Matrix Convex Functions With Applications to Weighted Centers for Semidefinite Programming

Jan Brinkhuis* Zhi-Quan Luo† Shuzhong Zhang‡

August 2005

Econometric Institute Report EI-2005-38

Abstract

In this paper, we develop various calculus rules for general smooth matrix-valued functions and for the class of matrix convex (or concave) functions first introduced by Löwner and Kraus in 1930s. Then we use these calculus rules and the matrix convex function $-\log X$ to study a new notion of weighted centers for semidefinite programming (SDP) and show that, with this definition, some known properties of weighted centers for linear programming can be extended to SDP. We also show how the calculus rules for matrix convex functions can be used in the implementation of barrier methods for optimization problems involving nonlinear matrix functions.

*Econometric Institute, Erasmus University.

†Department of Electrical and Computer Engineering, University of Minnesota. Research supported by National Science Foundation, grant No. DMS-0312416.

‡Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong. Research supported by Hong Kong RGC Earmarked Grant CUHK4174/03E.

Keywords: matrix monotonicity, matrix convexity, semidefinite programming.

AMS subject classification: 15A42, 90C22.

1 Introduction

For any real-valued function f, one can define a corresponding matrix-valued function $f(X)$ on the space of real symmetric matrices by applying f to the eigenvalues in the spectral decomposition of X. Matrix functions have played an important role in scientific computing and engineering. Well-known examples of matrix function include $\sqrt{X}$ (the square root function of a positive semidefinite matrix), and e^X (the exponential function of a square matrix). In this paper, we study calculus rules for general differentiable matrix valued functions and for a special class of matrix functions called *matrix convex* functions. Historically, Löwner [13] first introduced the notion of *matrix monotone* functions in 1934. Two years later, Löwner's student Kraus extended his work to matrix convex functions; see [11]. The standard matrix analysis books of Bhatia [1] and Horn and Johnson [10] contain more historical notes and related literature on this class of matrix functions.

Our interest in matrix convex functions is motivated by the study of weighted central paths for semidefinite programming (SDP). It is well known that many properties of interior point methods for linear programming (LP) readily extend to SDP. However, there are also exceptions, one of these being the notion of *weighted centers.* The latter is essential in the V-space interior-point algorithms for linear programming. Recall that, given any positive weight vector $w > 0$ and a LP

$$\min \langle c, x\rangle, \quad \text{s.t.} \;\; Ax = b, \;\; x \geq 0,$$

we can define the w-weighted primal center as the optimal solution of the following convex program:

$$\min \langle c, x\rangle - \langle w, \log x\rangle, \quad \text{s.t.}\; Ax = b, \;\; x \geq 0,$$

where $\log x := (\cdots, \log x_i, \cdots)^T$.[1] The dual weighted center can be defined similarly. For LP, it is well known that 1) each choice of weights uniquely determines a pair of primal-dual weighted centers, and 2) the set of all primal-dual weighted centers completely fills up the relative interior of the primal-dual feasible region. How can we extend the notion of weighted center and the associated properties to SDP? A natural approach would be to define a weighted barrier function similar to the function $-\langle w, \log x\rangle$ for the LP case. However, given a symmetric positive definite weight matrix $W \succ 0$, there is no obvious way to place the weights on the eigenvalues of the matrix variable X in the standard barrier function $-\log\det X$. This difficulty has led researchers [6,18] to define weighted centers for SDP using the weighted center equations rather than through an auxiliary SDP with an appropriately weighted objective (as is the case of LP). However, these existing approaches [6,18] not only lack an optimization interpretation but also can lead to complications of non-uniqueness of the primal-dual pair of weighted centers. In this paper, we propose to use $-\langle W, \log X\rangle$ as the weighted barrier function to define a W-weighted center for SDP. It is easy to verify that when W and X are

[1]Throughout this paper, log will represent the natural logarithm.

both diagonal and positive, then $-\langle W, \log X\rangle$ simply reduces to the usual barrier function $-\langle w, \log x\rangle$ for linear programming. To ensure the convexity and develop derivative formulas for the proposed barrier function $-\langle W, \log X\rangle$, we are led to study the calculus rules for the matrix function $-\log X$, which, by the theory of Löwner and Kraus, is matrix convex.

It turns out that the calculus rules for matrix-valued functions can be developed in two different ways by either using an integral representation or using eigenvalues of the matrix variable. The integral approach relies on a basic characterization result of Löwner and Kraus to develop the desired derivative formulas for matrix monotone functions, while the eigenvalue approach is based on the use of divided differences and is applicable to more general smooth matrix-valued functions; see Section 3. As an application of these calculus rules, we define the weighted center of an SDP using the barrier function $-\langle W, \log X\rangle$, and study various properties of the resulting notion of weighted center for SDP (Section 4). In particular, we show that for any $W \succ 0$ the W-center exists uniquely. However, the set of all weighted centers (as W varies in the set of positive definite matrices) do not fill up the primal-dual feasible set. Moreover, we will show how the calculus rules can be applied to matrix convex programming problems (Section 5).

Prior to our study, there has been extensive work on the analytic properties and calculus rules of a matrix-valued function. In the work of [4], it is shown that the matrix function $f(X)$ inherits from f the properties of continuity, (local) Lipschitz continuity, directional differentiability, Fréchet differentiability, continuous differentiability, as well as semismoothness. In contrast to our work, the focus of [4] and the related work [12, 19] is on the first order (directional) derivatives by using the nonsmooth analysis of matrix functions. The main applications of the resulting first order differential formula are in the smoothing/semismooth Newton methods for solving various complementarity problems. In addition, we remark that matrix functions have also played a significant role in quantum physics [8], quantum information theory [16] and in signal processing [7]. Analysis of smooth convex functions associated with the second-order cone can be found in [6] and [3].

