MATHEMATICS OF COMPUTATION
Volume 83, Number 287, May 2014, Pages 1143–1171
S 0025-5718(2013)02742-1
Article electronically published on July 15, 2013

ON THE GENERALIZED FISCHER-BURMEISTER MERIT FUNCTION FOR THE SECOND-ORDER CONE COMPLEMENTARITY PROBLEM

SHAOHUA PAN, SANGHO KUM, YONGDO LIM, AND JEIN-SHAN CHEN

Abstract. It has been an open question whether the family of merit functions ψ_p $(p > 1)$, the generalized Fischer-Burmeister (FB) merit function, associated to the second-order cone is smooth or not. In this paper we answer it partly, and show that ψ_p is smooth for $p \in (1, 4)$, and we provide the condition for its coerciveness. Numerical results are reported to illustrate the influence of p on the performance of the merit function method based on ψ_p.

1. Introduction

Given two continuously differentiable mappings $F, G\colon \mathbb{R}^n \to \mathbb{R}^n$, we consider the second-order cone complementarity problem (SOCCP): to seek a $\zeta \in \mathbb{R}^n$ such that

$$F(\zeta) \in \mathcal{K}, \quad G(\zeta) \in \mathcal{K}, \quad \langle F(\zeta), G(\zeta)\rangle = 0, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product that induces the norm $\|\cdot\|$, and $\mathcal{K}$ is the Cartesian product of a group of second-order cones (SOCs). In other words,

$$\mathcal{K} \;=\; \mathcal{K}^{n_1} \times \mathcal{K}^{n_2} \times \cdots \times \mathcal{K}^{n_m}, \tag{2}$$

where $n_1, \ldots, n_m \geq 1$, $n_1 + \cdots + n_m = n$, and $\mathcal{K}^{n_i}$ is the SOC in $\mathbb{R}^{n_i}$ defined by

$$\mathcal{K}^{n_i} := \left\{ (x_{i1}, x_{i2}) \in \mathbb{R} \times \mathbb{R}^{n_i - 1} \mid x_{i1} \geq \|x_{i2}\| \right\}.$$

As an extension of the nonlinear complementarity problem (NCP) over the nonnegative orthant cone $\mathbb{R}^n_+$ (see [13]), the SOCCP has important applications in engineering problems [21] and robust Nash equilibria [19]. In particular, it also

Received by the editor August 15, 2010, and in revised form, April 18, 2011 and August 7, 2012.

2010 *Mathematics Subject Classification.* Primary 90C33, 90C25.

Key words and phrases. Second-order cones, complementarity problem, generalized FB merit function.

The first author's work was supported by National Young Natural Science Foundation (No. 10901058) and the Fundamental Research Funds for the Central Universities (SCUT).

The second author's work was supported by Basic Science Research Program through NRF Grant No. 2012-0001740.

The third author's work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No.2012-005191).

Corresponding author: The fourth author is a member of the Mathematics Division, National Center for Theoretical Sciences, Taipei Office. The fourth author's work was supported by National Science Council of Taiwan.

arises from the suitable reformulation for the Karush-Kuhn-Tucker (KKT) optimality conditions of the nonlinear second-order cone programming (SOCP):

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \quad x \in \mathcal{K}, \end{array} \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function, A is an $m \times n$ real matrix with full row rank, and $b \in \mathbb{R}^m$. It is well known that the SOCP has very wide applications in engineering design, control, management science, and so on; see [1, 25] and the references therein.

In the past several years, there have various methods proposed for SOCPs and SOCCPs. They include the interior-point methods [2, 26, 28, 31, 33], the smoothing Newton methods [11, 15, 18], the semismooth Newton methods [22, 34], and the merit function methods [4, 12]. The merit function method aims to seek a smooth (continuously differentiable) function $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ satisfying

$$\psi(x, y) = 0 \iff x \in \mathcal{K}, \quad y \in \mathcal{K}, \quad \langle x, y\rangle = 0, \tag{4}$$

so that the SOCCP can be reformulated as an unconstrained minimization problem

$$\min_{\zeta \in \mathbb{R}^n} \Psi(\zeta) := \psi(F(\zeta), G(\zeta)) \tag{5}$$

in the sense that ζ^* is a solution to (1) if and only if it solves (5) with zero optimal value. We call such ψ a merit function associated with $\mathcal{K}$. Note that the smooth merit functions also play a key role in the globalization of semismooth and smoothing Newton methods.

This paper is concerned with the generalized Fischer-Burmeister (FB) merit function

$$\psi_p(x, y) \ := \ \frac{1}{2}\|\phi_p(x, y)\|^2, \tag{6}$$

where p is a fixed real number from $(1, +\infty)$, and $\phi_p : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$\phi_p(x, y) := \sqrt[p]{|x|^p + |y|^p} - (x + y) \tag{7}$$

with $|x|^p$ being the vector-valued function (or Löwner function) associated with $|t|^p$ $(t \in \mathbb{R})$ (see Section 2 for the definition). Clearly, when $p = 2$, ψ_p reduces to the FB merit function

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

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

$$\phi_{\text{FB}}(x, y) := \sqrt{x^2 + y^2} - (x + y),$$

with $x^2 = x \circ x$ being the Jordan product of x with itself, and $\sqrt{x}$ with $x \in \mathcal{K}$ being the unique vector such that $\sqrt{x} \circ \sqrt{x} = x$. The function ψ_{FB} is shown to be a smooth merit function with globally Lipschitz continuous derivative [10, 12]. Such a desirable property is also proved for the FB matrix-valued merit function [30, 32].

In this paper, we study the favorable properties of ψ_p. The motivations for us to study this family of merit functions are as follows. In the setting of NCPs, ψ_p is shown to share all favorable properties as the FB merit function holds (see [9, 8]), and the performance profile in [5] indicates that the semismooth Newton method based on ϕ_p with a smaller p has better performance than a larger p. Thus, it is very natural to ask whether ψ_p has the desirable properties of the FB merit function or not in the setting of SOCCPs, and what performance the merit function method

and the Newton-type methods based on ϕ_p display with respect to p. This work is the first step in resolving these questions. Although there are some papers [4, 6, 12] to study the smoothness of merit functions for the SOCCPs, the analysis techniques therein are not applicable for the general function ψ_p. We wish that the analysis technique of this paper would be helpful in handling general Löwner operators.

The main contribution of this paper is to show that ψ_p with $p \in (1,4)$ is a smooth merit function associated with $\mathcal{K}$, and to establish the coerciveness of $\Psi_p(\zeta) := \psi_p(F(\zeta), G(\zeta))$ under the uniform Jordan P-property and the linear growth of F.

Throughout this paper, we will focus on the case of $\mathcal{K} = \mathcal{K}^n$, and all the analysis can be carried over to the general case where $\mathcal{K}$ is the Cartesian product of $\mathcal{K}^{n_i}$. To this end, for any given $x \in \mathbb{R}^n$ with $n > 1$, we write $x = (x_1, x_2)$ where x_1 is the first component of x, and x_2 is the column vector consisting of the rest components of x; and let $\overline{x}_2 = \frac{x_2}{\|x_2\|}$ whenever $x_2 \neq 0$, and otherwise let $\overline{x}_2$ be an arbitrary vector in $\mathbb{R}^{n-1}$ with $\|\overline{x}_2\| = 1$. We denote $\mathbf{int}\mathcal{K}^n$, $\mathbf{bd}\mathcal{K}^n$ and $\mathbf{bd}^+\mathcal{K}^n$ by the interior, the boundary, and the boundary excluding the origin, respectively, of $\mathcal{K}^n$. For any $x, y \in \mathbb{R}^n$, $x \succeq_{\mathcal{K}^n} y$ means $x - y \in \mathcal{K}^n$; and $x \succ_{\mathcal{K}^n} y$ means $x - y \in \mathbf{int}\mathcal{K}^n$. For a real symmetric matrix A, we write $A \succeq 0$ (respectively, $A \succ 0$) to mean that A is positive semidefinite (respectively, positive definite). For a differentiable mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, $\nabla F(x)$ denotes the transposed Jacobian of F at x. For nonnegative α and β, $\alpha = O(\beta)$ means $\alpha \leq C\beta$ for some $C > 0$ independent of α and β. The notation I always represents an identity matrix of appropriate dimension.

2. Preliminaries

The Jordan product of any two vectors x and y associated with $\mathcal{K}^n$ (see [14]) is defined as

$$x \circ y := (\langle x, y\rangle,\ y_1 x_2 + x_1 y_2).$$

The Jordan product, 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 $e = (1, 0, \ldots, 0)^T \in \mathbb{R}^n$. For any given $x \in \mathbb{R}^n$, define $L_x : \mathbb{R}^n \to \mathbb{R}^n$ by

$$L_x y := \begin{bmatrix} x_1 & x_2^T \\ x_2 & x_1 I \end{bmatrix} y = x \circ y \quad \forall y \in \mathbb{R}^n.$$

Recall from [14] that each $x \in \mathbb{R}^n$ has a spectral factorization associated with $\mathcal{K}^n$:

$$x = \lambda_1(x) u_x^{(1)} + \lambda_2(x) u_x^{(2)}, \tag{8}$$

where $\lambda_i(x)$ and $u_x^{(i)}$ for $i = 1, 2$ are the spectral values of x and the corresponding spectral vectors, respectively, defined by

$$\lambda_i(x) := x_1 + (-1)^i \|x_2\| \ \text{ and } \ u_x^{(i)} := \frac{1}{2}\left(1,\ (-1)^i \overline{x}_2\right). \tag{9}$$

The factorization is unique when $x_2 \neq 0$. The following lemma states the relation between the spectral factorization of x and the eigenvalue decomposition of L_x.

Lemma 2.1 ([14, 15]). *For any given $x \in \mathbb{R}^n$, let $\lambda_1(x), \lambda_2(x)$ be the spectral values of x, and let $u_x^{(1)}, u_x^{(2)}$ be the corresponding spectral vectors. Then, we have*

$$L_x = U_x \mathrm{diag}\left(\lambda_2(x), x_1, \ldots, x_1, \lambda_1(x)\right) U_x^T$$

with $U_x = [\sqrt{2}u_x^{(2)}\ u^{(3)}\ \cdots\ u^{(n)}\ \sqrt{2}u_x^{(1)}] \in \mathbb{R}^{n\times n}$ *beings an orthogonal matrix, where* $u^{(i)} = (0, \overline{u}^i)$ *for* $i = 3, \ldots, n$ *with* $\overline{u}^3, \ldots, \overline{u}^n$ *being any unit vectors to span the linear subspace orthogonal to* x_2.

By Lemma 2.1, clearly, $L_x \succeq 0$ iff (if and only if) $x \succeq_{\mathcal{K}^n} 0$, $L_x \succ 0$ iff $x \succ_{\mathcal{K}^n} 0$, and L_x is invertible iff $x_1 \neq 0$ and $\det(x) := x_1^2 - \|x_2\|^2 \neq 0$. Also, if L_x is invertible,

$$L_x^{-1} = \frac{1}{\det(x)} \begin{bmatrix} x_1 & -x_2^T \\ -x_2 & \frac{\det(x)}{x_1} I + \frac{1}{x_1} x_2 x_2^T \end{bmatrix}. \tag{10}$$

Given a scalar function $g : \mathbb{R} \to \mathbb{R}$, define a vector function $g^{\mathrm{soc}} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$g^{\mathrm{soc}}(x) := g(\lambda_1(x))u_x^{(1)} + g(\lambda_2(x))u_x^{(2)}. \tag{11}$$

If g is defined on a subset of $\mathbb{R}$, then g^{soc} is defined on the corresponding subset of $\mathbb{R}^n$. The definition of g^{soc} is unambiguous whether $x_2 \neq 0$ or $x_2 = 0$. In this paper, we often use the vector-valued functions associated with $|t|^p$ $(t \in \mathbb{R})$ and $\sqrt[p]{t}$ $(t \geq 0)$, respectively, written as

$$|x|^p := |\lambda_1(x)|^p\ u_x^{(1)} + |\lambda_2(x)|^p\ u_x^{(2)} \quad \forall\, x \in \mathbb{R}^n,$$
$$\sqrt[p]{x} := \sqrt[p]{\lambda_1(x)}\ u_x^{(1)} + \sqrt[p]{\lambda_2(x)}\ u_x^{(2)} \quad \forall\, x \in \mathcal{K}^n.$$

The two functions show that ϕ_p in (7) is well defined for any $x, y \in \mathbb{R}^n$.

We next present four lemmas that will often be used in the subsequent analysis.

Lemma 2.2 ([23, 24])**.** *For any given* $0 \leq \rho \leq 1$, $\xi^\rho \succeq_{\mathcal{K}^n} \eta^\rho$ *when* $\xi \succeq_{\mathcal{K}^n} \eta \succeq_{\mathcal{K}^n} 0$.

Lemma 2.3. *For any nonnegative real numbers* a *and* b*, the following results hold:*

(a): $(a+b)^\rho \geq a^\rho + b^\rho$ *if* $\rho > 1$*, and the equality holds iff* $ab = 0$*;*
(b): $(a+b)^\rho \leq a^\rho + b^\rho$ *if* $0 < \rho < 1$*, and the equality holds iff* $ab = 0$.

Proof. Without loss of generality, we assume that $a \leq b$ and $b > 0$. Consider the function $h(t) = (t+1)^\rho - (t^\rho + 1)$ $(t \geq 0)$. It is easy to verify that h is increasing on $[0, +\infty)$ when $\rho > 1$. Hence, $h(a/b) \geq h(0) = 0$, i.e., $(a+b)^\rho \geq a^\rho + b^\rho$. Also, $h(a/b) = h(0)$ if and only if $a/b = 0$. That is, $(a+b)^\rho = a^\rho + b^\rho$ if and only if $ab = 0$. This proves part (a). Note that h is decreasing on $[0, +\infty)$ when $0 < \rho < 1$, and a similar argument leads to part (b). □

Lemma 2.4. *For any* $\xi, \eta \in \mathcal{K}^n$*, if* $\xi + \eta \in \mathbf{bd}\mathcal{K}^n$*, then one of the following cases must hold:* (i) $\xi = 0, \eta \in \mathbf{bd}\mathcal{K}^n$*;* (ii) $\xi \in \mathbf{bd}\mathcal{K}^n, \eta = 0$*;* (iii) $\xi = \gamma\eta$ *for some* $\gamma > 0$ *with* $\eta \in \mathbf{bd}^+\mathcal{K}^n$.

Proof. From $\xi, \eta \in \mathcal{K}^n$ and $\xi + \eta \in \mathbf{bd}\mathcal{K}^n$, we immediately obtain that

$$\|\xi_2\| + \|\eta_2\| \geq \|\xi_2 + \eta_2\| = \xi_1 + \eta_1 \geq \|\xi_2\| + \|\eta_2\|.$$

This shows that $\xi_2 = 0$, or $\eta_2 = 0$, or $\xi_2 = \gamma\eta_2 \neq 0$ for some $\gamma > 0$. Substituting $\xi_2 = 0$, or $\eta_2 = 0$, or $\xi_2 = \gamma\eta_2$ into $\|\xi_2 + \eta_2\| = \xi_1 + \eta_1$ yields the result. □

To close this section, we show that ϕ_p in (7) is an SOC complementarity function, and then its squared norm ψ_p is a merit function associated with $\mathcal{K}^n$.

Lemma 2.5. *Let* ϕ_p *be defined by* (7)*. Then, for any* $x, y \in \mathbb{R}^n$*, it holds that*

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

Proof. "$\Leftarrow$". From [16, Proposition 6], there exists a Jordan frame $\{u^{(1)}, u^{(2)}\}$ such that $x = \lambda_1 u^{(1)} + \lambda_2 u^{(2)}$ and $y = \mu_1 u^{(1)} + \mu_2 u^{(2)}$ with $\lambda_i, \mu_i \geq 0$ for $i = 1, 2$. Then,

$$\begin{aligned} (x+y)^p &= (\lambda_1+\mu_1)^p u^{(1)} + (\lambda_2+\mu_2)^p u^{(2)}, \\ x^p + y^p &= (\lambda_1^p+\mu_1^p) u^{(1)} + (\lambda_2^p+\mu_2^p) u^{(2)}. \end{aligned}$$

Since $0 = 2\langle x, y\rangle = \lambda_1\mu_1 + \lambda_2\mu_2$ implies $\lambda_1\mu_1 = \lambda_2\mu_2 = 0$, from the last two equalities and Lemma 2.3(a) we obtain $(x+y)^p = x^p + y^p$, and then $\phi_p(x, y) = 0$.

