A Note on the Lipschitz Continuity of the Gradient of the Squared Norm of the Matrix-valued Fischer-Burmeister Function

Chee-Khian Sim* Jie Sun* Danny Ralph†

November 12, 2004

Abstract

Based on a formula of Tseng, we show that the squared norm of the matrix-valued Fischer-Burmeister function has a Lipschitz continuous gradient.

1 Introduction

Active research have been done on the use of merit functions to solve nonlinear complementarity problem (NCP) and the solution methods thus developed, such as the Newton-type methods based on the Fischer-Burmeister function (FB function), appear to be effective. Motivated by these developments, Tseng [3] studied the extension of merit functions for NCP to its (symmetric) matrix-valued counterpart, the merit functions for the semidefinite complementarity problem (SDCP). In particular, he proved the differentiability of the squared norm of the FB function over the space of $n \times n$ symmetric matrices. However, it remains open whether the gradient function is continuous or is furthermore Lipschitz continuous, which are crucial from the algorithmic point of view. The same problems arise in the study of certain optimization algorithms over the Lorenz cone. Recently, Chen and Tseng [1] showed the continuous differentiability of the squared norm of the FB function for the Lorenz cone case. In this note, we show the Lipschtiz continuity of the gradient for the matrix case.

Let $S^n \subset \Re^{n\times n}$ denote the set of symmetric matrices of dimension n and S^n_+ (S^n_{++}) the set of symmetric positive semidefinite (symmetric positive definite) matrices of dimension n. For $X \in S^n$, let $\lambda_i(X)$ be the i^{th} eigenvalue of X with $\lambda_1(X) \geq \ldots \geq \lambda_n(X)$. Given $X \in \Re^{m\times n}$, X_{ij} denotes the entry of X at the (i,j) position and $\|\cdot\| : \Re^{m\times n} \to \Re_+$ stands for the Frobenius norm for matrices, i.e. $\|X\| = [\sum_{i,j} X_{ij}^2]^{1/2}$. A matrix $P \in \Re^{n\times n}$ is said

*Department of Decision Sciences, National University of Singapore, Singapore 117592. These authors' research was partially supported by Grants R314-000-057/042-112 of the National University of Singapore.
†Judge Institute of Management, University of Cambridge CB2 1AG, UK.

to be orthogonal if $PP^T = I_n$, where I_n is the identity matrix in $\Re^{n\times n}$. Also for $X \in S^n_+$, $X^{1/2}$ denotes its nonnegative square root. For $X_i \in S^{n_i}$, $i = 1, \dots, k$, $\mathrm{diag}\,(X_1, \dots, X_k)$ is the block diagonal square matrix with X_i as its i^{th} diagonal block. In what follows, we will use "$O(t)$" (respectively, "$o(t)$") as a shorthand to denote an element of S^n that depends on t and whose norm tends to 0 at least as far as (respectively, faster than) t, i.e., $\limsup_{t\to 0} \|O(t)\|/t < \infty$ (respectively, $\limsup_{t\to 0} \|o(t)\|/t = 0$).

2 Preliminaries

Definition 2.1 *The Fischer-Burmeister function (FB function) for SDCP is the function* $\Phi : S^n \times S^n \to S^n$ *defined by* $\Phi(X,Y) := (X^2+Y^2)^{1/2} - (X+Y)$.
The squared norm of the FB function for SDCP is the function $\Psi : S^n \times S^n \to \Re$ *defined by* $\Psi(X,Y) := \frac{1}{2}\|\Phi(X,Y)\|^2$

It was shown in [3] that Ψ is differentiable on $S^n \times S^n$ and a formula for its derivative with respect to X and Y was given. We will now elaborate more on the relevant results in that paper which we will use here.
First, for $C \in S^n_+$, let S_C denotes the subspace of S^n comprising those $X \in S^n$ whose nullspace contains the nullspace of C. It is readily seen that

$$S_C = \left\{ X \in S^n \ : \ PXP^T = \mathrm{diag}\,(\widetilde{X}, 0) \text{ for some submatrix } \widetilde{X} \in S^{|I|} \right\}, \tag{1}$$

for any P orthogonal matrix of dimension n and $I \subset \{1, \dots, n\}$ such that

