Optimization
Vol. 59, No. 5, July 2010, 661–676

Lipschitz continuity of the gradient of a one-parametric class of SOC merit functions

Jein-Shan Chen[a]* and Shaohua Pan[b]

[a]*Department of Mathematics, National Taiwan Normal University, Taipei, Taiwan;* [b]*School of Mathematical Sciences, South China University of Technology, Guangzhou, China*

(*Received 27 March 2007; final version received 11 March 2008*)

In this article, we show that a one-parametric class of SOC merit functions has a Lipschitz continuous gradient; and moreover, the Lipschitz constant is related to the parameter in this class of SOC merit functions. This fact will lay a building block when the merit function approach as well as the Newton-type method are employed for solving the second-order cone complementarity problem with this class of merit functions.

Keywords: second-order cone; merit function; spectral factorization; Lipschitz continuity

AMS Subject Classifications: 26B05; 26B35; 90C33; 65K05

1. Introduction

A well-known approach to solving the non-linear complementarity problem (NCP) is to reformulate it as the global minimization via a certain merit function over $\mathbb{R}^n$. For the approach to be effective, the choice of the merit function is crucial. A popular choice is the squared norm of the Fischer–Burmeister (FB) function $\psi_{\mathrm{FB}}: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ defined by

$$\psi_{\mathrm{FB}}(a, b) := \frac{1}{2}\sum_{i=1}^{n}[\phi_{\mathrm{FB}}(a_i, b_i)]^2 \tag{1}$$

for all $a = (a_1, \ldots, a_n)^{\mathrm{T}} \in \mathbb{R}^n$ and $b = (b_1, \ldots, b_n)^{\mathrm{T}} \in \mathbb{R}^n$, where $\phi_{\mathrm{FB}}: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the Fischer–Burmeister NCP function given as

$$\phi_{\mathrm{FB}}(a_i, b_i) = \sqrt{a_i^2 + b_i^2} - a_i - b_i. \tag{2}$$

It has been shown that ψ_{FB} enjoys many desirable properties [8,9], for example, smoothness (continuous differentiability). This merit function and its analysis were subsequently extended by Tseng [18] to the semidefinite complementarity problem (SDCP)

*Corresponding author. Email: jschen@math.ntnu.edu.tw

ISSN 0233–1934 print/ISSN 1029–4945 online

DOI: 10.1080/02331930802180319
http://www.informaworld.com

although only differentiability, not continuous differentiability, was established. In fact, the FB function for the SDCP is the matrix-valued function $\Phi_{\text{FB}}: \mathcal{S}^n \times \mathcal{S}^n \to \mathcal{S}^n$ defined by

$$\Phi_{\text{FB}}(X, Y) := (X^2 + Y^2)^{1/2} - (X + Y),$$

while the squared norm of the FB function for the SDCP is the function $\Psi_{\text{FB}}: \mathcal{S}^n \times \mathcal{S}^n \to \mathbb{R}_+$ given by

$$\Psi_{\text{FB}}(X, Y) := \frac{1}{2}\|\Phi(X, Y)\|^2,$$

where $\mathcal{S}^n$ denotes the set of real $n \times n$ symmetric matrices. The function Φ_{FB} has been proved to be strongly semismooth [17]. More recently, the squared norm of the matrix-valued FB function Ψ_{FB} was reported in [16] to be smooth and its gradient is Lipschitz continuous.

The *second-order cone* (SOC), also called the Lorentz cone, in $\mathbb{R}^n$ is defined as

$$\mathcal{K}^n := \left\{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x_2\| \le x_1\right\}, \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm. By definition, $\mathcal{K}^1$ is the set of non-negative reals $\mathbb{R}_+$. The second-order cone complementarity problem (SOCCP) is to find $x, y \in \mathbb{R}^n$ satisfying

$$x = F(\zeta), \quad y = G(\zeta), \quad \langle x, y\rangle = 0, \quad x \in \mathcal{K}^n, \quad y \in \mathcal{K}^n, \tag{4}$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product and $F, G: \mathbb{R}^n \to \mathbb{R}^n$ are continuous (possibly non-linear) mappings. The merit function approach based on reformulating the NCP as an equivalent unconstrained minimization can be extended to the SOCCP case [6]. This approach aims to find a smooth function $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ such that

$$\psi(x, y) = 0 \iff x \in \mathcal{K}^n, \quad y \in \mathcal{K}^n, \quad \langle x, y\rangle = 0. \tag{5}$$

We call such ψ a *SOC merit function*. Then the SOCCP can be expressed as an unconstrained smooth (global) minimization problem:

$$\min_{\zeta \in \mathbb{R}^n} f(\zeta) := \psi(F(\zeta), G(\zeta)). \tag{6}$$

Analogously, the squared norm of FB function can be considered in the SOCCP setting. We define $\psi_{\text{FB}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ associated with the second-order cone $\mathcal{K}^n$ as

$$_{\text{FB}}(x, y) := \frac{1}{2}\|\phi_{\text{FB}}(x, y)\|^2, \tag{7}$$

where $\phi_{\text{FB}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is the FB function defined by

$$\phi_{\text{FB}}(x, y) := (x^2 + y^2)^{1/2} - x - y. \tag{8}$$

More specifically, for any $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we define their *Jordan product* associated with $\mathcal{K}^n$ as

$$x \circ y := (\langle x, y\rangle,\ y_1x_2 + x_1y_2). \tag{9}$$

The Jordan product $\circ$, unlike scalar or matrix multiplication, is not associative, which is a main source of complication in the analysis of SOCCP. The identity element under this product is $e := (1, 0, \ldots, 0)^{\mathrm{T}} \in \mathrm{IR}^n$. We write x^2 to mean $x \circ x$ and write $x + y$ to mean the usual componentwise addition of vectors. It is known that $x^2 \in \mathcal{K}^n$ for all $x \in \mathrm{IR}^n$. Moreover, if $x \in \mathcal{K}^n$, then there exists a unique vector in $\mathcal{K}^n$, denoted by $x^{1/2}$, such that $(x^{1/2})^2 = x^{1/2} \circ x^{1/2} = x$. Thus, ϕ_{FB} defined as in (8) is well-defined for all $(x, y) \in \mathrm{IR}^n \times \mathrm{IR}^n$ and maps $\mathrm{IR}^n \times \mathrm{IR}^n$ to IR^n. It was shown in [10] that ϕ_{FB} satisfies the relation (5). Hence, ψ_{FB} as defined in (7) is a merit function for the SOCCP. In the recent manuscript [5], this SOC merit function was shown to be an LC^1 function (a smooth function with its gradient being locally Lipschitz continuous).

Another popular SOC merit function is the natural residual merit function

$$\psi_{\mathrm{NR}}(x, y) := \|\phi_{\mathrm{NR}}(x, y)\|^2, \tag{10}$$

which is induced by the natural residual function $\phi_{\mathrm{NR}} : \mathrm{IR}^n \times \mathrm{IR}^n \to \mathrm{IR}^n$

$$\phi_{\mathrm{NR}}(x, y) := x - [x - y]_+, \tag{11}$$

where $[\,\cdot\,]_+$ means the projection onto $\mathcal{K}^n$. The natural residual function ϕ_{NR} was studied in [10,11] which is involved in smoothing methods for the SOCCP. A drawback of ψ_{NR} is its non-differentiability compared to ψ_{FB}. Some other classes of SOC merit functions for the SOCCP are also recently studied in [1,2].