"$\Rightarrow$". Since $\phi_p(x, y) = 0$, we have $x = \sqrt{|x|^p + |y|^p} - y \succeq_{\mathcal{K}^n} |y| - y \in \mathcal{K}^n$, where the inequality is due to Lemma 2.2. Similarly, we have $y = \sqrt[p]{|x|^p + |y|^p} - x \succeq_{\mathcal{K}^n} |x| - x \in \mathcal{K}^n$. Now from $\phi_p(x, y) = 0$, we have $(x+y)^p = x^p + y^p$, and then

$$(\lambda_1(x+y))^p + (\lambda_2(x+y))^p = (\lambda_1(x))^p + (\lambda_2(x))^p + (\lambda_1(y))^p + (\lambda_2(y))^p.$$

Noting that $h(t) = (t_0+t)^p + (t_0-t)^p$ for a fixed $t_0 \geq 0$ is increasing on $[0, t_0]$, we also have

$$\begin{aligned} [\lambda_1(x+y)]^p + [\lambda_2(x+y)]^p &\geq (x_1+y_1-\|x_2\|+\|y_2\|)^p + (x_1+y_1+\|x_2\|-\|y_2\|)^p \\ &= (\lambda_1(x)+\lambda_2(y))^p + (\lambda_2(x)+\lambda_1(y))^p \\ &\geq (\lambda_1(x))^p + (\lambda_2(y))^p + (\lambda_2(x))^p + (\lambda_1(y))^p, \end{aligned} \tag{12}$$

where the last inequality is due to Lemma 2.3(a) and $x, y \in \mathcal{K}^n$. The last two equations imply that all the inequalities on the right-hand side of (12) become equalities. Therefore,

$$\|x_2+y_2\| = \|x_2\| - \|y_2\|, \quad \lambda_1(x)\lambda_2(y) = 0, \quad \lambda_2(x)\lambda_1(y) = 0. \tag{13}$$

Assume that $x_2 \neq 0$ and $y_2 \neq 0$. Since $x, y \in \mathcal{K}^n$, from the equalities in (13), we get $x_1 = \|x_2\|$, $y_1 = \|y_2\|$, and $x_2 = \widehat{\gamma} y_2$ for some $\widehat{\gamma} < 0$, which implies $\langle x, y\rangle = 0$. When $x_2 = 0$ or $y_2 = 0$, using the continuity of the inner product yields $\langle x, y\rangle = 0$. $\square$

3. Differentiability of ψ_p

Unless otherwise stated, in the rest of this paper, we assume that $p > 1$ with $q = (1-p^{-1})^{-1}$, and g^{soc} is the vector-valued function associated with $|t|^p$ ($t \in \mathbb{R}$), i.e., $g^{\mathrm{soc}}(x) = |x|^p$. For any $x, y \in \mathbb{R}^n$, we define

$$w = w(x, y) := |x|^p + |y|^p \quad \text{and} \quad z = z(x, y) := \sqrt[p]{|x|^p + |y|^p}. \tag{14}$$

By definitions of $|x|^p$ and $|y|^p$, clearly,

$$\begin{aligned} w_1 := w_1(x, y) &= \frac{|\lambda_2(x)|^p + |\lambda_1(x)|^p}{2} + \frac{|\lambda_2(y)|^p + |\lambda_1(y)|^p}{2}, \\ w_2 := w_2(x, y) &= \frac{|\lambda_2(x)|^p - |\lambda_1(x)|^p}{2}\overline{x}_2 + \frac{|\lambda_2(y)|^p - |\lambda_1(y)|^p}{2}\overline{y}_2, \end{aligned} \tag{15}$$

where $\overline{x}_2 = \frac{x_2}{\|x_2\|}$ if $x_2 \neq 0$, and otherwise $\overline{x}_2$ is an arbitrary vector in $\mathbb{R}^{n-1}$ with $\|\overline{x}_2\| = 1$, and $\overline{y}_2$ has a similar definition.

Noting that $z(x, y) = \sqrt[p]{w(x, y)}$, we have

$$\begin{aligned} z_1 = z_1(x, y) &= \frac{\sqrt[p]{\lambda_2(w)} + \sqrt[p]{\lambda_1(w)}}{2}, \\ z_2 = z_2(x, y) &= \frac{\sqrt[p]{\lambda_2(w)} - \sqrt[p]{\lambda_1(w)}}{2}\overline{w}_2, \end{aligned} \tag{16}$$

where $\overline{w}_2 = \frac{w_2}{\|w_2\|}$ if $w_2 \neq 0$, and otherwise $\overline{w}_2$ is an arbitrary vector in $\mathbb{R}^{n-1}$ with $\|\overline{w}_2\| = 1$.

To study the differentiability of ψ_p, we need the following two crucial lemmas. The first one gives the properties of the points (x, y) satisfying $w(x, y) \in \mathbf{bd}\mathcal{K}^n$, and the second one provides a sufficient characterization for the continuously differentiable points of $z(x, y)$.

Lemma 3.1. *For any (x, y) with $w(x, y) \in \mathbf{bd}\mathcal{K}^n$, we have the following equalities:*

$$w_1(x, y) = \|w_2(x, y)\| = 2^{p-1}(|x_1|^p + |y_1|^p),$$
$$x_1^2 = \|x_2\|^2,\ y_1^2 = \|y_2\|^2,\ x_1y_1 = x_2^Ty_2,\ x_1y_2 = y_1x_2. \tag{17}$$

If, in addition, $w_2(x, y) \neq 0$, the following equalities hold with $\overline{w}_2(x, y) = \frac{w_2(x,y)}{\|w_2(x,y)\|}$:

$$x_2^T\overline{w}_2(x, y) = x_1, \quad x_1\overline{w}_2(x, y) = x_2, \quad y_2^T\overline{w}_2(x, y) = y_1, \quad y_1\overline{w}_2(x, y) = y_2. \tag{18}$$

Proof. Fix any (x, y) with $w(x, y) \in \mathbf{bd}\mathcal{K}^n$. Since $|x|^p, |y|^p \in \mathcal{K}^n$, applying Lemma 2.4 with $\xi = |x|^p$ and $\eta = |y|^p$, we have $|x|^p \in \mathbf{bd}\mathcal{K}^n$ and $|y|^p \in \mathbf{bd}\mathcal{K}^n$. This means that $|\lambda_2(x)|^p \cdot |\lambda_1(x)|^p = 0$ and $|\lambda_2(y)|^p \cdot |\lambda_1(y)|^p = 0$. So, $x_1^2 = \|x_2\|^2$ and $y_1^2 = \|y_2\|^2$. Substituting this into $w_1(x, y)$, we readily obtain $w_1(x, y) = 2^{p-1}(|x_1|^p + |y_1|^p)$.

To prove other equalities in (17) and (18), we first consider the case where $x_1 + \|x_2\| = 0$ and $y_1 - \|y_2\| = 0$ with $x_2 \neq 0$ and $y_2 \neq 0$. Under this case,

$$w_1 = \frac{|\lambda_1(x)|^p + |\lambda_2(y)|^p}{2} = \left\| \frac{|\lambda_1(x)|^p}{2} \frac{x_2}{\|x_2\|} - \frac{|\lambda_2(y)|^p}{2} \frac{y_2}{\|y_2\|} \right\| = \|w_2\|,$$

which implies that $x_2^Ty_2 = -\|x_2\|\|y_2\| = x_1y_1$. Together with $x_1^2 = \|x_2\|^2$ and $y_1^2 = \|y_2\|^2$, we have that $x_1y_2 = y_1x_2$. From the definition of w_2, it follows that

$$x_2^Tw_2 = -\frac{|\lambda_1(x)|^p}{2}\|x_2\| + \frac{|\lambda_2(y)|^p}{2}\frac{x_1y_1}{\|y_2\|} = 2^{p-1}\left(|x_1|^p + |y_1|^p\right)x_1 = \|w_2\|x_1,$$
$$x_1w_2 = -\frac{|\lambda_1(x)|^p}{2}\frac{x_1x_2}{\|x_2\|} + \frac{|\lambda_2(y)|^p}{2}\frac{y_1x_2}{\|y_2\|} = 2^{p-1}\left(|x_1|^p + |y_1|^p\right)x_2 = \|w_2\|x_2.$$

Similarly, we also have $y_2^Tw_2 = \|w_2\|y_1$ and $y_1^Tw_2 = \|w_2\|y_2$. The above arguments show that equations (17) and (18) hold under the case where $x_1 = -\|x_2\|, y_1 = \|y_2\|$. Using the same arguments, we can prove that (17) and (18) hold under any one of the following cases: $x_1 = \|x_2\|, y_1 = \|y_2\|$; or $x_1 = -\|x_2\|, y_1 = \|y_2\|$; or $x_1 = -\|x_2\|, y_1 = -\|y_2\|$. □

Lemma 3.2. *$z(x, y)$ is continuously differentiable at (x, y) with $w(x, y) \in \mathbf{int}\mathcal{K}^n$, and*

$$\nabla_x z(x, y) = \nabla g^{\mathrm{soc}}(x)\nabla g^{\mathrm{soc}}(z)^{-1} \quad \text{and} \quad \nabla_y z(x, y) = \nabla g^{\mathrm{soc}}(y)\nabla g^{\mathrm{soc}}(z)^{-1},$$

where $\nabla g^{\mathrm{soc}}(z)^{-1} = (p\sqrt[q]{w_1})^{-1}I$ if $w_2 = 0$, and otherwise

$$\nabla g^{\mathrm{soc}}(z)^{-1} = \frac{1}{2p}\begin{bmatrix} \frac{1}{\sqrt[q]{\lambda_2(w)}} + \frac{1}{\sqrt[q]{\lambda_1(w)}} & \frac{\overline{w}_2^T}{\sqrt[q]{\lambda_2(w)}} - \frac{\overline{w}_2^T}{\sqrt[q]{\lambda_1(w)}} \\ \frac{\overline{w}_2}{\sqrt[q]{\lambda_2(w)}} - \frac{\overline{w}_2}{\sqrt[q]{\lambda_1(w)}} & \frac{2p(I-\overline{w}_2\overline{w}_2^T)}{a(z)} + \frac{\overline{w}_2\overline{w}_2^T}{\sqrt[q]{\lambda_2(w)}} + \frac{\overline{w}_2\overline{w}_2^T}{\sqrt[q]{\lambda_1(w)}} \end{bmatrix}.$$

Proof. Since $|t|^p$ $(t \in \mathbb{R})$ and $\sqrt[p]{t}$ $(t > 0)$ are continuously differentiable, by [15, Proposition 5.2] or [7, Proposition 5], the functions $g^{\mathrm{soc}}(x)$ and $\sqrt{x}$ are continuously differentiable in $\mathbb{R}^n$ and $\mathbf{int}\mathcal{K}^n$, respectively. This implies the first part of this

lemma. A simple calculation gives the expression of $\nabla z(x,y)$. By the formula in [15, Proposition 5.2],

$$\nabla g^{\mathrm{soc}}(x) = \begin{cases} p\ \mathrm{sign}(x_1)|x_1|^{p-1}I & \text{if } x_2 = 0; \\ \begin{bmatrix} b(x) & c(x)\overline{x}_2^T \\ c(x)\overline{x}_2 & a(x)I + (b(x)-a(x))\overline{x}_2\overline{x}_2^T \end{bmatrix} & \text{if } x_2 \neq 0, \end{cases} \tag{19}$$

where

$$\begin{aligned} \overline{x}_2 &= \frac{x_2}{\|x_2\|}, \quad a(x) = \frac{|\lambda_2(x)|^p - |\lambda_1(x)|^p}{\lambda_2(x) - \lambda_1(x)}, \\ b(x) &= \frac{p}{2}\left[\mathrm{sign}(\lambda_2(x))|\lambda_2(x)|^{p-1} + \mathrm{sign}(\lambda_1(x))|\lambda_1(x)|^{p-1}\right], \\ c(x) &= \frac{p}{2}\left[\mathrm{sign}(\lambda_2(x))|\lambda_2(x)|^{p-1} - \mathrm{sign}(\lambda_1(x))|\lambda_1(x)|^{p-1}\right]. \end{aligned} \tag{20}$$

We next derive the formula of $\nabla g^{\mathrm{soc}}(z)^{-1}$. When $w_2 = 0$, we have $\lambda_1(w) = \lambda_2(w) = w_1 > 0$, which by (16) implies $z_1 = \sqrt[p]{w_1}$ and $z_2 = 0$. From formula (19), it then follows that $\nabla g^{\mathrm{soc}}(z) = p|z_1|^{p-1}I = p\sqrt[q]{w_1}I$. Consequently, $\nabla g^{\mathrm{soc}}(z)^{-1} = \frac{1}{p\sqrt[q]{w_1}}I$. When $w_2 \neq 0$, since $\sqrt[p]{\lambda_2(w)} > \sqrt[p]{\lambda_1(w)}$, we have $z_2 \neq 0$ and $\overline{z}_2 = \frac{z_2}{\|z_2\|} = \overline{w}_2$ by (16). Using the expression of $\nabla g^{\mathrm{soc}}(z)$, it is easy to verify that $b(z) + c(z)$ and $b(z) - c(z)$ are the eigenvalues of $\nabla g^{\mathrm{soc}}(z)$ with $(1, \overline{w}_2)$ and $(1, -\overline{w}_2)$ being the corresponding eigenvectors, and $a(z)$ is the eigenvalue of multiplicity $n-2$ with corresponding eigenvectors of the form $(0, \overline{v}_i)$, where $\overline{v}_1, \ldots, \overline{v}_{n-2}$ are any unit vectors in $\mathbb{R}^{n-1}$ that span the subspace orthogonal to w_2. Hence,

$$\nabla g^{\mathrm{soc}}(z) = U \mathrm{diag}\left(b(z) - c(z), a(z), \ldots, a(z), b(z) + c(z)\right) U^T,$$

where $U = [u_1 \ \ v_1 \ \ \cdots \ \ v_{n-2} \ \ u_2] \in \mathbb{R}^{n\times n}$ is an orthogonal matrix with

$$u_1 = \begin{pmatrix} 1 \\ -\overline{w}_2 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix}, \quad v_i = \begin{pmatrix} 0 \\ \overline{v}_i \end{pmatrix} \quad \text{for} \quad i = 1, \ldots, n-2.$$

By this, we know that $\nabla g^{\mathrm{soc}}(z)^{-1}$ has the expression given as in the lemma. □

Now we are in a position to prove the following main result of this section.

Proposition 3.1. *The function ψ_p for $p \in (1,4)$ is differentiable everywhere. Also, for any given $x, y \in \mathbb{R}^n$, if $w(x,y) = 0$, then $\nabla_x \psi_p(x,y) = \nabla_y \psi_p(x,y) = 0$; if $w(x,y) \in \mathbf{int}\mathcal{K}^n$, then*

$$\begin{aligned} \nabla_x \psi_p(x,y) &= \left(\nabla g^{\mathrm{soc}}(x)\nabla g^{\mathrm{soc}}(z)^{-1} - I\right)\phi_p(x,y), \\ \nabla_y \psi_p(x,y) &= \left(\nabla g^{\mathrm{soc}}(y)\nabla g^{\mathrm{soc}}(z)^{-1} - I\right)\phi_p(x,y); \end{aligned} \tag{21}$$

and if $w(x,y) \in \mathbf{bd}^+\mathcal{K}^n$, then

$$\begin{aligned} \nabla_x \psi_p(x,y) &= \left(\frac{\mathrm{sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p + |y_1|^p}} - 1\right)\phi_p(x,y), \\ \nabla_y \psi_p(x,y) &= \left(\frac{\mathrm{sign}(y_1)|y_1|^{p-1}}{\sqrt[q]{|x_1|^p + |y_1|^p}} - 1\right)\phi_p(x,y). \end{aligned} \tag{22}$$

Proof. Fix any $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$. If $w(x,y) \in \mathbf{int}\mathcal{K}^n$, the result is implied by Lemma 3.2 since $\phi_p(x,y) = z(x,y) - (x+y)$. In fact, in this case, ψ_p is continuously differentiable at (x,y). Hence, it suffices to consider the cases $w(x,y) = 0$ and $w(x,y) \in \mathbf{bd}^+\mathcal{K}^n$. In the following arguments, x' and y' are arbitrary vectors in

$\mathbb{R}^n$, and $\mu_1(x',y'), \mu_2(x',y')$ are the spectral values of $w(x',y')$ with $\xi^{(1)}, \xi^{(2)} \in \mathbb{R}^n$ being the corresponding spectral vectors.

Case 1. $w(x,y)=0$. Note that $(x,y)=(0,0)$ in this case. Hence, we only need to prove, for any $x',y' \in \mathbb{R}^n$,

$$\psi_p(x',y') - \psi_p(0,0) = \frac{1}{2}\left\|z(x',y') - (x'+y')\right\|^2 = O(\|(x',y')\|), \tag{23}$$

which shows that ψ_p is differentiable at $(0,0)$ with $\nabla_x \psi_p(0,0) = \nabla_y \psi_p(0,0) = 0$. Indeed,

$$\begin{aligned} \|z(x',y') - (x'+y')\| &= \left\| \sqrt[p]{\mu_1(x',y')}\ \xi^{(1)} + \sqrt[p]{\mu_2(x',y')}\ \xi^{(2)} - (x'+y') \right\| \\ &\le \sqrt{2}\sqrt[p]{\mu_2(x',y')} + \|x'\| + \|y'\|. \end{aligned} \tag{24}$$

From the definition of $w_1(x,y)$ and $w_2(x,y)$, it is easy to obtain that

$$\mu_2(x',y') = w_1(x',y') + w_2(x',y') \le |\lambda_2(x')|^p + |\lambda_1(x')|^p + |\lambda_2(y')|^p + |\lambda_1(y')|^p.$$

Using the nondecreasing property of $\sqrt[p]{t}$ and Lemma 2.3(b), it then follows that

$$\begin{aligned} \sqrt[p]{\mu_2(x',y')} &\le \left(|\lambda_2(x')|^p + |\lambda_1(x')|^p + |\lambda_2(y')|^p + |\lambda_1(y')|^p\right)^{1/p} \\ &\le |\lambda_2(x')| + |\lambda_1(x')| + |\lambda_2(y')| + |\lambda_1(y')| \le 2(\|x'\| + \|y'\|). \end{aligned}$$

This, together with (24), implies that equation (23) holds.