$$PCP^T = \mathrm{diag}\,(\widetilde{C}, 0),$$

where $\widetilde{C} \in S^{|I|}_{++}$.
Define the linear mapping $\mathcal{L}_C : S_C \to S_C$ by $\mathcal{L}_C(X) := CX + XC$, where $X \in S_C$. It can be seen that $\mathcal{L}_C$ is positive definite on S_C and so has an inverse $\mathcal{L}_C^{-1}$ on S_C, that is, for any $X \in S_C$, $\mathcal{L}_C^{-1}(X)$ is the unique $U \in S_C$ satisfying $CU + UC = X$. $\mathcal{L}_C^{-1}$ can be further obtained in closed form by choosing P so that $\widetilde{C}$ is diagonal, so that

$$\mathcal{L}_C^{-1}(X) = P^T \begin{bmatrix} (\widetilde{X}_{ij}/(\widetilde{C}_{ii} + \widetilde{C}_{jj}))_{i,j\in I} & 0 \\ 0 & 0 \end{bmatrix} P, \tag{2}$$

where $\widetilde{X}$ is given in (1).
It was shown in [3, Lemma 6.3(b)] that Ψ is differentiable at every $(X,Y) \in S^n \times S^n$, with

$$\begin{aligned} \nabla_X \Psi(X,Y) &= \mathrm{sym}(\mathcal{L}_C^{-1}(C-X-Y)(X-C)), \\ \nabla_Y \Psi(X,Y) &= \mathrm{sym}(\mathcal{L}_C^{-1}(C-X-Y)(Y-C)), \end{aligned} \tag{3}$$

where $C := (X^2+Y^2)^{1/2}$ and $\mathrm{sym}(\cdot)$ is defined by $\mathrm{sym}(D) := D + D^T$ for $D \in \Re^{n\times n}$. Note that $\mathcal{L}_C^{-1}$ acting on $C - X - Y$ is well defined since $X, Y \in S_C$.

It is easy to see that

$$\Psi(X,Y) = \frac{1}{2}[\mathrm{Trace}(X^2+Y^2) - 2\mathrm{Trace}((X^2+Y^2)^{1/2}(X+Y)) + \mathrm{Trace}((X+Y)^2)]. \quad (4)$$

Since $\mathrm{Trace}(X^2+Y^2)$ and $\mathrm{Trace}((X+Y)^2)$ are infinitely differentiable, to show $\Psi(X,Y)$ has derivatives that are locally Lipschitz continuous, we need only show that the derivatives of

$$F(X,Y) := \mathrm{Trace}((X^2+Y^2)^{1/2}(X+Y))$$

are locally Lipschitz continuous. In fact, we shall show that they are globally Lipschitz continuous.
By (4) and (3), we have

$$\nabla_X F(X,Y) = (X^2+Y^2)^{1/2} + \mathcal{L}_C^{-1}(X+Y)X + X\mathcal{L}_C^{-1}(X+Y)$$

($\nabla_Y F(X,Y)$ is given by a similar formula). We will concentrate on showing $\nabla_X F(X,Y)$ is globally Lipschitz continuous over $S^n \times S^n$. The same analysis applies to $\nabla_Y F(X,Y)$. Since $(X^2+Y^2)^{1/2}$ is globally Lipschitz continuous [2], we need only show that $\mathcal{L}_C^{-1}(X+Y)X$ is globally Lipschitz continuous. A similar analysis can be made for $X\mathcal{L}_C^{-1}(X+Y)$.
Observe that, given any $(X,Y) \in S^n \times S^n$, $C = (X^2+Y^2)^{1/2}$ is not necessarily nonsingular. Hence, to make analysis simpler, we will look at a "smoothed" C, $C_\epsilon(X,Y) := (X^2+Y^2+\epsilon I)^{1/2}$ where $\epsilon > 0$. Observe that $C_\epsilon(X,Y)$ is positive definite for all $\epsilon > 0$. Hence $S_{C_\epsilon(X,Y)} = S^n$ and $\mathcal{L}_{C_\epsilon(X,Y)}$ is positive definite on S^n.