In this article, we consider the following one-parametric class of SOC merit functions which was originally proposed in [12] for the NCP case:

$$\psi_\tau(x, y) := \frac{1}{2}\|\phi_\tau(x, y)\|^2 \tag{12}$$

where $\phi_\tau : \mathrm{IR}^n \times \mathrm{IR}^n \to \mathrm{IR}^n$ is a family of functions associated with the SOC, defined by

$$\phi_\tau(x, y) := \left[(x - y)^2 + \tau(x \circ y)\right]^{1/2} - (x + y), \tag{13}$$

and τ is a fixed parameter such that $\tau \in (0, 4)$. It can be verified that for any $x,\ y \in \mathrm{IR}^n$

$$\begin{aligned}(x - y)^2 + \tau(x \circ y) &= \left(x + \frac{\tau - 2}{2}y\right)^2 + \frac{\tau(4 - \tau)}{4}y^2 \\ &= \left(y + \frac{\tau - 2}{2}x\right)^2 + \frac{\tau(4 - \tau)}{4}x^2 \\ &\succeq_{\mathcal{K}^n} 0,\end{aligned} \tag{14}$$

where the inequality holds because $\tau \in (0, 4)$. Therefore, ϕ_τ in (13) is well-defined. Notice that ϕ_τ reduces to the FB function ϕ_{FB} when $\tau = 2$, whereas it becomes a multiple of the natural residual function ϕ_{NR} when $\tau \to 0$. Thus, this class of SOC complementarity functions covers the current two most important SOC complementarity functions so that a closer look and study for this new class of functions is worthwhile.

In fact, as mentioned in [5], when solving the equivalent unconstrained minimization via SOC merit functions, it is very important to show that the gradient of the employed SOC merit function is sufficiently smooth so as to warrant the convergence of appropriate computational methods. Here, we are particularly concerned with the conjugate gradient method. The method generally requires the Lipschitz continuity of the gradient ($f \in LC^1$ by our notation). The main purpose of this article is to show that the function ψ_τ as defined in (12) has a globally Lipschitz continuous gradient. Thus, this article can be regarded as a follow-up of [5] since it extends the LC^1 property of $\psi_{\rm FB}$ to the class of SOC merit functions ψ_τ. Nonetheless, the technique used here is a bit different and the analysis is more tedious and subtle. In particular, from the extension work, we see that the Lipschitz continuity of the gradient of ψ_τ becomes worse when $\tau \to 0$. This fact will provide an instructional help for the design of algorithms with this class of SOC merit functions.

Throughout this article, $\mathbb{R}^n$ denotes the space of n-dimensional real column vectors and the supscript 'T' represents the transpose. For any differentiable function $f: \mathbb{R}^n \to \mathbb{R}$, $\nabla f(x)$ denotes the gradient of f at x. For any differentiable mapping $F=(F_1,\ldots,F_m)^{\rm T}: \mathbb{R}^n \to \mathbb{R}^m$, $\nabla F(x)=[\nabla F_1(x) \cdots \nabla F_m(x)]$ is a $n \times m$ matrix denoting the transposed Jacobian of F at x. For non-negative scalars α and β, we write $\alpha = O(\beta)$ to mean $\alpha \le C\beta$, with C independent of α and β.

2. Preliminaries

It is known that $\mathcal{K}^n$ is a closed convex self-dual cone with non-empty interior given as

$$\mathcal{K}^n := \left\{(x_1,x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x_2\| \le x_1\right\}.$$

For any $x=(x_1,x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we define the *determinant* and the *trace* of x as follows:

$$\det(x) := x_1^2 - \|x_2\|^2, \quad \operatorname{tr}(x) = 2x_1.$$

In general, $\det(x \circ y) \neq \det(x)\det(y)$ unless x and y are collinear, i.e. $x = \alpha y$ for some $\alpha \in \mathbb{R}$. A vector $x=(x_1,x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ is said to be *invertible* if $\det(x) \neq 0$. If x is invertible, then there exists a unique $y=(y_1,y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ satisfying $x \circ y = y \circ x = e$. We call this y the inverse of x and denote it by x^{-1}. In fact, we have

$$x^{-1} = \frac{1}{x_1^2 - \|x_2\|^2}(x_1, -x_2) = \frac{1}{\det(x)}\Big(\operatorname{tr}(x)e - x\Big).$$

It is not difficult to see that $x \in \operatorname{int}(\mathcal{K}^n)$ if and only if $x^{-1} \in \operatorname{int}(\mathcal{K}_n)$. For any $x=(x_1,x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, we define the matrix L_x by

$$L_x := \begin{bmatrix} x_1 & x_2^{\rm T} \\ x_2 & x_1 I \end{bmatrix}.$$

It is easily verified that $L_{x+y} = L_x + L_y$ and $x \circ y = L_x y$ for any $x, y \in \mathbb{R}^n$, and L_x is positive definite (and hence invertible) if and only if $x \in \operatorname{int}(\mathcal{K}^n)$. However, $L_x^{-1} y \neq x^{-1} \circ y$, for some $x \in \operatorname{int}(\mathcal{K}^n)$ and $y \in \mathbb{R}^n$, i.e. $L_x^{-1} \neq L_{x^{-1}}$.

We next recall from [10] that each $x=(x_1,x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ admits a spectral factorization, associated with $\mathcal{K}^n$, of the form

$$x = \lambda_1(x) u_x^{(1)} + \lambda_2(x) u_x^{(2)}, \tag{15}$$

where $\lambda_1(x)$, $\lambda_2(x)$ and $u_x^{(1)}$, $u_x^{(2)}$ are the *spectral values* and the associated *spectral vectors* of x, with respect to $\mathcal{K}^n$, given by

$$\lambda_i(x) = x_1 + (-1)^i \|x_2\|, \tag{16}$$

$$u_x^{(i)} = \begin{cases} \dfrac{1}{2}\left(1, (-1)^i \dfrac{x_2}{\|x_2\|}\right), & \text{if } x_2 \neq 0, \\ \dfrac{1}{2}\left(1, (-1)^i \bar{w}_2\right), & \text{if } x_2 = 0, \end{cases} \tag{17}$$

for $i = 1$, 2, with $\bar{w}_2$ being any vector in $\mathbb{R}^{n-1}$ satisfying $\|\bar{w}_2\| = 1$. The spectral factorization of x, x^2 and $x^{1/2}$ as well as the matrix L_x have various interesting properties (cf. [10]). We list some properties that we will use later.

Property 2.1 *For any $x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$ with spectral values $\lambda_1(x)$, $\lambda_2(x)$ and spectral vectors $u_x^{(1)}, u_x^{(2)}$, the following results hold.*

(a) $x^2 = \lambda_1^2(x)u_x^{(1)} + \lambda_2^2(x)u_x^{(2)} \in \mathcal{K}^n$.
(b) *If $x \in \mathcal{K}^n$, then $0 \le \lambda_1(x) \le \lambda_2(x)$ and $x^{1/2} = \sqrt{\lambda_1(x)}\, u_x^{(1)} + \sqrt{\lambda_2(x)}\, u_x^{(2)}$.*
(c) *If $x \in \operatorname{int}(\mathcal{K}^n$, then $0 < \lambda_1(x) \le \lambda_2(x)$, and L_x is invertible with*

$$L_x^{-1} = \frac{1}{\det(x)} \begin{bmatrix} x_1 & -x_2^{\mathrm{T}} \\ -x_2 & \dfrac{\det(x)}{x_1} I + \dfrac{1}{x_1} x_2 x_2^T \end{bmatrix}.$$

(d) *The determinant, the trace and the Euclidean norm of x can be denoted by $\lambda_1(x)$, $\lambda_2(x)$:*

$$\det(x) = \lambda_1(x)\lambda_2(x), \quad \operatorname{tr}(x) = \lambda_1(x) + \lambda_2(x), \quad \|x\|^2 = \frac{\lambda_1^2(x) + \lambda_2^2(x)}{2}.$$

Before giving out several technical lemmas that will be applied in the next section, we introduce some notations that will be frequently used in the subsequent analysis. Unless otherwise stated, in this article, we always write

$$w = w(x,y) := (x-y)^2 + \tau(x \circ y) \quad \text{and} \quad z = z(x,y) := [(x-y)^2 + \tau(x \circ y)]^{1/2}. \tag{18}$$

