continuous in μ. This extends analogous results for the CP case [5, 14, 15, 21, 28, 50]. In what follows, we denote $\phi_0(x, y) := \lim_{\mu\downarrow 0} \phi_\mu(x, y)$.

Lemma 1. *Let ϕ_μ be given by* (4) *with $g \in \mathcal{CM}$ or* (7). *Then, for any $x, y \in \mathcal{S}$ and any $\mu > \nu > 0$, we have*

$$\begin{aligned} \varrho(\mu - \nu)I &\succeq \phi_\nu(x, y) - \phi_\mu(x, y) \succ 0, \\ \varrho\mu I &\succeq \phi_0(x, y) - \phi_\mu(x, y) \succ 0, \end{aligned} \tag{10}$$

where $\varrho = g(0)$ and $\phi_0(x, y) = x - [x - y]_+$ if ϕ_μ is given by (4) *with $g \in \mathcal{CM}$, and $\varrho = \sqrt{2}$ and $\phi_0(x, y) = x + y - (x^2 + y^2)^{1/2}$ if ϕ_μ is given by* (7).

Proof. Fix any $x, y \in \mathcal{S}$ and any $\mu > \nu > 0$.

Suppose ϕ_μ is given by (4) with $g \in \mathcal{CM}$. Let $a := x - y$. Choose any $p \in \mathcal{O}$ and $\lambda_1, ..., \lambda_n \in \Re$ satisfying $a = p^T\mathrm{diag}[\lambda_1, ..., \lambda_n]p$. Then, $a/\mu = p^T\mathrm{diag}[\lambda_1/\mu, ..., \lambda_n/\mu]p$, so

$$g(a/\mu) = p^T\mathrm{diag}[g(\lambda_1/\mu), ..., g(\lambda_n/\mu)]p,$$

and analogously for $g(a/\nu)$. Thus,

$$\mu g(a/\mu) - \nu g(a/\nu) = p^T \operatorname{diag}[\mu g(\lambda_1/\mu) - \nu g(\lambda_1/\nu), ..., \mu g(\lambda_n/\mu) - \nu g(\lambda_n/\nu)]p.$$

It can be shown that $0 < \mu g(\lambda_i/\mu) - \nu g(\lambda_i/\nu) \le g(0)(\mu - \nu)$ for $i = 1, ..., n$ [50, Lemma 3.1], and hence

$$g(0)(\mu - \nu)I \succeq \mu g(a/\mu) - \nu g(a/\nu) \succ 0.$$

This together with (4) proves the first relation in (10), with $\varrho = g(0)$. Since this relation holds for any $\nu \in (0, \mu)$, taking $\nu \downarrow 0$ yields in the limit the second relation in (10). Also, using the fact $\lim_{\mu\downarrow 0} \mu g(\lambda_i/\mu) = \max\{0, \lambda_i\}$ for all i, we obtain

$$\lim_{\mu\downarrow 0} \mu g(a/\mu) = p^T \operatorname{diag}[\max\{0, \lambda_1\}, ..., \max\{0, \lambda_n\}]p.$$

By [49, Lemma 2.1], the right-hand side equals $[a]_+$. Thus, $\phi_0(x, y) = x - [x - y]_+$.

Suppose ϕ_μ is given by (7). Let $a = x^2 + y^2$ and choose any $p \in \mathcal{O}$ and $\lambda_1, ..., \lambda_n \in \Re$ satisfying $a = p^T \operatorname{diag}[\lambda_1, ..., \lambda_n]p$. Then,

$$\phi_\mu(x, y) = x + y - (a + 2\mu^2 I)^{1/2} = x + y - p^T \operatorname{diag}[(\lambda_1 + 2\mu^2)^{1/2}, ..., (\lambda_n + 2\mu^2)^{1/2}]p,$$

and similarly for $\phi_\nu(x, y)$. Thus,

$$\phi_\nu(x, y) - \phi_\mu(x, y) = p^T \operatorname{diag}[(\lambda_i + 2\mu^2)^{1/2} - (\lambda_i + 2\nu^2)^{1/2}]_{i=1}^n p.$$

Also, since $a \succeq 0$ so that $\lambda_i \ge 0$ for each i, we have

$$0 < (\lambda_i + 2\mu^2)^{1/2} - (\lambda_i + 2\nu^2)^{1/2} \le \sqrt{2}(\mu - \nu),$$

where the second inequality uses the observation that, for any $\lambda \in \Re_+$, the function $h(\mu) := (\lambda + 2\mu^2)^{1/2}$ is convex differentiable on $\Re_{++}$, with $h'(\mu) = 2(\lambda + 2\mu^2)^{-1/2}\mu \le \sqrt{2}$, so that $h(\mu) - h(\nu) \le h'(\mu)(\mu - \nu) \le \sqrt{2}(\mu - \nu)$. Thus,

$$\sqrt{2}(\mu - \nu)I \succeq \phi_\nu(x, y) - \phi_\mu(x, y) \succ 0.$$

This proves the first relation in (10) with $\varrho = \sqrt{2}$. Since this relation holds for any $\nu \in (0, \mu)$, taking $\nu \downarrow 0$ yields in the limit the second relation in (10). Also, we have

$$\phi_0(x, y) = \lim_{\mu\downarrow 0} \phi_\mu(x, y) = x + y - p^T \operatorname{diag}[\sqrt{\lambda_1}, ..., \sqrt{\lambda_n}]p = x + y - (x^2 + y^2)^{1/2}.$$

□

The constant ϱ in Lemma 1 depends on ϕ_μ. In fact, it can be seen that $\varrho = \|\phi_1(0, 0)\|/\sqrt{n}$. Using Lem. 1, we show below that ϕ_μ given by (4) or (7) satisfies (2).

Corollary 1. *Let ϕ_μ be given by* (4) *with $g \in \mathcal{CM}$ or* (7)*. Let ϱ be the constant given in Lemma 1. Then the following hold:*

(a) *For any* $x, y \in \mathcal{S}$ *and any* $\mu > \nu > 0$*, we have*

$$\|\phi_\nu(x, y) - \phi_\mu(x, y)\| \le \sqrt{n}\varrho(\mu - \nu) \quad \textit{and} \quad \|\phi_0(x, y) - \phi_\mu(x, y)\| \le \sqrt{n}\varrho\mu. \tag{11}$$

(b) ϕ_μ *satisfies* (2).

Proof. (a) Let $\lambda_1, ..., \lambda_n$ denote the eigenvalues of $\phi_\nu(x, y) - \phi_\mu(x, y)$. By (10) in Lemma 1, we have $\varrho(\mu - \nu) \ge \lambda_i > 0$, so

$$\|\phi_\nu(x, y) - \phi_\mu(x, y)\| = \sqrt{\lambda_1^2 + \cdots + \lambda_n^2} \le \sqrt{n}\varrho(\mu - \nu).$$

This proves the first inequality. An analogous argument yields the second inequality.

(b) Using (a), we have for any $a, b \in \mathcal{S}$ and any $\mu > 0$ that

$$\|\phi_0(a, b)\| \le \|\phi_\mu(a, b)\| + \|\phi_\mu(a, b) - \phi_0(a, b)\| \le \|\phi_\mu(a, b)\| + \sqrt{n}\varrho\mu.$$

Thus, as $(a, b, \mu) \to (x, y, 0)$ and $\|\phi_\mu(a, b)\| \to 0$, we have $\|\phi_0(a, b)\| \to 0$. If ϕ_μ is given by (4) with $g \in \mathcal{CM}$, then Lemma 1 implies $\phi_0(a, b) = a - [a - b]_+$ which is Lipschitz continuous in (a, b) (since $[\cdot]_+$ is nonexpansive), so $\phi_0(x, y) = 0$. By a known fact, e.g., [49, Lemma 2.1(b)], (x, y) satisfies the implications in (2). If ϕ_μ is given by (7), then $\phi_0(a, b) = a + b - (a^2 + b^2)^{1/2}$ which is continuous in (a, b) (since $(\cdot)^{1/2}$ is continuous on $\mathcal{K}$),[1] implying $\phi_0(x, y) = 0$. By [49, Lemma 6.1(b)], (x, y) satisfies the implications in (2). □

3. Differential properties of ϕ_μ

In this section we study the differential properties of the smoothing function ϕ_μ given by (4) or (7). For any $c \in \mathcal{S}_{++}$, define the linear mapping $L_c : \mathcal{S} \mapsto \mathcal{S}$ by

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

It can be seen that L_c is strictly monotone (i.e., $\langle x, L_c[x]\rangle = 2\mathrm{tr}[cx^2] > 0$ whenever $x \ne 0$) and so has an inverse L_c^{-1}, i.e., for any $x \in \mathcal{S}$, $L_c^{-1}[x]$ is the unique $d \in \mathcal{S}$ satisfying $cd + dc = x$. Moreover, $L_c^{-1}[x]$ is continuous in (x, c).[2]

Lemma 2. *Fix any* $\mu \in \Re_{++}$ *and any* $x, y, u, v \in \mathcal{S}$.
(a) *For* ϕ_μ *given by* (4) *with* $g \in \mathcal{CM}$*, we have that* ϕ_μ *is Fréchet-differentiable and*

$$\nabla\phi_\mu(x, y)(u, v) = u - p^T((p(u - v)p^T) \circ c)p, \tag{12}$$

where $p \in \mathcal{O}$ *and* $\lambda_1, ..., \lambda_n \in \Re$ *are such that* $p^T\mathrm{diag}[\lambda_1, ..., \lambda_n]p = (x - y)/\mu$*, and*

$$c_{ij} := \begin{cases} (g(\lambda_i) - g(\lambda_j))/(\lambda_i - \lambda_j) & \textit{if } \lambda_i \ne \lambda_j \\ g'(\lambda_i) & \textit{if } \lambda_i = \lambda_j. \end{cases} \tag{13}$$

[1] Suppose $y \succeq 0$ and $y^2 = x \to \bar{x}$. If $\|y\| \to \infty$, we would have $(y/\|y\|)^2 \to 0$, contradicting any cluster point of $y/\|y\|$ being nonzero. Thus $\|y\|$ is bounded. Then, any cluster point $\bar{y}$ of y satisfies $\bar{y} \succeq 0$ and $\bar{y}^2 = \bar{x}$, so $\bar{y} = \bar{x}^{1/2}$.

[2] Suppose $cy + yc = x$ and $(x, c) \to (\bar{x}, \bar{c}) \in \mathcal{S} \times \mathcal{S}_{++}$. If $\|y\| \to \infty$, we would have $c(y/\|y\|) + (y/\|y\|)c \to 0$ and so $\bar{c}(y/\|y\|) + (y/\|y\|)\bar{c} \to 0$, contradicting any cluster point of $y/\|y\|$ being nonzero. Thus $\|y\|$ is bounded. Then, any cluster point $\bar{y}$ of y satisfies $\bar{c}\bar{y} + \bar{y}\bar{c} = \bar{x}$, so $\bar{y} = L_{\bar{c}}^{-1}[\bar{x}]$.

(b) *For ϕ_μ given by* (4) *and* (5)*, we have that ϕ_μ is Fréchet-differentiable and*

$$2\nabla\phi_\mu(x, y)(u, v) = u + v - L_c^{-1}[(x - y)(u - v) + (u - v)(x - y)], \tag{14}$$

where $c := ((x - y)^2 + 4\mu^2 I)^{1/2}$.
(c) *For ϕ_μ given by* (7)*, we have that ϕ_μ is Fréchet-differentiable and*

$$\nabla\phi_\mu(x, y)(u, v) = u + v - L_c^{-1}[xu + ux + yv + vy], \tag{15}$$

where $c := (x^2 + y^2 + 2\mu^2 I)^{1/2}$.

Proof. In what follows, we will use "$O(t)$" to denote any nonnegative-valued function on $\Re_{++}$ (depending on ϕ_μ and (x, y)) such that that $\limsup_{\tau\downarrow 0} O(\tau)/\tau < \infty$.

(a) By a result of Dalecki and Krein (see Thm. V.3.3 and p. 150 of [4]), $g(a)$ is a Fréchet-differentiable function of $a \in \mathcal{S}$ and, for all $h \in \mathcal{S}$,

$$\nabla g(a)h = p^T((php^T) \circ c)p,$$

where $p \in \mathcal{O}$ and $\lambda_1, ..., \lambda_n \in \Re$ are such that $p^T \operatorname{diag}[\lambda_1, ..., \lambda_n]p = a$, and c is given by (13). [Here, $\nabla g(a)$ is the gradient of $g(a)$ as a function of $a \in \mathcal{S}$.] Then, using (4) and the chain rule, we obtain that ϕ_μ is Fréchet-differentiable and $\nabla\phi_\mu(x, y)(u, v)$ is given by (12).

(b) We have from (4) and (5) that $2\phi_\mu(x, y) = x + y - ((x - y)^2 + 4\mu^2 I)^{1/2}$. For any $u, v \in \mathcal{S}$, we have upon denoting $d := ((x - y + u - v)^2 + 4\mu^2 I)^{1/2}$ and using $d^2 - c^2 = [(d - c)c + c(d - c)] + (d - c)^2$ that

$$d - c + L_c^{-1}[(d - c)^2] = L_c^{-1}[d^2 - c^2] = L_c^{-1}[(x - y)(u - v) + (u - v)(x - y) + (u - v)^2].$$

Thus, $\|d - c\| = O(\|u - v\|)$ so $d - c = L_c^{-1}[(x - y)(u - v) + (u - v)(x - y)] + O(\|u - v\|^2)$. This and (5) yield

$$\begin{aligned} &2\phi_\mu(x, y) - 2\phi_\mu(x + u, y + v) \\ &\quad = d - c - (u + v) \ = \ L_c^{-1}[(x - y)(u - v) + (u - v)(x - y)] \\ &\qquad + O(\|u - v\|^2) - (u + v). \end{aligned}$$

The conclusion follows.

(c). For any $u, v \in \mathcal{S}$, we have upon denoting $d := ((x + u)^2 + (y + v)^2 + 2\mu^2 I)^{1/2}$ and using $d^2 - c^2 = [(d - c)c + c(d - c)] + (d - c)^2$ that

$$d - c + L_c^{-1}[(d - c)^2] = L_c^{-1}[d^2 - c^2] = L_c^{-1}[xu + ux + yv + vy + u^2 + v^2].$$

Thus, $\|d - c\| = O(\|(u, v)\|)$ so $d - c = L_c^{-1}[xu + ux + yv + vy] + O(\|(u, v)\|^2)$. This yields

$$\begin{aligned} \phi_\mu(x, y) - \phi_\mu(x + u, y + v) &= d - c - (u + v) \\ &= L_c^{-1}[xu + ux + yv + vy] + O(\|(u, v)\|^2) - (u + v). \end{aligned}$$

The conclusion follows. □

For our global convergence analysis (see Prop. 1), we need ϕ_1 to be continuously differentiable, i.e., for any $\bar{z} \in \mathcal{S} \times \mathcal{S}$, $\lim_{\|z-\bar{z}\| \to 0} |||\nabla\phi_1(z) - \nabla\phi_1(\bar{z})||| = 0$. To show this for ϕ_μ given by (4), we need the following lemma showing that the set of eigenvectors of $a \in \mathcal{S}$ has an upper Lipschitzian property.

Lemma 3. *For any $a \in \mathcal{S}$, there exist $\eta, \epsilon \in \Re_{++}$ such that*

$$\min_{\substack{p\in\mathcal{O} \\ p^T ap \in \mathcal{D}}} \|p-q\| \leq \eta\|a-b\| \quad \textit{whenever} \quad b \in \mathcal{S}, q \in \mathcal{O}, q^T bq \in \mathcal{D}, \|a-b\| \leq \epsilon, \tag{16}$$

where $\mathcal{D}$ denotes the space of $n \times n$ real diagonal matrices with decreasing diagonal entries.

Proof. See the Appendix.

By using Lemma 3, we obtain the following result showing that ϕ_μ given by (4), with $g \in \mathcal{CM}$, is continuously differentiable. If g is analytic (i.e., $g(\tau)$, with $\tau \in \mathbb{C}$, is defined and an analytic complex-valued function everywhere on $\mathbb{C}$), then ϕ_μ is infinitely differentiable. It can be seen that g given by (5) or (6) is analytic.

Lemma 4. *For any $g : \Re \mapsto \Re$, if g is continuously differentiable, then $g(a)$ is continuously differentiable in $a \in \mathcal{S}$. If g is analytic, then $g(a)$ is k-times continuously differentiable in $a \in \mathcal{S}$, for $k = 1, 2, ...$*

Proof. See the Appendix.

Corollary 2. *For ϕ_μ given by* (4) *with $g \in \mathcal{CM}$, we have that ϕ_μ is continuously differentiable. If g is analytic, then ϕ_μ is k-times continuously differentiable for $k = 1, 2, ...$*

Proof. This follows from Lemma 4 and the chain rule. □

For our global linear convergence analysis, we need $\nabla\phi_1$ to be Lipschitz continuous (on $\mathcal{S} \times \mathcal{S}$). The following lemma shows this to be true for two choices of ϕ_μ.

Lemma 5. *For ϕ_μ given by* (4)–(5) *or* (7), *$\nabla\phi_1$ is defined and Lipschitz continuous.*

Proof. See the Appendix.

4. Invertibility of Jacobian of H_μ

Using Lemma 2, we have the following two lemmas showing that F being monotone is sufficient for $\nabla H_\mu(z)$ to be invertible for all z and μ. The proof is based loosely on ideas from [48] on existence of search directions for interior-point methods. Moreover, if F is strongly monotone and $|||\nabla F(x)|||$ is uniformly bounded, then $|||\nabla H_\mu(z)^{-1}|||$ is uniformly bounded.

Lemma 6. *Suppose F is monotone and ϕ_μ is given by* (4) *with $g \in \mathcal{CM}$. Then $\nabla H_\mu(z)$ is invertible for all $z \in \mathcal{S} \times \mathcal{S}$ and $\mu > 0$. Moreover, if F is strongly monotone, then for any set $\mathcal{B} \subset \mathcal{S} \times \mathcal{S}$ such that $\sup_{z=(x,y)\in\mathcal{B}} |||\nabla F(x)||| < \infty$, we have, $\sup_{z\in\mathcal{B},\mu>0} |||\nabla H_\mu(z)^{-1}||| < \infty$.*

Proof. Fix any $x, y \in \mathcal{S}$ and $\mu \in \Re_{++}$. To show $\nabla H_\mu(x, y)$ is invertible, it suffices to show that, for any $(r, s) \in \mathcal{S} \times \mathcal{S}$, there is a unique $(u, v) \in \mathcal{S} \times \mathcal{S}$ satisfying the linear equation $\nabla H_\mu(x, y)(u, v) = (r, s)$, which by (3) is equivalent to

$$\nabla \phi_\mu(x, y)(u, v) = r, \qquad Mu - v = s, \tag{17}$$

with $M := \nabla F(x)$. Since F is monotone, M is a monotone linear mapping.

From Lemma 2(a), we have that $\nabla \phi_\mu(x, y)(u, v)$ is given by (12), where $p \in \mathcal{O}$ and $\lambda_1, ..., \lambda_n \in \Re$ are such that $p^T \mathrm{diag}[\lambda_1, ..., \lambda_n]p = (x - y)/\mu$, and c is given by (13). Then, defining the linear mapping $B : \mathcal{S} \mapsto \mathcal{S}$ by $Bu = p^T((pup^T) \circ c)p$ and letting $A := I - B$, we can rewrite (17) as

$$Au + Bv = r, \qquad Mu - v = s.$$

It can be verified that $B = B^*$ and hence $A = A^*$. Also, we obtain from (13) and the fact that $0 < g'(\tau) < 1$ for all $\tau \in \Re$ that $0 < c_{ij} < 1$ for all i, j. Thus, for any nonzero $u \in \mathcal{S}$, we have

$$\langle u, Bu \rangle = \mathrm{tr}[up^T((pup^T) \circ c)p] = \sum_{i,j=1}^{n} ((pup^T)_{ij})^2 c_{ij} > 0,$$

implying B is strongly monotone and hence invertible. Then, the above equation is equivalent to (upon eliminating v and setting $d := B^{-1}u$):

$$(AB + BMB)d = r + Bs. \tag{18}$$

For any $d \in \mathcal{S}$, we have from $A = A^*$ that

$$\begin{aligned}
\langle d, ABd \rangle &= \langle Ad, Bd \rangle \\
&= \langle (I - B)d, Bd \rangle \\
&= \mathrm{tr}[(d - p^T((pdp^T) \circ c)p)p^T((pdp^T) \circ c)p] \\
&= \mathrm{tr}[(pdp^T)((pdp^T) \circ c) - ((pdp^T) \circ c)^2] \\
&= \sum_{i,j=1}^{n} ((pdp^T)_{ij})^2 c_{ij}(1 - c_{ij}) \\
&\geq \|d\|^2 \min_{i,j} c_{ij}(1 - c_{ij}).
\end{aligned}$$

Thus AB is strongly monotone. Since M is monotone and $B = B^*$ so that BMB is monotone, this implies $AB + BMB$ is strongly monotone. Then (18) has a unique solution and so does (17).