Given $(X_1,Y_1),(X_2,Y_2) \in S^n \times S^n$, we want to show that $\|\mathcal{L}_{C_\epsilon(X_1,Y_1)}^{-1}(X_1+Y_1)X_1 - \mathcal{L}_{C_\epsilon(X_2,Y_2)}^{-1}(X_2+Y_2)X_2\| \le L(\|X_1-X_2\| + \|Y_1-Y_2\|)$, where L is independent of ϵ. Taking limit as ϵ approaches zero, we then show that $\mathcal{L}_C^{-1}(X+Y)X$ is globally Lipschitz continuous. For simplicity of notation, let

$$G_\epsilon(X,Y) := \mathcal{L}_{C_\epsilon(X,Y)}^{-1}(X+Y)X.$$

3 The proof

Lemma 3.1 *Let $W \in S^n$ be such that $C_\epsilon(X,Y)^2 + W \in S^n_+$. Let $Z := (C_\epsilon(X,Y)^2 + W)^{1/2} - C_\epsilon(X,Y)$. Then $Z = \mathcal{L}_{C_\epsilon(X,Y)}^{-1}(W) + o(\|W\|)$.*

Proof. We have $C_\epsilon(X,Y) + Z = (C_\epsilon(X,Y)^2 + W)^{1/2}$. Therefore, $(C_\epsilon(X,Y) + Z)^2 = C_\epsilon(X,Y)^2 + W$. Hence, $Z^2 + C_\epsilon(X,Y)Z + ZC_\epsilon(X,Y) = W$. Thus, we have $W = \mathcal{L}_{C_\epsilon(X,Y)}(Z) + Z^2$. The result then follows by the Inverse Function Theorem. **QED**

Note that the proof of Lemma 3.1 above follows the idea of the proof of Lemma 6.2 in [3]. Using Lemma 3.1, we compute $(G_\epsilon)'_X$ and $(G_\epsilon)'_Y$ in the following lemma.

Lemma 3.2 *Given* $U, V \in S^n$,

$$\begin{aligned}(G_\epsilon)'_X(X,Y)U &= \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)U + \\ &\quad \mathcal{L}^{-1}_{C_\epsilon(X,Y)}[U - sym(\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(XU+UX)]X, \\ (G_\epsilon)'_Y(X,Y)V &= \mathcal{L}^{-1}_{C_\epsilon(X,Y)}[V - sym(\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(YV+VY)]X.\end{aligned}$$

Proof. Given $U \in S^n$, one has

$$G_\epsilon(X+U,Y) = \mathcal{L}^{-1}_{C_\epsilon(X,Y)+Z}(X+U+Y)(X+U),$$

where

$$\begin{aligned}C_\epsilon(X,Y)+Z &= ((X+U)^2+Y^2+\epsilon I)^{1/2} \\ &= (X^2+XU+UX+U^2+Y^2+\epsilon I)^{1/2} \\ &= (C_\epsilon(X,Y)^2+XU+UX+U^2)^{1/2}.\end{aligned}$$

Setting $W = XU+UX+U^2$ in Lemma 3.1, there holds

$$Z = \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(XU+UX) + o(\|U\|). \tag{5}$$

Thus $Z \to 0$ as $U \to 0$ and $Z = O(\|U\|)$. Let

$$g := \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y) \quad \text{and} \quad g+h := \mathcal{L}^{-1}_{C_\epsilon(X,Y)+Z}(X+U+Y). \tag{6}$$

We will now express h in terms of g, U, Z and $C_\epsilon(X,Y)$. By (6),

$gC_\epsilon(X,Y)+C_\epsilon(X,Y)g = X+Y$ and $(g+h)(C_{\epsilon(X,Y)}+Z)+(C_\epsilon(X,Y)+Z)(g+h) = X+U+Y$.

Together, we have $\mathcal{L}_{C_\epsilon(X,Y)}(h) = U - gZ - Zg - hZ - Zh$. Therefore,

$$h = \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(U - gZ - Zg) - \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(hZ+Zh). \tag{7}$$

Now, observe that using (5), we have from (7), $\|h\|$ tends to zero as $\|U\|$ tends to zero. This together with (5) implies that $hZ+Zh = o(\|Z\|) = o(\|U\|)$. Hence, we have $\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(hZ+Zh) = o(\|U\|)$ and

$$h = \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(U - gZ - Zg) + o(\|U\|).$$