Since $(x - y)^2 + \tau(x \circ y) = x^2 + y^2 + (\tau - 2)\ (x \circ y) \in \mathcal{K}^n$ for any x, $y \in \mathbb{R}^n$, we have

$$w := \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} \|x\|^2 + \|y\|^2 + (\tau-2)x^{\mathrm{T}}y \\ 2(x_1x_2 + y_1y_2) + (\tau-2)(x_1y_2 + y_1x_2) \end{pmatrix} \in \mathcal{K}^n. \tag{19}$$

From this, it follows that the spectral values of w are given by

$$\begin{aligned} \lambda_1(w) &:= \|x\|^2 + \|y\|^2 + (\tau-2)x^{\mathrm{T}}y - \|2(x_1x_2 + y_1y_2) + (\tau-2)(x_1y_2 + y_1x_2)\|, \\ \lambda_2(w) &:= \|x\|^2 + \|y\|^2 + (\tau-2)x^{\mathrm{T}}y + \|2(x_1x_2 + y_1y_2) + (\tau-2)(x_1y_2 + y_1x_2)\|. \end{aligned} \tag{20}$$

By Property 2.1(b), the vector z has the spectral values $\sqrt{\lambda_1(w)}, \sqrt{\lambda_2(w)}$ and

$$z := (z_1, z_2) = \left(\frac{\sqrt{\lambda_1(w)} + \sqrt{\lambda_2(w)}}{2}, \frac{\sqrt{\lambda_2(w)} - \sqrt{\lambda_1(w)}}{2} \bar{w}_2 \right), \tag{21}$$

where $\bar{w}_2 := (w_2/\|w_2\|)$ if $w_2 \neq 0$ and otherwise $\bar{w}_2$ is any vector in $\mathbb{R}^{n-1}$ satisfying $\|\bar{w}_2\| = 1$.

The following four technical lemmas are crucial in proving our main results. Lemma 2.1 measures how close w comes to the boundary of $\mathcal{K}^n$, and Lemma 2.2 describes the behaviour of (x, y) when w lies on the boundary of $\mathcal{K}^n$. Lemma 2.3 talks about the differential rule for the Jordan product function. Lemma 2.4 gives the gradient of the function $z(x, y)$.

LEMMA 2.1 [4, Lemma 3.4] *For any* $x=(x_1,x_2),\ y=(y_1,y_2)\in \mathbb{R}\times\mathbb{R}^{n-1}$ *and* $\tau\in(0,4)$, *if* $w_2=2(x_1x_2+y_1y_2)+(\tau-2)(x_1y_2+y_1x_2)\neq 0$, *then we have*

$$\begin{aligned}
&\left[\left(x_1+\frac{\tau-2}{2}y_1\right)+(-1)^i\left(x_2+\frac{\tau-2}{2}y_2\right)^{\mathrm{T}}\frac{w_2}{\|w_2\|}\right]^2\\
&\le\left\|\left(x_2+\frac{\tau-2}{2}y_2\right)+(-1)^i\left(x_1+\frac{\tau-2}{2}y_1\right)\frac{w_2}{\|w_2\|}\right\|^2\\
&\le\|x\|^2+\|y\|^2+(\tau-2)\langle x,y\rangle+(-1)^i\|w_2\|\\
&\le\lambda_i(w)
\end{aligned}$$

for $i=1,\,2$, *and furthermore these relations also hold when interchanging* x *and* y.

LEMMA 2.2 [5, Lemma 3.2] *For any* $x=(x_1,x_2),\ y=(y_1,y_2)\in\mathbb{R}\times\mathbb{R}^{n-1}$ *and* $\tau\in(0,4)$, *if* $w=(x-y)^2+\tau(x\circ y)\notin\operatorname{int}(\mathcal{K}^n)$, *then there always holds that*

$$\begin{aligned}
x_1^2=\|x_2\|^2,\ y_1^2&=\|y_2\|^2,\ x_1y_1=x_2^Ty_2,\ x_1y_2=y_1x_2;\\
x_1^2+y_1^2+(\tau-2)x_1y_1&=\|x_1x_2+y_1y_2+(\tau-2)x_1y_2\|\\
&=\|x_2\|^2+\|y_2\|^2+(\tau-2)x_2^Ty_2.
\end{aligned}$$

If, in addition, $(x,y)\neq(0,0)$, *then* $w_2=2(x_1x_2+y_1y_2+(\tau-2)x_1y_2)\neq 0$, *and furthermore,*

$$x_2^T\frac{w_2}{\|w_2\|}=x_1,\quad x_1\frac{w_2}{\|w_2\|}=x_2,\quad y_2^T\frac{w_2}{\|w_2\|}=y_1,\quad y_1\frac{w_2}{\|w_2\|}=y_2.$$

LEMMA 2.3 [6, Lemma 3.1] *Let* $\omega:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^m$ *be given by* $\omega(x,y):=u(x,y)\circ v(x,y)$, *where* $u,\,v:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^m$ *are differentiable mappings. Then,* ω *is differentiable and*

$$\begin{aligned}
\nabla_x\omega(x,y)&=\nabla_xu(x,y)L_{v(x,y)}+\nabla_xv(x,y)L_{u(x,y)},\\
\nabla_y\omega(x,y)&=\nabla_yu(x,y)L_{v(x,y)}+\nabla_yv(x,y)L_{u(x,y)}.
\end{aligned}$$

In particular, when $\omega(x,y)=x\circ y$, *there holds*

$$\nabla_x\omega(x,y)=L_y,\quad \nabla_y\omega(x,y)=L_x;$$

and when $\omega(x,y)=x^2\circ y^2$, *there holds*

$$\nabla_x\omega(x,y)=2L_xL_{y^2},\quad \nabla_y\omega(x,y)=2L_yL_{x^2}.$$

LEMMA 2.4 *For any* $x,\ y\in\mathbb{R}^n$ *and* $\tau\in(0,4)$, *let* $z(x,y)$ *be defined as in* (18). *Then the function* $z(x,y)$ *is continuously differentiable at a point* (x,y) *satisfying* $(x-y)^2+\tau(x\circ y)\in\operatorname{int}(\mathcal{K}^n)$. *Moreover, we have that*

$$\begin{aligned}
\nabla_xz(x,y)&=\left(L_x+\frac{\tau-2}{2}L_y\right)L^{-1}_{z(x,y)},\\
\nabla_yz(x,y)&=\left(L_y+\frac{\tau-2}{2}L_x\right)L^{-1}_{z(x,y)}.
\end{aligned}$$

Proof The differentiability of $z(x,y)$ is an immediate consequence of [13], see also [3, Prop. 4]. Since $z^2(x,y)=(x-y)^2+\tau(x\circ y)$, applying Lemma 2.3 yields

$$2\nabla_xz(x,y)L_{z(x,y)}=2L_{x-y}+\tau L_y=2L_x+(\tau-2)L_y.$$

Hence, $\nabla_x z(x,y) = (L_x + (\tau - 2/2)L_y)L^{-1}_{z(x,y)}$. In view of the symmetry of x and y in the $z(x,y)$, there also holds that $\nabla_y z(x,y) = (L_y + (\tau - 2/2)L_x)L^{-1}_{z(x,y)}$. $\square$