Assume further F is strongly monotone, so that $\min_{\|d\|=1}\langle d, Md \rangle \geq \rho$ for some constant $\rho \in \Re_{++}$ independent of x. Then, we have from (17) and $A = I - B$ that

$$(I - B)u + Bv = r, \qquad Mu - v = s.$$

Since $Bu = p^T\left(c \circ (pup^T)\right)p$, multiplying the above equations on the left by p and on the right by p^T and letting $\tilde{u} := pup^T$, $\tilde{v} := pvp^T$, $\tilde{r} := prp^T$, $\tilde{s} := psp^T$, and defining $\tilde{M}$ by $\tilde{M}a := pM(p^T a)$ for any $a \in \mathcal{S}$, we have

$$(I - c\,\circ)\tilde{u} + c \circ \tilde{v} = \tilde{r}, \qquad \tilde{M}\tilde{u} - \tilde{v} = \tilde{s}.$$

Eliminating $\tilde{v}$ yields

$$(I - \ c \circ + c \circ \tilde{M})\tilde{u} = \tilde{r} + c \circ \tilde{s}.$$

Fix any $\alpha \in (0, 1]$ satisfying $\alpha < \rho$. Then, the above linear system can be written as

$$(I - (1 - \alpha)c\circ)\, \tilde{u} \ + \ c \circ (\tilde{M} - \alpha I)\tilde{u} = \tilde{r} + c \circ \tilde{s}.$$

Since $0 < c_{ij} < 1$ for all i, j, it is readily seen that $I - (1 - \alpha)c\, \circ$ is an invertible (in fact, strongly monotone) linear mapping. Then, the above linear system can be written as

$$\tilde{u} \ + \ (I - (1 - \alpha)c \circ)^{-1} c \circ (\tilde{M} - \alpha I)\tilde{u} = (I - (1 - \alpha)c \circ)^{-1}(\tilde{r} + c \circ \tilde{s}).$$

Taking the inner product of both sides with $(\tilde{M} - \alpha I)\tilde{u}$ yields

$$\begin{aligned} &\langle (\tilde{M} - \alpha I)\tilde{u}, \tilde{u} \ + \ (I - (1 - \alpha)c \circ)^{-1} c \ \circ (\tilde{M} - \alpha I)\tilde{u} \rangle \\ &= \langle (\tilde{M} - \alpha I)\tilde{u}, (I - (1 - \alpha)c \circ)^{-1}(\tilde{r} + c \ \circ \tilde{s}) \rangle. \end{aligned} \tag{19}$$

We now bound from below the left-hand side of (19) and bound from above the right-hand side of (19).

Using $0 < c_{ij} < 1$ for all i, j, we have

$$\begin{aligned} \|(I - (1 - \alpha)c \circ)^{-1}(\tilde{r} + c \circ \tilde{s})\|^2 &= \sum_{i,j=1}^{n} (\tilde{r}_{ij} + c_{ij}\tilde{s}_{ij})^2/(1 - (1 - \alpha)c_{ij})^2 \\ &\leq \sum_{i,j=1}^{n} (\tilde{r}_{ij} + c_{ij}\tilde{s}_{ij})^2/\alpha^2 \\ &\leq \sum_{i,j=1}^{n} 2\left((\tilde{r}_{ij})^2 + (c_{ij}\tilde{s}_{ij})^2 \right)/\alpha^2 \\ &\leq 2(\|\tilde{r}\|^2 + \|\tilde{s}\|^2)/\alpha^2 \\ &= 2\|(r, s)\|^2/\alpha^2. \end{aligned}$$

In addition,

$$\begin{aligned} \|(\tilde{M} - \alpha I)\tilde{u}\|^2 &= \|(M - \alpha I)u\|^2 \\ &= \|Mu\|^2 - 2\alpha\langle u, Mu\rangle + \alpha^2\|u\|^2 \\ &\leq \left(\| |M| \|^2 - 2\alpha\rho + \alpha^2 \right) \|u\|^2. \end{aligned} \tag{20}$$

$$\langle (\tilde{M} - \alpha I)\tilde{u}, \tilde{u} \rangle = \langle Mu, u\rangle - \alpha\|u\|^2 \geq (\rho - \alpha)\|u\|^2. \tag{21}$$

Also, letting $d := (I - (1 - \alpha)c \circ)^{-1}(\tilde{M} - \alpha I)\tilde{u}$ and using $0 < c_{ij} < 1$ for all i, j, we have

$$\begin{aligned} &\langle (\tilde{M} - \alpha I)\tilde{u}, (I - (1 - \alpha)c \circ)^{-1} c \circ (\tilde{M} - \alpha I)\tilde{u} \rangle \\ &= \langle (I - (1 - \alpha)c \circ)^{-1}(\tilde{M} - \alpha I)\tilde{u}, c \circ (\tilde{M} - \alpha I)\tilde{u} \rangle \\ &= \langle d, c \circ (I - (1 - \alpha)c \circ)d \rangle \\ &= \sum_{i,j=1}^{n} (d_{ij})^2 c_{ij}(1 - (1 - \alpha)c_{ij}) \\ &\geq 0. \end{aligned}$$

Using the above four inequalities, we obtain from (19) and the Cauchy-Schwartz inequality that

$$(\rho-\alpha)\|u\|^2 \le \sqrt{\||M\||^2 - 2\alpha\rho + \alpha^2}\ \|u\| \cdot \sqrt{2}\|(r,s)\|/\alpha$$

and hence

$$\|u\| \le C\|(r,s)\|, \qquad \text{where} \quad C := \sqrt{2}\,\frac{\sqrt{\||M\||^2 - 2\alpha\rho + \alpha^2}}{\alpha(\rho-\alpha)}.$$

Since $\|v\| = \|Mu - s\|$, this yields $\|(u,v)\|^2 \le \|u\|^2 + (\||M\||\|u\| + \|s\|)^2 \le C^2\|(r,s)\|^2 + (\||M\||C+1)^2\|(r,s)\|^2$, so that

$$\||\nabla H_\mu(z)^{-1}\|| \le \sqrt{C^2 + (\||M\||C+1)^2}.$$

Hence $\||\nabla H_\mu(z)^{-1}\||$ is uniformly bounded whenever $\||\nabla F(x)\||$ is uniformly bounded. □

Lemma 7. *Suppose F is monotone and ϕ_μ is given by* (7)*. Then $\nabla H_\mu(z)$ is invertible for all $z \in \mathcal{S}\times\mathcal{S}$ and $\mu > 0$. Moreover, if F is strongly monotone, then for any set $\mathcal{B} \subset \mathcal{S}\times\mathcal{S}$ such that* $\sup_{z=(x,y)\in\mathcal{B}} \||\nabla F(x)\|| < \infty$*, we have* $\sup_{z\in\mathcal{B},\mu>0} \||\nabla H_\mu(z)^{-1}\|| < \infty$.

Proof. Fix any $x, y \in \mathcal{S}$ and $\mu \in \Re_{++}$. To show $\nabla H_\mu(x,y)$ is invertible, it suffices to show that, for any $(r,s) \in \mathcal{S}\times\mathcal{S}$, there is a unique $(u,v) \in \mathcal{S}\times\mathcal{S}$ satisfying the linear equation $\nabla H_\mu(x,y)(u,v) = (r,s)$, which by (3) is equivalent to (17) with $M := \nabla F(x)$. Since F is monotone, M is a monotone linear mapping.

From Lemma 2(c), we have that $\nabla\phi_\mu(x,y)(u,v)$ is given by (15), where $c := (x^2+y^2+2\mu^2 I)^{1/2}$. Then, applying L_c to the first equation in (17) and rearranging terms yield

$$L_{c-x}[u] + L_{c-y}[v] = L_c[r], \qquad Mu - v = s.$$

Let $A := L_{c-x}$ and $B := L_{c-y}$. By observing that $c \succ x$ and $c \succ y$, we see that A, B are invertible. Also it is easily seen that $A = A^*$, $B = B^*$. Then, the above equation is equivalent to (upon eliminating v and setting $d := B^{-1}u$):

$$(AB + BMB)d = L_c[r] + Bs. \tag{22}$$

For any $d \in \mathcal{S}$, we have from $A = A^*$ that

$$\begin{aligned}
\langle d, ABd\rangle &= \langle Ad, Bd\rangle \\
&= \langle L_{c-x}[d], L_{c-y}[d]\rangle \\
&= \langle (c-x)d + d(c-x), (c-y)d + d(c-y)\rangle \\
&= \mathrm{tr}[(c-x)d(c-y)d + d^2(c-x)(c-y) \\
&\quad + d^2(c-y)(c-x) + d(c-x)d(c-y)] \\
&= \mathrm{tr}[2(c-x)d(c-y)d + d^2((c-x)(c-y) + (c-y)(c-x))] \\
&= 2\|(c-x)^{1/2}d(c-y)^{1/2}\|^2 + \mathrm{tr}[d^2((c-x-y)^2 + 2\mu^2 I)] \\
&\ge 2\|(c-x)^{1/2}d(c-y)^{1/2}\|^2 + 2\mu^2\|d\|^2,
\end{aligned} \tag{23}$$

where the inequality uses the fact $d^2 \succeq 0, (c-x-y)^2 \succeq 0$. Thus AB is strongly monotone. Since M is monotone and $B = B^*$ so that BMB is monotone, this implies $AB + BMB$ is strongly monotone. Then (22) has a unique solution and so does (17).

Assume further F is strongly monotone, so that $\min_{\|d\|=1}\langle d, Md\rangle \geq \rho$ for some constant $\rho \in \Re_{++}$ independent of x. Fix any $\alpha \in (0, 1]$ satisfying $\alpha < \rho$. Let $p \in \mathcal{O}$ and $\lambda_1, ..., \lambda_n \in \Re$ be such that $p^T \mathrm{diag}[\lambda_1, ..., \lambda_n]p = c - x + \alpha(c-y)$. Then, (22) can be written as

$$L_{\tilde{c}-\tilde{x}}[\tilde{u}] + L_{\tilde{c}-\tilde{y}}[\tilde{M}\tilde{u}] = L_{\tilde{c}}[\tilde{r}] + L_{\tilde{c}-\tilde{y}}[\tilde{s}],$$

where $\tilde{c} := pcp^T$, $\tilde{x} := pxp^T$, $\tilde{y} := pyp^T$, $\tilde{u} := pup^T$, $\tilde{v} := pvp^T$, $\tilde{r} := prp^T$, $\tilde{s} := psp^T$, and we define $\tilde{M}$ by $\tilde{M}a := pM(p^T a)$ for any $a \in \mathcal{S}$. Let $\Lambda := \mathrm{diag}[\lambda_1, ..., \lambda_n] = \tilde{c} - \tilde{x} + \alpha(\tilde{c} - \tilde{y})$. Then the above equation can be written as

$$L_\Lambda[\tilde{u}] + L_{\tilde{c}-\tilde{y}}[(\tilde{M} - \alpha I)\tilde{u}] = L_{\tilde{c}}[\tilde{r}] + L_{\tilde{c}-\tilde{y}}[\tilde{s}].$$

Since $\tilde{c} \succ \tilde{x}$ and $\tilde{c} \succ \tilde{y}$ so that $\Lambda \succ 0$, then L_Λ is invertible, so the above equation in turn can be written as

$$\tilde{u} + L_\Lambda^{-1}\left[L_{\tilde{c}-\tilde{y}}[(\tilde{M} - \alpha I)\tilde{u}]\right] = L_\Lambda^{-1}\left[L_{\tilde{c}}[\tilde{r}] + L_{\tilde{c}-\tilde{y}}[\tilde{s}]\right].$$

Taking the inner product of both sides with $(\tilde{M} - \alpha I)\tilde{u}$ yields

$$\langle \tilde{u}, (\tilde{M} - \alpha I)\tilde{u}\rangle + \langle d, L_\Lambda^{-1}\left[L_{\tilde{c}-\tilde{y}}[d]\right]\rangle = \langle (\tilde{M} - \alpha I)\tilde{u}, L_\Lambda^{-1}\left[L_{\tilde{c}}[\tilde{r}] + L_{\tilde{c}-y}[\tilde{s}]\right]\rangle, \quad (24)$$

where we let $d := (\tilde{M} - \alpha I)\tilde{u}$. We now bound from below the left-hand side of (24) and bound from above the right-hand side of (24).

By the definition of $\tilde{c}$, we have that

$$\tilde{c}^2 - (\tilde{x}\tilde{y} + \tilde{y}\tilde{x}) - 2\mu^2 I = (\tilde{x} - \tilde{y})^2 \succeq 0,$$

implying $\tilde{x}\tilde{y} + \tilde{y}\tilde{x} \preceq \tilde{c}^2$. Thus

$$(\tilde{x} + \tilde{y})^2 = \tilde{c}^2 + (\tilde{x}\tilde{y} + \tilde{y}\tilde{x}) - 2\mu^2 I \preceq 2\tilde{c}^2.$$

Therefore, by [49, Lemma 6.1(c)] (also see [4, Prop. V.1.8]),

$$\tilde{x} + \tilde{y} \preceq \sqrt{2}\tilde{c}. \quad (25)$$

By replacing $\tilde{y}$ with $-\tilde{y}$ in the above argument, we also have that

$$(\tilde{x} - \tilde{y})^2 \preceq 2\tilde{c}^2, \qquad \tilde{x} - \tilde{y} \preceq \sqrt{2}\tilde{c}. \quad (26)$$

Since $\tilde{c} - \tilde{x} \succ 0$ and $0 < \alpha \leq 1$, we have $\tilde{c} - \tilde{x} \succeq \alpha(\tilde{c} - \tilde{x})$. This together with (25) implies that

$$\Lambda = \tilde{c} - \tilde{x} + \alpha(\tilde{c} - \tilde{y}) \succeq \alpha(2\tilde{c} - \tilde{x} - \tilde{y}) \succeq \alpha(2 - \sqrt{2})\tilde{c}. \quad (27)$$

We also have $\tilde{x} + \alpha\tilde{y} = (1+\alpha)\tilde{c} - \Lambda$, so that squaring both sides and using $\alpha^2\tilde{y}^2 \preceq \tilde{y}^2$ yields

$$\begin{aligned}(1+\alpha)(\tilde{c}\Lambda + \Lambda\tilde{c}) - \Lambda^2 &= (1+\alpha)^2\tilde{c}^2 - (\tilde{x}^2 + \alpha^2\tilde{y}^2) - \alpha(\tilde{x}\tilde{y} + \tilde{y}\tilde{x}) \\ &\succeq (1+\alpha)^2\tilde{c}^2 - \tilde{c}^2 - \alpha\tilde{c}^2 \\ &= \alpha(1+\alpha)\tilde{c}^2.\end{aligned}$$

Comparing the diagonal entries, we obtain for all $i = 1, ..., n$ that

$$\alpha(1+\alpha)(\tilde{c}^2)_{ii} \le 2(1+\alpha)\tilde{c}_{ii}\lambda_i - \lambda_i^2 \le 2(1+\alpha)\frac{\lambda_i^2}{\alpha(2-\sqrt{2})} - \lambda_i^2,$$

where the second inequality uses (27). Thus,

$$\sum_{k=1}^{n}(\tilde{c}_{ik})^2 = (\tilde{c}^2)_{ii} \le \theta_1\lambda_i^2, \tag{28}$$

where $\theta_1 := \left(\frac{1}{\alpha(1-1/\sqrt{2})} - \frac{1}{1+\alpha}\right)\frac{1}{\alpha}$. Similarly, we have $(1+\alpha)(\tilde{c} - \tilde{y}) = (\tilde{x} - \tilde{y}) + \Lambda$, so squaring both sides yields

$$(1+\alpha)^2(\tilde{c} - \tilde{y})^2 = \Lambda^2 + ((\tilde{x} - \tilde{y})\Lambda + \Lambda(\tilde{x} - \tilde{y})) + (\tilde{x} - \tilde{y})^2.$$

Comparing the diagonal entries and using (26), we obtain for all $i = 1, ..., n$ that

$$\begin{aligned}(1+\alpha)^2((\tilde{c} - \tilde{y})^2)_{ii} &= \lambda_i^2 + 2(\tilde{x}_{ii} - \tilde{y}_{ii})\lambda_i + \left((\tilde{x} - \tilde{y})^2\right)_{ii} \\ &\le \lambda_i^2 + 2\sqrt{2}\tilde{c}_{ii}\lambda_i + 2(\tilde{c}^2)_{ii} \\ &\le \lambda_i^2 + 2\sqrt{2}\frac{\lambda_i^2}{\alpha(2-\sqrt{2})} + 2\theta_1\lambda_i^2,\end{aligned}$$

where the last inequality uses (27) and (28). Thus,

$$\sum_{k=1}^{n}(\tilde{c}_{ik} - \tilde{y}_{ik})^2 = \left((\tilde{c} - \tilde{y})^2\right)_{ii} \le \theta_2\lambda_i^2, \tag{29}$$

where $\theta_2 := \left(\frac{\sqrt{2}}{\alpha(1-1/\sqrt{2})} + 1 + 2\theta_1\right)\frac{1}{(1+\alpha)^2}$. Using (28) and the symmetry of $\tilde{c}$, we have

$$\begin{aligned}\left|\left(L_\Lambda^{-1}[L_{\tilde{c}}[\tilde{r}]]\right)_{ij}\right| &= |(\tilde{c}\tilde{r} + \tilde{r}\tilde{c})_{ij}|/(\lambda_i + \lambda_j) \\ &= \left|\sum_{k=1}^{n}(\tilde{c}_{ik}\tilde{r}_{kj} + \tilde{c}_{jk}\tilde{r}_{ik})\right|/(\lambda_i + \lambda_j) \\ &\le \sum_{k=1}^{n}(|\tilde{c}_{ik}||\tilde{r}_{kj}| + |\tilde{c}_{jk}||\tilde{r}_{ik}|)/(\lambda_i + \lambda_j)\end{aligned}$$

$$
\begin{aligned}
&\leq \sqrt{\sum_{k=1}^{n} |\tilde{c}_{ik}|^2 + |\tilde{c}_{jk}|^2} \sqrt{\sum_{k=1}^{n} |\tilde{r}_{kj}|^2 + |\tilde{r}_{ik}|^2/(\lambda_i + \lambda_j)} \\
&\leq \sqrt{\theta_1(\lambda_i^2 + \lambda_j^2)} \sqrt{\sum_{k=1}^{n} |\tilde{r}_{kj}|^2 + |\tilde{r}_{ik}|^2/(\lambda_i + \lambda_j)} \\
&\leq \sqrt{\theta_1} \sqrt{\sum_{k=1}^{n} |\tilde{r}_{kj}|^2 + |\tilde{r}_{ik}|^2}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\|L_\Lambda^{-1}[L_{\tilde{c}}[\tilde{r}]]\|^2 &= \sum_{i,j=1}^{n} \left| \left(L_\Lambda^{-1}[L_c[\tilde{r}]] \right)_{ij} \right|^2 \\
&\leq \sum_{i,j=1}^{n} \theta_1 \left(\sum_{k=1}^{n} |\tilde{r}_{kj}|^2 + |\tilde{r}_{ik}|^2 \right) \\
&= 2n\theta_1 \|r\|^2.
\end{aligned} \tag{30}
$$

By a similar argument using (29) in place of (28), we obtain that

$$
\|L_\Lambda^{-1}[L_{\tilde{c}-\tilde{y}}[\tilde{s}]]\|^2 \leq 2n\theta_2 \|s\|^2. \tag{31}
$$

Finally, we have

$$
\begin{aligned}
\langle d, L_\Lambda^{-1}[L_{\tilde{c}-\tilde{y}}[d]] \rangle &= \langle L_\Lambda^{-1}[d], L_{\tilde{c}-\tilde{y}}[d] \rangle \\
&= \langle h, L_{\tilde{c}-\tilde{y}} L_\Lambda[h] \rangle \\
&= \langle L_{\tilde{c}-\tilde{y}}[h], L_{\tilde{c}-\tilde{x}+\alpha(\tilde{c}-\tilde{y})}[h] \rangle \\
&= \alpha \|L_{\tilde{c}-\tilde{y}}[h]\|^2 + \langle L_{\tilde{c}-\tilde{y}}[h], L_{\tilde{c}-\tilde{x}}[h] \rangle \\
&\geq 0,
\end{aligned}
$$

where $h := L_\Lambda^{-1}[d]$ and the last inequality follows from (23). Using the above inequality together with (21), (30), (31), we obtain from (24) and the Cauchy-Schwartz inequality that

$$
\begin{aligned}
(\rho - \alpha)\|u\|^2 &\leq \langle (\tilde{M} - \alpha I)\tilde{u}, L_\Lambda^{-1} \left[L_{\tilde{c}}[\tilde{r}] + L_{\tilde{c}-\tilde{y}}[\tilde{s}] \right] \rangle \\
&\leq \|(\tilde{M} - \alpha I)\tilde{u}\| \left\| L_\Lambda^{-1} \left[L_{\tilde{c}}[\tilde{r}] + L_{\tilde{c}-\tilde{y}}[\tilde{s}] \right] \right\| \\
&\leq \|(M - \alpha I)u\| \sqrt{2n} (\sqrt{\theta_1}\|r\| + \sqrt{\theta_2}\|s\|) \\
&\leq |||M - \alpha I||| \|u\| \sqrt{2n} \sqrt{\theta_1 + \theta_2} \|(r, s)\|.
\end{aligned}
$$

Combining this with (20) and we obtain

$$
\|u\| \leq C\|(r, s)\|, \qquad \text{where} \quad C := \sqrt{|||M||| - 2\alpha\rho + \alpha^2}\,\sqrt{2n}\,\frac{\sqrt{\theta_1 + \theta_2}}{\rho - \alpha}.
$$

Since $\|v\| = \|Mu - s\|$, this yields $\|(u, v)\|^2 \le \|u\|^2 + (\||M\||\|u\| + \|s\|)^2 \le C^2\|(r, s)\|^2 + (\||M\||C + 1)^2\|(r, s)\|^2$, so that

$$\||\nabla H_\mu(z)^{-1}\|| \le \sqrt{C^2 + (\||M\||C + 1)^2}.$$

Hence $\||\nabla H_\mu(z)^{-1}\||$ is uniformly bounded whenever $\||\nabla F(x)\||$ is uniformly bounded. □

Notice that we can choose α in the proofs of Lemmas 6 and 7 to make the bounds on $\||\nabla H_\mu(z)^{-1}\||$ as sharp as possible. These bounds appear to be new even in the NCP case.