Consider $G_\epsilon(X+U,Y) - G_\epsilon(X,Y)$. We have

$$\begin{aligned}& G_\epsilon(X+U,Y) - G_\epsilon(X,Y) \\ =\ & \mathcal{L}^{-1}_{C_\epsilon(X,Y)+Z}(X+U+Y)(X+U) - \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)X \\ =\ & (g+h)(X+U) - gX \\ =\ & gX + gU + h(X+U) - gX \\ =\ & gU + (\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(U - gZ - Zg) + o(\|U\|))(X+U) \\ =\ & \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)U + \\ & \mathcal{L}^{-1}_{C_\epsilon(X,Y)}[U - \mathrm{sym}(\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(XU+UX)) + o(\|U\|)](X+U) \\ & + o(\|U\|)(X+U) \\ =\ & \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)U + \mathcal{L}^{-1}_{C_\epsilon(X,Y)}[U - \mathrm{sym}(\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(XU+UX))]X + \\ & o(\|U\|).\end{aligned}$$

This proves the formula of $(G_\epsilon)'_X(X,Y)U$. By similar argument, we can prove the formula of $(G_\epsilon)'_Y(X,Y)V$. **QED**

Lemma 3.3 *For any* $U, V \in S^n$, $\|\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(V)X\| \le \alpha\|V\|$, $\|\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(XU+UX)\| \le \beta\|U\|$ *and* $\|\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X+Y)\| \le \gamma$, *where* α, β *and* γ *are positive constants independent of* $\epsilon > 0$ *and* $X, Y \in S^n$.

Proof. First, let P be an orthogonal matrix such that $PCP^T = \mathrm{diag}(\widetilde{C}, 0)$ where $\widetilde{C} = \mathrm{diag}(\lambda_1(C), \ldots, \lambda_m(C)) \in S^m_{++}$. Therefore, $PXP^T = \mathrm{diag}(\widetilde{X}, 0)$ and $PYP^T = \mathrm{diag}(\widetilde{Y}, 0)$ where $\widetilde{X}, \widetilde{Y} \in S^m$. Also, $PC_\epsilon(X,Y)P^T = \mathrm{diag}(\sqrt{\lambda_1(C)^2+\epsilon}, \ldots, \sqrt{\lambda_m(C)^2+\epsilon}, \sqrt{\epsilon}, \ldots, \sqrt{\epsilon})$. Consider $\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(V)X$. We have

$$\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(V) = P^T \left[\frac{\widetilde{V}_{ij}}{\lambda_i(C_\epsilon(X,Y)) + \lambda_j(C_\epsilon(X,Y))} \right]_{1\le i\le n, 1\le j\le n} P,$$

where $PVP^T = \widetilde{V}$, by (2).
It can be easily seen that

$$\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(V)X = P^T \left[\left(\sum_{k=1}^{m} \frac{\widetilde{V}_{ik}\widetilde{X}_{kj}}{\lambda_i(C_\epsilon(X,Y)) + \lambda_k(C_\epsilon(X,Y))} \right)_{1\le i\le n, 1\le j\le m} \quad 0 \right] P.$$

For $1 \le k \le m, 1 \le i \le n, 1 \le j \le m$,

$$\begin{aligned} \left| \frac{\widetilde{X}_{kj}}{\lambda_i(C_\epsilon(X,Y)) + \lambda_k(C_\epsilon(X,Y))} \right| &\le \frac{|\widetilde{X}_{kj}|}{\lambda_k(C_\epsilon(X,Y))} \\ &= \frac{|\widetilde{X}_{kj}|}{\sqrt{\lambda_k(C)^2+\epsilon}} \\ &\le \frac{|\widetilde{X}_{kj}|}{\lambda_k(C)} \le 1, \end{aligned}$$

where the last inequality, which is independent of ϵ and $X, Y \in S^n$, follows since $\lambda_k(C)^2 = \sum_{j=1}^m (\widetilde{X}^2_{kj} + \widetilde{Y}^2_{kj})$.
Therefore, we have $\|\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(V)X\| \le \alpha\|V\|$, where α is a positive constant independent of ϵ and $X, Y \in S^n$. A value for α is n^2.
Next, consider $\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(XU+UX)$. We have

$$\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(XU+UX) = P^T \left[\frac{\sum_{k=1}^m (\widetilde{X}_{ik}\widetilde{U}_{kj} + \widetilde{U}_{ik}\widetilde{X}_{kj})}{\lambda_i(C_\epsilon(X,Y)) + \lambda_j(C_\epsilon(X,Y))} \right]_{1\le i\le n, 1\le j\le n} P.$$