Our notations are fairly standard. We will use $\mathcal{H}^n$, $\mathcal{H}^n_+$, and $\mathcal{H}^n_{++}$ to denote the set of $n \times n$ Hermitian matrices, Hermitian positive semidefinite matrices, and Hermitian positive definite matrices respectively. Similarly, $\mathcal{S}^n$, $\mathcal{S}^n_+$, and $\mathcal{S}^n_{++}$ will signify real symmetric $n \times n$ matrices, symmetric positive semidefinite matrices, and symmetric positive definite matrices respectively. For generality, we shall first use the Hermitian terms, and later for notational convenience restrict to the real case when we discuss the calculus rules. In addition, we use the notation $X \succeq Y$ ($X \succ Y$) to mean $X - Y \in \mathcal{H}^n_+$ ($X - Y \in \mathcal{H}^n_{++}$). For any interval $J \subseteq \Re$, we let $\mathcal{H}^n(J)$ denote the space of all Hermitian $n \times n$ matrices whose eigenvalues all fall within J. Clearly, $\mathcal{H}^n((-\infty, +\infty)) = \mathcal{H}^n$, $\mathcal{H}^n([0, +\infty)) = \mathcal{H}^n_+$, and $\mathcal{H}^n((0, +\infty)) = \mathcal{H}^n_{++}$.

2 Matrix functions

Consider a real function $f : J \mapsto \Re$. Now we will define the primary matrix function of f. For a given $Z \in \mathcal{H}^n(J)$, let its diagonal decomposition be $Z = Q^{\mathrm{H}} D Q$ where $Q^{\mathrm{H}} Q = I$ and D is a real-valued diagonal matrix. Since $D_{jj} \in J$, $j = 1, ..., n$, let $f(D) = \mathrm{diag}\ (f(D_{11}), ..., f(D_{nn})$. Then, the matrix function $f(Z)$ is defined as

$$f(Z) := Q^{\mathrm{H}} f(D) Q. \tag{1}$$

Although the matrix decomposition of Z may not be unique, the above matrix function is uniquely defined, i.e., it does not depend on the particular decomposition matrices Q and D. Clearly, $f(Z) \in \mathcal{H}^n$ for any $Z \in \mathcal{H}^n(J)$. The following definitions follow naturally.

Definition 2.1 A function $f : J \mapsto \Re$ is said to be a *matrix monotone function on* $\mathcal{H}_n(J)$ if

$$f(X) \succeq f(Y) \text{ whenever } X, Y \in \mathcal{H}_n(J) \text{ and } X \succeq Y.$$

Note that for $n = 1$ this corresponds to the usual concept monotonically *non-decreasing* function.

Definition 2.2 A function $f : J \mapsto \Re$ is said to be a matrix convex function on $\mathcal{H}_n(J)$ if

$$(1 - \alpha) f(X) + \alpha f(Y) \succeq f((1 - \alpha) X + \alpha Y))$$

for all $X, Y \in \mathcal{H}_n(J)$ and all $\alpha \in [0, 1]$.

Definition 2.3 A function $f : J \mapsto \Re$ is said to be a strictly matrix convex function on $\mathcal{H}_n(J)$ if

$$(1 - \alpha) f(X) + \alpha f(Y) \succ f((1 - \alpha) X + \alpha Y))$$

for all $X, Y \in \mathcal{H}_n(J)$ with $X - Y$ nonsingular, and all $\alpha \in (0, 1)$.

A function f is said to be *(strictly) matrix concave* whenever $-f$ is a *(strictly) matrix convex* function. The following fundamental characterization of matrix monotone functions is due to Löwner [13]. Chapter 6 of reference [10] contains more detailed discussions on this and other related results.

Theorem 2.4 *Let J be an open (finite or infinite) interval in $\Re$, and $f : J \mapsto \Re$. The primary matrix function of f on the set of Hermitian matrices with spectrum in J is monotone for each $n \geq 1$ if and only if f can be continued to an analytic function on the upper half of the complex plane that maps the upper half of the complex plane into itself. Moreover, these are precisely the functions $f : J \mapsto \Re$ that can be described explicitly in the following form:*

$$f(x) = \alpha x + \beta + \int_{\Re} \left[\frac{1}{u - x} - \frac{u}{u^2 + 1} \right] d\mu(u), \tag{2}$$

for all $x \in J$, where $\alpha, \beta \in \Re$ with $\alpha \geq 0$, and $d\mu$ is a positive Borel measure on $\Re$ that has no mass in the interval J and for which the integral

$$\int_{\Re} \frac{d\mu(u)}{1+u^2}$$

is finite.

Note that the requirement that $d\mu(u)$ has no mass in the interval J is natural, in view of the denominator $u-x$. For practical purposes, it is convenient to consider measures of the form $d\mu(u) = m(t)dt$ where $m(t) \geq 0$ for all $t \in \Re$ and $m(t) = 0$ for all $t \in J$. For instance, if $J = (0, \infty)$ and we choose $m(t) = 1$ for all $t \leq 0$ and $m(t) = 0$ for $t > 0$, then $f(x) = \alpha x + \beta + \log x$; if $J = (0, \infty)$ and we choose $m(t) = \sqrt{-t}/\pi$ for all $t \leq 0$ and $m(t) = 0$ for $t > 0$, then $f(x) = \alpha x + \beta + \sqrt{x} - 1/\sqrt{2}$. This in turn shows that both $\log x$ and $\sqrt{x}$ are matrix monotone functions. Similarly, one can show that x^α with $0 < \alpha < 1$ is matrix monotone in general. In fact, we shall see below that these functions are also matrix concave. In contrast to the ordinary functions, the monotonicity and the concavity for the matrix functions are closely related. Moreover, in his original paper [13], Löwner also established the connection between the monotonicity and the differentiability. Below is a direct proof of the matrix monotonicity and the matrix concavity of the functions $-1/x$ on $(0, \infty)$.