5. Boundedness of neighborhood

As is mentioned in Sec. 1, our method will generate a sequence of iterates $(z, \mu) \in \mathcal{N}_\beta$ with μ decreasing (see Algorithm 1). In this section we study sufficient conditions for z to be bounded. In the CP case, this has been much studied [9, 11, 14, 16, 28, 29, 50, 53], though their extension to the SDCP setting is not necessarily straightforward. We will consider each of the following three assumptions on F:

A1: The solution set $\mathcal{Z}$ is bounded and there exist $\eta > 0$ and $\gamma > 0$ such that

$$\min_{(\bar{x},\bar{y})\in\mathcal{Z}} \|x - \bar{x}\| \le \eta\|x - [x - F(x)]_+\|^\gamma \quad \forall x \in \mathcal{S}. \tag{32}$$

A2: F is Lipschitz continuous and a uniformly R_0-function in the sense that, for any sequence $x^k \in \mathcal{S}$, $k = 1, 2, ...,$

$$\|x^k\| \to \infty \text{ and } \lim_{k\to\infty} \frac{x^k}{\|x^k\|} \succeq 0 \text{ and } \lim_{k\to\infty} \frac{F(x^k)}{\|x^k\|} \succeq 0 \implies \lim_{k\to\infty} \inf \frac{\langle x^k, F(x^k)\rangle}{\|x^k\|^2} > 0. \tag{33}$$

A3: F is monotone and there exists $\bar{x} \in \mathcal{S}$ such that $\bar{x} \succ 0$, $F(\bar{x}) \succ 0$.

A1 is a global error bound condition based on the projection residual $x - [x - F(x)]_+$. In the CP case of $n_1 = \cdots = n_m = 1$, this condition has been much studied [34, 40]. In the general case, a result of Pang [39] showed that A1 holds with $\gamma = 1$ if F is Lipschitz continuous and strongly monotone. A2 is a generalization of the notion of R_0-matrix and uniformly R_0-function defined in the CP case [8, 19, 47]. It can be seen that A2 holds if F is Lipschitz continuous and strongly monotone or if F is affine and ∇F is representable by an R_0-matrix. A3 is a common assumption made for interior-point methods as well as for some non-interior continuation methods. In what follows, we denote for any $x \in \mathcal{S}$ with eigenvalues $\lambda_1, ..., \lambda_n$, $\min[x] = \min_i \lambda_i$.

Lemma 8. *For any $\beta \in \Re_{++}$ and $\mu_0 \in \Re_{++}$, the set $\{(z, \mu) \in \mathcal{N}_\beta : 0 < \mu \le \mu_0\}$ is bounded if any of the following conditions hold:*

B1: A1 holds and ϕ_μ is given by (4) with $g \in \mathcal{CM}$.

B2: A2 holds and ϕ_μ *satisfies* (2) *and* (8).
B3: A3 holds, and $\beta\mu_0 < \min[\bar{x}]$, $2\beta\mu_0 < \min[F(\bar{x})]$ *for some* $\bar{x} \in \mathcal{S}$, *and* ϕ_μ *is given by* (4) *with* $g \in \mathcal{CM}$ *or* (7).

Proof. Suppose B1 holds. For ϕ_μ given by (4) with $g \in \mathcal{CM}$, we have from Lemma 1 and Cor. 1(a) that, for any $(x, y, \mu) \in \mathcal{N}_\beta$ with $\mu \leq \mu_0$,

$$\begin{aligned}\|x - [x - F(x)]_+\| &= \|\phi_0(x, y) - \phi_\mu(x, y) + \phi_\mu(x, y) + [x - y]_+ - [x - F(x)]_+\| \\ &\leq \|\phi_0(x, y) - \phi_\mu(x, y)\| + \|\phi_\mu(x, y)\| + \|[x - y]_+ - [x - F(x)]_+\| \\ &\leq \sqrt{n}\varrho\mu + \|\phi_\mu(x, y)\| + \|y - F(x)\| \\ &\leq \sqrt{n}\varrho\mu + \beta\mu + \beta\mu \\ &\leq \sqrt{n}\varrho\mu_0 + \beta\mu_0 + \beta\mu_0,\end{aligned}$$

where ϱ is from Lem. 1, the second inequality uses the nonexpansive property of $[\cdot]_+$, and the third inequality uses (3) and (9). This and (32) imply $\min_{(\bar{x},\bar{y})\in\mathcal{Z}} \|x - \bar{x}\|$ is bounded. Since $\mathcal{Z}$ is bounded, then x is bounded and, by $\|y - F(x)\| \leq \beta\mu \leq \beta\mu_0$, y is bounded.

Suppose B2 holds. We argue by contradiction. Suppose there exist $\beta \in \Re_{++}$ and $\mu_0 \in \Re_{++}$ and a sequence $(x^k, y^k, \mu_k) \in \mathcal{N}_\beta$ with $0 < \mu_k \leq \mu_0$ for $k = 1, 2, \ldots$ and $\|x^k\| \to \infty$. Since F is Lipschitz continuous so that $\|F(x^k)\|/\|x^k\|$ is bounded, then, by passing to a subsequence if necessary, we can assume that $(x^k/\|x^k\|, F(x^k)/\|x^k\|)$ converges to some $(\bar{x}, \bar{y})$. Since $(x^k, y^k, \mu_k) \in \mathcal{N}_\beta$, we have (see (3) and (9))

$$\|\phi_{\mu_k}(x^k, y^k)\| \leq \beta\mu_k, \quad \|F(x^k) - y^k\| \leq \beta\mu_k \quad \forall k.$$

Dividing both sides of the second inequality by $\|x^k\|$ yields

$$\|F(x^k)/\|x^k\| - y^k/\|x^k\|\| \leq \beta\mu_k/\|x^k\| \to 0,$$

so $y^k/\|x^k\| \to \bar{y}$. Dividing both sides of the first inequality by $\|x^k\|$ and using the homogeneity property (8) yield

$$\|\phi_{\mu_k/\|x^k\|}(x^k/\|x^k\|, y^k/\|x^k\|)\| = \|\phi_{\mu_k}(x^k, y^k)\|/\|x^k\| \leq \beta\mu_k/\|x^k\| \to 0.$$

Since ϕ_μ satisfies (2) by Cor. 1, this gives in the limit $\bar{x} \succeq 0$, $\bar{y} \succeq 0$, $\langle\bar{x}, \bar{y}\rangle = 0$. Thus, $\{(x^k, y^k)\}$ satisfies the first three relations of (33), but violates the last relation. This contradicts A2.

Suppose B3 holds. Let $\bar{y} := F(\bar{x})$. Fix any $(x, y, \mu) \in \mathcal{N}_\beta$ with $0 < \mu \leq \mu_0$. The monotonicity of F implies

$$\begin{aligned}0 &\leq \langle\bar{x} - x, F(\bar{x}) - F(x)\rangle \\ &= \langle\bar{x} - x, \bar{y} - y + y - F(x)\rangle \\ &= \langle\bar{x}, \bar{y}\rangle - \langle x, \bar{y}\rangle - \langle\bar{x}, y\rangle + \langle x, y\rangle + \langle\bar{x} - x, y - F(x)\rangle \\ &\leq \langle\bar{x}, \bar{y}\rangle - \langle x, \bar{y}\rangle - \langle\bar{x}, y\rangle + \langle x, y\rangle + \|\bar{x} - x\|\beta\mu,\end{aligned}$$

where the last inequality uses (3) and (9). Thus,

$$\langle x, \bar{y}\rangle + \langle\bar{x}, y\rangle - \langle x, y\rangle - \|\bar{x} - x\|\beta\mu \leq \langle\bar{x}, \bar{y}\rangle. \tag{34}$$

Consider the case where ϕ_μ is given by (4) with $g \in \mathcal{CM}$. Fix any $\epsilon \in \Re_{++}$ satisfying $(2\beta + \epsilon)\mu_0 < \min[\bar{y}]$ and $(\beta + \epsilon)\mu_0 < \min[\bar{x}]$. We have that $g(\tau)$ increases from 0 to ∞ and $g(\tau) - \tau$ decreases from ∞ to 0 as τ goes from $-\infty$ to ∞, so there exist scalars τ_1 and τ_2 satisfying $g(\tau_1) = \epsilon$ and $g(\tau_2) - \tau_2 = \epsilon$ with $0 < g(\tau) < \epsilon$ for all $\tau < \tau_1$ and $0 < g(\tau) - \tau < \epsilon$ for all $\tau > \tau_2$. Moreover, the properties of g imply that, for any $\xi, \psi \in \Re$ satisfying $|\xi - g(\xi - \psi)| \le \beta$, we have $\xi > -\beta$, $\psi > -\beta$ and

$$\begin{aligned} \xi - \psi < \tau_1 &\Longrightarrow |\xi| < \beta + \epsilon, \\ \xi - \psi > \tau_2 &\Longrightarrow |\psi| < \beta + \epsilon, \\ \tau_1 \le \xi - \psi \le \tau_2 &\Longrightarrow |\xi| \le \beta + g(\tau_2) \text{ and } |\psi| \le \beta + g(\tau_1) - \tau_1. \end{aligned} \tag{35}$$

Consider any $p \in \mathcal{O}$ such that $p(x - y)p^T$ is diagonal. Let

$$\tilde{x} = pxp^T, \quad \tilde{y} = pyp^T, \quad \hat{x} = p\bar{x}p^T, \quad \hat{y} = p\bar{y}p^T.$$

Also, let $I := \{i \in \{1, ..., n\} : (\tilde{x}_{ii} - \tilde{y}_{ii})/\mu > \tau_2\}$, $J := \{i \in \{1, ..., n\} : (\tilde{x}_{ii} - \tilde{y}_{ii})/\mu < \tau_1\}$. By [24, Thm. 4.3.26], we have $\min[\bar{x}] = \min[\hat{x}] \le \min_i \hat{x}_{ii}$ and, similarly, $\min[\bar{y}] \le \min_i \hat{y}_{ii}$. Since $(x, y, \mu) \in \mathcal{N}_\beta$ so that

$$\begin{aligned} \beta^2 &\ge \|x/\mu - g((x - y)/\mu)\|^2 \\ &= \|\tilde{x}/\mu - \operatorname{diag}[g((\tilde{x}_{ii} - \tilde{y}_{ii})/\mu)]_{i=1}^n\|^2 \\ &= \sum_{i \ne j} (\tilde{x}_{ij}/\mu)^2 + \sum_{i=1}^n (\tilde{x}_{ii}/\mu - g((\tilde{x}_{ii} - \tilde{y}_{ii})/\mu))^2, \end{aligned}$$

we have (also using the fact $\tilde{x} - \tilde{y}$ is diagonal)

$$\sum_{i \ne j} (\tilde{x}_{ij})^2 = \sum_{i \ne j} (\tilde{y}_{ij})^2 \le \beta^2\mu^2, \qquad |\tilde{x}_{ii}/\mu - g((\tilde{x}_{ii} - \tilde{y}_{ii})/\mu)| \le \beta \ \forall i. \tag{36}$$

It follows from the last relation in (36) and (35) that $|\tilde{y}_{ii}|/\mu < \beta + \epsilon$ for $i \in I$, $|\tilde{x}_{ii}|/\mu < \beta + \epsilon$ for $i \in J$, and $|\tilde{x}_{ii}|/\mu \le \beta + g(\tau_2)$, $|\tilde{y}_{ii}|/\mu \le \beta + g(\tau_1) - \tau_1$ for $i \in \{1, ..., n\} \backslash (I \cup J)$. This, together with (34), $\|\bar{x} - x\| \le \|\bar{x}\| + \|x\| \le \|\bar{x}\| + \sum_{i,j} |\tilde{x}_{ij}|$, $\mu \le \mu_0$ and the first relation in (36), implies (after some algebra)

$$\sum_{i \in I} \tilde{x}_{ii}\hat{y}_{ii} + \sum_{i \in J} \hat{x}_{ii}\tilde{y}_{ii} - \sum_{i \in I \cup J} \tilde{x}_{ii}\tilde{y}_{ii} - \sum_{i \in I} |\tilde{x}_{ii}|\beta\mu_0 \le \eta$$

for some constant η (depending on $g, \beta, \mu_0, \|\bar{x}\|, \|\bar{y}\|$ only). Rewrite this as

$$\sum_{i \in I} \tilde{x}_{ii}(\hat{y}_{ii} - \tilde{y}_{ii} - \beta\mu_0) + \sum_{i \in J} (\hat{x}_{ii} - \tilde{x}_{ii})\tilde{y}_{ii} \le \eta + \sum_{i \in I} (|\tilde{x}_{ii}| - \tilde{x}_{ii})\beta\mu_0$$

and note that, for $i \in I$, we have from $|\tilde{y}_{ii}|/\mu < \beta + \epsilon$ and our choice of ϵ that

$$\hat{y}_{ii} - \tilde{y}_{ii} - \beta\mu_0 > \hat{y}_{ii} - (\beta + \epsilon)\mu_0 - \beta\mu_0 \ge \min[\bar{y}] - (2\beta + \epsilon)\mu_0 > 0$$

and, for $i \in J$, we have from $|\tilde{x}_{ii}|/\mu < \beta + \epsilon$ and our choice of ϵ that

$$\hat{x}_{ii} - \tilde{x}_{ii} > \hat{x}_{ii} - (\beta + \epsilon)\mu_0 \ge \min[\bar{x}] - (\beta + \epsilon)\mu_0 > 0.$$

Also, by (36) and (35), we have $\tilde{x}_{ii} > -\beta\mu_0$ (so $|\tilde{x}_{ii}| - \tilde{x}_{ii} < 2\beta\mu_0$) and $\tilde{y}_{ii} > -\beta\mu_0$ for all $i = 1, ..., n$. These together imply that $\tilde{x}_{ii}$, $i \in I$, and $\tilde{y}_{ii}$, $i \in J$, are bounded. We have already shown earlier that $\tilde{x}_{ii}$, $i \notin I$, and $\tilde{y}_{ii}$, $i \notin J$, are bounded, so all diagonal entries of $\tilde{x}$ and $\tilde{y}$ are bounded. By (36) and $\mu \le \mu_0$, the off-diagonal entries of $\tilde{x}$ and $\tilde{y}$ are also bounded. Thus, $\|\tilde{x}\| = \|x\|$ and $\|\tilde{y}\| = \|y\|$ are bounded (by constants depending on $g, \beta, \mu_0, \bar{x}, \bar{y}$ only).

Consider the case ϕ_μ is given by (7). Letting $\zeta_1 := \|x + y\|$ and $\zeta_2 := \|(x^2 + y^2 + 2\mu^2 I)^{1/2}\|$, we have

$$\beta\mu \ge \|x + y - (x^2 + y^2 + 2\mu^2 I)^{1/2}\| \ge |\zeta_1 - \zeta_2| = |\zeta_1^2 - \zeta_2^2|/(\zeta_1 + \zeta_2).$$

Since $\zeta_1^2 - \zeta_2^2 = \mathrm{tr}[(x + y)^2] - \mathrm{tr}[x^2 + y^2 + 2\mu^2 I] = 2\langle x, y\rangle - 2\mu^2 n$, then $\zeta_1 > \zeta_2$ would imply $\beta\mu \ge (\zeta_1^2 - \zeta_2^2)/(2\zeta_1) = (\langle x, y\rangle - \mu^2 n)/\|x + y\|$ while $\zeta_1 \le \zeta_2$ would imply $\langle x, y\rangle \le \mu^2 n$. Thus, in either case, we have

$$\langle x, y\rangle \le \beta\mu\|x + y\| + \mu^2 n.$$

Letting $x_+ := [x]_+$, $x_- := -[-x]_+$ and similarly for y_+, y_-, this together with (34) imply

$$\begin{aligned}
\langle \bar{x}, \bar{y}\rangle + \mu^2 n &\ge -\beta\mu\|x + y\| + \langle x, \bar{y}\rangle + \langle \bar{x}, y\rangle - \|\bar{x} - x\|\beta\mu \\
&= -\beta\mu\|x_+ + x_- + y_+ + y_-\| + \langle x_+, \bar{y}\rangle + \langle x_-, \bar{y}\rangle \\
&\quad + \langle \bar{x}, y_+\rangle + \langle \bar{x}, y_-\rangle - \|\bar{x} - x_+ - x_-\|\beta\mu \\
&\ge -\beta\mu(\|x_+\| + \|x_-\| + \|y_+\| + \|y_-\|) + \|x_+\| \min[\bar{y}] - \|x_-\|\|\bar{y}\| \\
&\quad + \min[\bar{x}]\|y_+\| - \|\bar{x}\|\|y_-\| - (\|\bar{x}\| + \|x_-\| + \|x_+\|)\beta\mu \\
&= \|x_+\|(\min[\bar{y}] - 2\beta\mu) + \|y_+\|(\min[\bar{x}] - \beta\mu) \\
&\quad - \|x_-\|\|\bar{y}\| - \|\bar{x}\|\|y_-\| - (\|\bar{x}\| + 2\|x_-\| + \|y_-\|)\beta\mu,
\end{aligned} \tag{37}$$

where the second inequality uses the Cauchy-Schwartz inequality and the fact $\langle a, b\rangle \ge \|b\| \min[a]$ for any $a \in \mathcal{S}_{++}$ and $b \in \mathcal{S}_+$ [48, Eq. (26)]. By [49, Lemma 6.1(c)], we see that $x - \phi_\mu(x, y) = (x^2 + y^2 + 2\mu^2 I)^{1/2} - y \in \mathcal{S}_+$, and hence

$$\beta\mu \ge \|\phi_\mu(x, y)\| = \|x - \phi_\mu(x, y) - x\| \ge \|x_+ - x\| = \|x_-\|,$$

where the second inequality uses the fact x_+ is the nearest-point projection of x onto $\mathcal{S}_+$. A symmetric argument shows that $\beta\mu \ge \|y_-\|$. This together with (37) yields

$$\langle \bar{x}, \bar{y}\rangle + \mu^2 n \ge \|x_+\|(\min[\bar{y}] - 2\beta\mu) + \|y_+\|(\min[\bar{x}] - \beta\mu) - (2\|\bar{x}\| + \|\bar{y}\| + 3\beta\mu)\beta\mu.$$

Since $\min[\bar{y}] - 2\beta\mu > 0$ and $\min[\bar{x}] - \beta\mu > 0$, this shows $\|x_+\|$ and $\|y_+\|$ are bounded. Since we already showed that $\|x_-\| \le \beta\mu$ and $\|y_-\| \le \beta\mu$, then $\|x\|$ and $\|y\|$ are bounded. □

In the CP case, ϕ_μ can also be given by (7) in B1; see [46, Lemma 3.3]. It is not known whether this extends to SDCP in general. By Cor. 1(b), B2 applies to ϕ_μ given by (4) with $g \in \mathcal{CM}$ or (7). Condition B3 is an extension of the assumption made in [50, Lemma 3.4] in the CP case (also see [9, Prop. 8] and [53, Lemma 2.4] for similar assumptions). It is known that (see [53, Example 2.5] and [50, pp. 392–393]) the assumption of $\beta\mu_0$ being sufficiently small cannot be removed from B3.

6. Algorithm description

Below, we formally describe our method, parameterized by $\beta, \varrho, \sigma \in \Re_{++}$, $\psi \in (0, 1)$, ϕ_μ and π. This method is a direct extension of a method studied in [50] for the CP case. Related non-interior path-following methods for the CP case are given in [5, 9, 14, 53].