Case 2. $w(x,y) \in \mathbf{bd}^+\mathcal{K}^n$. Now $w_1(x,y) = \|w_2(x,y)\| \neq 0$, and one of x_2 and y_2 is nonzero by (18). We proceed with the arguments in three steps, as shown below.

Step 1. We prove that $w_1(x',y')$ and $w_2(x',y')$ are $\lceil p \rceil$ times differentiable at $(x',y') = (x,y)$, where $\lceil p \rceil$ denotes the maximum integer not greater than p. Since one of x_2 and y_2 is nonzero, we prove this result by considering three possible cases: (i) $x_2 \neq 0$, $y_2 \neq 0$; (ii) $x_2 = 0$, $y_2 \neq 0$; and (iii) $x_2 \neq 0$, $y_2 = 0$. For case (i), since $\frac{x_2'}{\|x_2'\|}$, $\frac{y_2'}{\|y_2'\|}$, $\lambda_2(x'), \lambda_1(x'), \lambda_2(y')$, and $\lambda_1(y')$ are infinite times differentiable at (x,y), and $|t|^p$ is $\lceil p \rceil$ times continuously differentiable in $\mathbb{R}$, it follows that $w_1(x',y')$ and $w_2(x',y')$ are $\lceil p \rceil$ times differentiable at (x,y). Now assume that case (ii) is satisfied. From the arguments in case (i), we know that

$$\frac{|\lambda_2(y')|^p + |\lambda_1(y')|^p}{2} \quad \text{and} \quad \frac{|\lambda_2(y')|^p - |\lambda_1(y')|^p}{2}\frac{y_2'}{\|y_2'\|}$$

are $\lceil p \rceil$ times differentiable at (x,y). In addition, since $|\lambda_i(x')|^p \le 2^{\frac{p}{2}}\|x'\|^p$ for $i=1,2$, and $x=0$ in this case, we have that $|\lambda_2(x')|^p + |\lambda_1(x')|^p$ and $\frac{1}{2}(|\lambda_2(x')|^p - |\lambda_1(x')|^p)\overline{x}_2'$ are $\lceil p \rceil$ times differentiable at x with the first $\lceil p \rceil - 1$ order derivatives being zero. Thus, $w_1(x',y')$ and $w_2(x',y')$ are $\lceil p \rceil$ times differentiable at (x,y). By the symmetry of x',y' in $w(x',y')$ and the arguments in case (ii), the result also holds for case (iii).

Step 2. We show that ψ_p is differentiable at (x,y). By the definition of ψ_p, we have

$$2\psi_p(x',y') = \|x'+y'\|^2 + \|z(x',y')\|^2 - 2\langle z(x',y'), x'+y'\rangle.$$

Since $\|x'+y'\|^2$ is differentiable, it suffices to argue that the last two terms on the right-hand side are differentiable at (x,y). By formulas (8)-(9), it is not hard to

calculate that

$$2\left\|z(x',y')\right\|^2 = (\mu_2(x',y'))^{\frac{2}{p}} + (\mu_1(x',y'))^{\frac{2}{p}}, \tag{25}$$

$$\begin{aligned} 2\left\langle z(x',y'), x'+y'\right\rangle = {} & \sqrt[p]{\mu_2(x',y')}\left(x_1'+y_1'+\frac{(w_2(x',y'))^T(x_2'+y_2')}{\|w_2(x',y')\|}\right) \\ & + \sqrt[p]{\mu_1(x',y')}\left(x_1'+y_1'-\frac{(w_2(x',y'))^T(x_2'+y_2')}{\|w_2(x',y')\|}\right). \end{aligned} \tag{26}$$

Since $w_2(x,y) \neq 0$, $\mu_2(x,y) = \lambda_2(w) > 0$, and $w_1(x',y')$ and $w_2(x',y')$ are differentiable at (x,y), by Step 1 we have that $(\mu_2(x',y'))^{\frac{2}{p}}$ and the first term on the right-hand side of (26) is differentiable at (x,y). Thus, it suffices to prove that $(\mu_1(x',y'))^{\frac{2}{p}}$ and the last term on the right-hand side of (26) are differentiable at (x,y).

We first argue that $(\mu_1(x',y'))^{\frac{2}{p}}$ is differentiable at (x,y). Since $w_2(x,y) \neq 0$, and $w_1(x',y')$ and $w_2(x',y')$ are $\lceil p \rceil$ times differentiable at (x,y) by Step 1, the function $\mu_1(x',y')$ is $\lceil p \rceil$ times differentiable at (x,y). When $p < 2$, by the mean-value theorem and $\mu_1(x,y) = \lambda_1(w) = 0$, it follows that $\mu_1(x',y') = O(\|x'-x\| + \|y'-y\|)$ for any (x',y') sufficiently close to (x,y), and therefore $(\mu_1(x',y'))^{\frac{2}{p}} = O[(\|x'-x\| + \|y'-y\|)^{\frac{2}{p}}]$. This shows that $(\mu_1(x',y'))^{\frac{2}{p}}$ is differentiable at (x,y) with zero derivative. When $p \geq 2$, $\mu_1(x',y')$ is infinite times differentiable at (x,y), and its first derivative equals zero by the result in the Appendix. From the second-order Taylor expansion of $\mu_1(x',y')$ at (x,y), it follows that $(\mu_1(x',y'))^{\frac{2}{p}} = O[(\|x'-x\| + \|y'-y\|)^{\frac{4}{p}}]$. This implies that $(\mu_1(x',y'))^{\frac{2}{p}}$ is differentiable at (x,y) with zero gradient when $2 \leq p < 4$. Thus, we prove that $(\mu_1(x',y'))^{\frac{2}{p}}$ is differentiable at (x,y) with zero gradient when $p \in (1,4)$.

We next consider the last term on the right-hand side of (26). Observe that

$$x_1'+y_1'-\frac{(w_2(x',y'))^T(x_2'+y_2')}{\|w_2(x',y')\|}$$

is differentiable at (x,y), and its function value at (x,y) equals zero by (18). Hence, this term is $O(\|x'-x\| + \|y'-y\|)$, which, along with $\mu_1(x',y') = O(\|x'-x\| + \|y'-y\|)$, means that the last term of (26) is $O((\|x'-x\| + \|y'-y\|)^{1+\frac{1}{p}}) = o(\|x'-x\| + \|y'-y\|)$. This shows that the last term of (26) is differentiable at (x,y) with zero derivative.

Step 3. We derive the formula of $\nabla_x \psi_p(x,y)$. From Step 2, we see that $2\nabla\psi_p(x,y)$ equals the difference between the gradient of $\frac{1}{2}(\mu_2(x',y'))^{\frac{2}{p}} + \|x'+y'\|^2$ and that of the first term on the right-hand side of (26), evaluated at (x,y). By the Appendix, the gradients of $(\mu_2(x',y'))^{1/p}$ and $(\mu_2(x',y'))^{2/p}$ with respect to x', evaluated at $(x',y') = (x,y)$, are

$$\nabla_{x'}(\mu_2(x',y'))^{1/p}|_{(x',y')=(x,y)} = (\lambda_2(w))^{\frac{1}{p}-1}2^{p-1}\,\mathrm{sign}(x_1)|x_1|^{p-1}\left(\frac{1}{\overline{w}_2}\right), \tag{27}$$

$$\nabla_{x'}(\mu_2(x',y'))^{2/p}|_{(x',y')=(x,y)} = (\lambda_2(w))^{\frac{2}{p}-1}2^{p}\,\mathrm{sign}(x_1)|x_1|^{p-1}\left(\frac{1}{\overline{w}_2}\right). \tag{28}$$

By the product and quotient rules for differentiation, the gradient of $x_1' + y_1' + \frac{(w_2(x',y'))^T(x_2'+y_2')}{\|w_2(x',y')\|}$ with respect to x', evaluated at $(x',y') = (x,y)$, works out to be

$$\begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix} + \nabla_{x'} w_2(x',y')|_{(x',y')=(x,y)} \left(\frac{x_2+y_2}{\|w_2\|} - \frac{\overline{w}_2 \overline{w}_2^T (x_2+y_2)}{\|w_2\|} \right) = \begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix},$$

where the equality uses (18). Along with (27), the gradient of the first term on the right-hand side of (26) with respect to x', evaluated at $(x',y') = (x,y)$, is

$$(\lambda_2(w))^{\frac{1}{p}-1}(x_1+y_1)2^p \operatorname{sign}(x_1)|x_1|^{p-1} \begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix} + (\lambda_2(w))^{\frac{1}{p}} \begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix}. \tag{29}$$

In addition, the gradient of $\|x'+y'\|^2$ with respect to x', evaluated at $(x',y') = (x,y)$, is $2(x+y)$. Together with equations (28)-(29), we obtain that

$$\begin{aligned} 2\nabla_x \psi_p(x,y) = 2(x+y) &+ (\lambda_2(w))^{\frac{2}{p}-1} 2^{p-1} \operatorname{sign}(x_1)|x_1|^{p-1} \begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix} \\ &- (\lambda_2(w))^{\frac{1}{p}-1}(x_1+y_1)2^p \operatorname{sign}(x_1)|x_1|^{p-1} \begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix} - (\lambda_2(w))^{\frac{1}{p}} \begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix}. \end{aligned}$$

Since $\lambda_1(w) = 0$, from (16) it follows that

$$\phi_p(x,y) = z(x,y) - (x+y) = \frac{1}{2}(\lambda_2(w))^{\frac{1}{p}} \begin{pmatrix} 1 \\ \overline{w}_2 \end{pmatrix} - (x+y).$$

Combining the last two equations and using $x_1\overline{w}_2 = x_2$ and $y_1\overline{w}_2 = y_2$, we get

$$\begin{aligned} 2\nabla_x \psi_p(x,y) &= (\lambda_2(w))^{\frac{1}{p}-1} 2^p \operatorname{sign}(x_1)|x_1|^{p-1}(\phi_p(x,y) + (x+y)) \\ &\quad - (\lambda_2(w))^{\frac{1}{p}-1} 2^p \operatorname{sign}(x_1)|x_1|^{p-1}(x+y) - 2\phi_p(x,y) \\ &= 2\left[\frac{\operatorname{sign}(x_1)|x_1|^{p-1}}{(|x_1|^p + |y_1|^p)^{1/q}} - 1 \right] \phi_p(x,y), \end{aligned}$$

where the last equality is from $\lambda_2(w) = 2w_1 = 2^p(|x_1|^p + |y_1|^p)$. This proves the first equality in (22). By the symmetry of x and y in ψ_p, the second equality in (22) also holds. □

From the arguments in Step 2 of Case 2, we see that ψ_p will be differentiable for $p \geq 4$ if the first $\lceil \frac{p}{2} \rceil$-order derivatives of $\mu_1(x',y') = w_1(x',y') - \|w_2(x',y')\|$, evaluated at $(x',y') = (x,y)$, are equal to zero. We are not clear whether this holds or not.

4. Smoothness of ψ_p

In the last section ψ_p for $p \in (1,4)$ is proved to be differentiable everywhere. A natural question is whether $\nabla\psi_p$ is continuous or not. In this section we will provide an affirmative answer to it. For this purpose, we first establish three technical lemmas which hold for all $p > 1$.

Lemma 4.1. *There exists a constant c_1 such that for all (x,y) with $w(x,y) \in \mathbf{int}\mathcal{K}^n$,*

$$\left\| L_{|x|^{p-1}} L_{z^{p-1}}^{-1} \right\|_F \leq c_1 \quad \text{and} \quad \left\| L_{|y|^{p-1}} L_{z^{p-1}}^{-1} \right\|_F \leq c_1, \tag{30}$$

where c is independent of x and y, and $\|\cdot\|_F$ means the Frobenius norm.

Proof. Due to the symmetry of x and y in $z(x,y)$, it suffices to prove the first inequality. To this end, we first prove that for any (x,y) with $w(x,y) \in \mathbf{int}\mathcal{K}^n$,

$$0 \le \lambda\left(L_{|x|^{p-1}} L_{z^{p-1}}^{-1}\right) \le \mathbf{1}, \tag{31}$$

where, for a matrix $A \in \mathbb{R}^{n\times n}$, $\lambda(A) \in \mathbb{R}^n$ denotes the vector of eigenvalues of A, and $\mathbf{1}$ means a vector with all components being 1. Indeed, since $z \succ_{\mathcal{K}^n} 0$ and $|x|^{p-1} \succeq_{\mathcal{K}^n} 0$, we have $L_z \succ 0$ and $L_{|x|^{p-1}} \succeq 0$. Applying [20, Theorem 7.6.3] with $A = L_{z^{p-1}}^{-1}$ and $B = L_{|x|^{p-1}}$ yields that $\lambda(L_{z^{p-1}}^{-1} L_{|x|^{p-1}}) \ge 0$, and then $\lambda(L_{|x|^{p-1}} L_{z^{p-1}}^{-1}) \ge 0$. In addition, since $z^p \succeq_{\mathcal{K}^n} |x|^p$, from Lemma 2.2 it follows that $(z^p)^{\frac{p-1}{p}} \succeq_{\mathcal{K}^n} (|x|^p)^{\frac{p-1}{p}}$, i.e., $z^{p-1} \succeq_{\mathcal{K}^n} |x|^{p-1}$. Then $L_{z^{p-1}} - L_{|x|^{p-1}} \succeq 0$. Applying the result of Exercise 7 in [20, p. 468] with $A = L_{z^{p-1}}$ and $B = -L_{|x|^{p-1}}$, we have that $\lambda\left(-L_{z^{p-1}}^{-1} L_{|x|^{p-1}}\right) \ge -\mathbf{1}$. Consequently, $\lambda\left(L_{|x|^{p-1}} L_{z^{p-1}}^{-1}\right) \le \mathbf{1}$. Together with $\lambda(L_{|x|^{p-1}} L_{z^{p-1}}^{-1}) \ge 0$, we prove that (31) holds.

Next we prove that there exists a constant $c_1 > 0$ such that for all (x,y) satisfying $w(x,y) \in \mathbf{int}\mathcal{K}^n$, $\|L_{|x|^{p-1}} L_{z^{p-1}}^{-1}\|_F \le c_1$ where c_1 is independent of x and y. Suppose on the contrary that such c_1 does not exist. Then, there exists a sequence $\{(x^k, y^k)\} \subset \mathbb{R}^n \times \mathbb{R}^n$ with $w(x^k, y^k) \in \mathbf{int}\mathcal{K}^n$ such that $\|L_{|x^k|^{p-1}} L_{(z^k)^{p-1}}^{-1}\|_F$ is unbounded. We assume (taking a subsequence if necessary) that

$$\lim_{k\to\infty} \|L_{|x^k|^{p-1}} L_{(z^k)^{p-1}}^{-1}\|_F = +\infty.$$

For each k, let $A^k = L_{|x^k|^{p-1}}$ and $B^k = L_{(z^k)^{p-1}}^{-1}$. Subsequencing if necessary, we may assume that

$$\lim_{k\to\infty} \frac{A^k}{\|A^k\|_F} = A^* \quad \text{and} \quad \lim_{k\to\infty} \frac{B^k}{\|B^k\|_F} = B^*.$$

In the following arguments, for any $A, B \in \mathbb{R}^{n\times n}$ with all eigenvalues in $\mathbb{R}$, we let $\lambda^{\downarrow}(A)$ and $\lambda^{\uparrow}(A)$ be the vectors obtained by rearranging the coordinates of $\lambda(A)$ in the decreasing and increasing orders, respectively. That is, if $\lambda^{\downarrow}(A) = (\lambda_1^{\downarrow}(A), \ldots, \lambda_n^{\downarrow}(A))$, then $\lambda_1^{\downarrow}(A) \ge \cdots \ge \lambda_n^{\downarrow}(A)$. Similarly, if $\lambda^{\uparrow}(A) = (\lambda_1^{\uparrow}(A), \ldots, \lambda_n^{\uparrow}(A))$, then $\lambda_1^{\uparrow}(A) \le \cdots \le \lambda_n^{\uparrow}(A)$. We write $\lambda(A) \prec \lambda(B)$ if $\sum_{j=1}^{l} \lambda_j^{\downarrow}(A) \le \sum_{j=1}^{l} \lambda_j^{\downarrow}(B)$ for any $1 \le l \le n$ and $\sum_{j=1}^{n} \lambda_j^{\downarrow}(A) = \sum_{j=1}^{n} \lambda_j^{\downarrow}(B)$. Since $A^k \succ 0$ and $B^k \succeq 0$ for each k, applying the result of [3, Problem III.6.14] gives that

$$\lambda^{\downarrow}\left(\frac{A^k}{\|A^k\|_F}\right) \cdot \lambda^{\uparrow}\left(\frac{B^k}{\|B^k\|_F}\right) \prec \lambda\left(\frac{A^k B^k}{\|A^k\|_F \|B^k\|_F}\right) \prec \lambda^{\downarrow}\left(\frac{A^k}{\|A^k\|_F}\right) \cdot \lambda^{\downarrow}\left(\frac{B^k}{\|B^k\|_F}\right),$$

where "$\cdot$" denotes the componentwise product. Since $\lim_{k\to\infty} \|A^k\|_F \|B^k\|_F = +\infty$, taking the limit $k \to +\infty$ and using (31) and the continuity of $\lambda(\cdot)$, we get

$$\lambda^{\downarrow}(A^*) \cdot \lambda^{\uparrow}(B^*) \prec 0 \prec \lambda^{\downarrow}(A^*) \cdot \lambda^{\downarrow}(B^*). \tag{32}$$