Lemma 2.5 *The real valued function on $(0, \infty)$ defined as $x \mapsto -x^{-1}$ is both a matrix monotone function and a strictly matrix concave function.*

Proof. The monotonicity follows immediately from the following identity, which holds for positive definite $n \times n$ matrices X and Y:

$$X^{-1} - Y^{-1} = Y^{-1/2}(Y^{1/2}X^{-1}Y^{1/2})^{1/2}Y^{-1/2}(Y - X)Y^{-1/2}(Y^{1/2}X^{-1}Y^{1/2})^{1/2}Y^{-1/2}.$$

The matrix (strict) concavity follows from the following identity, which holds for $n \times n$ positive definite matrices X and Y with $0 \leq \alpha \leq 1$:

$$\begin{aligned} & \alpha X^{-1} + (1-\alpha)Y^{-1} - [\alpha X + (1-\alpha)Y]^{-1} \\ = \;& \alpha(1-\alpha)X^{-1}(Y - X)Y^{-1}[\alpha Y^{-1} + (1-\alpha)X^{-1}]^{-1}Y^{-1}(Y - X)X^{-1}. \end{aligned}$$

Q.E.D.

Lemma 2.6 *For all $u \leq 0$, the function $f_u(x) = 1/(u-x)$ is a monotone and strictly concave matrix function.*

Proof. This follows immediately from Lemma 2.5 by a change of variable: $f_u(x) = -\tilde{x}^{-1}$ if we put $\tilde{x} = x - u$. **Q.E.D.**

Therefore we can prove the following result:

Theorem 2.7 *If a function $f : (0, \infty) \to \Re$ is a monotone matrix function on $\mathcal{H}_+^n$ for all $n \geq 1$, then it is also a matrix concave function for all $n \geq 1$. Moreover, f is a strictly matrix concave function on $\mathcal{H}_+^n$ for all $n \geq 1$ provided the Borel measure $d\mu$ has positive mass.*

Proof. This is a consequence of Theorem 2.4, using Lemma 2.6 and noting that the matrix concavity is preserved under summation and multiplication of a nonnegative number. **Q.E.D.**

In particular, since

$$\log x = \int_{-\infty}^{0} \left[\frac{1}{u - x} - \frac{u}{u^2 + 1} \right] du, \tag{3}$$

where $x > 0$, it follows from Theorem 2.7 that the lg function is matrix monotone and strictly matrix concave. Moreover, we have the following explicit expression:

$$\log X = \int_{-\infty}^{0} \left[(uI - X)^{-1} - \frac{u}{u^2 + 1} I \right] du.$$

3 Calculating the derivatives of matrix monotone functions

In this section we discuss how to calculate the derivatives of the log matrix function. It turns out that there are two different ways to accomplish this goal: either using an integral representation (Subsection 3.1) or using a finite difference representation (Subsection 3.2). Although the two approaches are theoretically equivalent, they lead to distinct expressions which are useful in different application contexts.

3.1 An integral representation

We first introduce the following definition.

Definition 3.1 Let J be an open real interval and let $f : J \mapsto \Re$ be a three times continuously differentiable function; i.e., $f \in C^3(J)$. Then the first three derivatives are defined implicitly by the following Taylor expansion

$$f(X + H) = f(X) + f^{(1)}(X)[H] + f^{(2)}(X)[H, H] + f^{(3)}(X)[H, H, H] + o(\|H\|^3),$$

for each $X \in \mathcal{H}^n(J)$ and all $H \in \mathcal{H}^n$, where $f^{(1)}(X)$, $f^{(2)}(X)$, and $f^{(3)}(X)$ are Hermitian symmetric multi-linear mappings on the space $\mathcal{H}^n$.

We remark here that the kth derivative in the above definition differs from the conventional one by a factor of $1/k!$ ($k = 2, 3$), mainly for notational simplicity. The first simple observation is that the calculation of derivatives of matrix functions can be reduced to the case of diagonal matrices. This is summarized below.

Proposition 3.2 *Let J be an open real interval and let $f \in C^3(J)$. Let $X \in \mathcal{H}^n(J)$. Choose a diagonal decomposition $X = Q^{\mathrm{H}}DQ$. Then the following formulas hold true for all $H \in \mathcal{S}^n$ (with $K = QHQ^{\mathrm{H}}$):*

$$\begin{aligned} f^{(1)}(X)[H] &= Q^{\mathrm{H}}(f^{(1)}(D)[K])Q, \\ f^{(2)}(X)[H,H] &= Q^{\mathrm{H}}(f^{(2)}(D)[K,K])Q, \\ f^{(3)}(X)[H,H,H] &= Q^{\mathrm{H}}(f^{(3)}(D)[K,K,K])Q. \end{aligned}$$

Proof. The proposition follows immediately from the identity

$$f(X+H) - f(X) = Q^{\mathrm{H}}(f(D+K) - f(D))Q$$

and from the implicit definition of the derivatives of f at X and at D. **Q.E.D.**

In the remainder of the paper, we shall focus on the real case. Suppose that $W \in \mathcal{S}^n_{++}$. Let

$$b(X) = -\langle W, \log X\rangle \tag{4}$$

for $X \in \mathcal{S}^n_{++}$.

Theorem 3.3 *The following formulas hold true:*

$$\begin{aligned} b^{(1)}(X)[H] &= -\int_{-\infty}^{0} \langle W, (uI-X)^{-1}H(uI-X)^{-1}\rangle du, \\ b^{(2)}(X)[H,H] &= \int_{-\infty}^{0} \langle W, (uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}\rangle du, \\ b^{(3)}(X)[H,H,H] &= -\int_{-\infty}^{0} \langle W, (uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}\rangle du, \end{aligned}$$

for all $H \in \mathcal{S}^n$.

Before we prove Theorem 3.3, we comment that the expression for the first order derivative $b^{(1)}(X)[H]$ is well-known in various fields: for example, it has been used in signal processing [7], in the physics literature [8] and in quantum information theory [16]. To prove Theorem 3.3, let us first introduce two lemmas.