Algorithm 1. *Choose any* $(z^0, \mu_0) \in \mathcal{N}_\beta$*, any* $\psi \in (0, 1)$, $\sigma \in (0, \beta/(\beta + \sqrt{n}\varrho))$*, and any continuous* $\pi : \Re_{++} \mapsto \Re$ *satisfying* $0 < \pi(\mu) \leq (1 - \sigma)\mu$ *for all* $\mu \in \Re_{++}$*. For* $t = 0, 1, \ldots$*, we generate* (z^{t+1}, μ_{t+1}) *from* (z^t, μ_t) *as follows:*

Iteration. *Let* $w^t \in \mathcal{S} \times \mathcal{S}$ *satisfy*

$$\nabla H_{\mu_t}(z^t)w^t = -H_{\mu_t}(z^t). \tag{38}$$

Choose θ_t *to be the largest* $\theta \in \{1, \psi, \psi^2, \ldots\}$ *such that* $(z^t + \theta w^t, (1 - \sigma\theta)\mu_t) \in \mathcal{N}_\beta$*. Choose any* $\hat{z}^t \in \mathcal{S} \times \mathcal{S}$ *and any* $\nu^t \in \{0, 1, \ldots\}$*. Choose* α_t *to be the largest* $\alpha \in \{1, \psi, \psi^2, \ldots, \psi^{\nu^t}, 0\}$ *such that*

$$\left((z^t + \theta_t w^t)(1 - \alpha) + \hat{z}^t\alpha, (1 - \sigma\theta_t)\mu_t(1 - \alpha) + \pi(\mu_t)\alpha\right) \in \mathcal{N}_\beta, \tag{39}$$

and let

$$z^{t+1} := (z^t + \theta_t w^t)(1 - \alpha_t) + \hat{z}^t\alpha_t, \quad \mu_{t+1} := (1 - \sigma\theta_t)\mu_t(1 - \alpha_t) + \pi(\mu_t)\alpha_t. \tag{40}$$

Roughly speaking, at iteration t of Algorithm 1, we first compute a Newton direction w^t by solving the linear equations (38), and next we move (z^t, μ_t) in the direction $(w^t, -\sigma\mu_t)$ by as "large" a stepsize θ_t as possible while remaining in the neighborhood $\mathcal{N}_\beta$ (this is done using an Armijo-Goldstein-type line search), and lastly we move the resulting pair as near to $(\hat{z}^t, \pi(\mu_t))$ as possible while remaining in $\mathcal{N}_\beta$. [If $\alpha_t = 1$, then $(z^{t+1}, \mu_{t+1}) = (\hat{z}^t, \pi(\mu_t))$. The integer ν^t controls the accuracy and the work in computing α_t.] This last move is designed to accelerate the convergence of the method. Notice that, by choice of π, $\pi(\mu_t)$ is always below $1 - \sigma\theta_t$ and, assuming $\lim_{\mu\downarrow 0} \pi(\mu)/\mu = 0$, tends to zero superlinearly in μ_t. While $\hat{z}^t$ can be chosen arbitrarily without affecting the global convergence properties of the method, we would like $\hat{z}^t$ to be near the solution set $\mathcal{Z}$ for reasons of practical efficiency and improved convergence. In Sec. 8 we will consider a particular choice of $\hat{z}^t$.

In the next section, we will show that if $\nabla H_{\mu_t}(z^t)$ is invertible (so w^t is uniquely defined), then θ_t is defined and positive (due to our choice of σ) and, since α_t is well defined ($\alpha = 0$ always satisfies (39)), hence (z^{t+1}, μ_{t+1}) is well defined and, by $\sigma \in (0, 1)$ and $\pi(\mu_t) > 0$, μ_{t+1} is positive. We remark that, although our focus is on extending the method in [50], related continuation/smoothing methods (see, e.g., [5, 9, 10, 14–18, 28, 29, 52, 53]) may likely be similarly extended to the SDCP setting. Also, the convergence of our method can be further improved (in theory, at least) by letting $\mu_t[\theta] := \min\{\mu > 0 : (z^t + \theta w^t, \mu) \in \mathcal{N}_\beta\}$ and choosing $\theta_t \in \arg\min_{\theta>0} \mu_t[\theta]$, etc.

7. Global convergence analysis

In this section we use the results from Secs. 2–5 to analyze the global (linear) convergence of Algorithm 1. We begin with the following proposition showing that global convergence rests with $\nabla H_\mu(z)$ being invertible and $\|z\|$ being bounded for all $(z, \mu) \in \mathcal{N}_\beta$ with $\mu \le \mu_0$. If in addition $|||\nabla H_\mu(z)^{-1}|||$ is uniformly bounded and F and ϕ_μ are sufficiently smooth, then linear convergence is achieved. This result and its proof are based on [50, Prop. 3.1] for the CP case (also see [9, 14] for related results).

Proposition 1. *Fix any $\beta \in \Re_{++}$. Fix any ϕ_μ satisfying* (2), (8), *and* (11) *for some $\varrho > 0$, and with ϕ_1 continuously differentiable. Assume $\nabla H_\mu(z)$ is invertible for all $(z, \mu) \in \mathcal{N}_\beta$ with $\mu \le \mu_0 \in \Re_{++}$. Then $\{(z^t, \mu_t, \theta_t)\}_{t=0,1\ldots}$ generated by Algorithm 1 is well defined and satisfies $(z^t, \mu_t) \in \mathcal{N}_\beta$ and $\mu_{t+1} \le (1 - \sigma\theta_t)\mu_t$ for all t, where σ and ψ are chosen in the method. Moreover, the following holds.*

(a) *If $\{z^t\}$ has a convergent subsequence, then $\{\mu_t\} \downarrow 0$ and the limit is in $\mathcal{Z}$.*
(b) *If there exist $\kappa, \lambda_1 \in \Re_{++}$, $\lambda_2 \in \Re_+$, and $\gamma_1 > 1$, $\gamma_2 > 1$ such that $|||\nabla H_{\mu_t}(z^t)^{-1}||| \le \kappa$ for all t and $\|\phi_1(r + s) - \phi_1(r) - \nabla\phi_1(r)s\| \le \lambda_1\|s\|^{\gamma_1}$ and $\|F(x + u) - F(x) - \nabla F(x)u\| \le \lambda_2\|u\|^{\gamma_2}$ for all $r, s \in \mathcal{S} \times \mathcal{S}$ and $x, u \in \mathcal{S}$, then $\mu_{t+1} \le \omega\mu_t$ for all t, where $\omega \in (0, 1)$ depends on $\beta, \psi, \sigma, \gamma_1, \gamma_2, \sqrt{n}\varrho, \lambda_1(\kappa\beta)^{\gamma_1}, \lambda_2(\kappa\beta)^{\gamma_1}(\mu_0)^{\gamma_1-1}$ only.*

Proof. At the start of each iteration $t = 0, 1, \ldots$, we have $0 < \mu_t \le \mu_0$ (since μ_t is monotonically decreasing with t) and $(z^t, \mu_t) \in \mathcal{N}_\beta$, so $\nabla H_{\mu_t}(z^t)$ is invertible by assumption, implying w^t is well defined. We show below that θ_t is well defined and positive. Then, since α_t is well defined, so is (z^{t+1}, μ_{t+1}) given by (40). Moreover, our choice (39) of α_t ensures that $(z^{t+1}, \mu_{t+1}) \in \mathcal{N}_\beta$ and the property $\pi(\mu) \le (1-\sigma)\mu$ for all $\mu \in \Re_{++}$ ensures that $\mu_{t+1} \le (1 - \sigma\theta_t)\mu_t$.

(a) By assumption, there is some subsequence T of $\{0, 1, \ldots\}$ and some $z^\infty = (x^\infty, y^\infty) \in \mathcal{S} \times \mathcal{S}$ such that $\{z^t\}_{t\in T} \to z^\infty$. Since μ_t is monotonically decreasing, $\{\mu_t\} \to$ some $\mu^\infty \ge 0$. Since $(z^t, \mu_t) \in \mathcal{N}_\beta$ for all $t \in T$, then if $\mu^\infty = 0$, we would have $(z^t, \mu_t) \to (z^\infty, 0)$ and $\phi_{\mu_t}(x^t, y^t) \to 0$, $F(x^t) - y^t \to 0$ as $t \in T, t \to \infty$. Since F is continuous and ϕ_μ satisfies (2), we obtain in the limit that $z^\infty = (x^\infty, y^\infty)$ satisfies (1), so $z^\infty \in \mathcal{Z}$. Thus, it remains to consider the case $\mu^\infty > 0$. For any $(z, \mu) \in \mathcal{N}_\beta$ and $\theta \in [0, 1]$, $w := -\nabla H_\mu(z)^{-1}H_\mu(z)$ and $z^+ := z + \theta w$ satisfy

$$
\begin{aligned}
\|H_\mu(z^+)\| &= \|H_\mu(z + \theta w) - H_\mu(z) - \theta\nabla H_\mu(z)w + (1 - \theta)H_\mu(z)\| \\
&\le \|H_\mu(z + \theta w) - H_\mu(z) - \theta\nabla H_\mu(z)w\| + (1 - \theta)\|H_\mu(z)\| \\
&\le r(z, \mu, \theta w) + (1 - \theta)\beta\mu,
\end{aligned}
$$

where we denote the 1st-order remainder $r(z, \mu, a) := \|H_\mu(z+a) - H_\mu(z) - \nabla H_\mu(z)a\|$. Thus, for $\mu^+ := (1 - \sigma\theta)\mu$ and writing $z^+ = (x^+, y^+)$, we have

$$
\begin{aligned}
\|H_{\mu^+}(z^+)\| &\le \|H_\mu(z^+)\| + \|H_\mu(z^+) - H_{\mu^+}(z^+)\| \\
&= \|H_\mu(z^+)\| + \|\phi_\mu(x^+, y^+) - \phi_{\mu^+}(x^+, y^+)\| \\
&\le r(z, \mu, \theta w) + (1 - \theta)\beta\mu + \sqrt{n}\varrho(\mu - \mu^+),
\end{aligned}
$$

where the equality uses (3) and the last inequality also uses Cor. 1(a). Since the right-hand side is below $\beta\mu^+$ whenever

$$r(z,\mu,\theta w)/\theta + \sqrt{n}\varrho\sigma\mu \le (1-\sigma)\beta\mu,$$

which, by our choice of σ and the fact $r(z,\mu,\theta w)/\theta \to 0$ as $\theta \to 0$, occurs whenever θ is sufficiently small, it follows from our choice of θ_t that θ_t is well defined and positive for all t. Moreover, either $\theta_t = 1$ or else

$$r(z^t,\mu_t,(\theta_t/\psi)w^t)/(\theta_t/\psi) > ((1-\sigma)\beta - \sqrt{n}\varrho\sigma)\mu_t. \tag{41}$$

Since $\pi(\mu_t) \le (1-\sigma)\mu_t$ so that, by (40), $\mu_{t+1} \le (1-\sigma\theta_t)\mu_t$ for all t, we see from $\{\mu_t\} \to \mu^\infty > 0$ that $\{\theta_t\} \to 0$. Also, due to (8) and ϕ_1 and F being continuously differentiable, $\{\nabla H_{\mu_t}(z^t)\}_{t\in T}$ converges in operator norm to $\nabla H_{\mu^\infty}(z^\infty)$, which is invertible. Since $\{H_{\mu_t}(z^t)\}_{t\in T}$ converges, this and (38) imply $\{w^t\}_{t\in T}$ is bounded. These two observations, together with $\{z^t\}_{t\in T} \to z^\infty$ and $\{\mu_t\} \to \mu^\infty > 0$ (and ϕ_μ and F being continuously differentiable and satisfying (8)), implies the left-hand side of (41) tends to zero as $t \in T, t \to \infty$. On the other hand, by our choice of σ and $\mu_t \ge \mu^\infty > 0$, the right-hand side of (41) is bounded away from zero for all t, a contradiction.

(b) By (3), (8) and our assumptions on ϕ_1 and F, we have

$$\begin{aligned} r(z,\mu,a) &= \|(-\mu(\phi_1(r+s)-\phi_1(r)-\nabla\phi_1(r)s),\, F(x+u)-F(x)-\nabla F(x)u)\| \\ &\le \mu\|\phi_1(r+s)-\phi_1(r)-\nabla\phi_1(r)s\| + \|F(x+u)-F(x)-\nabla F(x)u\| \\ &\le \mu\lambda_1\|s\|^{\gamma_1} + \lambda_2\|u\|^{\gamma_2} \\ &\le \lambda_1\|a\|^{\gamma_1}/\mu^{\gamma_1-1} + \lambda_2\|a\|^{\gamma_2}, \end{aligned} \tag{42}$$

where for simplicity we write $z := (x,y), a := (u,v), r := z/\mu, s := a/\mu$, so that $\|s\| = \|a\|/\mu$. Also, by (38) and $(z^t,\mu_t) \in \mathcal{N}_\beta$, we have

$$\|w^t\| = \|\nabla H_{\mu_t}(z^t)^{-1}H_{\mu_t}(z^t)\| \le \||\nabla H_{\mu_t}(z^t)^{-1}\|\|\|H_{\mu_t}(z^t)\| \le \kappa\beta\mu_t.$$

This together with (41) and (42) yields

$$\begin{aligned} 0 < (1-\sigma)\beta - \sqrt{n}\varrho\sigma &< r(z^t,\mu_t,(\theta_t/\psi)w^t)/(\mu_t\theta_t/\psi) \\ &\le \lambda_1(\theta_t/\psi)^{\gamma_1-1}\|w^t\|^{\gamma_1}/(\mu_t)^{\gamma_1} + \lambda_2(\theta_t/\psi)^{\gamma_2-1}\|w^t\|^{\gamma_2}/\mu_t \\ &\le \lambda_1(\theta_t/\psi)^{\gamma_1-1}(\kappa\beta)^{\gamma_1} + \lambda_2(\theta_t/\psi)^{\gamma_2-1}(\kappa\beta)^{\gamma_2}(\mu_t)^{\gamma_2-1}, \end{aligned}$$

from which we obtain that θ_t is bounded below by a positive constant depending on $\beta, \psi, \sigma, \gamma_1, \gamma_2, \sqrt{n}\varrho, \lambda_1(\kappa\beta)^{\gamma_1}, \lambda_2(\kappa\beta)^{\gamma_2}(\mu_0)^{\gamma_2-1}$ only. Since $\mu_{t+1} \le (1-\sigma\theta_t)\mu_t$ for all t, the global linear convergence of $\{\mu_t\}$ follows. □

If in addition $\|\hat{z}^t - z^t\|$ is of the order of μ_t (such as when $\hat{z}^t$ is given by (43) and the assumptions of Prop. 1(b) hold), then the global linear convergence of μ_t in Prop. 1(b) yields, as a byproduct, that $\{z^t\}$ converges linearly in the root sense. Also, note that the convergence ratio ω depends on $\lambda_1, \lambda_2, \kappa, \mu_0$ through their respective products only and, in the case where F is affine (so $\lambda_2 = 0$), ω does not depend on μ_0.

By Cors. 1 and 2, ϕ_μ given by (4) with $g \in \mathcal{CM}$, together with ϱ given by Lemma 1, satisfies the assumption of Prop. 1. Combining this observation with Lemmas 5, 6, 8 and

Prop. 1 yield the following global (linear) convergence result for the case of ϕ_μ given by (4). Notice that $\nabla\phi_1$ being Lipschitz continuous implies ϕ_1 satisfies the assumption of Prop. 1(b) with $\gamma_1 = 2$.

Corollary 3. *Let ϕ_μ be given by* (4) *with $g \in \mathcal{CM}$. Assume F is monotone and satisfies either A1 or A2 or A3. In the case where A3 holds but not A1 or A2, assume $\beta\mu_0 < \min[\bar{x}]$, $2\beta\mu_0 < \min[F(\bar{x})]$ for some $\bar{x} \in \mathcal{S}$. Then $\{(z^t, \mu_t)\}_{t=0,1...}$ generated by Algorithm 1 is well defined and satisfies $(z^t, \mu_t) \in \mathcal{N}_\beta$ for all t. Moreover, the following hold:*
(a) *$\{z^t\}$ is bounded, $\{\mu_t\} \downarrow 0$, and every cluster point of $\{z^t\}$ is in $\mathcal{Z}$.*
(b) *If F is strongly monotone and ∇F is Lipschitz continuous and g is given by* (5), *then there exists $\omega \in (0, 1)$ such that $\mu_{t+1} \le \omega\mu_t$ for all t. If F is affine, then ω is independent of (z^0, μ_0).*

A similar reasoning using Lemma 7 shows that Cor. 3 still holds when ϕ_μ is instead given by (7), provided A1 is excluded from among the possible assumptions on F (since B1 in Lemma 8 excludes ϕ_μ given by (7)). Whether Cor. 3(a) holds under the assumption that F satisfies A1 is open when ϕ_μ is given by (7). For ϕ_μ given by (4)–(5) or (7), $\nabla\phi_1$ is Lipschitz continuous with constant $\lambda_1 = O(n)$, as is noted at the end of the Appendix. Thus, if F is monotone and affine and we set $\sigma := \frac{1}{2}\beta/(\beta + \sqrt{n}\varrho)$ in Algorithm 1, then the proof of Prop. 1(b) yields that $\theta_t > 1/(2\lambda_1\kappa) = 1/O(n\kappa)$ for all t, and hence

$$\mu_{t+1} \le (1 - \sigma\theta_t)\mu_t < \left(1 - \frac{1}{O(n^{1.5}\kappa)}\right)\mu_t$$

for all t, where $\kappa := \sup_{t\ge 0} |||\nabla H_{\mu_t}(z^t)^{-1}|||$. This result does not depend on $\{z^t\}$ being bounded. However, estimating κ can be difficult. If F is strongly monotone, the proofs of Lemmas 6 and 7 give an estimate of κ that depends only on $\sup_{t\ge 0} |||\nabla F(x^t)|||$ and the modulus ρ of strong monotonicity. In the case of LCP and ϕ_μ given by (4)–(5), Burke and Xu [7] derived an alternative estimate of κ by taking the minimum of moduli of all principal pivotal transforms of ∇F. It is not known whether their analysis can be extended to SDLCP.

8. Local superlinear convergence analysis

In this section we consider a special choice of $\hat{z}^t$ in Algorithm 1 given by

$$\nabla H_{\mu_t}(z^t)(\hat{z}^t - z^t) = -H_0(z^t), \tag{43}$$

which was considered in [9, 10, 14, 17] and is the analog of the predictor direction used in interior-point methods. An important feature of $\hat{z}^t$ given by (43) is that it can be computed relatively inexpensively since the left-hand linear mapping is the same as in (38).

For our analysis, we will need the following assumptions of strict complementarity and nondegeneracy.

C1: $\bar{x} + \bar{y} \succ 0$.

C2: The equations $\bar{x}v + u\bar{y} = 0$, $\nabla F(\bar{x})u = v$ have $(u, v) = 0$ as the only solution.

These assumptions were introduced in the cases of SDLP and monotone SDLCP by Kojima et al. [32] for the local superlinear convergence analysis of interior-point path-following methods using non-shrinking neighborhood [27, 32, 33]. It was shown by Haeberly (see [32, p. 144]) that, in the case of SDLP, C1 and C2 together are equivalent to the primal and dual nondegeneracy assumptions given by Alizadeh et al. [2].

The proof of our result is based in part on the connection between $\phi_0(x, y)$ and $\phi_\mu(x, y)$ (Cor. 1) and the observation that, under C1, $\bar{x} - \bar{y}$ is nonsingular, so that $\nabla\phi_\mu(x, y)$ is Lipschitz continuous in (x, y) for (x, y) near $(\bar{x}, \bar{y})$. In what follows, we denote for any $a \in \mathcal{S}$, $|a| = (a^2)^{1/2}$. It can be seen by diagonalizing a that $|a| - a = 2[-a]_+$ and $|a| + a = 2[a]_+$.

Lemma 9. *Let ϕ_μ be given by* (4)–(5) *or* (7) *and fix any $\bar{z} = (\bar{x}, \bar{y}) \in \mathcal{Z}$ satisfying C1 and C2. Fix any $\beta \in \Re_{++}$. Then, there exist scalars $\delta > 0$ and $\kappa_1 > 0$ such that for all $(z, \mu) \in \mathcal{N}_\beta$ satisfying $\|z - \bar{z}\| + \mu \le \delta$ and $w \in \mathcal{S} \times \mathcal{S}$ satisfying*

$$\nabla H_\mu(z)w = -H_0(z), \tag{44}$$

the following holds.

(a) $\|w\| \le \mu/\delta$.
(b) $\|\phi_0(z + w)\| \le \kappa_1\mu^2$.
(c) *If in addition $\beta > \sqrt{2n}\varrho$, where ϱ is the constant given in Lemma 1, and ∇F is Lipschitz continuous with constant $L \ge 0$ on the ball $\{x \in \mathcal{S} : \|x - \bar{x}\| \le \delta\}$, then $(z + w, \nu) \in \mathcal{N}_\beta$ whenever $\nu \ge \kappa_2\mu^2$, where $\kappa_2 := \max\{\kappa_1/(\beta/\sqrt{2} - \sqrt{n}\varrho), \sqrt{2}L/(\beta\delta^2)\}$.*

Proof. We give the proof for the case of ϕ_μ given by (4)–(5). The case of ϕ_μ given by (7) can be treated similarly using the fact $(\bar{x}^2 + \bar{y}^2)^{1/2} = \bar{x} + \bar{y} \succ 0$ and Lemma 2(c).