Since $A^* \succeq 0$ and $B^* \succeq 0$, each component of $\lambda^{\downarrow}(A^*)$ and $\lambda^{\uparrow}(B^*)$ is nonnegative, and the first relation of (32) then implies $\lambda^{\downarrow}(A^*) \cdot \lambda^{\uparrow}(B^*) = 0$. Note that for each

k, all eigenvalues of A^k and B^k, respectively are given as follows:

$$|\lambda_1(x^k)|^{p-1}, \underbrace{\left[|\lambda_1(x^k)|^{p-1}+|\lambda_2(x^k)|^{p-1}\right], \ldots, \left[|\lambda_1(x^k)|^{p-1}+|\lambda_2(x^k)|^{p-1}\right]}_{n-2}, |\lambda_2(x^k)|^{p-1},$$

$$\frac{1}{[\lambda_1(z^k)]^{p-1}}, \underbrace{\frac{1}{[\lambda_1(z^k)]^{p-1}+[\lambda_2(z^k)]^{p-1}}, \ldots, \frac{1}{[\lambda_1(z^k)]^{p-1}+[\lambda_2(z^k)]^{p-1}}}_{n-2}, \frac{1}{[\lambda_2(z^k)]^{p-1}},$$

which, by the positive homogeneousness of eigenvalue function, means that

$$\lambda_1^{\downarrow}(A^*) \geq \lambda_2^{\downarrow}(A^*) = \cdots = \lambda_{n-1}^{\downarrow}(A^*) \geq \lambda_n^{\downarrow}(A^*) \geq 0,$$
$$0 \leq \lambda_1^{\uparrow}(B^*) \leq \lambda_2^{\uparrow}(B^*) = \cdots = \lambda_{n-1}^{\uparrow}(B^*) \leq \lambda_n^{\uparrow}(B^*).$$

Then, from $\lambda^{\downarrow}(A^*) \cdot \lambda^{\uparrow}(B^*) = 0$, we deduce that $\lambda_1^{\uparrow}(B^*) = 0$ and $\lambda_1^{\downarrow}(A^*) > 0$. (*If not, we will have* $\lambda_1^{\downarrow}(A^*) = 0$, *which implies* $\lambda(A^*) = 0$, *and then* $A^* = 0$ *follows by the positive semidefiniteness of* A^*. *This contradicts the fact that* $\|A^*\|_F = 1$.) Similarly, we can deduce that $\lambda_n^{\uparrow}(B^*) > 0$ and $\lambda_n^{\downarrow}(A^*) = 0$. Also, either of $\lambda_2^{\downarrow}(A^*)$ and $\lambda_2^{\uparrow}(B^*)$ is zero. Without loss of generality, we assume that $\lambda_2^{\downarrow}(A^*) = 0$. Thus, the above arguments show that

$$\lambda_1^{\downarrow}(A^*) > \lambda_2^{\downarrow}(A^*) = 0 = \cdots = 0 = \lambda_n^{\downarrow}(A^*),$$
$$\lambda_n^{\uparrow}(B^*) \geq \lambda_{n-1}^{\uparrow}(B^*) = \cdots = \lambda_2^{\uparrow}(B^*) \geq \lambda_1^{\uparrow}(B^*) = 0 \text{ and } \lambda_n^{\uparrow}(B^*) > 0.$$

However, from the second relation of (32) and the last two equations, we have that

$$0 = \sum_{j=1}^{n} \lambda_j^{\downarrow}(A^*)\lambda_j^{\downarrow}(B^*) = \lambda_1^{\downarrow}(A^*)\lambda_n^{\uparrow}(B^*) + (n-1)\lambda_2^{\downarrow}(A^*)\lambda_2^{\uparrow}(B^*) + \lambda_n^{\downarrow}(A^*)\lambda_1^{\uparrow}(B^*)$$
$$= \lambda_1^{\downarrow}(A^*)\lambda_n^{\uparrow}(B^*) > 0,$$

which is clearly impossible. Thus, the constant c_1 satisfies the requirement. □

Lemma 4.1 generalizes the result of [12, Lemma 4] for $p = 2$, where it is achieved by direct computation, but we here adopt a different proof technique. Combining Lemma 4.1 with the expressions of $L_{|x|^{p-1}}L_{z^{p-1}}^{-1}$ and $L_{|y|^{p-1}}L_{z^{p-1}}^{-1}$, we have the following result.

Lemma 4.2. *For any* x, y *with* $w(x,y) \in \mathbf{int}\mathcal{K}^n$, *let* $\widetilde{x} = |x|^{p-1}$ *and* $\widetilde{y} = |y|^{p-1}$. *Then,*

$$\frac{\widetilde{x}_1 + (-1)^i \widetilde{x}_2^T \overline{w}_2}{\sqrt[q]{\lambda_i(w)}} = O(1), \qquad \frac{\widetilde{x}_2 + (-1)^i \widetilde{x}_1 \overline{w}_2}{\sqrt[q]{\lambda_i(w)}} = O(1),$$
$$\frac{\widetilde{y}_1 + (-1)^i \widetilde{y}_2^T \overline{w}_2}{\sqrt[q]{\lambda_i(w)}} = O(1), \qquad \frac{\widetilde{y}_2 + (-1)^i \widetilde{y}_1 \overline{w}_2}{\sqrt[q]{\lambda_i(w)}} = O(1) \tag{33}$$

for $i = 1, 2$, *where* $\overline{w}_2 = \frac{w_2(x,y)}{\|w_2(x,y)\|}$, *and* $O(1)$ *denotes a uniformly bounded term.*

Proof. Fix any (x, y) satisfying $w \in \mathbf{int}\mathcal{K}^n$. We write $\lambda_1 = \lambda_1(w)$ and $\lambda_2 = \lambda_2(w)$ for simplicity. From (15), we have $\sqrt[q]{\lambda_2} \geq \sqrt[q]{w_1} \geq \sqrt[q]{\frac{|\lambda_2(x)|^p + |\lambda_1(x)|^p}{2}}$. Note that

$$\sqrt[q]{\frac{|\lambda_2(x)|^p + |\lambda_1(x)|^p}{2}} \geq \left(\frac{\max(|\lambda_2(x)|, |\lambda_1(x)|)}{\sqrt[p]{2}}\right)^{\frac{p}{q}} \geq \left(\frac{\sqrt{|\lambda_2(x)|^2 + |\lambda_1(x)|^2}}{2^{\frac{1}{p}+1}}\right)^{\frac{p}{q}} = \frac{\|x\|^{\frac{p}{q}}}{2^{\frac{p+2}{2q}}}$$

for $p > 2$, and for $1 < p \le 2$,

$$\sqrt[q]{\frac{|\lambda_2(x)|^p + |\lambda_1(x)|^p}{2}} \ge \sqrt[q]{\frac{(|\lambda_2(x)|^2 + |\lambda_1(x)|^2)^{\frac{p}{2}}}{2}} = 2^{\frac{p-2}{2q}} \|x\|^{\frac{p}{q}}.$$

Therefore,

$$\sqrt[q]{\lambda_2} \ge \begin{cases} 2^{-\frac{p+2}{2q}} \|x\|^{\frac{p}{q}} & \text{if } p > 2; \\ 2^{\frac{p-2}{2q}} \|x\|^{\frac{p}{q}} & \text{if } p \in (1, 2]. \end{cases} \tag{34}$$

Since $\widetilde{x}_1 = \frac{1}{2}(|\lambda_2(x)|^{p-1} + |\lambda_1(x)|^{p-1})$ and $\widetilde{x}_2 = \frac{1}{2}(|\lambda_2(x)|^{p-1} - |\lambda_1(x)|^{p-1})$, we have

$$\begin{aligned} \widetilde{x}_1 &= \frac{1}{2}(|\lambda_2(x)|^{\frac{p}{q}} + |\lambda_1(x)|^{\frac{p}{q}}) \le \|x\|^{\frac{p}{q}}, \\ \|\widetilde{x}_2\| &\le \frac{1}{2}(|\lambda_2(x)|^{\frac{p}{q}} + |\lambda_1(x)|^{\frac{p}{q}}) \le \|x\|^{\frac{p}{q}}. \end{aligned} \tag{35}$$

Together with (34), we obtain the first two relations in (33) for $i = 2$. Notice that

$$z^{p-1} = \sqrt[q]{w} = \left(\frac{\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1}}{2}, \frac{\sqrt[q]{\lambda_2} - \sqrt[q]{\lambda_1}}{2} \overline{w}_2 \right).$$

By formula (10) and $\widetilde{x} = |x|^{p-1}$, we calculate that $L_{|x|^{p-1}} L^{-1}_{z^{p-1}}$ equals

$$\begin{bmatrix} \frac{\widetilde{x}_1 + \widetilde{x}_2^T \overline{w}_2}{2\sqrt[q]{\lambda_2}} + \frac{\widetilde{x}_1 - \widetilde{x}_2^T \overline{w}_2}{2\sqrt[q]{\lambda_1}} & \left(\frac{\widetilde{x}_1 \overline{w}_2^T}{2\sqrt[q]{\lambda_2}} - \frac{\widetilde{x}_1 \overline{w}_2^T}{2\sqrt[q]{\lambda_1}} \right) + \frac{2\widetilde{x}_2^T}{\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1}} + \frac{\frac{\sqrt[q]{\lambda_2}}{\sqrt[q]{\lambda_1}} - 2 + \frac{\sqrt[q]{\lambda_1}}{\sqrt[q]{\lambda_2}}}{2(\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1})} \widetilde{x}_2^T \overline{w}_2 \overline{w}_2^T \\ \frac{\widetilde{x}_2 + \widetilde{x}_1 \overline{w}_2}{2\sqrt[q]{\lambda_2}} + \frac{\widetilde{x}_2 - \widetilde{x}_1 \overline{w}_2}{2\sqrt[q]{\lambda_1}} & \left(\frac{\widetilde{x}_2 \overline{w}_2^T}{2\sqrt[q]{\lambda_2}} - \frac{\widetilde{x}_2 \overline{w}_2^T}{2\sqrt[q]{\lambda_1}} \right) + \frac{2\widetilde{x}_1 I}{\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1}} + \frac{\frac{\sqrt[q]{\lambda_2}}{\sqrt[q]{\lambda_1}} - 2 + \frac{\sqrt[q]{\lambda_1}}{\sqrt[q]{\lambda_2}}}{2(\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1})} \widetilde{x}_1 \overline{w}_2 \overline{w}_2^T \end{bmatrix}.$$

Substituting the first two relations in (33) for $i = 2$ into the last equation and noting that

$$\frac{\widetilde{x}_1 \overline{w}_2^T}{2\sqrt[q]{\lambda_2}}, \quad \frac{\widetilde{x}_2 \overline{w}_2^T}{2\sqrt[q]{\lambda_2}}, \quad \frac{\widetilde{x}_2^T}{\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1}}, \quad \text{and} \quad \frac{\widetilde{x}_1}{\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1}}$$

are all uniformly bounded by equations (34)-(35), we obtain that

$$\begin{aligned} L_{|x|^{p-1}} L^{-1}_{z^{p-1}} &= \begin{bmatrix} O(1) + \frac{\widetilde{x}_1 - \widetilde{x}_2^T \overline{w}_2}{2\sqrt[q]{\lambda_1}} & O(1) - \frac{\widetilde{x}_1 \overline{w}_2^T}{2\sqrt[q]{\lambda_1}} + \frac{\sqrt[q]{\lambda_2}}{2(\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1})\sqrt[q]{\lambda_1}} \widetilde{x}_2^T \overline{w}_2 \overline{w}_2^T \\ O(1) + \frac{\widetilde{x}_2 - \widetilde{x}_1 \overline{w}_2}{2\sqrt[q]{\lambda_1}} & O(1) - \frac{\widetilde{x}_2 \overline{w}_2^T}{2\sqrt[q]{\lambda_1}} + \frac{\sqrt[q]{\lambda_2}}{2(\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1})\sqrt[q]{\lambda_1}} \widetilde{x}_1 \overline{w}_2 \overline{w}_2^T \end{bmatrix} \\ &= \begin{bmatrix} O(1) + \frac{\widetilde{x}_1 - \widetilde{x}_2^T \overline{w}_2}{2\sqrt[q]{\lambda_1}} & O(1) - \frac{\widetilde{x}_1 \overline{w}_2^T}{2(\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1})} - \frac{\sqrt[q]{\lambda_2}(\widetilde{x}_1 - \widetilde{x}_2^T \overline{w}_2)}{2(\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1})\sqrt[q]{\lambda_1}} \overline{w}_2^T \\ O(1) + \frac{\widetilde{x}_2 - \widetilde{x}_1 \overline{w}_2}{2\sqrt[q]{\lambda_1}} & O(1) - \frac{\widetilde{x}_2 \overline{w}_2^T}{2(\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1})} + \frac{\sqrt[q]{\lambda_2}(\widetilde{x}_2 - \widetilde{x}_1 \overline{w}_2)}{2(\sqrt[q]{\lambda_2} + \sqrt[q]{\lambda_1})\sqrt[q]{\lambda_1}} \overline{w}_2^T \end{bmatrix}. \end{aligned}$$

This, by Lemma 4.1, implies that the first two relations in (33) hold for $i = 1$. By the symmetry of x and y in $w(x, y)$, the last two relations in (33) also hold. □

Remark 4.1. The first relation of (33) for $i = 1$ is equivalent to saying that

$$\frac{|\lambda_2(x)|^{p-1}(1 - \overline{x}_2^T \overline{w}_2) + |\lambda_1(x)|^{p-1}(1 + \overline{x}_2^T \overline{w}_2)}{\sqrt[q]{\lambda_1(w)}} = O(1), \tag{36}$$

whereas the second relation for $i = 1$ is equivalent to saying that

$$\left\| \frac{(|\lambda_2(x)|^{p-1} - |\lambda_1(x)|^{p-1})\overline{x}_2 - (|\lambda_2(x)|^{p-1} + |\lambda_1(x)|^{p-1})\overline{w}_2}{\sqrt[q]{\lambda_1(w)}} \right\|^2,$$

$$= \frac{|\lambda_2(x)|^{2p-2}(1 - \overline{x}_2^T\overline{w}_2) + |\lambda_1(x)|^{2p-2}(1 + \overline{x}_2^T\overline{w}_2)}{(\sqrt[q]{\lambda_1(w)})^2} = O(1). \tag{37}$$

Equations (36) and (37) play an important role in the proof of the following lemma.