Lemma 3.4 *The first three derivatives of the matrix function $f : (0, +\infty) \mapsto \Re$ defined by $f(x) = x^{-1}$ are given by the following formulas*

$$\begin{aligned} f^{(1)}(X)[H] &= -X^{-1}HX^{-1}, \\ f^{(2)}(X)[H,H] &= X^{-1}HX^{-1}HX^{-1}, \\ f^{(3)}(X)[H,H,H] &= -X^{-1}HX^{-1}HX^{-1}HX^{-1}, \end{aligned}$$

for all $H \in \mathcal{S}^n$.

Proof. We have, by definition,

$$(X+H)^{-1} = X^{-1} + f^{(1)}(X)[H] + f^{(2)}(X)[H,H] + f^{(3)}(X)[H,H,H] + o(\|H\|^3).$$

Multiplying both sides from the right with $(X + H)$, expanding brackets, and equating linear, quadratic and cubic functions of H respectively, gives the following three equations:

$$\begin{aligned} &X^{-1}H + f^{(1)}(X)[H]X = 0, \\ &f^{(1)}(X)[H]H + f^{(2)}(X)[H,H]X = 0, \\ &f^{(2)}(X)[H,H]H + f^{(3)}(X)[H,H,H]X = 0. \end{aligned}$$

These equations can be solved successively, starting with the first one, to give desired formulas. **Q.E.D.**

By shifting the variable, we obtain the derivative formulas for the function $f_u(x) = (u - x)^{-1}$

$$\begin{aligned} f_u^{(1)}(X)[H] &= (uI-X)^{-1}H(uI-X)^{-1}, \\ f_u^{(2)}(X)[H,H] &= -(uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}, \\ f_u^{(3)}(X)[H,H,H] &= (uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}, \end{aligned} \tag{5}$$

for all $H \in \mathcal{S}^n$.

Proof of Theorem 3.3: We start with the identity (3). This gives the following formula for the barrier:

$$b(X) = -\int_{-\infty}^{0} \left\langle W, \left[(uI-X)^{-1} - \frac{u}{u^2+1}I\right]\right\rangle du = \int_{0}^{\infty} \left\langle W, \left[(uI+X)^{-1} - \frac{u}{u^2+1}I\right]\right\rangle du.$$

Differentiating inside the integral and using (5) gives the required formulas for the first three derivatives of b. **Q.E.D.**

The ranges for the integrations can also be changed to $\Re_+$ for convenience, as we shall do in the next section; that is,

$$b^{(1)}(X)[H] = -\int_0^\infty \langle W, (uI+X)^{-1}H(uI+X)^{-1}\rangle du, \tag{6}$$

$$b^{(2)}(X)[H,H] \;=\; -\int_0^\infty \langle W, (uI+X)^{-1}H(uI+X)^{-1}H(uI+X)^{-1}\rangle du, \tag{7}$$

$$b^{(3)}(X)[H,H,H] \;=\; -\int_0^\infty \langle W, (uI+X)^{-1}H(uI+X)^{-1}H(uI+X)^{-1}H(uI+X)^{-1}\rangle du. \tag{8}$$

One immediate consequence of Theorem 3.3 is that $b(X)$ is indeed a matrix concave function. This is because formula (7) implies that for any $X \in \mathcal{S}^n_{++}$, $W \in \mathcal{S}^n_+$ and $H \in \mathcal{S}^n$ we always have $b^{(2)}(X)[H,H] \le 0$.

By a similar argument and using Löwner's theorem (Theorem 2.7), we can extend the derivative formulas for $b(X)$ to the general matrix monotone functions.

Theorem 3.5 *Let $f : (0,\infty) \mapsto \Re$ be a matrix monotone function, i.e., there is a Borel measure $d\mu(u)$ on $\Re_-$ such that*

$$f(x) = \alpha x + \beta + \int_{-\infty}^0 \left[\frac{1}{u-x} - \frac{u}{u^2+1}\right] d\mu(u),$$

where the integral

$$\int_{-\infty}^0 \frac{d\mu(u)}{1+u^2} < \infty.$$

Then, for $X \in \mathcal{S}^n_{++}$ and $H \in \mathcal{S}^n$, there holds

$$\begin{aligned}
f^{(1)}(X)(H) &= \int_{-\infty}^0 (uI-X)^{-1}H(uI-X)^{-1}d\mu(u),\\
f^{(2)}(X)[H,H] &= -\int_{-\infty}^0 (uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}d\mu(u),\\
f^{(3)}(X)[H,H,H] &= \int_{-\infty}^0 (uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}H(uI-X)^{-1}d\mu(u).
\end{aligned}$$

3.2 An eigenvalue representation

In this subsection we use an alternative way to compute the derivatives of the barrier function $b(X) = -\langle W, \log X\rangle$. We do so by means of divided differences. Let J be a real interval and let $f : J \to \Re$ be a k-times continuously differentiable function, that is, $f \in C^k(J)$. We define the divided differences $f^{[i]} : J^i \to \Re$, $0 \le i \le k$, of f to be the continuous functions defined recursively as follows:

$$\begin{aligned}
f^{[0]} &= f,\\
f^{[i+1]}(\lambda_1,\ldots,\lambda_{i+1}) &= \frac{f^{[i]}(\lambda_1,\ldots,\lambda_{i-1},\lambda_i) - f^{[i]}(\lambda_1,\ldots,\lambda_{i-1},\lambda_{i+1})}{\lambda_i - \lambda_{i+1}},\\
&\quad \text{for } i = 0,\ldots,k-1, \text{ if } \lambda_1,\ldots,\lambda_{i+1} \text{ are distinct.}
\end{aligned}$$

For other values of $\lambda_1, \ldots, \lambda_{i+1}$, $f^{[i+1]}$ is defined by continuity. For example,

$$f^{[1]}(\lambda,\lambda) = f'(\lambda), \quad f^{[2]}(\lambda,\lambda,\lambda) = \frac{1}{2}f''(\lambda), \quad f^{[i]}(\lambda,\ldots,\lambda) = \frac{1}{i!}\frac{d^i f}{dx^i}(\lambda).$$

These functions are symmetric, i.e., the value of the function is invariant with respect to the permutation of its entries.