(a) We argue this by contradiction. Suppose, for each integer $k > 0$, there exists $(z^k, \mu_k) = (x^k, y^k, \mu_k) \in \mathcal{N}_\beta$ such that $\|z^k - \bar{z}\| + \mu_k \le 1/k$ but $\|w^k\| > k\mu_k$, where $w^k = (u^k, v^k)$ satisfies $\nabla H_{\mu_k}(z^k)w^k = -H_0(z^k)$. Then $z^k = (x^k, y^k) \to \bar{z} = (\bar{x}, \bar{y})$, $\mu_k \to 0$ and $\mu_k/\|w^k\| \to 0$ and, using (3) and Lemmas 1, 2(b), we have

$$u^k + v^k - L_{c^k}^{-1}[(x^k - y^k)(u^k - v^k) + (u^k - v^k)(x^k - y^k)] = -2\phi_0(z^k),$$
$$\nabla F(x^k)u^k - v^k = y^k - F(x^k),$$

for all k, where we denote $c^k := ((x^k - y^k)^2 + 4(\mu_k)^2 I)^{1/2}$. Applying L_{c^k} to both sides of the first equation yields

$$\begin{aligned} c^k(u^k + v^k) + (u^k + v^k)c^k - ((x^k - y^k)(u^k - v^k) + (u^k - v^k)(x^k - y^k)) \\ = -2(c^k\phi_0(z^k) + \phi_0(z^k)c^k). \end{aligned}$$

Since $\|\phi_0(z^k)\| \le \|\phi_{\mu_k}(z^k)\| + \sqrt{n}\varrho\mu_k \le \beta\mu_k + \sqrt{n}\varrho\mu_k$ (see Cor. 1) and $\|y^k - F(x^k)\| \le \beta\mu_k$, we have upon dividing all sides by $\|w^k\|$ and using $\mu_k/\|w^k\| \to 0$ and $c^k \to ((\bar{x} - \bar{y})^2)^{1/2} = |\bar{x} - \bar{y}|$ that any cluster point $\bar{w} = (\bar{u}, \bar{v})$ of $\{w^k/\|w^k\|\}$ satisfies $\bar{w} \ne 0$ and

$$|\bar{x}-\bar{y}|(\bar{u}+\bar{v})+(\bar{u}+\bar{v})|\bar{x}-\bar{y}|-((\bar{x}-\bar{y})(\bar{u}-\bar{v})+(\bar{u}-\bar{v})(\bar{x}-\bar{y})) = 0, \quad \nabla F(\bar{x})\bar{u}-\bar{v} = 0.$$

Using the properties of $|\cdot|$ and $(\bar{x}, \bar{y}) \in \mathcal{Z}$, we have $|\bar{x}-\bar{y}|-(\bar{x}-\bar{y}) = 2[\bar{y}-\bar{x}]_+ = 2\bar{y}$ and $|\bar{x}-\bar{y}|+(\bar{x}-\bar{y}) = 2[\bar{x}-\bar{y}]_+ = 2\bar{x}$. Thus, the above equations yield

$$\bar{y}\bar{u} + \bar{u}\bar{y} + \bar{x}\bar{v} + \bar{v}\bar{x} = 0, \quad \nabla F(\bar{x})\bar{u} - \bar{v} = 0,$$

By C1, C2 and [33, Lemma 6.2], this implies $\bar{w} = 0$, contradicting $\bar{w} \neq 0$.

(b) Since $\bar{x} \succeq 0$, $\bar{y} \succeq 0$, $\langle \bar{x}, \bar{y}\rangle = 0$ and $\bar{x}+\bar{y} \succ 0$, it is well known and not difficult to show that, for some $p \in \mathcal{O}$ and $I \subseteq \{1, ..., n\}$ and $J = \{1, ..., n\}\backslash I$,

$$\bar{x} = p^T \begin{bmatrix} \tilde{x}_{II} & 0 \\ 0 & 0 \end{bmatrix} p, \quad \bar{y} = p^T \begin{bmatrix} 0 & 0 \\ 0 & \tilde{y}_{JJ} \end{bmatrix} p,$$

for some $\tilde{x}_{II} \succ 0$ and $\tilde{y}_{JJ} \succ 0$. Thus, $\bar{x}-\bar{y}$ is nonsingular and hence $|\bar{x}-\bar{y}| \succ 0$. Since, by (a), $w = (u, v)$ given by (44) satisfies $\|w\| \le \mu/\delta$ whenever $(z, \mu) \in \mathcal{N}_\beta$ and $\|z-\bar{z}\| + \mu \le \delta$, by taking δ sufficiently small, we can further assume that

$$\begin{aligned} &\|x-y\| \le \kappa, \ |||L^{-1}_{((x-y)^2+4\mu^2 I)^{1/2}}||| \le \kappa, \ |x-y+\tau(u-v)| \succ 0, \\ &|||L^{-1}_{|x-y+\tau(u-v)|}||| \le \kappa \ \forall \tau \in [0, 1], \end{aligned} \tag{45}$$

for some constant $\kappa > 0$ (depending on $|\bar{x}-\bar{y}|$ only).

Fix any $(z, \mu) \in \mathcal{N}_\beta$ satisfying $\|z-\bar{z}\| + \mu \le \delta$. By (a), $\|w\| \le \mu/\delta$ and by (45), $\|x-y\| \le \kappa$, $|x-y| \succ 0$, and $|||L^{-1}_{|x-y|}||| \le \kappa$. We have upon using the fact (see the proof of Lemma 2(b))

$$d - c + L_c^{-1}[(d-c)^2] = L_c^{-1}[(x-y)(u-v) + (u-v)(x-y) + (u-v)^2]$$

with $c := |x-y| \succ 0$ and $d := |x-y+u-v|$ that, analogous to (14), ϕ_0 is Fréchet-differentiable at z with

$$2\nabla\phi_0(x, y)(u, v) = u + v - L^{-1}_{|x-y|}[(x-y)(u-v) + (u-v)(x-y)].$$

In fact, $L^{-1}_{|x-y|}$ is continuous in (x, y), so $\nabla\phi_0(z)$ is continuous over the set of (z, μ) satisfying $\|z-\bar{z}\| + \mu \le \delta$.

Since $x-y+\tau(u-v)$ is nonsingular for all $\tau \in [0, 1]$ (see (45)), then ϕ_0 is continuously differentiable along this line segment, so the mean value theorem and (44) yield

$$\begin{aligned} \|\phi_0(z+w)\| &= \left\| \phi_0(z) + \int_0^1 \nabla\phi_0(z+\tau w)w \, d\tau \right\| \\ &= \left\| -\nabla\phi_\mu(z)w + \int_0^1 \nabla\phi_0(z+\tau w)w \, d\tau \right\| \\ &= \left\| \int_0^1 (\nabla\phi_0(z+\tau w)w - \nabla\phi_\mu(z)w) \, d\tau \right\| \\ &\le \int_0^1 \left\| \nabla\phi_0(z+\tau w)w - \nabla\phi_\mu(z)w \right\| \, d\tau. \end{aligned} \tag{46}$$

Letting $c := ((x-y)^2 + 4\mu^2 I)^{1/2}$, $d := |x - y + \tau(u - v)|$, and $e := (x-y)(u-v) + (u-v)(x-y)$, we have from (14) and the above formula for $\nabla\phi_0$ that, for any $\tau \in [0, 1]$,

$$\begin{aligned}
&2\nabla\phi_0(z+\tau w)w - 2\nabla\phi_\mu(z)w \\
&\quad = L_c^{-1}[(x-y)(u-v) + (u-v)(x-y)] \\
&\qquad - L_{|x-y+\tau(u-v)|}^{-1}[(x-y+\tau(u-v))(u-v) + (u-v)(x-y+\tau(u-v))] \\
&\quad = L_c^{-1}[e] - L_d^{-1}[e] - 2\tau L_d^{-1}[(u-v)^2]. \qquad (47)
\end{aligned}$$

Using

$$\begin{aligned}
d - c + L_c^{-1}[(d-c)^2] &= L_c^{-1}[d^2 - c^2] \\
&= L_c^{-1}[\tau(x-y)(u-v) + \tau(u-v)(x-y) + \tau^2(u-v)^2 + 4\mu^2 I]
\end{aligned}$$

and (45), we have

$$\begin{aligned}
\|d-c\| &\le \kappa\|\tau(x-y)(u-v) + \tau(u-v)(x-y) + \tau^2(u-v)^2 + 4\mu^2 I - (d-c)^2\| \\
&\le \kappa(2\|x-y\|\|u-v\| + \|u-v\|^2 + 4\sqrt{n}\mu^2 + \|d-c\|^2) \\
&\le \kappa(2\kappa\|u-v\| + \|u-v\|^2 + 4\sqrt{n}\mu^2) + \kappa\|d-c\|^2,
\end{aligned}$$

where the second inequality uses the fact that $\|ab\| \le \|a\|\|b\|$ for all $a, b \in \mathcal{S}$. Since $|\cdot|$ and $(\cdot)^{1/2}$ are continuous functions and $\|(u, v)\| \le \mu/\delta$ and $t \in [0, 1]$, by taking $\delta_1 \in (0, \delta]$ sufficiently small we can assume that $\kappa\|d - c\| \le 1/2$ and $\|u - v\| \le \kappa$ whenever $\|z - \bar{z}\| + \mu \le \delta_1$. Then, the above inequality and $\|w\| \le \mu/\delta$ yield

$$\begin{aligned}
\|d-c\| \le 2\kappa(2\kappa\|u-v\| + \|u-v\|^2 + 4\sqrt{n}\mu^2) &\le 2\kappa(3\kappa\|u-v\| \\
+4\sqrt{n}\mu^2) &\le 12\kappa^2\mu/\delta + 8\kappa\sqrt{n}\mu^2.
\end{aligned}$$

Thus, $\|d - c\| \le \kappa_0\mu$ for a suitable constant $\kappa_0 > 0$.

Letting $\Delta := L_c^{-1}[e] - L_d^{-1}[e]$, with c, d and e defined as above, we have

$$L_d[\Delta] = dL_c^{-1}[e] + L_c^{-1}[e]d - e = (d-c)L_c^{-1}[e] + L_c^{-1}[e](d-c),$$

where the second equality uses the fact $cL_c^{-1}[e] + L_c^{-1}[e]c = e$. Since $\|d - c\| \le \kappa_0\mu$, this together with (45) yields

$$\begin{aligned}
\|\Delta\| &= \|L_d^{-1}[L_d[\Delta]]\| \\
&\le \kappa\|L_d[\Delta]\| \\
&= \kappa\|(d-c)L_c^{-1}[e] + L_c^{-1}[e](d-c)\| \\
&\le 2\kappa\|d-c\|\|L_c^{-1}[e]\| \\
&\le 2\kappa_0\mu\kappa^2\|e\| \\
&\le 2\kappa_0\mu\kappa^2(2\kappa\|u-v\|) \\
&\le 8\kappa_0\mu^2\kappa^3/\delta.
\end{aligned}$$

Combining this with (47), we have

$$\begin{aligned}2\|\nabla\phi_0(z+\tau w)w - \nabla\phi_\mu(z)w\| &= \|\Delta - 2\tau L_d^{-1}[(u-v)^2]\|\\ &\le 8\kappa_0\mu^2\kappa^3/\delta + 2\tau\kappa\|u-v\|^2\\ &\le 8\kappa_0\mu^2\kappa^3/\delta + 2\tau\kappa(2\mu/\delta)^2.\end{aligned}$$

This holds for each $\tau \in [0, 1]$, which together with (46) and (45) yields

$$\|\phi_0(z+w)\| \le \int_0^1 4\kappa_0\mu^2\kappa^3/\delta + 4\tau\kappa\mu^2/\delta^2\, dt = \kappa_1\mu^2,$$

where we let $\kappa_1 := 4\kappa_0\kappa^3/\delta + 2\kappa/\delta^2$.

(c) By Cor. 1 and part (b), for any $\nu \ge \kappa_1\mu^2/(\beta/\sqrt{2} - \sqrt{n}\varrho)$, we have

$$\begin{aligned}\|\phi_\nu(z+w)\| &\le \|\phi_0(z+w)\| + \sqrt{n}\varrho\nu \le \kappa_1\mu^2 + \sqrt{n}\varrho\nu\\ &\le (\beta/\sqrt{2} - \sqrt{n}\varrho)\nu + \sqrt{n}\varrho\nu = \beta\nu/\sqrt{2}.\end{aligned}$$

Also, we have from (44) that $F(x) + \nabla F(x)u = y + v$, which together with part (a) and $\nu \ge \sqrt{2}L\mu^2/(\beta\delta^2)$ yields

$$\|F(x+u)-(y+v)\| = \|F(x+u)-F(x)-\nabla F(x)u\| \le L\|u\|^2 \le L\mu^2/\delta^2 \le \beta\nu/\sqrt{2}.$$

Thus, by (3),

$$\|H_\nu(z+w)\|^2 = \|\phi_\nu(z+w)\|^2 + \|F(x+u)-(y+v)\|^2 \le (\beta\nu)^2/2 + (\beta\nu)^2/2 = (\beta\nu)^2,$$

so $(z+w, \nu) \in \mathcal{N}_\beta$. □

In Lemma 9, if we assume F to be only continuously differentiable near $\bar{x}$, then we obtain $\nu \ge o(\mu)$ instead, with $\lim_{\mu\downarrow 0} o(\mu)/\mu = 0$. Notice that if F is affine, then $L = 0$. Also, it can be shown that part (a) of Lemma 9 holds for ϕ_μ given by (4) for any $g \in \mathcal{CM}$ (the proof uses Lemmas 2(a) and 3), but it is not known whether part (b) holds similarly. By using Lemma 9, we obtain the following local superlinear convergence result for Algorithm 1 with $\hat{z}^t$ chosen according to (43).

Proposition 2. *Let ϕ_μ be given by* (4)–(5) *or* (7)*, and fix any $\bar{z} = (\bar{x}, \bar{y}) \in \mathcal{Z}$ satisfying C1 and C2 and such that ∇F is Lipschitz continuous with constant $L \ge 0$ on some ball of radius $\delta_0 > 0$ around $\bar{x}$. Fix any $\beta > \sqrt{2n}\varrho$, where ϱ is given in Lemma 1. Assume $\nabla H_\mu(z)$ is invertible for all $(z, \mu) \in \mathcal{N}_\beta$ with $\mu \le \mu_0 \in \Re_{++}$. Let $\{(z^t, \mu_t)\}_{t=0,1\ldots}$ be generated by Algorithm 1 with $\hat{z}^t$ chosen by* (43) *and π chosen such that $\pi(\mu) \ge \kappa_2\mu^2$ for all $\mu > 0$ sufficiently small, where κ_2 is the constant given in Lemma 9. Then there exists $\delta \in \Re_{++}$ such that if $\|z^{\bar{t}} - \bar{z}\| + \mu^{\bar{t}} \le \delta$ for some $\bar{t}$, then*

$$\|z^t - \bar{z}\| + \mu_t \le \delta \quad \textit{and} \quad \mu_{t+1} \le \pi(\mu_t) \le (1-\sigma)\mu_t \quad \forall t = \bar{t}, \bar{t}+1, \ldots$$

Proof. By Prop. 1, $\{(z^t, \mu_t)\}_{t=0,1\ldots}$ is well defined and satisfies $(z^t, \mu_t) \in \mathcal{N}_\beta$ for all t. Then, by Lemma 9, there exist $\delta_2 \in (0, \delta_0]$ and κ_1 which, together with κ_2, have the property that

$$\|\hat{z}^t - z^t\| \le \mu_t/\delta_2,\ (\hat{z}^t, \nu) \in \mathcal{N}_\beta \text{ whenever } \|z^t - \bar{z}\| + \mu_t \le \delta_2 \text{ and } \nu \ge \kappa_2(\mu_t)^2. \tag{48}$$

By our choice of π, there exists $\delta \in (0, \delta_2]$ such that $\pi(\mu) \ge \kappa_2\mu^2$ whenever $\mu \le \delta$. Then (48) implies

$$(\hat{z}^t, \pi(\mu_t)) \in \mathcal{N}_\beta \text{ whenever } \|z^t - \bar{z}\| + \mu_t \le \delta,$$

in which case our choice of α_t in Algorithm 1 yields $\alpha_t = 1, z^{t+1} = \hat{z}^t$ and $\mu_{t+1} \le \pi(\mu_t)$. This together with (48) yields

$$\|z^{t+1} - z^t\| \le \mu_t/\delta_2,\ \mu_{t+1} \le \pi(\mu_t) \le (1-\sigma)\mu_t \text{ whenever } \|z^t - \bar{z}\| + \mu_t \le \delta\ . \tag{49}$$

Then if there exists an index $\bar{t}$ such that $\|z^{\bar{t}} - \bar{z}\| \le \delta/3$ and $\mu^{\bar{t}} \le \delta/3$ and $\mu^{\bar{t}}/(\sigma\delta_2) \le \delta/3$, a simple induction argument using (49) yields

$$\|z^{t+1} - z^t\| \le (1-\sigma)^{t-\bar{t}}\frac{\mu^{\bar{t}}}{\delta_2}, \quad \|z^t - \bar{z}\| \le \delta/3 + \frac{1-(1-\sigma)^{t-\bar{t}}}{\sigma}\frac{\mu^{\bar{t}}}{\delta_2} \le 2\delta/3, \quad \mu_t \le \delta/3,$$

for all $t = \bar{t}, \bar{t}+1, \ldots$ Thus $\|z^t - \bar{z}\| + \mu_t \le \delta$ for all $t \ge \bar{t}$, which together with (49) yields $\mu_{t+1} \le \pi(\mu_t) \le (1-\sigma)\mu_t$ for all $t \ge \bar{t}$. □

By Lemmas 6 and 7, F being a monotone mapping is sufficient for $\nabla H_\mu(z)$ to be invertible. Prop. 2 says that the rate of local convergence depends on the rate at which $\pi(\mu) \to 0$ as $\mu \downarrow 0$. If we choose, say,

$$\pi(\mu) = \mu^2|\ln(\mu)|,$$

which satisfies the assumption of Prop. 2, then Algorithm 1 would achieve a convergence rate very close to quadratic.

9. SDLP and generalized SDCP

As was remarked in Sec. 1, (1) includes SDLP as a special case. However, casting an SDLP in the form (1) requires introducing auxiliary variables which is impractical for computation. In this section we consider a generalization of (1) for which the analyses and algorithm of previous sections can be readily extended and into which an SDLP can be cast without increasing the problem dimension.

The generalized SDCP is to find, for given mappings $F : \mathcal{S} \mapsto \mathcal{S}$ and $G : \mathcal{S} \mapsto \mathcal{S}$, an $(x, y, \zeta) \in \mathcal{S} \times \mathcal{S} \times \mathcal{S}$ satisfying

$$x \in \mathcal{S}_+, \quad y \in \mathcal{S}_+, \quad \langle x, y\rangle = 0, \quad F(\zeta) - y = 0, \quad G(\zeta) - x = 0. \tag{50}$$

We assume that F and G are continuously differentiable and satisfy

$$\nabla F(\zeta)\Delta\zeta = \nabla G(\zeta)\Delta\zeta = 0 \Rightarrow \Delta\zeta = 0 \quad \text{and} \quad \lim_{\|\zeta\|\to\infty} \|(F(\zeta), G(\zeta))\| = \infty. \tag{51}$$

Clearly (1) is a special case of this problem in which G is the identity mapping. We say that F and G are *relatively monotone* if

$$\langle F(\zeta) - F(\zeta'), G(\zeta) - G(\zeta') \rangle \geq 0 \quad \forall \zeta, \zeta' \in \mathcal{S}.$$

Relative strong monotonicity is similarly defined; also see [49].

Consider an SDLP in the standard form

$$\min \ \langle c, x \rangle \ \text{ subject to } \ x \in \mathcal{S}_+, \ [\langle a^i, x \rangle]_{i=1}^{\ell} = b, \tag{52}$$

where $l \geq 1$ and $c, a^1, ..., a^\ell \in \mathcal{S}$ and $b = [b_i]_{i=1}^\ell \in \Re^\ell$ are given. We assume that $a^1, ..., a^\ell$ are linearly independent and there exists $d \in \mathcal{S}$ satisfying $[\langle a^i, d \rangle]_{i=1}^\ell = b$. The optimality condition for this problem is

$$x \in \mathcal{S}_+, \ y \in \mathcal{S}_+, \ \langle x, y \rangle = 0, \ [\langle a^i, x \rangle]_{i=1}^\ell = b, \ y + \sum_{i=1}^{\ell} a^i \zeta_i = c \text{ for some } \zeta_i \in \Re. \tag{53}$$

Let $a^{\ell+1}, ..., a^\nu \in \mathcal{S}$ be a basis for the subspace of $\mathcal{S}$ orthogonal to $a^1, ..., a^\ell$, where $\nu := \frac{1}{2} \sum_{k=1}^m n_k(n_k + 1)$ is the dimension of $\mathcal{S}$. Then, by letting $\zeta := (\cdots, \zeta_i, \cdots)_{1 \leq i \leq \nu}^T \in \Re^\nu$ and identifying $\Re^\nu$ with $\mathcal{S}$, we see that (53) is a special case of (50) with

$$F(\zeta) := c - \sum_{i=1}^{\ell} a^i \zeta_i, \qquad G(\zeta) := d - \sum_{i=\ell+1}^{\nu} a^i \zeta_i. \tag{54}$$

Moreover, F and G are relatively monotone and satisfy (51).[3]

Analogous to (3), consider

$$H_\mu(x, y, \zeta) := (\phi_\mu(x, y), F(\zeta) - y, G(\zeta) - x).$$

Then Lemmas 6 and 7 can be readily extended to this general problem whereby (strong) monotonicity of F is replaced by relative (strong) monotonicity of F and G. Lemmas 8, 9, Algorithm 1, Props. 1, 2, Cor. 3 can be similarly extended. We note that an alternative geometric formulation of SDCP has been proposed by Shida and Shindoh [41], into which an SDLP can be cast without increasing problem dimension. However, this formulation does not appear to lend itself easily to the adaptation of Algorithm 1 and the associated global convergence and local superlinear convergence analysis.