Here, $\widetilde{X}_{ik} = \widetilde{X}_{kj} = 0$, for $m+1 \le i \le n, m+1 \le j \le n$.
For $1 \le k \le m, 1 \le i \le m, 1 \le j \le m$, as above,

$$\left| \frac{\widetilde{X}_{ik}}{\lambda_i(C_\epsilon(X,Y)) + \lambda_j(C_\epsilon(X,Y))} \right|, \left| \frac{\widetilde{X}_{kj}}{\lambda_i(C_\epsilon(X,Y)) + \lambda_j(C_\epsilon(X,Y))} \right| \le 1$$

independent of $\epsilon > 0$ and $X, Y \in S^n$. Therefore, $\|\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(XU + UX)\| \le \beta\|U\|$, where β is a positive constant independent of ϵ and $X, Y \in S^n$. A value for β is $2n^2$.
Similarly, $\|\mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X + Y)\| \le \gamma$, where γ is a positive constant independent of $\epsilon > 0$ and $X, Y \in S^n$. A value for γ is $2n^2$. **QED**

Theorem 3.1 $\mathcal{L}_C^{-1}(X + Y)X$ *is globally Lipschitz continuous for all* $(X, Y) \in S^n \times S^n$.

Proof. We first consider $G_\epsilon(X, Y) = \mathcal{L}^{-1}_{C_\epsilon(X,Y)}(X + Y)X$. Let $U, V \in S^n$. We have

$$G_\epsilon(X+U, Y+V) - G_\epsilon(X, Y) = G_\epsilon(X+U, Y+V) - G_\epsilon(X, Y+V) + G_\epsilon(X, Y+V) - G_\epsilon(X, Y).$$

It is clear that $(G_\epsilon)'_X$ is continuous with respect to X, therefore, by the Mean Value Theorem,

$$G_\epsilon(X + U, Y + V) - G_\epsilon(X, Y + V) = \int_0^1 (G_\epsilon)'_X(X + tU, Y + V)U dt.$$

Now,

$$(G_\epsilon)'_X(X + tU, Y + V)U = \mathcal{L}^{-1}_{C_\epsilon(X+tU,Y+V)}(X + tU + Y + V)U + \mathcal{L}^{-1}_{C_\epsilon(X+tU,Y+V)}[U - \text{sym}(\mathcal{L}^{-1}_{C_\epsilon(X+tU,Y+V)}(X + tU + Y + V)\mathcal{L}^{-1}_{C_\epsilon(X+tU,Y+V)}((X + tU)U + U(X + tU))](X + tU).$$

By Lemma 3.3, each of the terms on the right hand side of the above equation is bounded by a constant multiplied by $\|U\|$, where the constant is independent of $\epsilon > 0$, $X, Y, U, V \in S^n$ and $t \in [0, 1]$. Therefore, $\|G_\epsilon(X + U, Y + V) - G_\epsilon(X, Y + V)\| \le L_1\|U\|$ where L_1 is independent of $\epsilon > 0$, X, Y, U, and V.
By similar argument, we also have $\|G_\epsilon(X, Y + V) - G_\epsilon(X, Y)\| \le L_2\|V\|$ where L_2 is independent of $\epsilon > 0$, $X, Y, U, V \in S^n$.
Hence, letting ϵ approaches zero in $\|G_\epsilon(X + U, Y + V) - G_\epsilon(X, Y)\| \le L_1\|U\| + L_2\|V\|$, we then show that $\mathcal{L}_C^{-1}(X + Y)X$ is globally Lipschitz continuous for all $(X, Y) \in S^n \times S^n$. **QED**
We have shown that $\nabla_X F(X, Y)$ is globally Lipschitz continuous over $S^n \times S^n$, which implies the (local) Lipschitz continuity of the gradient of $\Psi(X, Y)$, the squared norm of the matrix-valued F-B function.

References

[1] Jein-Shan Chen and Paul Tseng, *An Unconstrained Smooth Minimization Reformulation of the Second-Order Cone Complementarity Problem,* Preprint, July 10, 2004.

[2] Defeng Sun and Jie Sun, *Strong Semismoothness of the Fischer-Burmeister SDC and SOC Complementarity Functions,* Preprint, December 26, 2002, Revised: May 12, 2004.

[3] Paul Tseng, *Merit Functions for Semi-definite Complementarity Problems,* Math. Program., Series A, No. 83, pp. 159 – 185, 1998.