Lemma 4.3. *There exists a positive constant c_2 (independent of x and y) such that for all (x, y) with $w(x, y) \in \mathbf{int}\mathcal{K}^n$,*

$$\left\|\nabla g^{\mathrm{soc}}(x)\nabla g^{\mathrm{soc}}(z)^{-1}\right\|_F \le c_2 \quad \text{and} \quad \left\|\nabla g^{\mathrm{soc}}(y)\nabla g^{\mathrm{soc}}(z)^{-1}\right\|_F \le c_2.$$

Proof. By the symmetry of x and y in $\nabla g^{\mathrm{soc}}(z)$, it suffices to prove the first inequality. Fix any (x, y) with $w = w(x, y) \in \mathbf{int}\mathcal{K}^n$. Suppose $w_2 \neq 0$ and $x_2 \neq 0$. By the expressions of $\nabla g^{\mathrm{soc}}(x)$ and $\nabla g^{\mathrm{soc}}(z)^{-1}$ given by Lemma 3.2, it is not hard to calculate that

$$2p\nabla g^{\mathrm{soc}}(x)\nabla g^{\mathrm{soc}}(z)^{-1} = \begin{bmatrix} a_1(x,z) & a_2^T(x,z) \\ b_2(x,z) & A_1(x,z) \end{bmatrix},$$

where

$$\begin{aligned}
a_1(x,z) &= \frac{1}{\sqrt[q]{\lambda_2(w)}}\left(b(x) + c(x)\overline{x}_2^T\overline{w}_2\right) + \frac{1}{\sqrt[q]{\lambda_1(w)}}\left(b(x) - c(x)\overline{x}_2^T\overline{w}_2\right), \\
a_2(x,z) &= \frac{\left(b(x)+c(x)\overline{x}_2^T\overline{w}_2\right)\overline{w}_2}{\sqrt[q]{\lambda_2(w)}} - \frac{\left(b(x)-c(x)\overline{x}_2^T\overline{w}_2\right)\overline{w}_2}{\sqrt[q]{\lambda_1(w)}} + \frac{2pc(x)\left(\overline{x}_2 - \overline{x}_2^T\overline{w}_2\overline{w}_2\right)}{a(z)}, \\
b_2(x,z) &= \frac{1}{\sqrt[q]{\lambda_2(w)}}\left[c(x)\overline{x}_2 + a(x)\overline{w}_2 + (b(x) - a(x))\overline{x}_2^T\overline{w}_2\overline{x}_2\right] \\
&\quad + \frac{1}{\sqrt[q]{\lambda_1(w)}}\left[c(x)\overline{x}_2 - a(x)\overline{w}_2 - (b(x) - a(x))\overline{x}_2^T\overline{w}_2\overline{x}_2\right], \\
A_1(x,z) &= \frac{1}{\sqrt[q]{\lambda_2(w)}}\left[c(x)\overline{x}_2\overline{w}_2^T + a(x)\overline{w}_2\overline{w}_2^T + (b(x) - a(x))\overline{x}_2^T\overline{w}_2\overline{x}_2\overline{w}_2^T\right] \\
&\quad - \frac{1}{\sqrt[q]{\lambda_1(w)}}\left[c(x)\overline{x}_2\overline{w}_2^T - a(x)\overline{w}_2\overline{w}_2^T - (b(x) - a(x))\overline{x}_2^T\overline{w}_2\overline{x}_2\overline{w}_2^T\right] \\
&\quad + \frac{2p}{a(z)}\left[a(x)(I - \overline{w}_2\overline{w}_2^T) + (b(x) - a(x))\left(\overline{x}_2\overline{x}_2^T - \overline{x}_2^T\overline{w}_2\overline{x}_2\overline{w}_2^T\right)\right].
\end{aligned}$$

From the definitions of $a(x)$, $b(x)$ and $c(x)$ in (20), it follows that

$$\max(|b(x)|, |c(x)|) \le \frac{p}{2}\left(|\lambda_2(x)|^{p-1} + |\lambda_1(x)|^{p-1}\right),$$

$$|a(x)| = p\left|t_1\lambda_2(x) + (1 - t_1)\lambda_1(x)\right|^{p-1} \le p\max(|\lambda_2(x)|^{p-1}, |\lambda_1(x)|^{p-1}) \tag{38}$$

for some $t_1 \in (0, 1)$, where the equality is using the mean-value theorem. Therefore,

$$|a(x)| \le p\|x\|^{\frac{p}{q}}, \quad |b(x)| \le p\|x\|^{\frac{p}{q}} \quad \text{and} \quad |c(x)| \le p\|x\|^{\frac{p}{q}}. \tag{39}$$

Noting that $0 \le \lambda_1(w)/\lambda_2(w) < 1$ and $\sqrt[p]{\lambda_1(w)/\lambda_2(w)} \ge \lambda_1(w)/\lambda_2(w)$, we have

$$a(z) = \frac{\lambda_2(w) - \lambda_1(w)}{\sqrt[p]{\lambda_2(w)} - \sqrt[p]{\lambda_1(w)}} = \sqrt[q]{\lambda_2(w)}\frac{1 - \lambda_1(w)/\lambda_2(w)}{1 - \sqrt[p]{\lambda_1(w)/\lambda_2(w)}} \ge \sqrt[q]{\lambda_2(w)}. \tag{40}$$

By (39), (40), and (34), we simplify $a_1(x,z), a_2(x,z), b_1(x,z)$, and $A_1(x,z)$ as

$$
\begin{aligned}
a_1(x,z) &= O(1) + \frac{1}{\sqrt[q]{\lambda_1(w)}}\left(b(x) - c(x)\overline{x}_2^T\overline{w}_2\right),\\
a_2(x,z) &= O(1) - \frac{1}{\sqrt[q]{\lambda_1(w)}}\left(b(x) - c(x)\overline{x}_2^T\overline{w}_2\right)\overline{w}_2, \qquad (41)\\
b_2(x,z) &= O(1) + \frac{1}{\sqrt[q]{\lambda_1(w)}}\left[\left(c(x) - b(x)\overline{x}_2^T\overline{w}_2\right)\overline{x}_2 + a(x)(\overline{x}_2^T\overline{w}_2\overline{x}_2 - \overline{w}_2)\right],\\
A_1(x,z) &= O(1) - \frac{1}{\sqrt[q]{\lambda_1(w)}}\left[\left(c(x) - b(x)\overline{x}_2^T\overline{w}_2\right)\overline{x}_2\overline{w}_2^T + a(x)(\overline{x}_2^T\overline{w}_2\overline{x}_2\overline{w}_2^T - \overline{w}_2\overline{w}_2^T)\right].
\end{aligned}
$$

By the definitions of $b(x)$ and $c(x)$, it is easy to verify that

$$
\begin{aligned}
\left|b(x) - c(x)\overline{x}_2^T\overline{w}_2\right| &\le \frac{p}{2}\left[|\lambda_2(x)|^{p-1}(1 - \overline{x}_2^T\overline{w}_2) + |\lambda_1(x)|^{p-1}(1 + \overline{x}_2^T\overline{w}_2)\right],\\
\left|c(x) - b(x)\overline{x}_2^T\overline{w}_2\right| &\le \frac{p}{2}\left[|\lambda_2(x)|^{p-1}(1 - \overline{x}_2^T\overline{w}_2) + |\lambda_1(x)|^{p-1}(1 + \overline{x}_2^T\overline{w}_2)\right],
\end{aligned}
$$

which, together with (36), implies that

$$
\frac{b(x) - c(x)\overline{x}_2^T\overline{w}_2}{\sqrt[q]{\lambda_1(w)}} = O(1), \quad \frac{c(x) - b(x)\overline{x}_2^T\overline{w}_2}{\sqrt[q]{\lambda_1(w)}} = O(1). \qquad (42)
$$

In addition, it is easy to compute that

$$
\begin{aligned}
\|a(x)(\overline{x}_2^T\overline{w}_2\overline{x}_2 - \overline{w}_2)\|^2 &= a^2(x)(1 - \overline{x}_2^T\overline{w}_2)(1 + \overline{x}_2^T\overline{w}_2),\\
\|a(x)(\overline{x}_2^T\overline{w}_2\overline{x}_2\overline{w}_2^T - \overline{w}_2\overline{w}_2^T)\|_F^2 &\le a^2(x)(1 - \overline{x}_2^T\overline{w}_2)(1 + \overline{x}_2^T\overline{w}_2).
\end{aligned}
$$

By equation (38), we have $a^2(x) \le p^2\max(|\lambda_2(x)|^{2p-2}, |\lambda_1(x)|^{2p-2})$. Using (37) and noting that $0 \le 1 - \overline{x}_2^T\overline{w}_2 \le 2$ and $0 \le 1 + \overline{x}_2^T\overline{w}_2 \le 2$, we may obtain that

$$
\frac{\|a(x)(\overline{x}_2^T\overline{w}_2\overline{x}_2 - \overline{w}_2)\|}{\sqrt[q]{\lambda_1(w)}} = O(1), \quad \frac{\|a(x)(\overline{x}_2^T\overline{w}_2\overline{x}_2\overline{w}_2^T - \overline{w}_2\overline{w}_2^T)\|_F}{\sqrt[q]{\lambda_1(w)}} = O(1). \qquad (43)
$$

By (41)-(43), $a_1(x,z), a_2(x,z), b_2(x,z)$, and $A_1(x,z)$ are all uniformly bounded, and hence there exists a constant $C_2 > 0$ such that $\left\|\nabla g^{\rm soc}(x)[\nabla g^{\rm soc}(z)]^{-1}\right\|_F \le C_2$.

Suppose that $x_2 = 0$ or $w_2 = 0$. Then there exists a sequence $\{(x^k, y^k)\} \subset \mathbb{R}^n \times \mathbb{R}^n$ with $x_2^k \ne 0$, $w_2(x^k, y^k) \ne 0$ and $w(x^k, y^k) \in \mathbf{int}\mathcal{K}^n$ for all k such that $x^k \to x$ and $w^k \to w$ as $k \to \infty$. From the above result, $\|\nabla g^{\rm soc}(x^k)\nabla g^{\rm soc}(z^k)^{-1}\|_F \le c_2$ for all k. Noting that $\nabla g^{\rm soc}(x)$ is continuous since $|t|^p$ is continuously differentiable, and $\nabla g^{\rm soc}(z)^{-1}$ is continuous at any $z(x,y) \in \mathbf{int}\mathcal{K}^n$, we get $\|\nabla g^{\rm soc}(x)\nabla g^{\rm soc}(z)^{-1}\|_F \le c_2$. The proof is complete. □

Now we are in a position to prove the continuity of the gradient function $\nabla\psi_p$.

Proposition 4.1. *The function ψ_p with $p \in (1,4)$ is smooth everywhere on $\mathbb{R}^n \times \mathbb{R}^n$.*

Proof. By Proposition 3.1 and the symmetry of x and y in $\nabla\psi_p$, it suffices to prove that $\nabla_x\psi_p$ is continuous at every $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$. Choose a point $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$ arbitrarily. When $w(x,y) \in \mathbf{int}\mathcal{K}^n$, the conclusion was shown in Proposition 3.1. We next consider the other two cases where $w(x,y) = 0$ and $w(x,y) \in \mathbf{bd}^+\mathcal{K}^n$.

Case 1. $w(x,y) = 0$. Now we have $(x,y) = (0,0)$ and $\nabla_x\psi_p(0,0) = 0$ by Proposition 3.1 it suffices to show that $\nabla_{x'}\psi_p(x',y') \to 0$ as $(x',y') \to (0,0)$. If $w(x',y') \in \mathbf{int}\mathcal{K}^n$, then $\nabla_{x'}\psi_p(x',y')$ is given by (21); and if $w(x',y') \in \mathbf{bd}^+\mathcal{K}^n$, then

$\nabla_{x'}\psi_p(x',y')$ is given by (22). Since $\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}-I$ and $\frac{{\rm sign}(x_1')|x_1'|^{p-1}}{\sqrt[q]{|x_1'|^p+|y_1'|^p}}$ are uniformly bounded, where the uniform boundedness of the former is due to Lemma 4.3, using the continuity of ϕ_p and noting that $\phi_p(0,0)=0$ immediately yields that $\nabla_{x'}\psi_p(x',y')\to 0$ as $(x',y')\to(0,0)$.

Case 2. $w(x,y)\in\mathbf{bd}^+\mathcal{K}^n$. For any (x',y') sufficiently close to (x,y), in order to prove that $\nabla_x\psi_p(x',y')\to\nabla_x\psi_p(x,y)$, we only need to consider the cases where $w(x',y')\in\mathbf{int}\mathcal{K}^n$ and $w(x',y')\in\mathbf{bd}^+\mathcal{K}^n$. When $w(x',y')\in\mathbf{bd}^+\mathcal{K}^n$, $\nabla_x\psi_p(x',y')$ has an expression of (22) which is continuous at (x,y) since $|x_1|^p+|y_1|^p>0$ by Lemma 3.1, and then $\nabla_x\psi_p(x',y')\to\nabla_x\psi_p(x,y)$. We next concentrate on the case $w(x',y')\in\mathbf{int}\mathcal{K}^n$, for which case

$$\begin{aligned}\nabla_x\psi_p(x',y') =&\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}z'\\ &-\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}(x'+y')-\phi_p(x',y').\end{aligned}\tag{44}$$

We next proceed with the arguments for the following two subcases: $x_2\neq 0$ and $x_2=0$.

Subcase (2.1). $x_2\neq 0$. By the expression of $\nabla_x\psi_p(x,y)$ in (22), we have

$$\begin{aligned}\nabla_x\psi_p(x,y) &= \frac{{\rm sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p+|y_1|^p}}\phi_p(x,y)-\phi_p(x,y)\\ &= \frac{{\rm sign}(x_1)|x_1|^{p-1}\sqrt[p]{|x_1|^p+|y_1|^p}}{\sqrt[q]{|x_1|^p+|y_1|^p}}\begin{pmatrix}1\\ \overline{w}_2\end{pmatrix}\\ &\quad-\frac{{\rm sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p+|y_1|^p}}(x+y)-\phi_p(x,y),\end{aligned}$$

where the second equality is by $z(x,y)=\sqrt[p]{|x_1|^p+|y_1|^p}\begin{pmatrix}1\\ \overline{w}_2\end{pmatrix}$. Comparing it with (44), we see that to prove $\nabla_x\psi_p(x',y')\to\nabla_x\psi_p(x,y)$ as $(x',y')\to(x,y)$, it suffices to argue that, when $(x',y')\to(x,y)$,

$$\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}z'\to\frac{{\rm sign}(x_1)|x_1|^{p-1}\sqrt[p]{|x_1|^p+|y_1|^p}}{\sqrt[q]{|x_1|^p+|y_1|^p}}\begin{pmatrix}1\\ \overline{w}_2\end{pmatrix}\tag{45}$$

and

$$\begin{aligned}\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}x' &\to \frac{{\rm sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p+|y_1|^p}}x,\\ \nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}y' &\to \frac{{\rm sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p+|y_1|^p}}y.\end{aligned}\tag{46}$$

First of all, we prove (45). Since $w(x,y)\in\mathbf{bd}^+\mathcal{K}^n$ implies $|x|^p\in\mathbf{bd}\mathcal{K}^n$, we have

$$a(x)=2^{p-1}{\rm sign}(x_1)|x_1|^{p-1},\ b(x)=2^{p-2}p\ {\rm sign}(x_1)|x_1|^{p-1},\ c(x)=2^{p-2}p|x_1|^{p-1}.\tag{47}$$

Since $w_2(x,y)\neq 0$ (if not, $w(x,y)=0$), we have $w_2'=w_2(x',y')\neq 0$. By the expressions of $\nabla g^{\rm soc}(x')$ and $\nabla g^{\rm soc}(z')^{-1}$, it is not hard to calculate that $\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}z'$ equals

$$\frac{\lambda_2(w')^{\frac{1}{p}-\frac{1}{q}}}{2p}\nabla g^{\rm soc}(x')\begin{pmatrix}1\\ \overline{w}_2'\end{pmatrix}+\frac{\sqrt[p]{\lambda_1(w')}}{2}\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}\begin{pmatrix}1\\ -\overline{w}_2'\end{pmatrix},$$

where $z' = z(x', y'), w' = w(x', y')$, and $\overline{w}_2' = \frac{w_2'}{\|w_2'\|}$. Note that $\lambda_1(w') \to 0$ and $\overline{w}_2' \to \overline{w}_2$ with $\overline{w}_2 = \frac{w_2}{\|w_2\|}$ as $(x', y') \to (x, y)$. By Lemma 4.3, the last term on the right-hand side tends to 0, whereas by the continuity of $\nabla g^{\rm soc}$ the first term approaches $\frac{1}{2p}\lambda_2(w)^{\frac{1}{p}-\frac{1}{q}}\nabla g^{\rm soc}(x)\begin{pmatrix}1\\ \overline{w}_2\end{pmatrix}$. Thus, together with $\lambda_2(w) = 2w_1 = 2^p(|x_1|^p + |y_1|^p)$, it holds that as $(x', y') \to (x, y)$,

$$\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}z' \to p^{-1}2^{-\frac{p}{q}}(|x_1|^p + |y_1|^p)^{\frac{1}{p}-\frac{1}{q}}\nabla g^{\rm soc}(x)\begin{pmatrix}1\\ \overline{w}_2\end{pmatrix}. \tag{48}$$

In addition, using equations (47) and (19), we readily obtain that

$$\nabla g^{\rm soc}(x) = 2^{p-2}p\ {\rm sign}(x_1)|x_1|^{p-1}\begin{bmatrix}1 & \frac{x_2^T}{x_1}\\ \frac{x_2}{x_1} & \frac{2}{p}I + \left(1-\frac{2}{p}\right)\frac{x_2x_2^T}{x_1^2}\end{bmatrix}.$$

This along with (48) means that as $(x', y') \to (x, y)$, $\nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}z'$ approaches

$$\begin{aligned}&\frac{1}{2}{\rm sign}(x_1)|x_1|^{p-1}(|x_1|^p + |y_1|^p)^{\frac{1}{p}-\frac{1}{q}}\begin{bmatrix}1 & x_2^T/x_1\\ \frac{x_2}{x_1} & \frac{2}{p}I + \left(1-\frac{2}{p}\right)\frac{x_2x_2^T}{x_1^2}\end{bmatrix}\begin{pmatrix}1\\ \overline{w}_2\end{pmatrix},\\ &= {\rm sign}(x_1)|x_1|^{p-1}(|x_1|^p + |y_1|^p)^{\frac{1}{p}-\frac{1}{q}}\begin{pmatrix}1\\ \overline{w}_2\end{pmatrix}\end{aligned}$$

where the equality is using Lemma 3.1. This shows that equation (45) holds.