In the SDLP case where F and G are given by (54), we can work directly with $[\langle a^i, x \rangle]_{i=1}^\ell = b$ instead of $G(\zeta) - x = 0$. This avoids computing G explicitly. In particular, for a given (x, y) and $[\zeta_i]_{i=1}^\ell$, the Newton direction (u, v) and $[\Delta\zeta_i]_{i=1}^\ell$ is the unique solution of the linear equations

$$\nabla\phi_\mu(x, y)(u, v) = -\phi_\mu(x, y), \quad [\langle a^i, u \rangle]_{i=1}^\ell = r, \quad v + \sum_{i=1}^{\ell} a^i \Delta\zeta_i = s, \tag{55}$$

[3] We can also write (53) as the horizontal SDLCP: $x \in \mathcal{S}_+, \ y \in \mathcal{S}_+, \ \langle x, y \rangle = 0, \ Mx + Ny = Nc + Md$, where M and N denote the orthogonal projections onto the subspaces spanned by, respectively, $a^1, ..., a^\ell$ and $a^{\ell+1}, ..., a^\nu$.

where $r := b - [\langle a^i, x\rangle]_{i=1}^{\ell}$ and $s := c - y - \sum_{i=1}^{\ell} a^i \zeta_i$. For the "pure" Newton direction, which is analogous to the predictor direction in interior-point methods, the linear equations are identical except the right-hand term $\phi_\mu(x, y)$ is replaced with $\phi_0(x, y)$. Analogously, we work with $\|H_\mu(x, y, \zeta)\| = \sqrt{\|\phi_\mu(x, y)\|^2 + \|s\|^2 + \|r\|^2}$ in the definition of the neighborhood $\mathcal{N}_\beta$ (9).

For ϕ_μ given by (4)–(5), by using Lemma 2(b), (55) reduces to

$$\sum_{j=1}^{\ell}\langle a^i, L_{c-w}^{-1}\left[L_{c+w}[a^j]\right]\rangle\Delta\zeta_j = \langle a^i, L_{c-w}^{-1}\left[2L_c[\phi_\mu(x, y)] + L_{c+w}[s]\right]\rangle + r_i, \quad i = 1, ..., \ell, \tag{56}$$

where $w := x - y$ and $c := (w^2 + 4\mu^2 I)^{1/2}$. To compute the coefficients in (56), we first find an eigenvalue factorization (i.e., spectral decomposition) of w, i.e., a $p \in \mathcal{O}$ such that $\tilde{w} := pwp^T$ is diagonal. Then, $\tilde{c} := pcp^T = (\tilde{w}^2 + 4\mu^2 I)^{1/2}$ is also diagonal and the coefficients in (56) can be written as

$$\begin{aligned} B_{ij} &:= \langle \tilde{a}^i, L_{\tilde{c}-\tilde{w}}^{-1}\left[L_{\tilde{c}+\tilde{w}}[\tilde{a}^j]\right]\rangle, \quad i, j = 1, ..., \ell, \\ h_i &:= \langle \tilde{a}^i, L_{\tilde{c}-\tilde{w}}^{-1}\left[2L_{\tilde{c}}[p\phi_\mu(x, y)p^T] + L_{\tilde{c}+\tilde{w}}[psp^T]\right]\rangle + r_i, \quad i = 1, ..., \ell, \end{aligned}$$

where $\tilde{a}^i := pa^i p^T$. It was shown by Monteiro and Zanjácomo [36, Appendix 10] that $5\ell n^3/3$ flops suffice to compute $\tilde{a}^i$ for $i = 1, ..., \ell$, assuming that a reverse Cholesky factorization of $a^i + \upsilon_i I$, with $\upsilon_i \in \Re_+$ chosen so that $a^i + \upsilon_i I \succ 0$, is available. Since $\tilde{c}$ and $\tilde{w}$ are diagonal, computing $L_{\tilde{c}-\tilde{w}}^{-1}\left[L_{\tilde{c}+\tilde{w}}[\tilde{a}^j]\right]$ for $j = 1, ..., \ell$, requires only $O(\ell n^2)$ flops. Since computing $\langle a, b\rangle$ requires $n^2 + n$ flops for any $a, b \in \mathcal{S}$ (using the symmetry of a, b) and $B_{ij} = B_{ji}$ for all i, j, then $\ell(\ell + 1)/2 \cdot (n^2 + n)$ additional flops suffice to compute B_{ij} for $1 \le i \le j \le \ell$. A similar analysis shows that $\ell n^2 + O(n^3)$ additional flops suffice to compute h_i for $i = 1, ..., \ell$. Thus, the work in computing the coefficients is

$$1\frac{2}{3}\ell n^3 + \frac{1}{2}\ell^2 n^2 + O(\ell n^2 + \ell^2 n + n^3) \quad \text{flops.}$$

Given the coefficients, $2\ell^3/3 + O(\ell n^2)$ flops suffice to solve for $[\Delta\zeta_i]_{i=1}^{\ell}$ and (u, v). The total work is less than that for the AHO and X-MT directions, but more than that for the NT direction in interior-point methods [36, Sec. 3.7]. The work can be more or less than that for the S-Ch-MT and HRVW/KSH/M directions, depending on the values of ℓ and n. However, these estimates of work do not take into account the exploitation of sparsity in $a^1, ..., a^\ell$.

For ϕ_μ given by (7), by using Lemma 2(c), (55) reduces to

$$\sum_{j=1}^{\ell}\langle a^i, L_{c-x}^{-1}\left[L_{c-y}[a^j]\right]\rangle\Delta\zeta_j = \langle a^i, L_{c-x}^{-1}\left[L_c[\phi_\mu(x, y)] + L_{c-y}[s]\right]\rangle + r_i, \ i = 1, ..., \ell, \tag{57}$$

where $c := (x^2 + y^2 + 2\mu^2 I)^{1/2}$. Forming this equation requires two eigenvalue factorization, one of $x^2 + y^2$ to compute c and another of $c - x$ to evaluate L_{c-x}^{-1}. This contrasts with interior-point methods which require either a Cholesky factorization or a single eigenvalue factorization of y to form the Newton equation [2, 36, 45]. Using an eigenvalue factorization of $c - x$, i.e., a $p \in \mathcal{O}$ such that $\tilde{c} - \tilde{x}$ is diagonal, where $\tilde{c} := pcp^T$, $\tilde{x} := pxp^T$, $\tilde{y} := pyp^T$, the coefficients in (57) can be written as

$$B_{ij} := \langle \tilde{a}^i, L^{-1}_{\tilde{c}-\tilde{x}} \left[L_{\tilde{c}-\tilde{y}}[\tilde{a}^j] \right] \rangle, \quad i, j = 1, ..., \ell,$$
$$h_i := \langle \tilde{a}^i, L^{-1}_{\tilde{c}-\tilde{x}} \left[L_{\tilde{c}}[p\phi_\mu(x, y)p^T] + L_{\tilde{c}-\tilde{y}}[psp^T] \right] \rangle + r_i, \quad i = 1, ..., \ell,$$

where $\tilde{a}^i := pa^i p^T$. As was discussed above, $5\ell n^3/3$ flops suffice to compute $\tilde{a}^i$ for $i = 1, ..., \ell$. However, $\tilde{c} - \tilde{y}$ may not be diagonal and $[B_{ij}]_{i,j=1}^{\ell}$ may not be symmetric. As a result, the work in computing the coefficients is more than that for (56). In particular, computing $L_{\tilde{c}-\tilde{y}}[\tilde{a}^j]$, for $j = 1, ..., \ell$, requires $\ell(2n^3 + n^2)$ flops and using these to compute B_{ij}, for $i, j = 1, ..., \ell$, requires $\ell^2(n^2 + n)$ additional flops. Thus, the work in computing the coefficients is

$$3\frac{2}{3}\ell n^3 + \ell^2 n^2 + O(\ell n^2 + \ell^2 n + n^3) \quad \text{flops.}$$

Given the coefficients, $2\ell^3/3 + O(\ell n^2 + n^3)$ flops suffice to solve for $[\Delta\zeta_i]_{i=1}^{\ell}$ and (u, v). The total work is comparable to that for the AHO direction [36, Sec. 3.7].

Notice that the above two Newton directions are *Q-scale invariant* in the sense of Todd [44, Sec. 6]. In particular, if $x, y, s, a^1, ..., a^\ell$ are each multiplied, respectively, on the left and on the right by p and p^T for any $p \in \mathcal{O}$, then the Newton directions u, v would be similarly transformed. However, neither direction is *P-scale invariant.*

10. Preliminary computational experience

To gain some understanding of the numerical behavior/performance of Algorithm 1 with $\hat{z}^t$ chosen as described in Prop. 2, we implemented this method in Matlab to solve the SDLP (52), with adaptations as described in Sec. 9. In this section we describe the implementation and report our preliminary numerical experience with it. We chose SDLP for its availability of test problems and Matlab solvers.

In our Matlab implementation of Algorithm 1, we choose ϕ_μ given by (7) with $\varrho = \sqrt{2}$, and set $\psi = .9$, $\sigma = \min\{.3, \hat{\sigma}\}$ with $\hat{\sigma} := \beta/(\beta + \sqrt{2n}\varrho)$, and $\pi(\mu) = \min\{(1-\sigma)\mu, \mu^{1.5}\}$. The choice of β will be clarified below. We also experimented with ϕ_μ given by (4)–(5) and, while the results are quite similar in terms of iteration count and solution accuracy, there are some differences in implementation and CPU time which we will comment on. We choose $\hat{z}^t$ by (43) whenever $\mu_t < 0.1$ (otherwise we set $\alpha_t = 0$) and set $\nu^t = 0$ for all t (corresponding to α_t being either 0 or 1). The above parameter choices, though reasonable, were made without much fine tuning and can conceivably be improved. To further accelerate the method, we replace the term $(1 - \sigma\theta_t)\mu_t$ in (39) and (40) with the smallest $\mu \in \{(.7)^j(1 - \sigma\theta_t)\mu_t\}_{j=0,1,...}$ such that $(z^t + \theta_t w^t, \mu) \in \mathcal{N}_\beta$. This μ is easy to compute since an eigenvalue factorization of $(x^t)^2 + (y^t)^2$ is already available from computing w^t. In particular, w^t is obtained by solving a reduced system of linear equations in $\Delta\zeta_j \in \Re$, $j = 1, ..., \ell$, of the form (57), with (x, y) and μ indexed by t. The equation for $\hat{z}^t$ is identical except the right-hand term

$\phi_\mu(x, y)$ is replaced with $\phi_0(x, y)$. Finally, to improve the primal feasibility at termination, we employ a projection technique used in SDPT^3 [45]: After computing the Newton direction $w = (u, v)$, check if $\|[\langle a^i, x + u\rangle]_{i=1}^{\ell} - b\| > \|[\langle a^i, x\rangle]_{i=1}^{\ell} - b\|$ and if yes, replace u by its orthogonal projection on to the null space $\{u \in \mathcal{S} : [\langle a^i, u\rangle]_{i=1}^{\ell} = 0\}$.

To simplify the programming and testing, we implemented Algorithm 1 by borrowing the data structure, problem input, and linear algebra routines from the SDPT^3 (version 1.3) Matlab code of Toh, Todd, and Tütüncü [45]. In particular, the Newton equations (56) and (57) are formed and solved much like the AHO direction in SDPT^3. SDPT^3 implements a primal-dual Mehrotra-predictor-corrector interior-point method and is linked to a set of eight test problems. For comparison purposes, we use the same initial $x^0 \in \mathcal{S}$ and $\zeta^0 \in \Re^\ell$ as given by SDPT^3. [SDPT^3 can initialize with either a feasible or an infeasible $x^0 \in \mathcal{S}_{++}$ and ζ^0. The results below are obtained with the infeasible initialization. Qualitatively similar results are obtained with the feasible initialization.] Then we initialize the remaining parameters according to:

$$y^0 = c - \sum_{i=1}^{\ell} a^i \zeta_i^0, \quad \mu_0 = \|H_0(z^0)\|/4, \quad \beta = 1.5\|H_{\mu_0}(z^0)\|/\mu_0.$$

We note that y^0 need not be positive definite. For example, on random, ETP and LogCheby problems of Table 2, y^0 has negative eigenvalue ranging from -2 to -600. We use the same termination criterion as in SDPT^3, namely, terminate the method when "relative duality gap" and "relative primal and dual infeasibility", as defined in [45], are below a specified threshold. In our testing, we set the threshold to $3 \cdot 10^{-9}$.

Tables 1 and 2 tabulate the single-run iteration count (niter) for SDPT^3 (using three different choices of search directions: AHO, HKM and NT) and Algorithm 1 on, respectively, small and medium-sized problems from the SDPT^3 test set. Since Algorithm 1 does not maintain x and y to be positive semidefinite, we also report the minimum eigenvalue of, respectively, x and y on termination (minx and miny). [On some problems, both x and y had non-negligible negative eigenvalues at the early stages of the method.] As can be seen from these tables, Algorithm 1 has comparable average iteration counts as the interior-point methods on the small problems, but has higher average iteration counts

Table 1. Iteration counts for small SDLP problems.

		AHO	**HKM**	**NT**	**Alg. 1**
Problem	n, l	**niter**	**niter**	**niter**	**niter/minx/miny**
random	10,10	9	13	13	$17/{-6.5} \cdot 10^{-11}/{-1.5} \cdot 10^{-12}$
Norm min	20,6	12	14	15	$12/{-1.6} \cdot 10^{-11}/{-2.7} \cdot 10^{-14}$
Cheby	20,11	12	14	13	$11/{-1.2} \cdot 10^{-12}/{-8.9} \cdot 10^{-11}$
Maxcut	10,10	10	11	11	$10/{-4.7} \cdot 10^{-09}/{-1.4} \cdot 10^{-11}$
ETP	20,10	12	14	13	$17/{-3.8} \cdot 10^{-12}/{-1.0} \cdot 10^{-15}$
Lovasz	10,22	11	13	11	$12/{-1.2} \cdot 10^{-09}/{-5.6} \cdot 10^{-09}$
LogCheby	60,6	13	15	13	$17/{-4.7} \cdot 10^{-08}/{-3.0} \cdot 10^{-09}$
ChebyC	40,11	11	12	13	$11/{-6.7} \cdot 10^{-17}/{-1.1} \cdot 10^{-12}$

Table 2. Iteration counts for medium-sized SDLP problems.

		AHO	HKM	NT	Alg. 1
Problem	n, l	**niter**	**niter**	**niter**	**niter/minx/miny**
random	20,20	11	12	12	$21/-1.5 \cdot 10^{-07}/-1.3 \cdot 10^{-10}$
Norm min	40,11	15	17	16	$12/-1.6 \cdot 10^{-11}/-1.4 \cdot 10^{-13}$
Cheby	40,21	16	16	16	$13/-9.0 \cdot 10^{-12}/-2.5 \cdot 10^{-11}$
Maxcut	21,21	10	11	11	$12/-3.8 \cdot 10^{-08}/-9.9 \cdot 10^{-12}$
ETP	41,21	14	16	16	$23/-7.2 \cdot 10^{-13}/-1.0 \cdot 10^{-16}$
Lovasz	21,88	14	16	16	$20/-2.3 \cdot 10^{-10}/-1.2 \cdot 10^{-07}$
LogCheby	120,11	16	17	15	$22/-4.5 \cdot 10^{-08}/-1.3 \cdot 10^{-10}$
ChebyC	80,21	12	13	12	$12/-6.7 \cdot 10^{-17}/-4.2 \cdot 10^{-15}$

on four of the medium-sized problems. On the remaining four problems (namely, Norm min, Cheby, Maxcut, ChebyC), Algorithm 1 has comparable iteration counts. Thus, for certain classes of SDLP, a non-interior method like Algorithm 1 may provide a viable alternative to interior-point methods. In general, the use of $\hat{z}^t$ significantly improves the local convergence of Algorithm 1, enabling μ_t to decrease rapidly in the last few iterations before termination. Figures 1 and 2 plot, using the plotting feature of SDPT^3, the duality gap and infeasibility trajectories for all methods on the small test problems. The trajectories for the medium-sized problems are qualitatively similar and are omitted for brevity. Notice that Algorithm 1 decreases the infeasibility faster than interior-point methods at the early stages, but near the end the infeasibility increases. The reason for this is not well understood, as the condition number for the Newton equation (57) is not much worse than that for the interior-point methods. In any case, this shows that further improvements are needed if Algorithm 1 is to solve problems to high accuracy.

In terms of CPU times (on a Dec Alpha workstation), Algorithm 1 is typically 10%-30% slower when the FB function (7) is used than when the CHKS function (4)–(5) is used, even though the iteration counts are comparable. This seems to be mostly due to the more efficient computation of the Newton direction for the CHKS function, requiring one eigenvalue factorization rather than two. The CPU times are about 2-4 times those of interior-point methods, even though the iteration counts are comparable. The

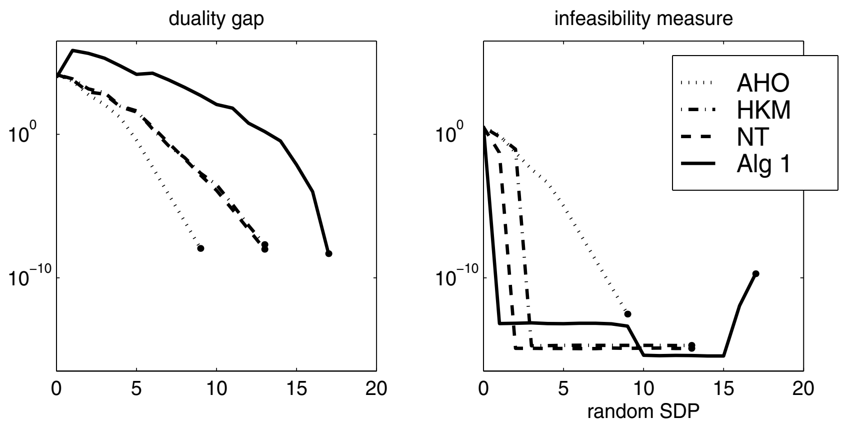

Fig. 1. Duality gap and infeasibility trajectories for problems 1–4 of Table 1.