3. Main results

In this section, we present the proof showing that the gradient function of ψ_τ is Lipschitz continuous. By the notation in Section 2, the function ψ_τ can be rewritten as

$$\begin{aligned}
(x,y) &= \frac{1}{2}\left\|[(x-y)^2 + \tau(x\circ y)]^{1/2} - (x+y)\right\|^2 \\
&= \frac{1}{2}\left\|z(x,y)\right\|^2 - (x+y)^T[(x-y)^2 + \tau(x\circ y)]^{1/2} + \frac{1}{2}\|x+y\|^2 \\
&= \frac{1}{2}\left[\frac{\lambda_2(w)+\lambda_1(w)}{2} + \|x+y\|^2\right] - (x+y)^T[(x-y)^2 + \tau(x\circ y)]^{1/2} \\
&= \frac{1}{2}\left[2\|x\|^2 + 2\|y\|^2 + \tau(x\circ y)\right] - (x+y)^T[(x-y)^2 + \tau(x\circ y)]^{1/2} \\
&= \|x\|^2 + \|y\|^2 + \frac{\tau}{2}(x\circ y) - (x+y)^T[(x-y)^2 + \tau(x\circ y)]^{1/2},
\end{aligned} \tag{22}$$

where the third equality is due to Property 2.1(d). Clearly, the gradient of the function $\|x\|^2 + \|y\|^2 + (\tau/2)(x\circ y)$ is globally Lipschitz continuous. Therefore, to show that the gradient of ψ_τ is globally Lipschitz continuous, we only need to show that following function

$$F(x,y) := (x+y)^T[(x-y)^2 + \tau(x\circ y)]^{1/2} \tag{23}$$

has a Lipschitz continuous gradient. The following lemma states the gradient of $F(x,y)$.

Lemma 3.1 *For any x, $y \in \mathbb{R}^n$ and $\tau \in (0,4)$, let $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be defined as in* (23). *Then, the function $F(x,y)$ is continuously differentiable everywhere. Moreover, $\nabla_x F(0,0) = \nabla_y F(0,0) = 0$. If $(x,y) \neq (0,0)$ and $(x-y)^2 + \tau(x\circ y) \in \mathrm{int}(\mathcal{K}^n)$, then*

$$\begin{aligned}
\nabla_x F(x,y) &= z(x,y) + \left(L_x + \frac{\tau-2}{2}L_y\right)L^{-1}_{z(x,y)}(x+y) \\
\nabla_y F(x,y) &= z(x,y) + \left(L_y + \frac{\tau-2}{2}L_x\right)L^{-1}_{z(x,y)}(x+y).
\end{aligned} \tag{24}$$

If $(x,y) \neq (0,0)$ and $(x-y)^2 + \tau(x\circ y) \notin \mathrm{int}(\mathcal{K}^n)$, then

$$\begin{aligned}
\nabla_x F(x,y) &= z(x,y) + \frac{x_1 + \frac{\tau-2}{2}y_1}{\sqrt{x_1^2 + y_1^2 + (\tau-2)x_1y_1}}(x+y) \\
\nabla_y F(x,y) &= z(x,y) + \frac{y_1 + \frac{\tau-2}{2}x_1}{\sqrt{x_1^2 + y_1^2 + (\tau-2)x_1y_1}}(x+y).
\end{aligned} \tag{25}$$

Proof *Case 1* $x = y = 0$. For any $h, k \in \mathbb{R}^n$, let $\mu_1 \le \mu_2$ be the spectral values and $v^{(1)}$ and $v^{(2)}$ be the corresponding vectors of $(h-k)^2 + \tau(h\circ k)$. Then, by Property 2.1(b),

$$\|[(h-k)^2 + \tau(h\circ k)]^{1/2}\| = \|\sqrt{\mu_1}v^{(1)} + \sqrt{\mu_2}v^{(2)}\| \le (\sqrt{\mu_1} + \sqrt{\mu_2})/\sqrt{2} \le \sqrt{2\mu_2}.$$

Therefore,

$$\begin{aligned} F(h,k) - F(0,0) &= (h+k)^{\mathrm{T}}[(h-k)^2 + \tau(h \circ k)]^{1/2} \\ &\le \|[(h-k)^2 + \tau(h \circ k)]^{1/2}\| \|h+k\| \\ &\le \sqrt{2\mu_2} \cdot (\|h\| + \|k\|) \\ &\le \sqrt{2(\|h\|^2 + \|k\|^2 + (\tau-2)h^{\mathrm{T}}k)} \cdot (\|h\| + \|k\|) \\ &= O(\|h\|^2 + \|k\|^2). \end{aligned}$$

This shows that $F(x,y)$ is differentiable at $(0,0)$ with $\nabla_x F(0,0) = \nabla_y F(0,0) = 0$.

Case 2 $(x,y) \neq (0,0)$ and $(x-y)^2 + \tau(x \circ y) \in \mathrm{int}(\mathcal{K}^n)$. Using Lemma 2.4, we readily obtain the formula in (24). Clearly, in this case, $\nabla_x F(x,y)$ and $\nabla_y F(x,y)$ are continuous, and consequently, $F(x,y)$ is continuously differentiable at such points.

Case 3 $(x,y) \neq (0,0)$ and $(x-y)^2 + \tau(x \circ y) \notin \mathrm{int}(\mathcal{K}^n)$. Since $\lambda_1(w) = 0$, now it follows from (18) and (21) that

$$z(x,y) = [(x-y)^2 + \tau(x \circ y)]^{1/2} = \frac{1}{2}\left(\sqrt{\lambda_2(w)}, \sqrt{\lambda_2(w)}\bar{w}_2\right)^{\mathrm{T}},$$

Moreover, by Lemma 2.2, we can compute that

$$w_2 = 2(x_1x_2 + y_1y_2 + (\tau-2)x_1y_2), \quad \lambda_2(w) = 4(x_1^2 + y_1^2 + (\tau-2)x_1y_1) = 2\|w_2\|.$$

Therefore, under this case, we have that

$$z(x,y) = [(x-y)^2 + \tau(x \circ y)]^{1/2} = \begin{bmatrix} \sqrt{x_1^2 + y_1^2 + (\tau-2)x_1y_1} \\ \dfrac{x_1x_2 + y_1y_2 + (\tau-2)x_1y_2}{\sqrt{x_1^2 + y_1^2 + (\tau-2)x_1y_1}} \end{bmatrix}.$$

By this, it is easy to verify that the formula (25) holds.

Note that $z(x,y)$ is a continuous function. This together with the proof of Case (i) and Case (iii) of [4, Proposition 3.3] means that the gradient functions $\nabla_x F(x,y)$ and $\nabla_y F(x,y)$ are continuous at every $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$. Hence, $F(x,y)$ is continuously differentiable everywhere. Thus, we complete the proof. □

For the symmetry of x and y in $\nabla_x F(x,y)$ and $\nabla_y F(x,y)$, in the rest of this section, we concentrate on the proof of globally Lipschitz continuity of $\nabla_x F(x,y)$. We first define a smooth approximation of $\nabla_x F(x,y)$. For any $\epsilon > 0$, we let

$$\begin{aligned} \hat{w} = \hat{w}(x,y,\epsilon) &:= (x-y)^2 + \tau(x \circ y) + \epsilon e, \\ \hat{z} = \hat{z}(x,y,\epsilon) &:= [(x-y)^2 + \tau(x \circ y) + \epsilon e]^{1/2}, \end{aligned} \tag{26}$$

where e is the identity element under the Jordan product. It is not hard to see that

$$\hat{w}_1 = w_1 + \epsilon, \quad \hat{w}_2 = w_2, \quad \lambda_1(\hat{w}) = \lambda_1(w) + \epsilon, \quad \lambda_2(\hat{w}) = \lambda_2(w) + \epsilon, \tag{27}$$

and furthermore,

$$\hat{z} = (\hat{z}_1, \hat{z}_2) = \frac{1}{2}\left(\sqrt{\lambda_2(\hat{w})} + \sqrt{\lambda_1(\hat{w})}, \left(\sqrt{\lambda_2(\hat{w})} - \sqrt{\lambda_1(\hat{w})}\right)\bar{w}_2\right) \tag{28}$$

where $\bar{w}_2 := (w_2/\|w_2\|)$ if $w_2 \neq 0$ and otherwise $\bar{w}_2$ is any vector in $\mathbb{R}^{n-1}$ satisfying $\|\bar{w}_2\| = 1$. We define the mapping $G(\cdot, \cdot, \epsilon): \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$G(x, y, \epsilon) = \hat{z}(x, y, \epsilon) + \left(L_x + \frac{\tau - 2}{2} L_y\right) L^{-1}_{\hat{z}(x,y,\epsilon)}(x + y). \tag{29}$$

By Lemmas 3.2 and 3.4 below, $G(x, y, \epsilon)$ is actually a smooth approximation of $\nabla_x F(x, y)$. Based on the relation, in the sequel, we will prove the Lipschitz continuity of $\nabla_x F(x, y)$ through arguing that $G(x, y, \epsilon)$ is globally Lipschitz continuous.