For any $1 \leq i \leq n$, we write $E_{i,i}$ for the diagonal $n \times n$-matrix which has 1 on the (i,i)-place and zero everywhere else. Below is our main result.

Theorem 3.6 *Let J be a real interval and let $f : J \mapsto \Re$ be a function. Consider a diagonal matrix $X = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ whose spectrum is contained in J.*

1. *For any $H \in \mathcal{S}^n$ and $f \in C^1(J)$,*

$$f^{(1)}(X)[H] = \sum_{i,j=1}^{n} f^{[1]}(\lambda_i, \lambda_j) E_{i,i} H E_{j,j}.$$

2. *For any $H \in \mathcal{S}^n$ and $f \in C^2(J)$,*

$$f^{(2)}(X)[H,H] = \sum_{i,j,k=1}^{n} f^{[2]}(\lambda_i, \lambda_j, \lambda_k) E_{i,i} H E_{j,j} H E_{k,k}.$$

3. *For any $H \in \mathcal{S}^n$ and $f \in C^3(J)$,*

$$f^{(3)}(X)[H,H,H] = \sum_{i,j,k,l=1}^{n} f^{[3]}(\lambda_i, \lambda_j, \lambda_k, \lambda_l) E_{i,i} H E_{j,j} H E_{k,k} H E_{l,l}.$$

Notice that the formula for the first derivative can be simplified using the Hadamard product of two matrices: writing $f^{[1]}(X)$ for the $n \times n$-symmetric matrix whose (i,j)-entry is $f^{[1]}(\lambda_i, \lambda_j)$, we obtain

$$f^{(1)}(X)[H] = f^{[1]}(X) \circ H. \tag{9}$$

To illustrate Theorem 3.6 let us take a few steps. First we introduce the following result.

Lemma 3.7 *If $f(x) = x^r$, then*

1. $f^{[1]}(\kappa, \lambda) = \sum_{\substack{k+l=r-1 \\ k,l \geq 0, \text{ integers}}} \kappa^k \lambda^l$, *whenever $r \geq 1$.*

2. $f^{[2]}(\kappa,\lambda,\mu) = \sum_{\substack{k+l+m=r-2 \\ k,l,m\geq 0,\text{integers}}} \kappa^k\lambda^l\mu^m$, *whenever* $r \geq 2$.

3. $f^{[3]}(\kappa,\lambda,\mu,\nu) = \sum_{\substack{k+l+m+p=r-3 \\ k,l,m,p\geq 0,\text{integers}}} \kappa^k\lambda^l\mu^m\nu^p$, *whenever* $r \geq 3$.

Proof. The first formula follows from the definition

$$f^{[1]}(\kappa,\lambda) = \frac{\kappa^r - \lambda^r}{\kappa - \lambda},$$

for $\kappa \neq \lambda$, which, by the formula for geometric progression, equals $\sum_{k,l} \kappa^k\lambda^l$, where the summation is over nonnegative integers k, l with sum $r-1$. The second formula is due to

$$\begin{aligned} f^{[2]}(\kappa,\lambda,\mu) &= \frac{f^{[1]}(\kappa,\lambda) - f^{[1]}(\kappa,\mu)}{\lambda-\mu} \\ &= \frac{\sum_{\substack{k+l=r-1 \\ k,l\geq 0,\text{integers}}} \left(\kappa^k\lambda^l - \kappa^k\mu^l\right)}{\lambda-\mu} \\ &= \sum_{\substack{k+l=r-1 \\ k,l\geq 0,\text{integers}}} \kappa^k \sum_{\substack{p+q=l-1 \\ p,q\geq 0,\text{integers}}} \lambda^p\mu^q \\ &= \sum_{\substack{k+l+m=r-2 \\ k,l,m\geq 0,\text{integers}}} \kappa^k\lambda^l\mu^m. \end{aligned}$$

The last formula can be established in a similar way. **Q.E.D.**

Lemma 3.8 *Theorem 3.6 holds true for power functions* $f(x) = x^r$, *with* r *a nonnegative integer.*

Proof. Notice that

$$f(X+H) - f(X) = (X+H)^r - X^r = \sum_{\substack{k+l=r-1 \\ k,l\geq 0,\text{integers}}} X^kHX^l + o(\|H\|).$$

Writing the diagonal matrix X as $X = \sum_{i=1}^{n} \lambda_i E_{i,i}$ in the above expression and expanding the products yields

$$f(X+H) - f(X) = \sum_{i,j=1}^{n} \sum_{\substack{k+l=r-1 \\ k,l\geq 0,\text{integers}}} \lambda_i^k\lambda_j^l {E_{i,i}}^k H {E_{j,j}}^l + o(\|H\|)$$

$$= \sum_{i,j=1}^{n} \sum_{\substack{k+l=r-1 \\ k,l\geq 0,\text{integers}}} \lambda_i^k \lambda_j^l E_{i,i} H E_{j,j} + o(\|H\|)$$

$$= \sum_{i,j} f^{[1]}(\lambda_i, \lambda_j) E_{i,i} H E_{j,j} + o(\|H\|),$$

where first step is due to $E_{i,i}E_{j,j} = 0$ whenever $i \neq j$, the second step follows from $E_{i,i}^k = E_{i,i}$, $E_{j,j}^l = E_{j,j}$, and the last step is due to Lemma 3.7. This proves the first formula. The other two formulas can be established in a similar manner. **Q.E.D.**

Lemma 3.8 shows that Theorem 3.6 holds for power functions. Taking linear combinations, we see immediately that the theorem holds for polynomials. This further suggests that the theorem holds for general functions with sufficient smoothness. A rigorous proof of Theorem 3.6 requires a careful analysis of the local behavior of $f(X+H)$ using the Lipschitzian continuity of the eigen-decomposition of $X+H$. Our proof is an extension of the first order differentiability argument used in [4]. Since the complete proof is tedious, we relegate it to an appendix at the end of the paper.