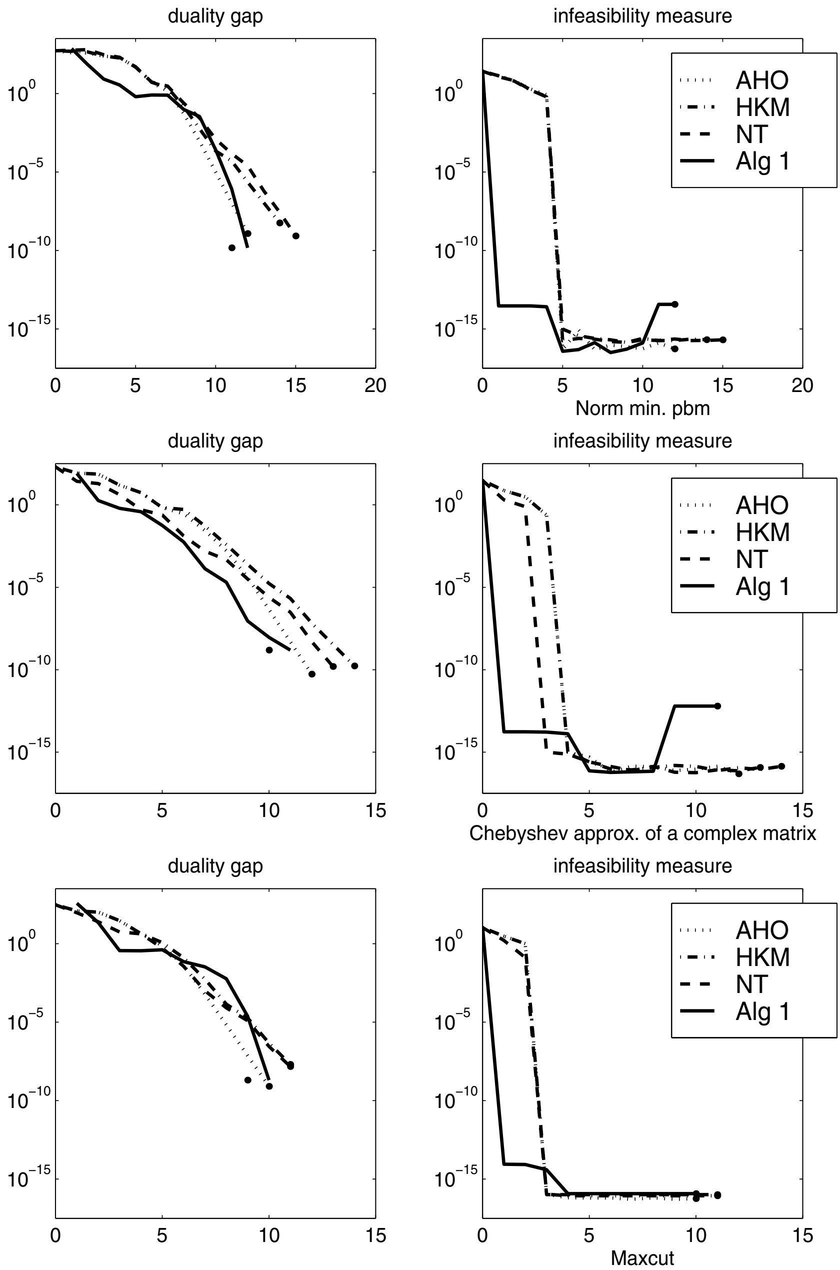

Fig. 1. (Continued)

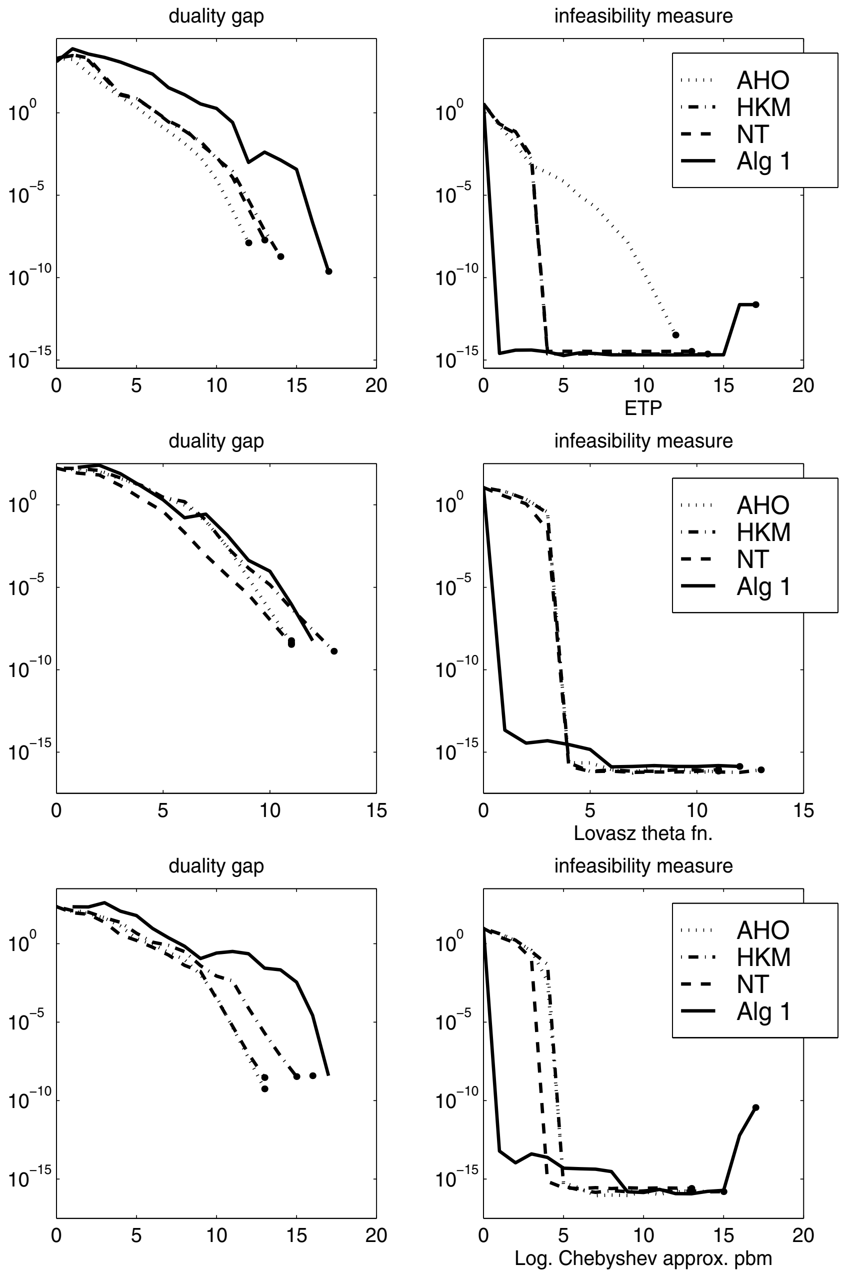

Fig. 2. Duality gap and infeasibility trajectories for problems 5–8 of Table 1.

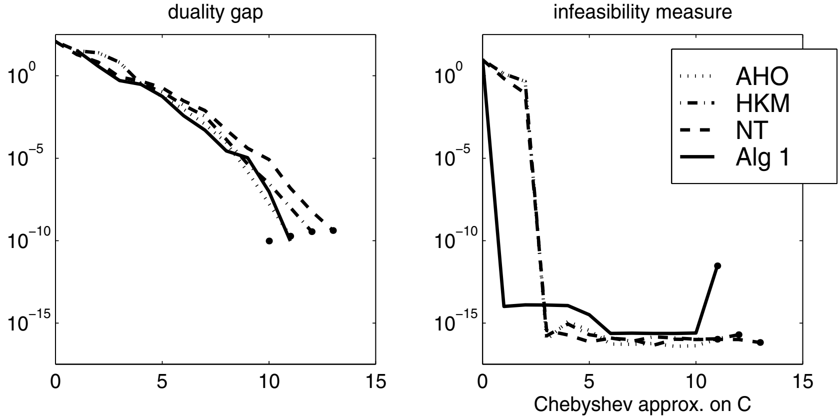

Fig. 2. (Continued)

greater time is due to the cost of the Armijo-Goldstein-type line search to find θ_t. In particular, in the earlier iterations, θ_t can be quite small (e.g., below 10^{-4}) so a large number of evaluations of $\|H_{(1-\sigma\theta)\mu_t}(z^t + \theta w^t)\|$, with θ successively decreased from 1 by a factor of $\psi = .9$, is needed to find θ_t. Each evaluation requires one eigenvalue factorization, so the line search can be very expensive. When we change ψ to .2, the iteration count increases but the CPU time typically decreases due to less time being spent in the line search. For the CHKS function, this typically decreases the CPU times by 10%-50%, though they are still about twice that for the interior-point method using the AHO direction. An exception is the ChebyC problem, for which the CPU times are about equal. In general, for Algorithm 1 to be competitive with interior-point methods in terms of CPU times, it seems necessary to find a more efficient line search strategy.

We also performed some testing with warm start, as one referee suggested. After solving the SDLP, we perturbed b and c and resolved the problem starting at the solution of the unperturbed problem. In general, interior-point methods were faster at resolving the perturbed problem than Algorithm 1. An exception is the random problem of Table 2, where Algorithm 1 was able to resolve the perturbed problem while the interior point methods quit after 1 iteration without resolving the problem. However, we caution that these results are very preliminary and further studies are needed to draw any reasonable conclusion.

11. Possible extensions

In this paper we studied a non-interior continuation method for SDCP and analyzed its global (linear) convergence and local superlinear convergence. There is a number of directions in which our work can be extended. One direction is the extension of other smoothing functions for CP, e.g., [30]. Another direction is the extension of nonsmooth methods to SDCP. A third direction is a more comprehensive computational study that considers further implementation issues such as problem preprocessing, non-monotone line search, and uses a broader set of test problems. There are also some specific open questions. For example, can Lemma 5 be extended to ϕ_μ given by (4) with arbitrary $g \in \mathcal{CM}$ having Lipschitz continuous derivative? We hope these and related issues will be further studied in the future.

Lastly, it was suggested to us by Takashi Tsuchiya that the scaling technique used in interior-point methods for the Monteiro-Zhang and Monteiro-Tsuchiya families of search directions (see [35] and references therein) might be adapted to our non-interior point approach. This seems to be possible. In particular, for any nonsingular $q \in \mathcal{X}$, we can define the scaled function

$$\hat{\phi}_\mu(x, y) = \phi_\mu(qxq^T, q^{-T}yq^{-1})$$

and consider the corresponding Newton equation analogous to (17):

$$\nabla\hat{\phi}_\mu(x, y)(u, v) = r, \qquad Mu - v = s,$$

with $M := \nabla F(x)$ and $(r, s) \in \mathcal{S} \times \mathcal{S}$. [Here q^{-T} is an abbreviation for $(q^{-1})^T$.] It is not difficult to verify that

$$\nabla\hat{\phi}_\mu(x, y)(u, v) = \nabla\phi_\mu(\hat{x}, \hat{y})(\hat{u}, \hat{v}),$$

where we make the change of variable: $\hat{x} = qxq^T$, $\hat{y} = q^{-T}yq^{-1}$, $\hat{u} = quq^T$, $\hat{v} = q^{-T}vq^{-1}$. Then, the above Newton equation may be rewritten as

$$\nabla\phi_\mu(\hat{x}, \hat{y})(\hat{u}, \hat{v}) = r, \qquad \hat{M}\hat{u} - \hat{v} = \hat{s},$$

where $\hat{s} := q^{-T}sq^{-1}$ and $\hat{M}$ is the linear mapping defined by $\hat{M}\hat{u} := q^{-T}(M(q^{-1}\hat{u}q^{-T}))q^{-1}$. It can be verified that if M is monotone, then so is $\hat{M}$, and hence, as was argued in the proof of Lemmas 6 and 7, this Newton equation has a unique solution $(\hat{u}, \hat{v})$. The global convergence result given in Cor. 3(a) would still hold when the scaled Newton direction $(q^{-1}\hat{u}q^{-T}, q^T\hat{v}q)$ is used (possibly with a different q at each iteration), provided that $\|q\|$ and $\|q^{-1}\|$ are uniformly bounded.

Acknowledgements. We thank Defeng Sun and the three referees for their helpful suggestions and comments on the original version of this paper. We particularly thank one of the referees who provided numerous detailed and insightful comments.

12. Appendix

Proof of Lemma 3. By considering each diagonal block of $\mathcal{X}$ separately, we can without loss of generality assume the elements of $\mathcal{X}$ have one diagonal block, i.e., $m = 1$. Fix any $a \in \mathcal{S}$ with eigenvalues of, respectively, $\lambda_1 \geq \cdots \geq \lambda_n$. If $\lambda_1 = \cdots = \lambda_n$, then $a = \lambda_1 I$, so $p^T ap \in \mathcal{D}$ for any $p \in \mathcal{O}$ and (16) holds for any $\eta, \epsilon \in \Re_{++}$. Thus, in what follows, we consider the case where $n \geq 2$ and $\lambda_{i-1} > \lambda_i$ for some $2 \leq i \leq n$. Then, there exist unique $r \in \{2, ..., n\}$ and indices $l_0 = 1 < l_1 < \cdots < l_r = n + 1$ such that $\lambda_i = \lambda_{i+1}$ for $l_{k-1} \leq i < l_k - 1$ and $\lambda_{l_k - 1} > \lambda_{l_k}$, for $k = 1, .., r - 1$. Let

$$\delta := \min_{k=1,\ldots,r-1} (\lambda_{l_k - 1} - \lambda_{l_k})/2.$$

Consider any $b \in \mathcal{S}$ and $q \in \mathcal{O}$ such that $\|a-b\| \le \delta$ and $q^T bq = \operatorname{diag}[\mu_1, ..., \mu_n]$ for some $\mu_1 \ge \cdots \ge \mu_n \in \Re$. By an inequality of Weyl [4, p. 63], [24, p. 367],

$$|\lambda_i - \mu_i| \le \|a-b\| \quad \forall i = 1, ..., n.$$

Thus, for each $k = 1, .., r-1$, we have for $i < l_{k-1}$ that

$$\lambda_i \ge \lambda_{l_{k-1}-1} \ge \lambda_{l_{k-1}} + \|a-b\| + \delta \ge \mu_{l_{k-1}} + \delta \ge \mu_j + \delta \quad \forall j = l_{k-1}, ..., l_k - 1,$$

and for $i \ge l_k$ that

$$\lambda_i \le \lambda_{l_k} \le \lambda_{l_k-1} - (\|a-b\| + \delta) \le \mu_{l_k-1} - \delta \le \mu_j - \delta \quad \forall j = l_{k-1}, ..., l_k - 1.$$

Thus, for the index set $\mathcal{I}_k := \{l_{k-1}, ..., l_k - 1\}$, we have

$$\min\{|\lambda_i - \mu_j| : i \in \{1, ..., n\} \backslash \mathcal{I}_k, j \in \mathcal{I}_k\} \ge \delta.$$

Let $m_k := l_k - l_{k-1}$ and let $q_k \in \Re^{n \times m_k}$ denote the submatrix of q comprising the columns indexed by $i \in \mathcal{I}_k$. Let $\mathcal{Q}_k \subset \Re^n$ denote the range space of q_k. Similarly, let $\mathcal{P}_k \subset \Re^n$ denote the eigenspace of a corresponding to λ_i, $i \in \mathcal{I}_k$. By Thm. 3.4 of [43, p. 250] and Thm. 4.5 of [43, p. 92], we have

$$\begin{aligned}
\max\left\{ \sup_{\substack{\|p\|=1 \\ p \in \mathcal{P}_k}} \inf_{q \in \mathcal{Q}_k} \|p - q\|, \sup_{\substack{\|q\|=1 \\ q \in \mathcal{Q}_k}} \inf_{p \in \mathcal{P}_k} \|q - p\| \right\} &\le \|a q_k - q_k \operatorname{diag}[\mu_i]_{i \in \mathcal{I}_k}\| / \delta \\
&= \|(a-b) q_k\| / \delta \\
&\le \|a-b\| \|q_k\| / \delta \;=\; \|a-b\| / \delta,
\end{aligned}$$

where the first equality follows from the fact $bq_k = q_k \operatorname{diag}[\mu_i]_{i \in \mathcal{I}_k}$. The above inequality implies that, for each column $\tilde{q}_i$ of q_k indexed by $i \in \mathcal{I}_k$, there exists $\tilde{p}_i \in \mathcal{P}_k$ such that $\|\tilde{p}_i - \tilde{q}_i\| \le \|a-b\|/\delta$. Let $p_k \in \Re^{n \times m_k}$ be the matrix comprising $\tilde{p}_i$, $i \in \mathcal{I}_k$, for its columns. Then, letting $d_k := p_k - q_k$, we have

$$\|d_k\|^2 = \sum_{i \in \mathcal{I}_k} \|\tilde{p}_i - \tilde{q}_i\|^2 \le m_k \|a-b\|^2 / \delta^2. \tag{58}$$

Also, using $q_k^T q_k = I$, we have

$$p_k^T p_k = (q_k + d_k)^T (q_k + d_k) = I + d_k^T q_k + q_k^T d_k + d_k^T d_k.$$

Thus, further assuming $\sqrt{m_k} \|a-b\|/\delta \le 1/3$, we have $\|d_k\| \le 1/3$ and so

$$\|d_k^T q_k + q_k^T d_k + d_k^T d_k\| \le \|d_k^T q_k\| + \|q_k^T d_k\| + \|d_k^T d_k\| \le 2\|d_k\| + \|d_k\|^2 \le 7/9 < 1,$$

where the second inequality uses the facts $\|q_k\| = 1$ and $\|q^T d\| \le \|q\|\|d\|$ for any $q, d \in \Re^{n\times m_k}$. Hence $p_k^T p_k$ is nonsingular, so p_k has full column rank. Then, letting $o_k := (p_k^T p_k)^{-1/2}$, we have $(p_k o_k)^T (p_k o_k) = I$ and

$$\begin{aligned}
\|o_k - I\|^2 &= \|(p_k^T p_k)^{-1/2} - I\|^2 \\
&= \|(I + q_k^T d_k + d_k^T q_k + d_k^T d_k)^{-1/2} - I\|^2 \\
&= \sum_{i=1}^n ((1+\xi_i)^{-1/2} - 1)^2 \\
&< \sum_{i=1}^n \left(\xi_i \frac{27}{4}\right)^2 \\
&= \|q_k^T d_k + d_k^T q_k + d_k^T d_k\|^2 \left(\frac{27}{4}\right)^2 \\
&\le (2\|d_k\| + \|d_k\|^2)^2 \left(\frac{27}{4}\right)^2 \\
&\le \left(\frac{7}{3}\|d_k\|\right)^2 \left(\frac{27}{4}\right)^2,
\end{aligned}$$

where $\xi_1, ..., \xi_n$ denote the eigenvalues of $q_k^T d_k + d_k^T q_k + d_k^T d_k$, and the first inequality uses the facts that $|\xi_i| \le 7/9$ for all i and $\max_{|\xi|\le 7/9} |\frac{d}{d\xi}[(1+\xi)^{-1/2}]| = 27/(2\sqrt{8}) \le 27/4$. Then, if we let $\hat{p}_k := p_k o_k$, this together with $\|d_k\| \le 1/3$ yields

$$\begin{aligned}
\|\hat{p}_k - q_k\| &= \|(q_k + d_k)(o_k - I) + d_k\| \\
&\le (\|q_k\| + \|d_k\|)\|o_k - I\| + \|d_k\| \\
&\le (1 + \frac{1}{3})(\frac{7}{3}\|d_k\|)\frac{27}{4} + \|d_k\| \;=\; 22\|d_k\|,
\end{aligned}$$

for $k = 1, ..., r$. Since $(\hat{p}_k)^T \hat{p}_k = I$ and the columns of $\hat{p}_k$ span the eigenspace of a corresponding to $\lambda_{l_{k-1}} = \cdots = \lambda_{l_k - 1}$ (and eigenspace corresponding to distinct eigenvalues of a are orthogonal), we see that $\hat{p} := [\hat{p}_1 \;\cdots\; \hat{p}_r]$ is an $n \times n$ real orthogonal matrix, and $\hat{p}^T a \hat{p} = \operatorname{diag}[\lambda_1, ..., \lambda_n]$. This together with $q = [q_1 \;\cdots\; q_r]$ and (58) yields

$$\|\hat{p} - q\|^2 = \sum_{k=1}^r \|\hat{p}_k - q_k\|^2 \le \sum_{k=1}^r (22)^2 m_k \|a - b\|^2/\delta^2.$$

□

Proof of Lemma 4. Suppose g is continuously differentiable. For any $a \in \mathcal{S}$, let $\lambda_1 \ge \cdots \ge \lambda_n \in \Re$ be the eigenvalues of a in decreasing order and let c be given by (13). Since $0 < g'(\tau) < 1$ for all $\tau \in \Re$, then $0 < c_{ij} < 1$ for all i, j. For any other $\tilde{a} \in \mathcal{S}$, let $\mu_1 \ge \cdots \ge \mu_n \in \Re$ be the eigenvalues of $\tilde{a}$ in decreasing order and let $\tilde{c}$ be given by a formula analogous to (13) but with "μ" replacing "λ". Then, for any $p, q \in \mathcal{O}$ such that

$p^T \mathrm{diag}[\lambda_1, ..., \lambda_n]p = a$, $q^T \mathrm{diag}[\mu_1, ..., \mu_n]q = \tilde{a}$, we have from a result of Dalecki and Krein (see the proof of Lemma 2(a)) that, for all $e \in \mathcal{S}$,

$$\begin{aligned}\nabla g(a)e - \nabla g(\tilde{a})e &= p^T((pep^T) \circ c)p - q^T((qeq^T) \circ \tilde{c})q \\ &= d^T((pep^T) \circ c)p + q^T((dep^T) \circ c)p + q^T((qed^T) \circ c)p \\ &+ q^T((qeq^T) \circ c)d + q^T((qeq^T) \circ (c - \tilde{c}))q,\end{aligned}$$

where we define $d := p - q$, and the second equality is obtained by substituting p with $q + d$ and expanding the terms. Thus,

$$\begin{aligned}\|\nabla g(a)e - \nabla g(\tilde{a})e\| &\le \|d^T((pep^T) \circ c)p\| + \|q^T((dep^T) \circ c)p\| + \|q^T((qed^T) \circ c)p\| \\ &+ \|q^T((qeq^T) \circ c)d\| + \|q^T((qeq^T) \circ (c - \tilde{c}))q\| \\ &\le 4\|d\|\|e\| + \|e\|\|c - \tilde{c}\|,\end{aligned}$$

where the second inequality uses the facts $p, q \in \mathcal{O}$ and $0 < c_{ij} < 1$ for all i, j, so that $\|d^T((pep^T) \circ c)p\| = \|d^T((pep^T) \circ c)\| \le \|d\|\|(pep^T) \circ c\| \le \|d\|\|pep^T\| = \|d\|\|e\|$, and similarly for the other terms. Thus,

$$\begin{aligned}|||\nabla g(a) - \nabla g(\tilde{a})||| &= \max_{\|e\|=1} \|\nabla g(a)e - \nabla g(\tilde{a})e\| \\ &\le (4\|p - q\| + \|c - \tilde{c}\|) \max_{\|e\|=1} \|e\| \\ &\le 4\|p - q\| + \|c - \tilde{c}\|.\end{aligned}$$

By an inequality of Weyl [4, p. 63], [24, p. 367], we have $|\lambda_i - \mu_i| \le \|a - \tilde{a}\|$ for all i, so the continuity of g' and the definition of $c, \tilde{c}$ imply that $\|c - \tilde{c}\| \to 0$ whenever $\|a - \tilde{a}\| \to 0$. Also, by Lemma 3, we could have chosen $p \in \mathcal{O}$ (depending on q) so that $\|p - q\| \to 0$ whenever $\|a - \tilde{a}\| \to 0$. This then yields that

$$|||\nabla g(a) - \nabla g(\tilde{a})||| \le 4 \min_{\substack{p,q \in \mathcal{O} \\ pap^T \in \mathcal{D} \\ q\tilde{a}q^T \in \mathcal{D}}} \|p - q\| + \|c - \tilde{c}\| \to 0 \ \text{ as } \ \|a - \tilde{a}\| \to 0,$$

where $\mathcal{D}$ is as defined in Lemma 3. Thus, ∇g is continuous at a.