Lemma 3.2 *For any $x, y \in \mathbb{R}^n$ and $\epsilon > 0$, let $G(x, y, \epsilon)$ be defined as in* (29). *Then,*

$$\lim_{\epsilon \to 0^+} G(x, y, \epsilon) = \nabla_x F(x, y).$$

Proof If $(x, y) = (0, 0)$, then $G(x, y, \epsilon) = (\epsilon e)^{1/2}$ for any $\epsilon > 0$. Therefore,

$$\lim_{\epsilon \to 0^+} G(0, 0, \epsilon) = \nabla_x F(0, 0) = 0.$$

If $(x, y) \neq (0, 0)$ and $(x - y)^2 + \tau(x \circ y) \in \operatorname{int}(\mathcal{K}^n)$, then by (27), (28) and Property 2.1(c), it is easy to verify that $\lim_{\epsilon \to 0^+} L^{-1}_{\hat{z}(x,y,\epsilon)} = L^{-1}_{z(x,y)}$. This together with $\lim_{\epsilon \to 0^+} L_{\hat{z}(x,y,\epsilon)} = L_{z(x,y)}$ implies that the conclusion holds.

Next, we consider $(x, y) \neq (0, 0)$ and $(x - y)^2 + \tau(x \circ y) \notin \operatorname{int}(\mathcal{K}^n)$. For convenience, let

$$g := x + \frac{\tau - 2}{2} y, \;\; h := y + \frac{\tau - 2}{2} x, \;\; u(x, y, \epsilon) := L^{-1}_{\hat{z}(x,y,\epsilon)} g, \;\; v(x, y, \epsilon) := L^{-1}_{\hat{z}(x,y,\epsilon)} h. \tag{30}$$

By Property 2.1(c), it is easy to compute that

$$\begin{aligned} u = u(x, y, \epsilon) &= \frac{1}{\det(\hat{z})} \begin{bmatrix} \hat{z}_1 & -\hat{z}_2 \\ -\hat{z}_2 & \dfrac{\det(\hat{z})}{\hat{z}_1} I + \dfrac{1}{\hat{z}_1} \hat{z}_2 \hat{z}_2 \end{bmatrix} [g_1 g_2] \\ &= \frac{1}{\det(\hat{z})} \begin{bmatrix} g_1 \hat{z}_1 - g_2 \hat{z}_2 \\ -g_1 \hat{z}_2 + \dfrac{\det(\hat{z})}{\hat{z}_1} g_2 + \dfrac{g_2 \hat{z}_2}{\hat{z}_1} \hat{z}_2 \end{bmatrix} \\ &:= \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \end{aligned} \tag{31}$$

$$\begin{aligned} v = v(x, y, \epsilon) &= \frac{1}{\det(\hat{z})} \begin{bmatrix} \hat{z}_1 & -\hat{z}_2^{\mathrm{T}} \\ -\hat{z}_2 & \dfrac{\det(\hat{z})}{\hat{z}_1} I + \dfrac{1}{\hat{z}_1} \hat{z}_2 \hat{z}_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\ &= \frac{1}{\det(\hat{z})} \begin{bmatrix} h_1 \hat{z}_1 - h_2^{\mathrm{T}} \hat{z}_2 \\ -h_1 \hat{z}_2 + \dfrac{\det(\hat{z})}{\hat{z}_1} h_2 + \dfrac{h_2^{\mathrm{T}} \hat{z}_2}{\hat{z}_1} \hat{z}_2 \end{bmatrix} \\ &:= \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \end{aligned} \tag{32}$$

Using (28) and Lemma 2.2, we have that

$$
\begin{aligned}
u_1 &= \frac{1}{\sqrt{\lambda_1(\hat{w})\lambda_2(\hat{w})}}\left[\frac{\sqrt{\lambda_1(\hat{w})}+\sqrt{\lambda_2(\hat{w})}}{2}g_1 + \frac{\sqrt{\lambda_1(\hat{w})}-\sqrt{\lambda_2(\hat{w})}}{2}g_2^{\mathrm{T}}\bar{w}_2\right] \\
&= \frac{1}{2\sqrt{\lambda_2(\hat{w})}}\Big(g_1 + g_2^{\mathrm{T}}\bar{w}_2\Big) + \frac{1}{2\sqrt{\lambda_1(\hat{w})}}\Big(g_1 - g_2^{\mathrm{T}}\bar{w}_2\Big) \\
&= \frac{g_1}{\sqrt{\lambda_2(w)+\epsilon}}, \quad (\text{since} \quad g_2^{\mathrm{T}}\bar{w}_2 = g_1)
\end{aligned} \tag{33}
$$

$$
\begin{aligned}
u_2 &= \frac{1}{\sqrt{\lambda_1(\hat{w})\lambda_2(\hat{w})}}\left[\frac{\sqrt{\lambda_1(\hat{w})}-\sqrt{\lambda_2(\hat{w})}}{2}g_1\bar{w}_2 + \frac{2\sqrt{\lambda_1(\hat{w})}\sqrt{\lambda_2(\hat{w})}}{\sqrt{\lambda_1(\hat{w})}+\sqrt{\lambda_2(\hat{w})}}g_2\right. \\
&\qquad\left. + \frac{\Big(\sqrt{\lambda_1(\hat{w})}-\sqrt{\lambda_2(\hat{w})}\Big)^2}{2\Big(\sqrt{\lambda_1(\hat{w})}+\sqrt{\lambda_2(\hat{w})}\Big)}g_2^{\mathrm{T}}\bar{w}_2\bar{w}_2\right] \\
&= \frac{(g_1 + g_2^{\mathrm{T}}\bar{w}_2)\bar{w}_2}{2\sqrt{\lambda_2(\hat{w})}} - \frac{(g_1 - g_2^{\mathrm{T}}\bar{w}_2)\bar{w}_2}{2\sqrt{\lambda_1(\hat{w})}} + \frac{2g_2 - 2g_2^{\mathrm{T}}\bar{w}_2\bar{w}_2}{\sqrt{\lambda_1(\hat{w})}+\sqrt{\lambda_2(\hat{w})}} \\
&= \frac{g_2}{\sqrt{\lambda_2(w)+\epsilon}}. \quad (\text{since } g_1\bar{w}_2 = g_2,\ g_2^{\mathrm{T}}\bar{w}_2 = g_1)
\end{aligned} \tag{34}
$$

Similarly, we can also obtain that

$$
v_1 = \frac{h_1}{\sqrt{\lambda_2(w)+\epsilon}}, \quad v_2 = \frac{h_2}{\sqrt{\lambda_2(w)+\epsilon}}.
$$

From the above expressions of u and v, it then follows that

$$
\begin{aligned}
\lim_{\epsilon\to 0^+} L^{-1}_{\hat{z}(x,y,\epsilon)}(x+y) &= \frac{2}{\tau}\lim_{\epsilon\to 0^+} L^{-1}_{\hat{z}(x,y,\epsilon)}\Big(x + \frac{\tau-2}{2}y + y + \frac{\tau-2}{2}x\Big) \\
&= \frac{2}{\tau}\lim_{\epsilon\to 0^+}\Big(u(x,y,\epsilon) + v(x,y,\epsilon)\Big) \\
&= \frac{1}{\sqrt{\lambda_2(w)}}\Big(x_1 + y_1, x_2 + y_2\Big)^{\mathrm{T}}.
\end{aligned}
$$