We emphasize that Theorem 3.6 is applicable to general (smooth) functions. In this sense, it is much more general than the corresponding expressions in Theorem 3.5 which are valid only for matrix monotone functions. Now we apply Theorem 3.6 to the barrier function $b(X) = -\langle W, \log X\rangle$.

Theorem 3.9 *For any $H \in \mathcal{S}^n$, the following formulas hold true for the barrier function $b(X) = -\langle W, \log X\rangle$ at a positive definite diagonal matrix $X = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$:*

1. $b^{(1)}(X)[H] = -\sum_{i,j=1}^{n} \log^{[1]}(\lambda_i, \lambda_j)\langle W, E_{i,i} H E_{j,j}\rangle$;
2. $b^{(2)}(X)[H,H] = -\sum_{i,j,k=1}^{n} \log^{[2]}(\lambda_i, \lambda_j, \lambda_k)\langle W, E_{i,i} H E_{j,j} H E_{k,k}\rangle$;
3. $b^{(3)}(X)[H,H,H] = -\sum_{i,j,k,l=1}^{n} \log^{[3]}(\lambda_i, \lambda_j, \lambda_k, \lambda_l)\langle W, E_{i,i} H E_{j,j} H E_{k,k} H E_{l,l}\rangle$.

By combining Theorem 3.9 and Proposition 3.2, we can derive similar derivative formulas for general matrices admitting a diagonal decomposition. Notice that the derivative formulas above require the divided differences of $\log x$. Unfortunately, this is not so easy to compute in a direct way. However, it is possible to do this indirectly by computing the divided differences first for the function $x \mapsto x^{-1}$, and then for the functions $x \mapsto f_u(x) = (u-x)^{-1}$, and finally using relation (3). This leads to the same formula as in the previous section. We will only display here the formulas for the derivatives of the primary matrix functions of the function $x \mapsto -x^{-1}$.

Proposition 3.10 *Let* $f(x) = -x^{-1}$ *for* $x \in (0, \infty)$. *The following formulas hold true for the first three derivatives of the primary matrix function* $f(X)$ *at any positive diagonal matrix* $X = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$*:*

1. $f^{(1)}(X)[H] = \sum_{k,l=1}^{n} \lambda_k^{-1}\lambda_l^{-1} E_{k,k} H E_{l,l}$,

2. $f^{(2)}(X)[H, H] = \sum_{k,l,m=1}^{n} \lambda_k^{-1}\lambda_l^{-1}\lambda_m^{-1} E_{k,k} H E_{l,l} H E_{m,m}$,

3. $f^{(3)}(X)[H, H, H] = \sum_{k,l,m,n=1}^{n} \lambda_k^{-1}\lambda_l^{-1}\lambda_m^{-1}\lambda_n^{-1} E_{k,k} H E_{l,l} H E_{m,m} H E_{n,n}$,

for all $H \in \mathcal{S}^n$.

Proof. We only need to compute the divided differences for the function $f(x) = -x^{-1}$. We claim

1. $f^{[1]}(\kappa, \lambda) = (\kappa\lambda)^{-1}$,
2. $f^{[2]}(\kappa, \lambda, \mu) = (\kappa\lambda\mu)^{-1}$,
3. $f^{[3]}(\kappa, \lambda, \mu, \nu) = (\kappa\lambda\mu\nu)^{-1}$,

for all $\kappa, \lambda, \mu, \nu \geq 0$. To see the first formula, we note $f^{[1]}(\kappa, \lambda) = \frac{-\kappa^{-1}+\lambda^{-1}}{\kappa-\lambda} = (\kappa\lambda)^{-1}$, for $\kappa \neq \lambda$, as desired. Continuing in the same way we can verify the remaining two formulas. **Q.E.D.**

As a remark, we notice that the formula for the first derivative can also be rewritten as

$$f^{(1)}(X)[H] = f^{[1]}(X) \circ H.$$

while the second derivative can be written alternatively as

$$f^{(2)}(X)[H, H] = \sum_{k,l,m=1}^{n} \lambda_k^{-1}\lambda_l^{-1}\lambda_m^{-1} h_{kl} h_{lm}.$$

Finally, we can use Theorem 3.9 to derive some simple properties for matrix convex functions.

Proposition 3.11 *For any matrix convex function* $f(X)$ *and any* $1 \leq j \leq n$, *there holds*

$$M_j := \left(f^{[2]}(\lambda_i, \lambda_j, \lambda_k)\right)_{n\times n} \succeq 0$$

for all $X \in \mathcal{S}^n_{++}$ *and* $H \in \mathcal{S}^n$.

Proof. In light of Proposition 3.2, we only need to consider the case where $X = \text{diag}\ (\lambda_1, \lambda_2, ..., \lambda_n)$ is a positive diagonal matrix. Then, Theorem 3.6 asserts that

$$f^{(2)}(X)[H,H] = \left(\sum_{j=1}^{n} f^{[2]}(\lambda_i, \lambda_j, \lambda_k) h_{ij} h_{jk} \right)_{n \times n}.$$

Take any $x, y \in \Re^n$ with $y_i \neq 0$, $i = 1, ..., n$. Let $w = x \circ y^{-1}$ and $H = yy^T$. We have

$$\begin{aligned} w^T f^{(2)}(X)[H,H]w &= \sum_{j=1}^{n} \sum_{i,k=1}^{n} f^{[2]}(\lambda_i, \lambda_j, \lambda_k) \frac{x_i}{y_i} y_i y_j y_j y_k \frac{x_k}{y_k} \\ &= \sum_{j=1}^{n} \left(\sum_{i,k=1}^{n} f^{[2]}(\lambda_i, \lambda_j, \lambda_k) x_i x_k \right) y_j^2 \\ &= \sum_{j=1}^{n} (x^T M_j x) y_j^2 \\ &\geq 0, \end{aligned}$$

where the last step follows from matrix convexity which implies $f^{(2)}(X)[H,H] \succeq 0$. This shows that $M_j \succeq 0$ for all $j = 1, ..., n$. **Q.E.D.**

Let us now specialize Theorem 3.6 to the matrix exponential function e^X (which is known not to be matrix convex so Theorem 3.5 does not apply).