Suppose g is analytic. For any $a \in \mathcal{S}$, $g(a)$ can be expressed by the Cauchy integral formula

$$g(a) = \frac{1}{2\pi} \oint_\Gamma g(\tau)(\tau I - a)^{-1}\, d\tau,$$

where Γ is any simple closed rectifiable curve (say, a circle) in $\mathbb{C}$ that strictly encloses all the eigenvalues of a [25, p. 427]. Using this formula, one obtains that, for all $d \in \mathcal{S}$,

$$\nabla g(a)d = \frac{1}{2\pi} \oint_\Gamma g(\tau)(\tau I - a)^{-1} d(\tau I - a)^{-1}\, d\tau,$$

e.g., [25, p. 521]. Fix any $\delta \in \Re_{++}$ such that Γ strictly encloses all the eigenvalues of b for all $b \in \mathcal{S}$ with $\|a - b\| \le \delta$. Then, $\nabla g(b)d$ may be expressed by the same formula except with a replaced by b. Then

$$\begin{aligned}
&|||\nabla g(b) - \nabla g(a)||| \\
&\quad = \sup_{\|d\|=1} \|\nabla g(b)d - \nabla g(a)d\| \\
&\quad = \frac{1}{2\pi} \sup_{\|d\|=1} \left\| \oint_\Gamma g(\tau) \left((\tau I - b)^{-1} d (\tau I - b)^{-1} - (\tau I - a)^{-1} d (\tau I - a)^{-1} \right) d\tau \right\| \\
&\quad = \frac{1}{2\pi} \sup_{\|d\|=1} \left\| \oint_\Gamma g(\tau) \left((\tau I - b)^{-1} d \left((\tau I - b)^{-1} - (\tau I - a)^{-1} \right) \right.\right. \\
&\qquad \left.\left. + \left((\tau I - b)^{-1} - (\tau I - a)^{-1} \right) d (\tau I - a)^{-1} \right) d\tau \right\| \\
&\quad \le \frac{1}{2\pi} \oint_\Gamma \|g(\tau)\| (\|(\tau I - b)^{-1}\| + \|(\tau I - a)^{-1}\|) d\tau \\
&\quad \cdot \max_{\tau \in \Gamma} \|(\tau I - b)^{-1} - (\tau I - a)^{-1}\|.
\end{aligned}$$

The right-hand side tends to zero as b approaches a, i.e., $\|b - a\| \to 0$. Thus, $g(a)$ is continuously differentiable in a. By using the above integral formula for $\nabla g(a)d$ we can similarly show that $g(a)$ is twice continuously differentiable in a and so on. By such an inductive argument, we find that $g(a)$ is k-times continuously differentiable in a, for $k = 1, 2, ...,$ with kth-order derivative

$$\nabla^k g(a)[d_1, ..., d_k] = \frac{1}{2\pi} \oint_\Gamma g(\tau)(\tau I - a)^{-1} \left(\sum_\sigma \prod_{i=1}^k (d_{\sigma(i)} (\tau I - a)^{-1}) \right) d\tau,$$

where the summation is taken over all permutations $\sigma : \{1, ..., k\} \mapsto \{1, ..., k\}$. □

Proof of Lemma 5. Suppose ϕ_μ is given by (7). Then (15) yields

$$\nabla \phi_1(x, y)(u, v) = u + v - L_c^{-1}[xu + ux + yv + vy] \quad \text{with } c = (x^2 + y^2 + 2I)^{1/2}.$$

Since $c \succ 0$ and is continuous in (x, y) so that L_c^{-1} is continuous in (x, y), it is readily seen that $\nabla \phi_1(x, y)$ is continuous in (x, y). Now we show $\nabla \phi_1$ is Lipschitz continuous. Fix any $x, y, \bar{x}, \bar{y} \in \mathcal{S}$. For any $u, v \in \mathcal{S}$ with $\|(u, v)\| = 1$, we have from (15) that

$$\begin{aligned}
&\nabla \phi_1(x, y)(u, v) - \nabla \phi_1(\bar{x}, \bar{y})(u, v) \\
&\quad = L_{\bar{c}}^{-1}[\bar{x}u + u\bar{x} + \bar{y}v + v\bar{y}] - L_c^{-1}[xu + ux + yv + vy] \\
&\quad = L_{\bar{c}}^{-1}[(\bar{x} - x)u + u(\bar{x} - x) + (\bar{y} - y)v + v(\bar{y} - y)] + \bar{s} - s,
\end{aligned} \tag{59}$$

where we let $c := (x^2 + y^2 + 2I)^{1/2}$, $\bar{c} := (\bar{x}^2 + \bar{y}^2 + 2I)^{1/2}$, and $s := L_c^{-1}[xu + ux + yv + vy]$, $\bar{s} := L_{\bar{c}}^{-1}[xu + ux + yv + vy]$. For any $a \in \mathcal{S}$ and $b := L_{\bar{c}}^{-1}[a]$, we have $\bar{c}b + b\bar{c} = a$ and hence

$$\begin{aligned}
\|a\|^2 = \|\bar{c}b + b\bar{c}\|^2 &= 2\mathrm{tr}[\bar{c}^2 b^2 + \bar{c}b\bar{c}b] \\
&= 2(\langle \bar{x}^2 + \bar{y}^2, b^2 \rangle + 2\|b\|^2 + \|\bar{c}^{1/2} b \bar{c}^{1/2}\|^2) \ge 4\|b\|^2,
\end{aligned}$$

where the last inequality uses $\bar{x}^2 + \bar{y}^2 \succeq 0$, $b^2 \succeq 0$ so their inner product is nonnegative. Thus, the first term on the right-hand side of (59) can be bounded as follows:

$$\begin{aligned}
&\|L_{\bar{c}}^{-1}[(\bar{x}-x)u + u(\bar{x}-x) + (\bar{y}-y)v + v(\bar{y}-y)]\| \\
&\quad \le \|(\bar{x}-x)u + u(\bar{x}-x) + (\bar{y}-y)v + v(\bar{y}-y)\|/2 \\
&\quad \le \|\bar{x}-x\|\|u\| + \|\bar{y}-y\|\|v\| \\
&\quad \le \|(\bar{x}-x, \bar{y}-y)\|, \qquad (60)
\end{aligned}$$

where the last inequality uses the arithmetic identity $(\alpha_1\beta_1 + \alpha_2\beta_2)^2 \le (\alpha_1^2 + \alpha_2^2)(\beta_1^2 + \beta_2^2)$ and $\|(u, v)\| \le 1$. Now we bound the second term on the right-hand side of (59). We have $cs + sc = xu + ux + yv + vy = \bar{c}\bar{s} + \bar{s}\bar{c}$ and hence

$$(c-\bar{c})s + s(c-\bar{c}) = \bar{c}(\bar{s}-s) + (\bar{s}-s)\bar{c} = L_{\bar{c}}[\bar{s}-s].$$

We showed earlier that, for any $a \in \mathcal{S}$ and $b := L_{\bar{c}}^{-1}[a]$, we have $\|a\|^2 \ge 4\|b\|^2$. Thus the above equation implies

$$\|\bar{s}-s\| \le \|(c-\bar{c})s + s(c-\bar{c})\|/2 \le \|c-\bar{c}\|\|s\|. \qquad (61)$$

Now we bound $\|s\|$ and $\|c-\bar{c}\|$. Choose $p \in \mathcal{O}$ such that $pcp^T = \mathrm{diag}[\lambda_1, ..., \lambda_n]$ for some $\lambda_1 \ge \cdots \ge \lambda_n > 0$. Then, $(\tilde{x}^2)_{ii} + (\tilde{y}^2)_{ii} + 2 = \lambda_i^2$ for $i = 1, ..., n$ and $\tilde{s}_{ij} = (\tilde{x}\tilde{u} + \tilde{u}\tilde{x} + \tilde{y}\tilde{v} + \tilde{v}\tilde{y})_{ij}/(\lambda_i + \lambda_j)$ for $i, j = 1, ..., n$, where $\tilde{s} := psp^T$, $\tilde{x} := pxp^T$, $\tilde{y} := pyp^T$, $\tilde{u} := pup^T$, $\tilde{v} := pvp^T$. Denoting column i of $a \in \mathcal{S}$ by a_i, we thus obtain

$$\begin{aligned}
|\tilde{s}_{ij}| &= |\tilde{x}_i^T\tilde{u}_j + \tilde{u}_i^T\tilde{x}_j + \tilde{y}_i^T\tilde{v}_j + \tilde{v}_i^T\tilde{y}_j|/(\lambda_i + \lambda_j) \\
&\le (|\tilde{x}_i^T\tilde{u}_j| + |\tilde{u}_i^T\tilde{x}_j| + |\tilde{y}_i^T\tilde{v}_j| + |\tilde{v}_i^T\tilde{y}_j|)/(\lambda_i + \lambda_j) \\
&\le (\|\tilde{x}_i\|\|\tilde{u}_j\| + \|\tilde{u}_i\|\|\tilde{x}_j\| + \|\tilde{y}_i\|\|\tilde{v}_j\| + \|\tilde{v}_i\|\|\tilde{y}_j\|)/(\lambda_i + \lambda_j) \\
&\le \|\tilde{u}_j\| + \|\tilde{u}_i\| + \|\tilde{v}_j\| + \|\tilde{v}_i\|,
\end{aligned}$$

where the last inequality uses $\|\tilde{x}_i\|^2 + \|\tilde{y}_i\|^2 = (\tilde{x}^2)_{ii} + (\tilde{y}^2)_{ii} = \lambda_i^2 - 2 \le \lambda_i^2$ for all i. Hence

$$\begin{aligned}
\|s\|^2 &= \|\tilde{s}\|^2 \\
&= \sum_{i,j} |\tilde{s}_{ij}|^2 \\
&\le \sum_{i,j} (\|\tilde{u}_j\| + \|\tilde{u}_i\| + \|\tilde{v}_j\| + \|\tilde{v}_i\|)^2 \\
&\le \sum_{i,j} 4(\|\tilde{u}_j\|^2 + \|\tilde{u}_i\|^2 + \|\tilde{v}_j\|^2 + \|\tilde{v}_i\|^2) \\
&= 8n(\|\tilde{u}\|^2 + \|\tilde{v}\|^2) \;=\; 8n\|(u, v)\|^2 \;=\; 8n. \qquad (62)
\end{aligned}$$

Define $\psi : \mathcal{S} \times \mathcal{S} \mapsto \mathcal{S}$ by $\psi(x, y) := (x^2 + y^2 + 2I)^{1/2}$. Since $\phi_1(x, y) = x + y - \psi(x, y)$ and we already proved that ϕ_1 is continuously differentiable, then so is ψ. Moreover, by (15), $\nabla\psi(x, y)(u, v) = L_c^{-1}[xu + ux + yv + vy]$, with $c = (x^2 + y^2 + 2I)^{1/2}$.

This together with (62) implies $\|\nabla\psi(x, y)(u, v)\| \le \sqrt{8n}\|(u, v)\|$, so the mean value theorem yields

$$\begin{aligned}\|\bar{c} - c\| &= \|\psi(\bar{x}, \bar{y}) - \psi(x, y)\| \\ &= \left\| \int_0^1 \nabla\psi((x, y) + \tau(\bar{x} - x, \bar{y} - y))(\bar{x} - x, \bar{y} - y)d\tau \right\| \\ &\le \int_0^1 \|\nabla\psi((x, y) + \tau(\bar{x} - x, \bar{y} - y))(\bar{x} - x, \bar{y} - y)\|d\tau \\ &\le \sqrt{8n}\|(\bar{x} - x, \bar{y} - y)\|.\end{aligned}$$

Combining (59), (60), (61), (62) yields

$$\begin{aligned}\||\nabla\phi_1(x, y) - \nabla\phi_1(\bar{x}, \bar{y})\|| &= \max_{\|(u,v)\|=1} \|\nabla\phi_1(x, y)(u, v) - \nabla\phi_1(\bar{x}, \bar{y})(u, v)\| \\ &\le (1 + 8n)\|(\bar{x} - x, \bar{y} - y)\|.\end{aligned}$$

For ϕ_μ given by (4)–(5), we have from Cor. 2 that $\nabla\phi_1$ is defined and continuous. The proof that $\nabla\phi_1$ is Lipschitz continuous is very similar to the one given above, but using (14) in place of (15). The corresponding Lipschitz constant is $(1+4n)/\sqrt{2}$, instead of $1 + 8n$. For brevity, the details are omitted. □

References

[1] Alizadeh, F.: Interior point methods in semidefinite programming with application to combinatorial optimization . SIAM J. Optim. **5**, 13–51 (1995)

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

[3] Auslender, A.: Penalty and barrier methods: a unified framework. SIAM J. Optim. **10**, 211–230 (1999)

[4] Bhatia, R.: *Matrix Analysis*, Springer-Verlag, New York, 1997

[5] Burke, J., Xu, S.: The global linear convergence of a non-interior path-following algorithm for linear complementarity problems. Math. Oper. Res. **23**, 719–734 (1998)

[6] Burke, J., Xu, S.: A Non-interior predictor-corrector path-following method for LCP. In: *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods* M. Fukushima, L. Qi (ed), Kluwer Academic Publishers, Boston, pp. 45–63, 1999

[7] Burke, J., Xu, S.: The complexity of a non-interior path following method for the linear complementarity problem, Report, Department of Mathematics, University of Washington, Seattle (September 2000; revised April 2001). J. Optim. Theory Appl. to appear

[8] Chen, B.: Error bounds for R_0-type and monotone nonlinear complementarity problems. J. Optim. Theory Appl. **108**, 297–316 (2001)

[9] Chen, B., Chen, X.-J.: A global linear and local quadratic continuation smoothing method for variational inequalities with box constraints. Comput. Optim. Appl. **17**, 131–158 (2000)

[10] Chen, B., Chen, X.-J.: A global and local superlinear continuation-smoothing method for P_0 and R_0 NCP or monotone NCP. SIAM J. Optim. **9**, 624–645 (1999)

[11] Chen, B., Harker, P.T.: A non-interior-point continuation method for linear complementarity problems. SIAM J. Matrix Anal. Appl. **14**, 1168–1190 (1993)

[12] Chen, B., Harker, P.T.: A continuation method for monotone variational inequalities. Math. Program. **69**, 237–253 (1995)

[13] Chen, B., Harker, P.T.: Smooth approximations to nonlinear complementarity problems. SIAM J. Optim. **7**, 403–420 (1997)

[14] Chen, B., Xiu, N.: A global linear and local quadratic non-interior continuation method for nonlinear complementarity problems based on Chen-Mangasarian smoothing function. SIAM J. Optim. **9**, 605–623 (1999)

[15] Chen, C., Mangasarian, O.L.: Smoothing methods for convex inequalities and linear complementarity problems. Math. Program. **71**, 51–69 (1995)

[16] Chen, C., Mangasarian, O.L.: A class of smoothing functions for nonlinear and mixed complementarity problems. Comput. Optim. Appl. **5**, 97–138 (1996)
[17] Chen, X.-J., Qi, L., Sun, D.: Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities. Math. Comput. **67**, 519–540 (1998)
[18] Chen, X.-J., Ye, Y.: On homotopy-smoothing methods for box-constrained variational inequalities. SIAM J. Control Optim. **37**, 589–616 (1999)
[19] Cottle, R.W., Pang, J.-S., Stone, R.E.: *The Linear Complementarity Problem*, Academic Press, New York, New York, 1992
[20] Ferris, M.C., Pang, J.-S. (eds): *Complementarity and Variational Problems: State of the Art*, SIAM Publishing, Philadelphia, Pennsylvania, 1997
[21] Fukushima, M., Luo, Z.-Q., Pang, J.-S.: A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints. Comput. Optim. Appl. **10**, 5–34 (1998)
[22] Fukushima, M., Qi, L. (eds): *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Academic Publishers, Boston, 1999
[23] Gabriel, S.A., Moré, J.J.: Smoothing of mixed complementarity problems. In: *Complementarity and Variational Problems: State of the Art*, M.C. Ferris, J.-S. Pang (eds), SIAM Publishing, Philadelphia, Pennsylvania, pp. 105–116, 1997
[24] Horn, R.A., Johnson, C.R.: *Matrix Analysis*, Cambridge University Press, Cambridge, United Kingdom, 1985
[25] Horn, R.A., Johnson, C.R.: *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, United Kingdom, 1991
[26] Hotta, K., Yoshise, A.: Global convergence of a class of non-interior point algorithms using Chen-Harker-Kanzow-Smale functions for nonlinear complementarity problems. Math. Program. **86**, 105–133 (1999)
[27] Ji, J., Potra, F., Sheng, R.: On the local convergence of a predictor-corrector method for semidefinite programming. SIAM J. Optim. **10**, 195–210 (1999)
[28] Kanzow, C.: Some noninterior continuation methods for linear complementarity problems. SIAM J. Matrix Anal. Appl. **17**, 851–868 (1996)
[29] Kanzow, C.: A new approach to continuation methods for complementarity problems with uniform P-functions. Oper. Res. Letters, **20**, 85–92 (1997)
[30] Kanzow, C., Yamashita, N., Fukushima, M.: New NCP-functions and their properties. J. Optim. Theory Appl. **94**, 115–135 (1997)
[31] Kojima, M., Shindoh S., Hara, S.: Interior-point methods for the monotone semidefinite linear complementarity problems. SIAM J. Optim. **7**, 86–125 (1997)
[32] Kojima, M., Shida, M., Shindoh, S.: Local convergence of predictor-corrector infeasible interior-point algorithms for SDPs and SDLCPs. Math. Program. **80**, 129–160 (1998)
[33] Kojima, M., Shida, M., Shindoh S.: A predictor-corrector interior-point algorithms for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction. SIAM J. Optim. **9**, 444–465 (1999)
[34] Luo, Z.-Q., Tseng, P.: A new class of merit functions for the nonlinear complementarity problem, In: *Complementarity and Variational Problems: State of the Art*, M.C. Ferris, J.-S.Pang (eds), SIAM Publishing, Philadelphia, Pennsylvania, pp. 204–225, 1997
[35] Monteiro, R.D.C., Tsuchiya, T.: Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming. SIAM J. Optim. **9**, 551–577 (1999)
[36] Monteiro, R.D.C., Zanjácomo, P.: Implementation of primal-dual methods for semidefinite programming based on Monteiro and Tsuchiya Newton directions and their variants. Optim. Methods Software **11/12**, 91–140 (1999)
[37] Nesterov, Y.E., Nemirovskii, A.S.: *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*, SIAM, Philadelphia, 1993
[38] Nesterov, Y.E., Todd, M.J.: Primal-dual interior-point methods for self-scaled cones. SIAM J. Optim. **8**, 324–364 (1998)
[39] Pang, J.-S.: A Posteriori error bounds for the linearly–constrained variational inequality problem. Math. Oper. Res. **12**, 474–484 (1987)
[40] Pang, J.-S.: Error bounds in mathematical programming. Math. Program. **79**, 299–332 (1997)
[41] Shida, M., Shindoh, S.: Monotone semidefinite complementarity problems, Research Report 312, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo (March 1996; revised April 1996)
[42] Smale, S.: Algorithms for solving equations. In: *Proceedings of the International Congress of Mathematicians*, A.M. Gleason (eds), American Mathematical Society, Providence, Rhode Island pp. 172-195, 1987

[43] Stewart, G.W., Sun, J.-G.: *Matrix Perturbation Theory*, Academic Press, New York, 1990
[44] Todd, M.J.: A study of search directions in primal-dual interior-point methods for semidefinite programming. Optim. Methods Software **11**, 1–46 (1999)
[45] Toh, K.C., Todd, M.J., Tütüncü, R.H.: SDPT^3 - a Matlab software package for semidefinite programming. Optim. Methods Software **11**, 545–581 (1999)
[46] Tseng, P.: Growth behavior of a class of merit functions for the nonlinear complementarity problem. J. Optim. Theory Appl. **89**, 17–37 (1996)
[47] Tseng, P.: An infeasible path-following method for monotone complementarity problems. SIAM J. Optim. **7**, 386–402 (1997)
[48] Tseng, P.: Search directions and convergence analysis of some infeasible path-following methods for the monotone semi-definite LCP. Optim. Methods Software **9**, 245–268 (1998)
[49] Tseng, P.: Merit functions for semi-definite complementarity problems. Math. Program. **83**, 159–185 (1998)
[50] Tseng, P.: Analysis of a non-interior continuation method based on Chen-Mangasarian smoothing functions for complementarity problems. In: *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, M. Fukushima, L. Qi (eds), Kluwer Academic Publishers, Boston, pp. 381–404, 1999
[51] Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Review **38**, 49–95 (1996)
[52] Xu, S.: The global linear convergence an infeasible non-interior path-following algorithm for complementarity problems with uniform P-functions. Math. Program. **87**, 501–517 (2000)
[53] Xu, S.: The global linear convergence and complexity of a non-interior path-following algorithm for monotone LCP based on Chen-Harker-Kanzow-Smale smoothing functions, Preprint, Department of Mathematics, University of Washington, Seattle, Washington (February 1997)
[54] Yamashita, N., Fukushima, M.: A new merit function and a descent method for semidefinite complementarity problems. In: *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, M. Fukushima, L. Qi (eds), Kluwer Academic Publishers, Boston, pp. 405–420, 1999