This together with Lemma 2.2 yields that

$$
\begin{aligned}
\lim_{\epsilon\to 0^+} G(x,y,\epsilon) &= \lim_{\epsilon\to 0^+}\hat{z}(x,y,\epsilon) + \Big(L_x + \frac{\tau-2}{2}L_y\Big)\lim_{\epsilon\to 0^+} L^{-1}_{\hat{z}(x,y,\epsilon)}(x+y) \\
&= z(x,y) + \begin{bmatrix} x_1 + \frac{\tau-2}{2}y_1 & x_2^T + \frac{\tau-2}{2}y_2^{\mathrm{T}} \\ x_2 + \frac{\tau-2}{2}y_2 & \Big(x_1 + \frac{\tau-2}{2}y_1\Big)I \end{bmatrix}\begin{pmatrix} \frac{x_1+y_1}{\sqrt{\lambda_2(w)}} \\ \frac{x_2+y_2}{\sqrt{\lambda_2(w)}} \end{pmatrix} \\
&= z(x,y) + \frac{1}{\sqrt{\lambda_2(w)}}\begin{pmatrix} 2\Big(x_1 + \frac{\tau-2}{2}y_1\Big)(x_1+y_1) \\ 2\Big(x_2 + \frac{\tau-2}{2}y_2\Big)(x_1+y_1) \end{pmatrix}
\end{aligned}
$$

$$= z(x,y) + \frac{1}{\sqrt{\lambda_2(w)}}\begin{pmatrix} 2\left(x_1 + \frac{\tau-2}{2}y_1\right)(x_1+y_1) \\ 2\left(x_1 + \frac{\tau-2}{2}y_1\right)(x_2+y_2) \end{pmatrix}$$
$$= z(x,y) + \frac{x_1 + \frac{\tau-2}{2}y_1}{\sqrt{x_1^2+y_1^2+\tau(x_1y_1)}}(x+y).$$

Thus, we complete the proof. □

In what follows, we argue that the gradient function of $G(x,y,\epsilon)$ is uniformly bounded, and then, by applying the mean-value theorem for vector-valued functions, we conclude that $G(x,y,\epsilon)$ is globally Lipschitz continuous. The following two lemmas are crucial in proving our main result.

Lemma 3.3 *For any x, $y \in \mathbb{R}^n$ and $\epsilon > 0$, let $\hat{z} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ be given as in* (26). *Then, the function $\hat{z}(x,y,\epsilon)$ is continuously differentiable everywhere. Moreover, there exists a constant C independent of x, y and ϵ, τ such that*

$$\|\nabla_x\hat{z}(x,y,\epsilon)\| = \left\|\left(L_x + \frac{\tau-2}{2}L_y\right)L^{-1}_{\hat{z}(x,y,\epsilon)}\right\| \le C,$$
$$\|\nabla_y\hat{z}(x,y,\epsilon)\| = \left\|\left(L_y + \frac{\tau-2}{2}L_x\right)L^{-1}_{\hat{z}(x,y,\epsilon)}\right\| \le C.$$

Proof The first part of the conclusion follows directly from [10, Proposition 5.2]. The second part is implied by the proof of [15, Proposition 3.1]. □

Lemma 3.4 *For any x, $y \in \mathbb{R}^n$ and $\epsilon > 0$, let $G : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ be defined as in* (29). *Then, the function $G(x,y,\epsilon)$ is continuously differentiable everywhere. Moreover, there exists a constant C such that $\|\nabla_x G(x,y,\epsilon)\| \le C(1+\tau^-1)$ and $\|\nabla_y G(x,y,\epsilon)\| \le C(1+\tau^{-1})$.*

Proof The first part of the conclusion is due to [10, Proposition 5.2]. For the second part, by Lemma 3.3, it suffices to prove that the gradient of the following function:

$$H(x,y,\epsilon) := \left(L_x + \frac{\tau-2}{2}L_y\right)L^{-1}_{\hat{z}(x,y,\epsilon)}(x+y)$$

is uniformly bounded. From the definition of $H(x,y,\epsilon)$, we notice that

$$H(x,y,\epsilon) = \frac{2}{\tau}\left(L_x + \frac{\tau-2}{2}L_y\right)L^{-1}_{\hat{z}(x,y,\epsilon)}\left(x + \frac{\tau-2}{2}y + y + \frac{\tau-2}{2}x\right)$$
$$= \frac{2}{\tau}\left[x\circ(u+v) + \frac{\tau-2}{2}y\circ(u+v)\right]$$

where u and v are defined as in (30). Therefore, applying Lemma 2.3 yields that

$$\begin{aligned}\nabla_x H(x,y,\epsilon) &= \frac{2}{\tau}\left[L_{u+v} + \left(\nabla_x u(x,y,\epsilon) + \nabla_x v(x,y,\epsilon)\right)\left(L_x + \frac{\tau-2}{2}L_y\right)\right],\\ \nabla_y H(x,y,\epsilon) &= \frac{2}{\tau}\left[\frac{\tau-2}{2}L_{u+v} + \left(\nabla_y u(x,y,\epsilon) + \nabla_y v(x,y,\epsilon)\right)\left(L_x + \frac{\tau-2}{2}L_y\right)\right].\end{aligned} \tag{35}$$

To show that $\|\nabla_x H(x,y,\epsilon)\|$ is uniformly bounded, we shall verify that both $\|L_{u+v}\|$ and $\|(\nabla_x u(x,y,\epsilon)+\nabla_x v(x,y,\epsilon))(L_x+(\tau-2/2)L_y)\|$ are uniformly bounded.

(i) To prove that $\|L_{u+v}\|$ is uniformly bounded, it is sufficient to argue that $|u_1|$, $\|u_2\|$ and $|v_1|$, $\|u_2\|$ are both uniformly bounded. First, we argue that $|u_1|$ and $|v_1|$ are uniformly bounded. From (33), we have that

$$u_1 = \frac{1}{2\sqrt{\lambda_2(\hat{w})}}\left(g_1 + g_2^{\mathrm{T}}\bar{w}_2\right) + \frac{1}{2\sqrt{\lambda_1(\hat{w})}}\left(g_1 - g_2^{\mathrm{T}}\bar{w}_2\right),$$
$$v_1 = \frac{1}{2\sqrt{\lambda_2(\hat{w})}}\left(h_1 + h_2^{\mathrm{T}}\bar{w}_2\right) + \frac{1}{2\sqrt{\lambda_1(\hat{w})}}\left(h_1 - h_2^{\mathrm{T}}\bar{w}_2\right).$$

Note that $g_1 = x_1 + (\tau-2/2)y_1, g_2 = x_2 + (\tau-2/2)y_2$ and $h_1 = y_1 + (\tau-2/2)x_1$, $h_2 = y_2 + (\tau-2/2)x_2$, and $\bar{w}_2 = (w_2/\|w_2\|)$. Therefore, applying Lemma 2.1 yields that

$$\left|g_1 - g_2^{\mathrm{T}}\bar{w}_2\right| \le \sqrt{\lambda_1(w)} \le \sqrt{\lambda_1(\hat{w})}, \quad |g_1 + g_2^{\mathrm{T}}\bar{w}_2| \le \sqrt{\lambda_2(w)} \le \sqrt{\lambda_1(\hat{w})} \tag{36}$$

and

$$\left|h_1 - h_2^{\mathrm{T}}\bar{w}_2\right| \le \sqrt{\lambda_1(w)} \le \sqrt{\lambda_1(\hat{w})}, \quad |h_1 + h_2^{\mathrm{T}}\bar{w}_2| \le \sqrt{\lambda_2(w)} \le \sqrt{\lambda_1(\hat{w})}. \tag{37}$$

Combing with the expressions of u_1 and v_1 given as above, we get $|u_1| \le 1$ and $|v_1| \le 1$.