Next, we prove the second relation of (46), and an analogous argument can be used to prove the first relation. Let $(\zeta_1, \zeta_2) := \nabla g^{\rm soc}(x')\nabla g^{\rm soc}(z')^{-1}y'$. We only need to establish, when $(x', y') \to (x, y)$,

$$\zeta_1 \to \frac{{\rm sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p + |y_1|^p}}y_1 \quad\text{and}\quad \zeta_2 \to \frac{{\rm sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p + |y_1|^p}}y_2. \tag{49}$$

Note that $x_2' \neq 0$ for (x', y') sufficiently close to (x, y). By equation (19) and the expression of $\nabla g^{\rm soc}(z')^{-1}$ in Lemma 3.2, a direct calculation yields that

$$\begin{aligned}2p\zeta_1 = &\frac{1}{\sqrt[q]{\lambda_2(w')}}\left[b(x') + c(x')(\overline{x}_2')^T\overline{w}_2'\right]\left[y_1' + (\overline{w}_2')^Ty_2'\right]\\ &+ \frac{1}{\sqrt[q]{\lambda_1(w')}}\left[b(x') - c(x')(\overline{x}_2')^T\overline{w}_2'\right]\left[y_1' - (\overline{w}_2')^Ty_2'\right]\\ &+ \frac{2p}{a(z')}c(x')\left[(\overline{x}_2')^Ty_2' - (\overline{x}_2')^T\overline{w}_2'(\overline{w}_2')^Ty_2'\right],\end{aligned} \tag{50}$$

(51)

$$
\begin{aligned}
2p\zeta_2 = & \frac{1}{\sqrt[q]{\lambda_2(w')}}\left[c(x')\overline{x}_2' + a(x')\overline{w}_2' + (b(x')-a(x'))(\overline{x}_2')^T\overline{w}_2'\overline{x}_2'\right] y_1' \\
& + \frac{1}{\sqrt[q]{\lambda_2(w')}}\left[c(x')\overline{x}_2'(\overline{w}_2')^T + a(x')\overline{w}_2'(\overline{w}_2')^T + (b(x')-a(x'))(\overline{x}_2')^T\overline{w}_2'\overline{x}_2'(\overline{w}_2')^T\right] y_2' \\
& + \frac{2p}{a(z')}\left[a(x')(I-\overline{w}_2'(\overline{w}_2')^T) + (b(x')-a(x'))\left(\overline{x}_2'(\overline{x}_2')^T - (\overline{x}_2')^T\overline{w}_2'\overline{x}_2'(\overline{w}_2')^T\right)\right] y_2' \\
& + \frac{1}{\sqrt[q]{\lambda_1(w')}}\left[c(x')\overline{x}_2' - a(x')\overline{w}_2' - (b(x')-a(x'))(\overline{x}_2')^T\overline{w}_2'\overline{x}_2'\right] y_1', \\
& - \frac{1}{\sqrt[q]{\lambda_1(w')}}\left[c(x')\overline{x}_2'(\overline{w}_2')^T - a(x')\overline{w}_2'(\overline{w}_2')^T - (b(x')-a(x'))(\overline{x}_2')^T\overline{w}_2'\overline{x}_2'(\overline{w}_2')^T\right] y_2'
\end{aligned}
$$

where $a(x'), b(x')$ and $c(x')$ are defined as in (20) with x replaced by x'. Since $\sqrt[q]{\lambda_2(w')}$, $b(x'), c(x'), \overline{x}_2'$ and $\overline{w}_2'$ are continuous at (x,y), it follows that

$$
\begin{aligned}
\sqrt[q]{\lambda_2(w')} \ &\to \ \sqrt[q]{2w_1} = 2^{\frac{p}{q}}(|x_1|^p + |y_1|^p)^{\frac{1}{q}}, \\
\left[b(x') + c(x')(\overline{x}_2')^T\overline{w}_2'\right]\left[y_1' + (\overline{w}_2')^T y_2'\right] &\to \left(b(x) + c(x)\overline{x}_2^T\overline{w}_2\right)\left(y_1 + y_2^T\overline{w}_2\right)
\end{aligned}
$$

as $(x',y') \to (x,y)$. This, along with Lemma 3.1 and equation (47), implies that the first term on the right-hand side of (50) tends to $\frac{2p\ \mathrm{sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p+|y_1|^p}}y_1$. Since

$$
\begin{aligned}
y_1' - (\overline{w}_2')^T y_2' \ &\to \ y_1 - y_2^T\overline{w}_2 = 0, \\
(\overline{x}_2')^T y_2' - (\overline{x}_2')^T\overline{w}_2'(\overline{w}_2')^T y_2' \ &\to \ \overline{x}_2^T y_2 - \overline{x}_2^T\overline{w}_2\overline{w}_2^T y_2 = 0,
\end{aligned}
$$

whereas $(b(x') - c(x')(\overline{x}_2')^T\overline{w}_2')/\sqrt[q]{\lambda_1(w')}$ and $2pc(x')/a(z')$ are uniformly bounded by the proof of Lemma 4.2, the last two terms of (50) tend to 0 as $(x',y') \to (x,y)$, and we prove the first relation in (49). We next prove the second relation of (49). From the above discussions, the first three terms on the right-hand side of (51), respectively, tend to

$$
\begin{aligned}
& \frac{1}{\sqrt[q]{2w_1}}\left[c(x)\overline{x}_2 + a(x)\overline{w}_2 + (b(x)-a(x))\overline{x}_2^T\overline{w}_2\overline{x}_2\right] y_1, \\
& \frac{1}{\sqrt[q]{2w_1}}\left[c(x)\overline{x}_2 + a(x)\overline{w}_2 + (b(x)-a(x))\overline{x}_2^T\overline{w}_2\overline{x}_2\right] \overline{w}_2^T y_2, \\
& \frac{2p}{a(z)}\left[a(x)(I-\overline{w}_2\overline{w}_2^T) + (b(x)-a(x))\left(\overline{x}_2\overline{x}_2^T - \overline{x}_2^T\overline{w}_2\overline{x}_2\overline{w}_2^T\right)\right] y_2,
\end{aligned}
$$

as $(x',y') \to (x,y)$, whose sum, by Lemma 3.1 and formula (47), can be simplified as

$$
\frac{2}{\sqrt[q]{2w_1}}\left[\mathrm{sign}(x_1)c(x) + b(x)\right] y_2 = \frac{2p\ \mathrm{sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p+|y_1|^p}}y_2.
$$

Observe that the sum of the last two terms on the right-hand side of (51) can be rewritten as

$$
\frac{1}{\sqrt[q]{\lambda_1(w')}}\left[c(x')\overline{x}_2' - a(x')\overline{w}_2' - (b(x')-a(x'))(\overline{x}_2')^T\overline{w}_2'\overline{x}_2'\right]\left(y_1' - (\overline{w}_2')^T y_2'\right),
$$

which clearly tends to zero as $(x',y') \to (x,y)$, since the first term is uniformly bounded by the proof of Lemma 4.2, whereas the term $y_1' - (\overline{w}_2')^T y_2' \to y_1 - \overline{w}_2^T y_2 = 0$. Thus, we complete the proof of the second relation in (49). Consequently, the second relation in (45) follows. This shows that $\nabla_{x'}\psi_p(x',y') \to \nabla_x\psi_p(x,y)$ as $(x',y') \to (x,y)$.

Subcase (2.2). $x_2 = 0$. Now we have $x = 0$ from $|x|^p \in \mathbf{bd}\mathcal{K}^n$, and $\nabla g^{\mathrm{soc}}(x) = 0$. Hence,

$$\nabla_x \psi_p(x,y) = \frac{\mathrm{sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{|x_1|^p + |y_1|^p}}\phi_p(x,y) - \phi_p(x,y) = -\phi_p(0,y). \tag{52}$$

On the other hand, since $\nabla g^{\mathrm{soc}}(x) = 0$, it follows from (48) that

$$\nabla g^{\mathrm{soc}}(x')\nabla g^{\mathrm{soc}}(z')^{-1}z' \to 0 \quad \text{as} \quad (x',y') \to (x,y);$$

while using Lemma 4.3 and $x = 0$, we have

$$\nabla g^{\mathrm{soc}}(x')\nabla g^{\mathrm{soc}}(z')^{-1}x' \to 0 \quad \text{as} \quad (x',y') \to (x,y);$$

and from the continuity of ϕ_p and $x = 0$, it follows that

$$\phi_p(x',y)' \to \phi_p(0,y) \quad \text{as} \quad (x',y') \to (x,y).$$

Using the last three equations and comparing (44) with (52), we see, in order to prove that $\nabla_x \psi_p(x',y') \to \nabla_x \psi_p(x,y)$ as $(x',y') \to (x,y)$, it suffices to show that

$$\nabla g^{\mathrm{soc}}(x')\nabla g^{\mathrm{soc}}(z')^{-1}y' \to 0 \quad \text{as} \quad (x',y') \to (x,y). \tag{53}$$

Next, for any (x',y') sufficiently close to (x,y), we write

$$(\zeta_1,\zeta_2) := \nabla g^{\mathrm{soc}}(x')\nabla g^{\mathrm{soc}}(z')^{-1}y'.$$

If $x_2' \neq 0$, then $2p\zeta_1$ and $2p\zeta_2$ are given by (50) and (51), respectively. Using the same arguments as Subcase 2.1, we have that the second term of (50) and the sum of the last two terms of (51) tend to 0 as $(x',y') \to (x,y)$. Since $\sqrt[q]{\lambda_2(w')}, a(z'), b(x'), c(x')$, and $\overline{w}_2'$ are continuous at (x,y), $\|\overline{x}_2'\| = 1$, and $b(x) = c(x) = 0$ from (47) and $x = 0$, the first term and the third term of (50) also tend to 0. This proves that $2p\zeta_1 \to 0$ as $(x',y') \to (x,y)$. We next prove that the first three terms of (51) also tend to 0. From the mean-value theorem, $|a(x')| = p\,|t_1\lambda_2(x') + (1-t_1)\lambda_1(x')|^{p-1}$ for some $t_1 \in (0,1)$. Note that the function $|t|^{p-1}$ $(p > 1)$ is continuous on $\mathbb{R}$, whereas $\lambda_2(x') \to 0$ and $\lambda_1(x') \to 0$ as $(x',y') \to (x,y)$. So, $|a(x')| \to 0$ when $(x',y') \to (x,y)$. In addition, as $(x',y') \to (x,y)$,

$$b(x') \to 0, \quad c(x') \to 0, \quad \sqrt[q]{\lambda_2(w')} \to \sqrt[q]{2w_1} > 0, \quad \text{and} \quad a(z') \to a(z) > 0.$$

This implies that the first three terms of (51) also tend to 0. Consequently, $2p\zeta_2 \to 0$ as $(x',y') \to (x,y)$. Thus, (53) holds for this case.

If $x_2' = 0$, then using (19) and the expression of $\nabla g^{\mathrm{soc}}(z')^{-1}$ in Lemma 3.2, we have

$$\begin{aligned}
\zeta_1 &= \frac{p\mathrm{sign}(x_1')|x_1'|^{p-1}}{\sqrt[q]{\lambda_2(w')}}\left[y_1' + (\overline{w}_2')^T y_2'\right] + \frac{p\mathrm{sign}(x_1')|x_1'|^{p-1}}{\sqrt[q]{\lambda_1(w')}}\left[y_1' - (\overline{w}_2')^T y_2'\right], \\
\zeta_2 &= \frac{p\mathrm{sign}(x_1')|x_1'|^{p-1}}{\sqrt[q]{\lambda_2(w')}}\left[y_1' + (\overline{w}_2')^T y_2'\right]\overline{w}_2' - \frac{p\mathrm{sign}(x_1')|x_1'|^{p-1}}{\sqrt[q]{\lambda_1(w')}}\left[y_1' - (\overline{w}_2')^T y_2'\right]\overline{w}_2' \\
&\quad + \frac{2p^2\,\mathrm{sign}(x_1')|x_1'|^{p-1}}{a(z')}\left[y_2' - \overline{w}_2'(\overline{w}_2')^T y_2'\right].
\end{aligned}$$

Since $\mathrm{sign}(x_1')|x_1'|^{p-1}$ is continuous and $\sqrt[q]{\lambda_2(w)} > 0$, we have, as $(x',y') \to (x,y)$,

$$\frac{p\mathrm{sign}(x_1')|x_1'|^{p-1}}{\sqrt[q]{\lambda_2(w')}}[y_1' + (\overline{w}_2')^T y_2'] \;\to\; \frac{p\mathrm{sign}(x_1)|x_1|^{p-1}}{\sqrt[q]{\lambda_2(w)}}(y_1 + \overline{w}_2^T y_2) = 0.$$

In addition, $|x_1'|^{p-1}/\sqrt[q]{\lambda_1(w')}$ is bounded with the bound independent of x' and y' because $\lambda_1(w') = w_1' - \|w_2'\| \geq |x_1'|^p$ by (15) when $x_2' = 0$. Besides, $y_1' - (\overline{w}_2')^T y_2' \to y_1 - \overline{w}_2^T y_2 = 0$ as $(x', y') \to (x, y)$, where the equality is due to Lemma 3.1. Hence we have

$$\frac{p\mathrm{sign}(x_1')|x_1'|^{p-1}}{\sqrt[q]{\lambda_1(w')}}(y_1' - (\overline{w}_2')^T y_2') \to 0 \quad \text{as} \quad (x', y') \to (x, y).$$

Thus, we prove that $\zeta_1 \to 0$ and the first two terms of ζ_2 tend to 0 as $(x', y') \to (x, y)$. Since $a(z') = \frac{\lambda_2(w') - \lambda_1(w')}{\sqrt[p]{\lambda_2(w')} - \sqrt[p]{\lambda_1(w')}}$ and $(y_2' - \overline{w}_2'(\overline{w}_2')^T y_2')$ are continuous, we have

$$a(z') \to \frac{\lambda_2(w) - \lambda_1(w)}{\sqrt[p]{\lambda_2(w)} - \sqrt[p]{\lambda_1(w)}} = \sqrt[q]{\lambda_2(w)}$$

and

$$\left[y_2' - \overline{w}_2'(\overline{w}_2')^T y_2'\right] \to y_2 - \overline{w}_2 \overline{w}_2^T y_2 = 0$$

as $(x', y') \to (x, y)$, where the last equality is due to Lemma 3.1. This, together with $\mathrm{sign}(x_1')|x_1'|^{p-1} \to \mathrm{sign}(x_1)|x_1|^{p-1} = 0$, means that the last term of ζ_2 also tends to 0. Thus, we show that ζ_2 tends to 0 as $(x', y') \to (x, y)$. Consequently, (53) holds in this case. □

Remark 4.2. Note that the proof of Proposition 4.1 is also suitable for $p \geq 4$. Hence, if ψ_p with $p \geq 4$ is differentiable, then it must be continuously differentiable.

To close this section, we present a property of the partial gradients $\nabla_x \psi_p$ and $\nabla_y \psi_p$. Since the proof is easy by a direct calculation and Lemma 3.1, we omit it.

Lemma 4.4. *For any $x, y \in \mathbb{R}^n$, it always holds that*

$$\langle x, \nabla_x \psi_p(x, y)\rangle + \langle y, \nabla_y \psi_p(x, y)\rangle = \|\phi_p(x, y)\|^2.$$

5. Coerciveness of the function Ψ_p

In this section, we study under what conditions the merit function Ψ_p is coercive, i.e., $\limsup_{\|x\|\to\infty} \Psi_p(x) = \infty$, which plays a crucial role in analyzing the global convergence of merit function methods and equation-based methods based on ϕ_p. The following two technical lemmas are needed.

Lemma 5.1. *Let ϕ_p and ψ_p be given by* (7) *and* (6)*, respectively. Then,*

$$4\psi_p(x, y) \geq 2\left\|[\phi_p(x, y)]_+\right\|^2 \geq \max\left(\|(-x)_+\|^2, \|(-y)_+\|^2\right) \quad \forall x, y \in \mathbb{R}^n,$$

where $(\cdot)_+$ means the minimum Euclidean distance projection onto $\mathcal{K}^n$.

Proof. The first inequality is direct, and the second one is due to [12, Lemma 7(c)] and the fact that $\sqrt[p]{|x|^p + |y|^p} - x \in \mathcal{K}^n$ and $\sqrt[p]{|x|^p + |y|^p} - y \in \mathcal{K}^n$. □

Lemma 5.2. *Assume that $\{(x^k, y^k)\} \subseteq \mathbb{R}^n \times \mathbb{R}^n$ satisfies either of the conditions*

- **(a):** *$\lambda_1(x^k) \to -\infty$ or $\lambda_1(y^k) \to -\infty$;*
- **(b):** *$\{\lambda_1(x^k)\}$ and $\{\lambda_1(y^k)\}$ are bounded below, $\lambda_2(x^k), \lambda_2(y^k) \to +\infty$, and $\langle \frac{x^k}{\|x^k\|}, \frac{y^k}{\|y^k\|}\rangle \nrightarrow 0$.*

Then, when p is a rational number, it holds that $\limsup_{k\to\infty} \psi_p(x^k, y^k) = +\infty$.

Proof. If $\{(x^k, y^k)\}$ satisfies (a), the result follows from Lemma 5.1 and the fact

$$2\|(-x^k)_+\|^2 = \min\left(0, \lambda_1(x^k)\right)^2 + \min\left(0, \lambda_2(x^k)\right)^2.$$

It remains to consider the case where $\{(x^k, y^k)\}$ satisfies (b). Now from the given assumptions we have (taking a subsequence if necessary), $x_1^k \to +\infty$ and $y_1^k \to +\infty$. Without loss of generality, we assume (subsequencing if necessary) that

$$\lim_{k\to\infty} x^k/\|x^k\| = x^* \quad \text{and} \quad \lim_{k\to\infty} y^k/\|y^k\| = y^*. \tag{54}$$

Since $\{\lambda_1(x^k)\}$ and $\{\lambda_1(y^k)\}$ are bounded below, there exists a fixed element $d \in \mathbb{R}^n$ such that $x^k - d \in \mathbf{int}\mathcal{K}^n$ and $y^k - d \in \mathbf{int}\mathcal{K}^n$ for each k. (*Indeed, letting γ be the lower bound of $\{\lambda_1(x^k)\}$ and $\{\lambda_1(y^k)\}$, we have $x^k - (\gamma-1)e \in \mathbf{int}\mathcal{K}^n$ and $y^k - (\gamma-1)e \in \mathbf{int}\mathcal{K}^n$ since $\lambda_1(z^k - (\gamma-1)e) \geq \lambda_1(z^k) + \lambda_1((1-\gamma)e) \geq \gamma + 1 - \gamma = 1$ for $z^k = x^k$ or y^k.*) So, $\frac{x^k - d}{\|x^k\|} \in \mathbf{int}\mathcal{K}^n$ and $\frac{y^k - d}{\|y^k\|} \in \mathbf{int}\mathcal{K}^n$ for each k. This implies that $\frac{x^k}{\|x^k\|} \in \mathcal{K}^n$ and $\frac{y^k}{\|y^k\|} \in \mathcal{K}^n$, and consequently $x^k \in \mathcal{K}^n$ and $y^k \in \mathcal{K}^n$, for all sufficiently large k. We will continue the arguments using three cases as shown below, where all k are assumed to be sufficiently large.