Proposition 3.12 *For any symmetric X and H, there holds*

$$(e^X)^{(1)}[H] = \int_0^1 e^{(1-u)X} H e^{uX} du \tag{10}$$

$$(e^X)^{(2)}[H,H] = \int_0^1 \int_0^1 (1-u) e^{uX} H e^{v(1-u)X} H e^{(1-v)(1-u)X} dudv \tag{11}$$

Proof. We only need to prove the proposition for diagonal matrix $X = \text{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. Assume that λ_i's are all distinct. In light of (9), we can compute the (i,j)-th entry $(i \neq j)$ of the matrix differential $(e^X)^{(1)}[H]$:

$$\begin{aligned} (e^X)^{(1)}[H]_{i,j} &= \left[(e^X)^{[1]} \circ H \right]_{i,j} \\ &= \frac{e^{\lambda_i} - e^{\lambda_j}}{\lambda_i - \lambda_j} H_{i,j} \\ &= \int_0^1 e^{\lambda_i u} e^{(1-u)\lambda_j} H_{i,j} du \\ &= \left[\int_0^1 e^{uX} H e^{(1-u)X} du \right]_{i,j}, \end{aligned}$$

where the third equality follows from the identity

$$\frac{e^{\lambda_i}-e^{\lambda_j}}{\lambda_i-\lambda_j}=\int_0^1 e^{\lambda_i u}e^{(1-u)\lambda_j}du. \tag{12}$$

This proves (10) for the case of $i\neq j$. The case of $i=j$ can be considered in a similar fashion.

Now we prove the second order differential formula (11). Consider the (i,j)-th entry $(i\neq j)$ of the second order matrix differential $(e^X)^{(2)}[H,H]_{i,j}$:

$$\begin{aligned}
(e^X)^{(2)}[H,H]_{i,j} &= \sum_k (e^X)^{[2]}(\lambda_i,\lambda_k,\lambda_j)H_{i,k}H_{k,j}\\
&= \sum_k (e^X)^{[2]}(\lambda_i,\lambda_k,\lambda_j)H_{i,k}H_{k,j}\\
&= \sum_k \frac{\frac{e^{\lambda_i}-e^{\lambda_k}}{\lambda_i-\lambda_k}-\frac{e^{\lambda_i}-e^{\lambda_j}}{\lambda_i-\lambda_j}}{\lambda_k-\lambda_j}H_{i,k}H_{k,j}\\
&= \sum_k \frac{\int_0^1 e^{\lambda_i u}e^{(1-u)\lambda_k}du-\int_0^1 e^{\lambda_i u}e^{(1-u)\lambda_j}du}{\lambda_k-\lambda_j}H_{i,k}H_{k,j}\\
&= \sum_k \int_0^1 e^{\lambda_i u}\frac{e^{(1-u)\lambda_k}-e^{(1-u)\lambda_j}}{\lambda_k-\lambda_j}H_{i,k}H_{k,j}du.
\end{aligned}$$

Now we use the identity (12) to obtain

$$\begin{aligned}
(e^X)^{(2)}[H,H]_{i,j} &= \sum_k \int_0^1 (1-u)e^{\lambda_i u}\int_0^1 e^{v(1-u)\lambda_k}e^{(1-v)(1-u)\lambda_j}H_{i,k}H_{k,j}dvdu\\
&= \int_0^1\int_0^1 (1-u)\left[e^{uX}He^{v(1-u)X}He^{(1-v)(1-u)X}\right]_{i,j}dvdu,
\end{aligned}$$

which establishes (11). **Q.E.D.**

Notice that the first order derivative formula (10) for the matrix exponential function e^X has been used extensively in the physics literature [8] and in applied mathematics [15].

4 Weighted centers for semidefinite programming

Consider the following standard semidefinite programming (SDP) problem

$$\begin{array}{lll}
(P) & \text{minimize} & \langle C,X\rangle\\
 & \text{subject to} & \langle A_i,X\rangle=b_i,\ i=1,...,m\\
 & & X\succeq 0,
\end{array}$$

and its dual

$$
\begin{array}{lll}
(D) & \text{maximize} & b^T y \\
& \text{subject to} & \sum_{i=1}^{m} y_i A_i + Z = C \\
& & Z \succeq 0.
\end{array}
$$

The study of various aspects of SDP can be found in [20]. It is well known that many properties of the interior point methods for linear programming (LP) readily extend to SDP. However, one exception is the notion of *weighted centers*. Sturm and Zhang [18] proposed to define the weighted centers of the SDP problems (P) and (D) based on the eigenvalues of the product of a pair of primal-dual feasible solutions XZ. However, such pair may not be unique. Chua [6] proposed the weighted centers based on a diagonal and positive weight matrix W. Since the $\log X$ is a matrix function, it is now natural to define the weighted centers by means of the barrier function $b(X) = -\langle W, \log X \rangle$. To be specific, given any weight matrix $W \succ 0$, let us consider

$$
\begin{array}{lll}
(P_w) & \text{minimize} & \langle C, X \rangle - \langle W, \log X \rangle \\
& \text{subject to} & \langle A_i, X \rangle = b_i, \; i = 1, ..., m.
\end{array}
$$

We shall first establish the existence of a primal weighted center based on (P_w). Note the following lemmas.