Second, we argue that $\|u_2\|$ and $\|v_2\|$ are also uniformly bounded. By (34),

$$u_2 = \frac{\left(g_1 + g_2^{\mathrm{T}}\bar{w}_2\right)\bar{w}_2}{2\sqrt{\lambda_2(\hat{w})}} - \frac{\left(g_1 - g_2^{\mathrm{T}}\bar{w}_2\right)\bar{w}_2}{2\sqrt{\lambda_1(\hat{w})}} + \frac{2g_2 - 2g_2^{\mathrm{T}}\bar{w}_2\bar{w}_2}{\sqrt{\lambda_1(\hat{w})} + \sqrt{\lambda_2(\hat{w})}},$$
$$v_2 = \frac{\left(h_1 + h_2^{\mathrm{T}}\bar{w}_2\right)\bar{w}_2}{2\sqrt{\lambda_2(\hat{w})}} - \frac{\left(h_1 - h_2^{\mathrm{T}}\bar{w}_2\right)\bar{w}_2}{2\sqrt{\lambda_1(\hat{w})}} + \frac{2h_2 - 2h_2^{\mathrm{T}}\bar{w}_2\bar{w}_2}{\sqrt{\lambda_1(\hat{w})} + \sqrt{\lambda_2(\hat{w})}}.$$

Using (36) and (37) and the fact that $\|\bar{w}_2\| = 1$, we obtain that

$$\left\| \frac{\left(g_1 + g_2^{\mathrm{T}}\bar{w}_2\right)\bar{w}_2}{2\sqrt{\lambda_2(\hat{w})}} - \frac{\left(g_1 - g_2^{\mathrm{T}}\bar{w}_2\right)\bar{w}_2}{2\sqrt{\lambda_1(\hat{w})}} \right\| \le \frac{1}{2}\|\bar{w}_2\| + \frac{1}{2}\|\bar{w}_2\| = 1,$$
$$\left\| \frac{\left(h_1 + h_2^{\mathrm{T}}\bar{w}_2\right)\bar{w}_2}{2\sqrt{\lambda_2(\hat{w})}} - \frac{\left(h_1 - h_2^{\mathrm{T}}\bar{w}_2\right)\bar{w}_2}{2\sqrt{\lambda_1(\hat{w})}} \right\| \le \frac{1}{2}\|\bar{w}_2\| + \frac{1}{2}\|\bar{w}_2\| = 1,$$

and

$$\left\| \frac{2g_2 - 2g_2^{\mathrm{T}}\bar{w}_2\bar{w}_2^{\mathrm{T}}}{\sqrt{\lambda_1(\hat{w})} + \sqrt{\lambda_2(\hat{w})}} \right\| \le \frac{4\|g_2\|}{\sqrt{\lambda_2(\hat{w})}} \le \frac{4\|g_2\|}{\sqrt{\|x\|^2 + \|y\|^2 + (\tau-2)x^{\mathrm{T}}y + \epsilon}} \le 4,$$
$$\left\| \frac{2h_2 - 2h_2^{\mathrm{T}}\bar{w}_2\bar{w}_2^{\mathrm{T}}}{\sqrt{\lambda_1(\hat{w})} + \sqrt{\lambda_2(\hat{w})}} \right\| \le \frac{4\|h_2\|}{\sqrt{\lambda_2(\hat{w})}} \le \frac{4\|h_2\|}{\sqrt{\|x\|^2 + \|y\|^2 + (\tau-2)x^{\mathrm{T}}y + \epsilon}} \le 4.$$

The above inequalities imply that $\|u_2\|$ and $\|v_2\|$ are uniformly bounded. This together with the uniform boundedness of $|u_1|$ and $|v_1|$ implies that

$$\|L_{u+v}\| = \left\| \begin{bmatrix} u_1 + v_1 & (u_2 + v_2)^{\mathrm{T}} \\ u_2 + v_2 & (u_1 + v_1)I \end{bmatrix} \right\|$$

is also uniformly bounded.

(ii) Now, it comes to show that $\|(\nabla_x u(x,y,\epsilon) + \nabla_x v(x,y,\epsilon))(L_x + (\tau - 2/2)L_y)\|$ is uniformly bounded. From the definition of $u(x,y,\epsilon)$ and $v(x,y,\epsilon)$ as given in (30), we have

$$\hat{z}(x,y,\epsilon) \circ (u(x,y,\epsilon) + v(x,y,\epsilon)) = \frac{\tau(x+y)}{2}.$$

Applying Lemma 2.3 then gives that

$$\nabla_x \hat{z}(x,y,\epsilon) L_{u+v} + (\nabla_x u(x,y,\epsilon) + \nabla_x v(x,y,\epsilon)) L_{\hat{z}(x,y,\epsilon)} = (\tau/2)I.$$

This is equivalent to saying that

$$\begin{aligned}(\nabla_x u(x,y,\epsilon) + \nabla_x v(x,y)) L_{\hat{z}(x,y,\epsilon)} &= (\tau/2)I - \nabla_x \hat{z}(x,y,\epsilon) L_{u+v} \\ &= \frac{\tau}{2} I - \left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} L_{u+v},\end{aligned}$$

where the second equality is due to Lemma 2.4. Therefore,

$$\begin{aligned}&\Big(\nabla_x u(x,y,\epsilon) + \nabla_x v(x,y,\epsilon)\Big)\left(L_x + \frac{\tau-2}{2} L_y \right) \\ &= \left[\frac{\tau}{2} I - \left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} L_{u+v} \right] L^{-1}_{\hat{z}(x,y,\epsilon)} \left(L_x + \frac{\tau-2}{2} L_y \right) \\ &= \frac{\tau}{2} L^{-1}_{\hat{z}(x,y,\epsilon)} \left(L_x + \frac{\tau-2}{2} L_y \right) - \left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} L_{u+v} L^{-1}_{\hat{z}(x,y,\epsilon)} \left(L_x + \frac{\tau-2}{2} L_y \right) \\ &= \frac{\tau}{2} \left[\left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} \right]^{\mathrm{T}} \\ &\quad - \left[\left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} \right] L_{u+v} \left[\left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} \right]^{\mathrm{T}}.\end{aligned}$$

Now we have that

$$\begin{aligned}&\left\| \Big(\nabla_x u(x,y,\epsilon) + \nabla_x v(x,y,\epsilon)\Big)\left(L_x + \frac{\tau-2}{2} L_y \right) \right\| \\ &\le \frac{\tau}{2} \left\| \left[\left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} \right]^{\mathrm{T}} \right\| \\ &\quad + \left\| \left[\left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} \right] \right\| \cdot \|L_{u+v}\| \cdot \left\| \left[\left(L_x + \frac{\tau-2}{2} L_y \right) L^{-1}_{\hat{z}(x,y,\epsilon)} \right]^{\mathrm{T}} \right\|.\end{aligned}$$

From Lemma 3.3, $\|[(L_x + (\tau - 2/2)L_y) L^{-1}_{\hat{z}(x,y,\epsilon)}]^{\mathrm{T}}\|$ is uniformly bounded. This together with the uniform boundedness of $\|L_{u+v}\|$ yields that $\|(\nabla_x u(x,y,\epsilon) + \nabla_x v(x,y,\epsilon))(L_x + (\tau - 2/2)L_y)\|$ is uniformly bounded.