Case 1. The sequence $\{\|x^k\|/\|y^k\|\}$ is unbounded. Since p is a rational number, we may write $p = n/m$ with n, m being natural numbers and $n > m$. Suppose that the conclusion does not hold, i.e., $\{\phi_p(x^k, y^k)\}$ is bounded. From the definition of ϕ_p and $x^k, y^k \in \mathcal{K}^n$, we have $(x^k)^{\frac{n}{m}} + (y^k)^{\frac{n}{m}} = \left[x^k + y^k + \phi_p(x^k, y^k)\right]^{\frac{n}{m}}$, which is equivalent to

$$\left[(x^k)^{\frac{n}{m}} + (y^k)^{\frac{n}{m}}\right]^m = \left[x^k + y^k + \phi_p(x^k, y^k)\right]^n. \tag{55}$$

Since $\{\|x^k\|/\|y^k\|\}$ is unbounded, $\|x^k\| \to +\infty$, $\|y^k\| \to +\infty$, and $n > m$, by expanding $\left[(x^k)^{\frac{n}{m}} + (y^k)^{\frac{n}{m}}\right]^m = \underbrace{\left[(x^k)^{\frac{n}{m}} + (y^k)^{\frac{n}{m}}\right] \circ \cdots \circ \left[(x^k)^{\frac{n}{m}} + (y^k)^{\frac{n}{m}}\right]}_{m}$, we obtain that the left-hand side of (55) is $(x^k)^n + (y^k)^n + o(\|x^k\|^{n-1}\|y^k\|)$, whereas by expanding $\left[x^k + y^k + \phi_p(x^k, y^k)\right]^n$ and noting that $\{\phi_p(x^k, y^k)\}$ is bounded and $\{\|x^k\|/\|y^k\|\}$ is unbounded, the right-hand side of (55) is $(x^k + y^k)^n + o(\|x^k\|^{n-1}\|y^k\|)$, which can be further written as

$$\begin{aligned}
&(x^k)^n + (y^k)^n + (x^k)^{n-1} \circ y^k + (x^k) \circ ((x^k)^{n-2} \circ y^k) \\
&\quad + \cdots + x^k \circ (x^k \circ (\cdots ((x^k)^2 \circ y^k \underbrace{)\cdots))}_{n-3}) + 2x^k \circ (x^k \circ (x^k \circ (\cdots (x^k \circ y^k \underbrace{)\cdots))}_{n-2}) \\
&\quad + o(\|x^k\|^{n-1}\|y^k\|).
\end{aligned}$$

Here, $o(\|x^k\|^{n-1}\|y^k\|)$ denotes the term e_k satisfying $\lim_{k\to\infty} \frac{\|e_k\|}{\|x^k\|^{n-1}\|y^k\|} = 0$. Therefore,

$$\begin{aligned}
&(x^k)^{n-1} \circ y^k + (x^k) \circ ((x^k)^{n-2} \circ y^k) + \cdots + x^k \circ (x^k \circ (\cdots ((x^k)^2 \circ y^k \underbrace{)\cdots))}_{n-3}) \\
&+ 2x^k \circ (x^k \circ (x^k \circ (\cdots (x^k \circ y^k \underbrace{)\cdots))}_{n-2}) = o(\|x^k\|^{n-1}\|y^k\|).
\end{aligned}$$

Making the inner product with the unit element e for the both sides then gives

$$n\langle (x^k)^{n-1}, y^k\rangle = o(\|x^k\|^{n-1}\|y^k\|).$$

Dividing the both sides by $\|x^k\|^{n-1}\|y^k\|$ and taking the limit $k \to \infty$, we get $\langle (x^*)^{n-1}, y^* \rangle = 0$. Noting that $x^*, y^* \in \mathcal{K}^n$ and $\|x^*\| = \|y^*\| = 1$, from $\langle (x^*)^{n-1}, y^* \rangle = 0$ we deduce that

$$y_1^* = \|y_2^*\| \ \text{ and } \ (x^*)^{n-1} = \alpha(y_1^*, -y_2^*) \ \text{ for some } \alpha > 0.$$

From this, it is easy to get $\langle x^*, y^* \rangle = 0$, which by (54) contradicts the given condition that $\langle \frac{x^k}{\|x^k\|}, \frac{y^k}{\|y^k\|} \rangle \nrightarrow 0$. Thus, we prove that the conclusion $\limsup_{k\to\infty} \psi_p(x^k, y^k) = +\infty$ holds.

Case 2. The sequence $\{\|y^k\|/\|x^k\|\}$ is unbounded. By the symmetry of x and y in $\phi_p(x, y)$, using the same arguments as in Case 1 leads to the desired result.

Case 3. The sequences $\{\|y^k\|/\|x^k\|\}$ and $\{\|x^k\|/\|y^k\|\}$ are bounded. In this case, taking subsequences of $\{x^k\}$ and $\{y^k\}$ if necessary, we may assume that $\lim_{k\to\infty} \frac{\|y^k\|}{\|x^k\|} = c$ for some $0 < c < +\infty$. By the definition of ϕ_p and $x^k, y^k \in \mathcal{K}^n$, we have

$$(x^k)^p + (y^k)^p = \left[x^k + y^k + \phi_p(x^k, y^k)\right]^p.$$

Suppose that the conclusion does not hold. Then, dividing the both sides of the last equality by $\|x^k\|$ and taking the limit $k \to \infty$, it is not hard to obtain

$$(x^*)^p + (cy^*)^p = (x^* + cy^*)^p,$$

which is equivalent to saying that $\phi_p(x^*, cy^*) = 0$ since $x^*, y^* \in \mathcal{K}^n$. Therefore, $x^* \circ cy^* = 0$. This clearly contradicts the given condition $\langle \frac{x^k}{\|x^k\|}, \frac{y^k}{\|y^k\|} \rangle \nrightarrow 0$, and the result follows. □

Remark 5.1. So far, we cannot prove the result of Lemma 5.2 for an irrational number p by using the dense of rational numbers in $\mathbb{R}$, although numerical test shows that it is correct.

In the following, assume that $\mathcal{K}$ has the Cartesian structure in (2). Recall that $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is said to have the *uniform Jordan P-property* if there exists a constant $\varrho > 0$ such that, for any $\zeta, \xi \in \mathbb{R}^n$, there is a $\nu \in \{1, \ldots, m\}$ such that

$$\lambda_2[(\zeta_\nu - \xi_\nu) \circ (F_\nu(\zeta) - F_\nu(\xi))] \geq \varrho\|\zeta - \xi\|^2,$$

where $\lambda_2[(\zeta_\nu - \xi_\nu) \circ (F_\nu(\zeta) - F_\nu(\xi))]$ means the second spectral value of $(\zeta_\nu - \xi_\nu) \circ (F_\nu(\zeta) - F_\nu(\xi))$; and F is said to have the *linear growth* if there is a constant $c > 0$ such that $\|F(\zeta)\| \leq \|F(0)\| + c\|\zeta\|$.

Proposition 5.1. *Suppose that G is an identity mapping and F has the uniform Jordan P-property and satisfies the linear growth. Then, $\Psi_p(\zeta)$ is coercive for a rational p.*

Proof. Suppose on the contrary that there is a constant $\gamma > 0$ and a sequence $\{\zeta^k\} \subset \mathbb{R}^n$ with $\|\zeta^k\| \to \infty$ such that $\Psi_p(\zeta^k) \leq \gamma$ for all k. Let $\zeta^k = (\zeta_1^k, \ldots, \zeta_m^k)$ with $\zeta_i^k \in \mathbb{R}^{n_i}$ for $i = 1, \ldots, m$. Let $\mathcal{I} := \left\{i \in \{1, \ldots, m\} \mid \{\zeta_i^k\} \text{ is unbounded}\right\}$. Clearly, $\mathcal{I} \neq \emptyset$. Define

$$\xi_i^k = \begin{cases} 0 & \text{if } i \in \mathcal{I}, \\ \zeta_i^k & \text{otherwise}, \end{cases} \qquad i = 1, 2, \ldots, m.$$

Then, the sequence $\{\xi^k\}$ is bounded. Since F has the uniform Jordan P-property,

$$\lambda_2\left[(\zeta^k - \xi^k) \circ (F(\zeta^k) - F(\xi^k))\right] \geq \varrho\|\zeta^k - \xi^k\|^2$$

for some $\varrho > 0$. Let $z^k = (\zeta^k - \xi^k) \circ (F(\zeta^k) - F(\xi^k))$ for each k. Suppose that each z^k has the spectral decomposition $\lambda_1(z^k)u_1^k + \lambda_2(z^k)u_2^k$. Then, we obtain that

$$\begin{aligned} \varrho\|\zeta^k - \xi^k\|^2 \le 2\langle z^k, u_2^k\rangle &= 2\left\langle (\zeta^k - \xi^k) \circ (F(\zeta^k) - F(\xi^k)), u_2^k \right\rangle \\ &\le 2\|\zeta^k - \xi^k\|\|F(\zeta^k) - F(\xi^k)\|. \end{aligned} \tag{56}$$

This implies that $\|F(\zeta^k)\| \to \infty$. Since $\Psi_p(\zeta^k) \le \gamma$ for each k, from Lemma 5.2(a) it follows that $\{\lambda_1(\zeta^k)\}$ and $\{\lambda_1(F(\zeta^k))\}$ are bounded below. Together with $\|\zeta^k\| \to \infty$ and $\|F(\zeta^k)\| \to \infty$, we obtain $\lambda_2(\zeta^k), \lambda_2(F(\zeta^k)) \to +\infty$. In addition, from (56) and the linear growth of F, we necessarily have $\lim_{k\to\infty} \frac{\zeta^k}{\|\zeta^k\|} \circ \frac{F(\zeta^k)}{\|F(\zeta^k)\|} \neq 0$. If not, on one hand, from the boundedness of $\{\xi^k\}$ we have $\lim_{k\to\infty} \frac{(\zeta^k - \xi^k)\circ(F(\zeta^k) - F(\xi^k))}{\|\zeta^k\|\|F(\zeta^k)\|} = 0$; and on the other hand,

$$\lim_{k\to\infty} \frac{\varrho\|\zeta^k - \xi^k\|^2}{\|\zeta^k\|\|F(\zeta^k)\|} \ge \lim_{k\to\infty} \frac{\varrho\|\zeta^k - \xi^k\|^2}{\|\zeta^k\|(\|F(0)\| + c\|\zeta^k\|)} = \frac{\varrho}{c} > 0,$$

which is impossible by (56). By Lemma 5.2(b), $\limsup_{k\to\infty} \|\psi_p(\zeta^k, F(\zeta^k))\| = \infty$. This gives a contradiction to $\Psi_p(\zeta^k) \le \gamma$ for all k. Hence, Ψ_p is coercive. □

6. Numerical results

In this section, we check the numerical performance of the merit function Ψ_p corresponding to different $p \in (1, 4]$ † by solving the unconstrained minimization reformulation $\min_{\zeta\in\mathbb{R}^n} \Psi_p(\zeta)$ for convex SOCPs in the form of (3), whose KKT conditions can be rewritten as (1) with

$$F(\zeta) := \widehat{x} + (I - A^T(AA^T)^{-1}A)\zeta \quad \text{and} \quad G(\zeta) := \nabla f(F(\zeta)) - A^T(AA^T)^{-1}A\zeta, \tag{57}$$

where $\widehat{x} \in \mathbb{R}^n$ satisfies $A\widehat{x} = b$. All tests were done on a PC using a Pentium 4 with 2.8GHz CPU and 512MB memory, and the codes were all written in Matlab 6.5.

During the testing, we used the Cholesky factorization of AA^T to evaluate F and G in (57), which was completed via Matlab routine `chol`. For the vector $\widehat{x}$ satisfying $A\widehat{x} = b$, we computed it with Matlab's solver "\", i.e., $\widehat{x} = A\backslash b$. We employed the L-BFGS method, a limited-memory quasi-Newton method, with five limited-memory vector-updates to solve the minimization problem $\min_{\zeta\in\mathbb{R}^n} \Psi_p(\zeta)$. For the scaling matrix $H^0 = \gamma I$ in the L-BFGS, we adopted $\gamma = \overline{\zeta}^T\overline{\eta}/\overline{\eta}^T\overline{\eta}$ as recommended by [27, p. 226], where $\overline{\zeta} := \zeta - \zeta^{\text{old}}$ and $\overline{\eta} := \nabla\Psi_p(\zeta) - \nabla\Psi_p(\zeta^{\text{old}})$. To ensure convergence, we reverted to the steepest descent direction $-\nabla\Psi_p(\zeta)$ whenever $\overline{\zeta}^T\overline{\eta} \le 10^{-5}\|\overline{\zeta}\| \cdot \|\overline{\eta}\|$. We adopted a nonmonotone line search in [17] to seek a suitable step-length, i.e., we computed the smallest nonnegative integer l such that

$$\Psi_p(\zeta^k + \rho^l d^k) \le \mathcal{W}_k + \sigma\rho^l \nabla\Psi_p(\zeta^k)^T d^k,$$

where d^k means the direction in kth iteration generated by L-BFGS, $\rho, \sigma \in (0, 1)$ are given parameters, and $\mathcal{W}_k = \max_{j=k-m_k,\ldots,k} \Psi_p(\zeta^j)$ where, for given nonnegative integers $\widehat{m}$ and s,

$$m_k = \begin{cases} 0 & \text{if } k \le s, \\ \min\{m_{k-1} + 1, \widehat{m}\} & \text{otherwise.} \end{cases}$$

†For $p = 4$, we use the gradient formula in Proposition 3.1, though its differentiability is not established.

Throughout the experiments, we chose the following parameters for the algorithm:

$$\rho = 0.5, \quad \sigma = 10^{-4}, \quad \widehat{m} = 5, \quad \text{and} \quad s = 5.$$

The initial point was set to be $\zeta^0 = 0$, and the algorithm was terminated whenever $\max\{\Psi_p(\zeta^k), |\langle F(\zeta^k), G(\zeta^k)\rangle|\} \leq 10^{-6}$, or the step-length is less than 10^{-12}.

The first group of test instances is the linear SOCPs with sparse A from [29]. Numerical results are reported in Table 1, where the first row gives the name of problems and the dimension (m, n) of A, and in the second row, $\Psi_p(\zeta)$ means the value of Ψ_p at the final iteration, **Gap** denotes the value of $|\langle G(\zeta), F(\zeta)\rangle|$ at the final iteration, **NF** and **Cpu** records the number of function evaluations and the CPU time in seconds for solving each test problem, and "$-$" means that the method fails due to too small a step-length.

Table 1 shows that the merit function method based on Ψ_p with $p \in [1.1, 4]$ can solve "nb-L2" and "nb-L2-bessel" successfully, but for problem "nb", it fails due to too small a step-length when $p > 3.5$ or $p < 1.1$, and it cannot yield desirable decrease for Ψ_p even for $p = 2$. The convergence process plotted in Figure 1 for "nb-L2-bessel" shows that the method with a smaller p has a faster reduction of Ψ_p at the beginning of the iterations, and when the value of $\Psi_p(\zeta)$ is less than 10^{-5}, the method with a larger p has a faster reduction of Ψ_p. This implies that the method with a larger p seems to have a better convergence rate, which coincides with the numerical performance of generalized FB NCP functions; see [8].