From (i), (ii) and (35), we conclude that $\|\nabla_x H(x,y,\epsilon)\|$ is uniformly bounded with the bound related to τ^{-1}. Using similar arguments, we can prove that $\|\nabla_y H(x,y,\epsilon)\|$ is uniformly bounded with the bound related to τ^{-1}. Combining with Lemma 3.3, we then show that there exists a constant C such that $\|\nabla_x G(x,y,\epsilon)\| \le C(1+\tau^{-1})$ and $\|\nabla_y G(x,y,\epsilon)\| \le C(1+\tau^{-1})$. □

Theorem 3.1 *For any x, $y \in \mathbb{R}^n$ and $\tau \in (0,4)$, let $F(x,y)$ be given as in (23). Then, their gradient functions $\nabla_x F(x,y)$ and $\nabla_y F(x,y)$ are globally Lipschitz continuous, i.e. there exists a constant C such that*

$$\begin{aligned} \|\nabla_x F(x,y) - \nabla_x F(a,b)\| &\le C(1+\tau^{-1})\|(x,y)-(a,b)\|, \\ \|\nabla_y F(x,y) - \nabla_y F(a,b)\| &\le C(1+\tau^{-1})\|(x,y)-(a,b)\| \end{aligned} \tag{38}$$

for all (x,y), $(a,b) \in \mathbb{R}^n \times \mathbb{R}^n$.

Proof We first prove that the function $G(x,y,\epsilon)$ defined by (29) is globally Lipschitz continuous for any $\epsilon > 0$. For any $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$ and $(a,b) \in \mathbb{R}^n \times \mathbb{R}^n$, we have that

$$G(x,y,\epsilon) - G(a,b,\epsilon) = G(x,y,\epsilon) - G(a,y,\epsilon) + G(a,y,\epsilon) - G(a,b,\epsilon).$$

From Lemma 3.4, we know that $G(x,y,\epsilon)$ is continuously differentiable everywhere. Hence, from the mean-value theorem, it follows that

$$\begin{aligned} G(x,y,\epsilon) - G(a,y,\epsilon) &= \int_0^1 \nabla_x G(a+t(x-a),y,\epsilon)(x-a)\mathrm{d}t, \\ G(a,y,\epsilon) - G(a,b,\epsilon) &= \int_0^1 \nabla_y G(a,b+t(y-b),\epsilon)(y-b)\mathrm{d}t \end{aligned}$$

Combining the last two equations and using Lemma 3.4, we then obtain that

$$\begin{aligned} \|G(x,y,\epsilon) - G(a,b,\epsilon)\| &\le \left\| \int_0^1 \nabla_x G(a+t(x-a),y,\epsilon)(x-a)\mathrm{d}t \right\| \\ &\quad + \left\| \int_0^1 \nabla_y G(a,b+t(y-b),\epsilon)(y-b)\mathrm{d}t \right\| \\ &\le \int_0^1 \|\nabla_x G(a+t(x-a),y,\epsilon)\|\,\|x-a\|\mathrm{d}t \\ &\quad + \int_0^1 \|\nabla_x G(a+t(x-a),y,\epsilon)\|\,\|(x-a)\|\mathrm{d}t \\ &\le C(1+\tau^{-1})\|(x,y)-(a,b)\|, \end{aligned} \tag{39}$$

where C is a constant independent of x, y and ϵ, τ. From Lemma 3.2, we know that

$$\lim_{\epsilon \to 0^+} G(x,y,\epsilon) = \nabla_x F(x,y)$$

for any x, $y \in \mathbb{R}^n \times \mathbb{R}^n$. This together with (39) immediately yields that

$$\begin{aligned} \|\nabla_x F(x,y) - \nabla_x F(a,b)\| &= \| \lim_{\epsilon \to 0^+} G(x,y,\epsilon) - \lim_{\epsilon \to 0^+} G(a,b,\epsilon)\| \\ &= \lim_{\epsilon \to 0^+} \|G(x,y,\epsilon) - G(a,b,\epsilon)\| \\ &\le C(1+\tau^{-1})\|(x,y)-(a,b)\|. \end{aligned}$$

Thus, we prove that $\nabla_x F(x, y)$ is globally Lipschitz continuous. Similarly, we may prove that $\nabla_y F(x, y)$ is also globally Lipchitz continuous. $\square$

From the above theorem, we immediately obtain the following corollary.

COROLLARY 3.1 *Let ψ_τ with $\tau \in (0, 4)$ be defined as in* (12). *Then ψ_τ is an LC^1 function*, *i.e. the gradient functions $\nabla_x \psi_\tau(x, y)$ and $\nabla_y \psi_\tau(x, y)$ are globally Liptchitz continuous with the Lipschitz constant being $O(1 + \tau^{-1})$.*

Note

1. Member of Mathematics Division, National Center for Theoretical Sciences, Taipei Office. The author's work is partially supported by National Science Council of Taiwan. E-mail: jschen@math.ntnu.edu.tw.

References

[1] J.-S. Chen, *A new merit function and its related properties for the second-order cone complementarity problem*, Pacific J. Optim. 2 (2006), pp. 167–179.
[2] ———, *Two classes of merit functions for the second-order cone complementarity problem*, Math. Methods Oper. Res. 64 (2006), pp. 495–519.
[3] J.-S. Chen, X. Chen, and P. Tseng, *Analysis of nonsmooth vector-valued functions associated with second-order cones*, Math. Progr. 101 (2004), pp. 95–117.
[4] J.-S. Chen and S.-H. Pan, *A one-parametric class of merit functions for the second-order cone complementarity problem*, Comput. Optimi. Appl. (2008), doi: 10.1007/s10589-008-9182-9.
[5] J.-S. Chen, D. Sun, and J. Sun, *The SC^1 property of the squared norm of the SOC Fischer-Burmeister function*, Oper. Res. Lett 36 (2008), pp. 385–392.
[6] J.-S. Chen and P. Tseng, *An unconstrained smooth minimization reformulation of the second-order cone complementarity problem*, Math. Progr. 104 (2005), pp. 293–327.
[7] U. Faraut and A. Korányi, *Analysis on Symmetric Cones*. Oxford Mathematical Monographs, Oxford University Press, New York, 1994.
[8] A. Fischer, *A special Newton-type optimization methods*, Optimization 24 (1992), pp. 269–284.
[9] ———, *An NCP-function and its use for the solution of complementarity problem,* in Recent Advances in Nonsmooth Optimization, D.-Z. Du, L.-Q. Qi, and R. Womersley, eds., World Scientific, Singapore, 1995, pp. 88–105.
[10] M. Fukushima, Z.-Q. Luo, and P. Tseng, *Smoothing functions for second-order cone complementarity problems*, SIAM J. Optim. 12 (2002), pp. 436–460.
[11] S. Hayashi, N. Yamashita, and M. Fukushima, *A combined smoothing and regularization method for monotone second-order cone complementarity problems*, SIAM J. Optim. 15 (2005), pp. 593–615.
[12] H. Kleinmichel and C. Kanzow, *A new class of semismooth Newton-type methods for nonlinear complementarity problems*, Comput. Optim. Appl. 11 (1998), pp. 227–251.
[13] A. Korányi, *Monotone functions on formally real Jordan algebras*, Math. Ann. 269 (1984), pp. 73–76.
[14] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. *SIAM Classics Appl. Math.*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2000.
[15] S.-H. Pan and J.-S. Chen, *A semismooth Newton method for SOCCPs based on one-parametric class of SOC complementarity functions*. Comput. Optimi. Appl. (2008), doi: 10.1007/s10589-008-9166-9.

[16] C.-K. Sim, J. Sun, and D. Ralph, *A note on the Lipschitz continuity of the gradient of the squared norm of the matrix-valued Fischer-Burmeister function*, Math. Progr. 107 (2006), pp. 547–553.
[17] D. Sun and J. Sun, *Strong semismoothness of the Fischer-Burmeister SDC and SOC complementarity functions*, Math. Progr. 103 (2005), pp. 575–581.
[18] P. Tseng, *Merit function for semidefinite complementarity problems*, Math. Progr. 83 (1998), pp. 159–185.