TABLE 1. Numerical results with different p for three DIMACS problems

	nb $(123, 2383)$				nb-L2 $(123, 4195)$				nb-L2-bessel $(123, 2641)$			
p	$\Psi_p(\zeta)$	**NF**	**Gap**	**Cpu**	$\Psi_p(\zeta)$	**NF**	**Gap**	**Cpu**	$\Psi_p(\zeta)$	**NF**	**Gap**	**Cpu**
4.0	–	–	–	–	1.2e–8	742	1.85e–7	329.2	1.7e–8	359	8.73e–7	147.5
3.5	–	–	–	–	9.9e–13	502	1.49e–7	233.9	2.6e–9	188	7.57e–7	87.3
3.0	3.8e-8	6551	2.04e–5	2622.4	4.3e–13	763	9.54e–7	305.2	2.0e–9	248	1.79e–8	106.1
2.5	6.8e-7	2920	9.47e–8	1148.1	2.5e–13	505	7.97e–7	241.8	7.3e–10	268	8.85e–7	108.6
2.0	5.7e-7	3297	2.52e–7	1211.9	2.8e–10	665	7.97e–7	245.9	1.2e–10	316	9.40e–10	112.8
1.8	3.2e–8	5676	2.86e–5	2459.9	2.9e–11	548	3.97e–7	225.2	5.3e–10	174	4.54e–7	80.68
1.5	7.8e–7	1895	1.78e–7	712.7	9.7e–11	526	5.45e–7	224.3	3.1e–10	204	8.45e–7	88.10
1.3	2.5e–7	4807	7.39e–7	1869.9	4.8e–13	644	3.41e–7	301.5	1.2e–9	339	3.15e–7	137.6
1.1	–	–	–	–	3.5e–11	1471	7.42e–7	640.8	7.2e–13	956	9.67e–7	405.8

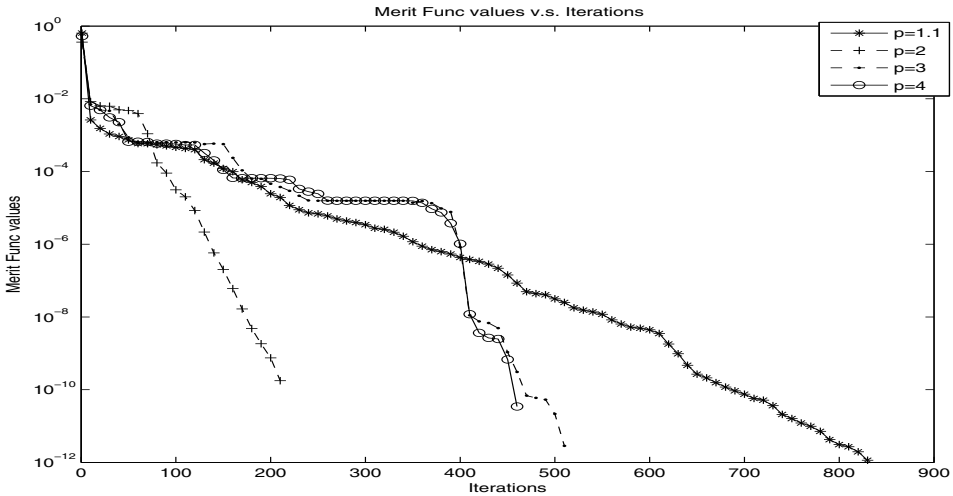

FIGURE 1. Numerical results with different p for the problem "nb-L2-bessel"

The second group of test problems is the convex SOCPs with dense A. To generate such test problems randomly, we considered the problem of minimizing a sum of the k largest Euclidean norms with a convex regularization term: $\min_{u\geq 0}\sum_{i=1}^{k}\|s_{[i]}\| + h(u)$, where $\|s_{[1]}\|,\ldots,\|s_{[r]}\|$ are the norms $\|s_1\|,\ldots,\|s_r\|$ sorted in nonincreasing order with $r \geq k$ and $s_i = b_i - A_i x$ for $i = 1,\ldots,r$ with $A_i \in \mathbb{R}^{m_i\times l}$ and $b_i \in \mathbb{R}^{m_i}$, and $h : \mathbb{R}^l \to \mathbb{R}$ is a twice continuously differentiable convex function. The problem can be converted into

$$\begin{array}{ll}
\min & (1-k/r)\sum_{i=1}^{r} v_i + (k/r)\sum_{i=1}^{r} w_i + h(u) \\
\text{s.t.} & A_i u + s_i = b_i, \quad i = 1,2,\ldots,r, \\
& (w_1 - v_1) - (w_2 - v_2) = 0, \\
& \qquad\vdots \\
& (w_1 - v_1) - (w_r - v_r) = 0, \\
& u \geq 0, \quad v_i \geq 0, \quad (w_i, s_i) \in \times\mathcal{K}^{m_i+1}, \quad i = 1,2,\ldots,r.
\end{array}$$

In our tests, we set $h(u) := \frac{1}{3}\|u\|_3^3$ with $\|\cdot\|_3$ being the 3-norm, and generated each m_i randomly from $\{2,3,\ldots,10\}$, and each entry of A_i and b_i randomly by a uniform distribution from the interval $[-1,1]$ and $[-5,5]$, respectively. Thus, if $d \geq m = m_1 + \cdots + m_r$, the constraint matrix is dense. The problem parameters and the numerical results are reported in Table 2, in which the first row lists several groups of different (l,r,k), and the second row gives the corresponding dimension (m,n) of A. From Table 2, we see that for the test problems with dense A, Ψ_p displays a similar performance to those with sparse A.

TABLE 2. Numerical results with different p for convex SOCPs with dense A

	$(500, 50, 1)$			$(1000, 50, 5)$			$(1000, 100, 5)$			$(2000, 100, 5)$		
	$(353, 904)$			$(374, 1425)$			$(671, 1772)$			$(727, 2828)$		
p	**NF**	**Gap**	**Cpu**	**NF**	**Gap**	**Cpu**	**NF**	**Gap**	**Cpu**	**NF**	**Gap**	**Cpu**
4.0	4000	9.96e–7	287.5	268	7.80e–7	42.32	1237	3.23e–7	322.1	–	–	–
3.5	2328	8.29e–7	203.0	327	7.67e–7	54.28	1407	6.13e–7	351.0	–	–	–
3.0	2215	2.52e–7	189.6	319	9.14e–7	51.61	1224	6.69e–7	311.7	200	9.12e–7	101.8
2.5	2292	8.76e–7	200.6	293	9.07e–7	48.31	1483	1.01e–7	381.4	125	7.77e–7	60.57
2.0	2258	6.28e–7	180.0	340	1.69e–7	49.53	1012	7.34e–7	225.9	144	9.84e–7	67.55
1.8	2482	4.88e–7	208.4	357	9.71e–7	51.56	1257	6.07e–7	305.2	168	9.41e–7	81.87
1.5	3180	5.09e–7	269.5	245	1.48e–7	39.37	1595	3.72e–7	376.7	150	2.24e–8	61.96
1.3	2309	8.19e–7	192.8	371	7.97e–7	55.20	1296	8.67e–7	316.9	212	8.77e–7	95.42
1.1	–	–	–	2999	6.08e–7	189.9	3968	6.11e–7	963.9	576	9.20e–7	245.2

7. CONCLUSIONS

We established the smoothness of the generalized FB merit function Ψ_p with $p \in (1, 4)$, and studied its coerciveness under some mild conditions, which partially generalized the results of the FB merit function for the SOCCP [12] and the generalized FB merit functions for the NCPs [9]. There are some topics worthy of further study, for example, the stationary point conditions of Ψ_p, the strong semismoothness of ϕ_p, the characterization of directional derivative of ϕ_p, and the characterization of the B-subdifferential of ϕ_p.

APPENDIX

Lemma 1. *Assume that $p \geq 2$. Let $x, y \in \mathbb{R}^n$ satisfy $w(x, y) \in \mathbf{bd}^+\mathcal{K}^n$. Then, we have*

$$\begin{aligned} \nabla_{x'} w_1(x', y')|_{(x',y')=(x,y)} &= \nabla_{x'} \|w_2(x', y')\||_{(x',y')=(x,y)} \\ &= \begin{pmatrix} 2p^{-2}\mathrm{sign}(x_1)|x_1|^{p-1} \\ 2^{p-2}p\mathrm{sign}(x_1)|x_1|^{p-1}\overline{w}_2 \end{pmatrix}. \end{aligned} \tag{58}$$

Proof. Assume that $x_2 \neq 0$. By the exprssions of $w_1(x', y')$ and $w_2(x', u')$, we calculate

$$\begin{aligned}
&\nabla_{x'} w_1(x', y')|_{(x',y')=(x,y)} \\
&\quad = \frac{p}{2}\begin{pmatrix} \operatorname{sign}(\lambda_2(x))|\lambda_2(x)|^{p-1} + \operatorname{sign}(\lambda_1(x))|\lambda_1(x)|^{p-1} \\ \operatorname{sign}(\lambda_2(x))|\lambda_2(x)|^{p-1} - \operatorname{sign}(\lambda_1(x))|\lambda_1(x)|^{p-1}\frac{x_2}{\|x_2\|} \end{pmatrix}, \\
&\nabla_{x'} \|w_2(x', y')\||_{(x',y')=(x,y)} \\
&\quad = \frac{p}{2}\begin{pmatrix} \operatorname{sign}(\lambda_2(x))|\lambda_2(x)|^{p-1} - \operatorname{sign}(\lambda_1(x))|\lambda_1(x)|^{p-1} \\ \operatorname{sign}(\lambda_2(x))|\lambda_2(x)|^{p-1} - \operatorname{sign}(\lambda_1(x))|\lambda_1(x)|^{p-1}\frac{x_2}{\|x_2\|} \end{pmatrix} \frac{x_2^T w_2}{\|x_2\|\|w_2\|} \\
&\qquad + \frac{|\lambda_2(x)|^p - |\lambda_1(x)|^p}{2}\begin{pmatrix} 0 \\ \frac{w_2}{\|x_2\|\|w_2\|} - \frac{x_2 x_2^T w_2}{\|x_2\|^3\|w_2\|} \end{pmatrix}.
\end{aligned}$$

Using the equalities in (18), the last two equalities can be simplified as

$$\begin{aligned}
\nabla_{x'} w_1(x', y')|_{(x',y')=(x,y)} &= \nabla_{x'} \|w_2(x', y')\| \,|_{(x',y')=(x,y)} \\
&= \begin{pmatrix} 2p^{-2}\operatorname{sign}(x_1)|x_1|^{p-1} \\ 2^{p-2}p\operatorname{sign}(x_1)|x_1|^{p-1}\overline{w}_2 \end{pmatrix}.
\end{aligned}$$

If $x_2 = 0$, using the result for $x_2 \neq 0$ and the continuity of $\nabla w_1(x', y')$ and $\nabla\|w_2(x', y')\|$ at (x, y), we easily obtain equation (58). □

References

1. F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Mathematical Programming, vol. 95, pp. 3-51, 2003. MR1971381 (2004j:90060)
2. E. D. Andersen, C. Roos, and T. Terlaky, *On implementing a primal-dual interior-point method for conic quadratic optimization*, Mathematical Programming, vol. 95, pp. 249-277, 2003. MR1976481 (2004d:90130)
3. R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1997. MR1477662 (98i:15003)
4. J.-S. Chen, *Two classes of merit functions for the second-order cone complementarity problem*, Mathematical Methods of Operations Research, vol. 64, pp. 495-519, 2006. MR2264798 (2007j:90078)
5. J.-S. Chen, S.-H. Pan, and C.-Y. Yang, *Numerical comparisons of two effective methods for complementarity problems*, Journal of Computational and Applied Mathematics, vol. 234, pp. 667-683, 2010. MR2606654 (2011d:65159)
6. J.-S. Chen and S.-H. Pan, *A one-parametric class of merit functions for the second-order cone complementarity problem*, Computational Optimization and Applications, vol. 45, pp. 581-606, 2010. MR2600897 (2011c:90105)
7. J.-S. Chen, X. Chen, and P. Tseng, *Analysis of nonsmooth vector-valued functions associated with second-order cone*, Mathematical Programming, vol. 101, pp. 95-117, 2004. MR2085260 (2005e:49029)
8. J.-S. Chen, H.-T. Gao and S. Pan, *A R-linearly convergent derivative-free algorithm for the NCPs based on the generalized Fischer-Burmeister merit function*, Journal of Computational and Applied Mathematics, vol 232, pp. 455-471, 2009. MR2555414 (2010h:90137)
9. J.-S. Chen and S. Pan, *A family of NCP-functions and a descent method for the nonlinear complementarity problem*, Computational Optimization and Applications, vol. 40, pp. 389-404, 2008. MR2411201 (2009b:90158)
10. J.-S. Chen, D.-F. Sun, and J. Sun, *The SC^1 property of the squared norm of the SOC Fischer-Burmeister function*, Operations Research Letters, vol. 36, pp. 385-392, 2008. MR2424468 (2009g:90123)
11. X.-D. Chen, D. Sun, and J. Sun, *Complementarity functions and numerical experiments for second-order cone complementarity problems*, Computational Optimization and Applications, vol. 25, pp. 39-56, 2003. MR1996662 (2004f:90142)

12. J.-S. Chen and P. Tseng, *An unconstrained smooth minimization reformulation of the second-order cone complementarity problem*, Mathematical Programming, vol. 104, pp. 293-327, 2005. MR2179239 (2006g:90123)
13. F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Volumes I and II, Springer-Verlag, New York, 2003.
14. J. Faraut and A. Korányi, *Analysis on Symmetric Cones*, Oxford Mathematical Monographs, Oxford University Press, New York, 1994. MR1446489 (98g:17031)
15. M. Fukushima, Z.-Q. Luo, and P. Tseng, *Smoothing functions for second-order cone complementarity problems*, SIAM Journal on Optimization, vol. 12, pp. 436-460, 2002. MR1885570 (2002m:90084)
16. M. S. Gowda, R. Sznajder, and J. Tao, *Some P-properties for linear transformations on Euclidean Jordan algebras*, Linear Algebra and its Applications, vol. 393, pp. 203-232, 2004. MR2098615 (2005h:17059)
17. L. Grippo, F. Lampariello and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM Journal on Numerical Analysis, vol. 23, pp. 707–716, 1986. MR849278 (87g:90105)
18. S. Hayashi, N. Yamashita, and M. Fukushima, *A combined smoothing and regularization method for monotone second-order cone complementarity problems*, SIAM Journal of Optimization, vol. 15, pp. 593–615, 2005. MR2144183 (2006f:90081)
19. S. Hayashi, N. Yamashita, and M. Fukushima, *Robust Nash equilibria and second-order cone complementarity problems*, Journal of Nonlinear and Convex Analysis, vol. 6, pp. 283–296, 2005. MR2159841
20. R. A. Horn and C. R. Johnson, *Matrix Analyis*. Cambridge University Press, Cambridge, 1986.
21. Y. Kanno, J. A. C. Martins and A. Pinto da Costa, *Three-dimensional quasi-static frictional contact by using second-order cone linear complementarity problem*. International Journal for Numerical Methods in Engineering, vol. 65, pp. 62–83, 2006. MR2185946 (2006k:74078)
22. C. Kanzow, I. Ferenczi and M. Fukushima, *On the local convergence of semismooth Newton methods for linear and nonlinear second-order cone programs without strict complementarity*, SIAM Journal on Optimization, vol. 20, pp. 297–320, 2009. MR2496902 (2010e:90136)
23. A. Korányi, *Monotone functions on formally real Jordan algebras*, Mathematische Annalen, vol. 269, pp. 73–76, 1984. MR756776 (85i:17028)
24. Y.-D. Lim, *Applications of geometric means on symmetric cones*, Mathematische Annalen, vol. 319, pp. 457–468, 2001. MR1819877 (2002a:17023)
25. M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, *Application of second-order cone programming*, Linear Algebra and its Applications, vol. 284, pp. 193–228, 1998. MR1655138 (99k:90103)
26. R. D. C. Monteiro and T. Tsuchiya, *Polynomial convergence of primal-dual algorithms for the second-order cone programs based on the MZ-family of directions*, Mathematical Programming, vol. 88, pp. 61-83, 2000. MR1765893 (2001d:90128)
27. J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999. MR1713114 (2001b:90002)
28. J. Peng, C. Roos and T. Terlaky, *Primal-dual interior-point methods for second-order cone optimization based on self-regular proximities*, SIAM Journal on Optimization, vol. 13, pp. 179-203, 2002. MR1922760 (2003h:90110)
29. G. Pataki and S. Schmieta, *The DIMACS library of semidefinite-quadratic-linear programs*, Computational Optimization Research Center, Columbia University, New York. http://dimacs.rutgers.edu/Challenges/
30. 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*, Mathematical Programming, vol. 107, pp. 547-553, 2006. MR2221830 (2007b:90085)
31. T. Tsuchiya, *A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming*, Optimization Methods and Software, vol. 11, pp. 141-182, 1999. MR1777456 (2001d:90134)
32. P. Tseng, *Merit functions for semidefinite complementarity problems*, Mathematical Programming, vol. 83, pp. 159-185, 1998. MR1647837 (99m:90147)

33. A. Yoshise, *Interior point trajectories and a homogenous model for nonlinear complementarity problems over symmetric cones*, SIAM Journal on Optimization, vol. 17, pp. 1129-1153, 2006. MR2274506 (2007k:90137)
34. Y. Wang and L. W. Zhang, *Properties of equation reformulation of the Karush-Kuhn-Tucker condition for nonlinear second-order cone optimization problems*, Mathematical Methods of Operations Research, vol. 70, pp. 195-218, 2009. MR2558410 (2011b:90179)

School of Mathematical Sciences, South China University of Technology, Guangzhou 510640, People's Republic of China
E-mail address: shhpan@scut.edu.cn

Department of Mathematics Education, Chungbuk National University, Cheongju 361-763, South Korea
E-mail address: shkum@chungbuk.ac.kr

Department of Mathematics, Sungkyunkwan University, Suwan 440-746, South Korea
E-mail address: ylim@knu.ac.kr

Department of Mathematics, National Taiwan Normal University, Taipei 11677, Taiwan
E-mail address: jschen@math.ntnu.edu.tw