Lemma 4.1 *For any $X \succ 0$ and $t > 0$ it holds that $b(tX) = b(X) + (\log t)\operatorname{tr} W$.*

Proof. Let the orthonormal decomposition of X be $X = P^T D P$ where P is an orthonormal matrix and D is positive diagonal. Then

$$
\log(tX) = P^T (\log(tD))\, P = P^T (\log D + (\log t)\, I)\, P = X + (\log t)\, I,
$$

and so

$$
b(tX) = \langle W, \log(tX) \rangle = b(X) + (\log t)\operatorname{tr} W.
$$

Q.E.D.

Lemma 4.2 *Let $\mathcal{K} \subseteq \Re^n$ be a closed convex cone, $\mathcal{K}^*$ be its dual cone, and $\mathcal{L} \subseteq \Re^n$ be a subspace. Let $c \in \Re^n$ be a given vector. Suppose that* $\operatorname{int} \mathcal{K}^* \cap (c + \mathcal{L}^\perp) \neq \emptyset$. *In that case, if there is any $0 \neq x \in \mathcal{K} \cap \mathcal{L}$ then it must follow that $c^T x > 0$.*

This result is also known as the extended Farkas lemma; see e.g. [17] for discussions.

Theorem 4.3 *Suppose that both (P) and (D) satisfy the Slater condition. Then for any symmetric $W \succ 0$ there exists a unique optimal solution for (P_w).*

Proof. Let X^k be a sequence of feasible solutions for (P_w) such that $\langle C, X^k\rangle - \langle W, \log X^k\rangle$ converges to the optimal value of (P_w). First we see that $\|X^k\|$ must be bounded, for otherwise we may assume without loss of generality that $\lim_{k\to\infty} \|X^k\| = \infty$ and

$$\lim_{k\to\infty} \frac{X^k}{\|X^k\|} = \hat{X}.$$

In that case, since by Lemma 4.2 we know that $\langle C, \hat{X}\rangle > 0$, and also using Lemma 4.1 it follows that

$$\langle C, X^k\rangle - \left\langle W, \log X^k \right\rangle = \|X^k\| \left\langle C, \frac{X^k}{\|X^k\|} \right\rangle - \left\langle W, \log \frac{X^k}{\|X^k\|} \right\rangle - \log \|X^k\| \operatorname{tr} W \to \infty,$$

which is impossible. This shows that (P_w) must indeed have attainable optimal solution. Due to the strict convexity of the objective function, such optimal solution is unique. **Q.E.D.**

Let X^p_w be the optimal solution for (P_w). Using Theorem 3.3 we obtain the following Karush-Kuhn-Tucker optimality condition for X^p_w: $\exists\, y^p \in \Re^m$ such that

$$C - \sum_{i=1}^m y_i^p A_i - \int_0^\infty (uI + X_w^p)^{-1} W (uI + X_w^p)^{-1} du = 0. \tag{13}$$

Let us define a matrix mapping $F_W : \mathcal{S}^n_+ \mapsto \mathcal{S}^n_+$:

$$F_W(X) := \int_0^\infty (uI + X)^{-1} W (uI + X)^{-1} du.$$

Obviously, (13) induces a dual solution

$$Z^p_w = C - \sum_{i=1}^m y_i^p A_i = F_W(X^p_w). \tag{14}$$

For the same weight matrix $W \succ 0$, we can also consider the barrier problem for the dual:

$$(D_w) \quad \text{maximize} \quad b^T y + \left\langle W, \log \left(C - \sum_{i=1}^m y_i A_i \right) \right\rangle.$$

Similar to Theorem 4.3, we can show (D_w) has a unique optimal solution, which we denote by y^d. Again, by Theorem 3.3, the KKT optimality condition for (D_w) reduces to

$$b_i - \left\langle A_i, \int_0^\infty \left(vI + C - \sum_{i=1}^m y_i^d A_i \right)^{-1} W \left(vI + C - \sum_{i=1}^m y_i^d A_i \right)^{-1} dv \right\rangle = 0,\ i = 1, 2, ..., m. \tag{15}$$

The condition (15) induces a primal solution

$$X^d_w = \int_0^\infty \left(vI + C - \sum_{i=1}^m y_i^d A_i \right)^{-1} W \left(vI + C - \sum_{i=1}^m y_i^d A_i \right)^{-1} dv = F_W \left(C - \sum_{i=1}^m y_i^d A_i \right). \tag{16}$$

It is well known that, for linear programming, the weighted center pairs $\{X_w^p, Z_w^p\}$, $\{X_w^d, Z_w^d\}$ coincide; furthermore, both pairs of centers are diagonal and therefore they commute and satisfy $X_w^p Z_w^p = X_w^d Z_w^d = W$. Interestingly, in the SDP case, the two pairs of centers $\{X_w^p, Z_w^p\}$, $\{X_w^d, Z_w^d\}$ do not coincide and the commutability fails to hold in general. This can be seen from the following simple 2×2 example: let

$$A_1 = \frac{1}{2}E_{1,1}, \quad A_2 = E_{2,2}, \quad A_3 = E_{1,2}, \quad b_1 = b_2 = b_3 = 1, \quad C = E_{1,1} + E_{2,2} + E_{1,2}, \quad W = C + E_{1,1},$$

where $E_{i,j}$ denotes the symmetric matrix with all entries zero except at (i,j)- and (j,i)-th entries which equal 1. In this case, there is a unique primal feasible matrix which is also equal to the W-center: $X_w^p = \mathrm{Diag}\{2,1\}$. The corresponding dual center is

$$Z_w^p = F_W(X_w^p) = \log^{[1]}(2,1) \circ W = \begin{bmatrix} 1 & \log 2 \\ \log 2 & 1 \end{bmatrix}.$$

Clearly, the matrices X_w^p and Z_w^p do not commute. Moreover, we can directly compute the dual weighted center pair $\{X_w^d, Z_w^d\}$ to verify that $X_w^d = X_w^p = \mathrm{Diag}\{2,1\}$, and $Z_w^d \neq Z_w^p$. Alternatively, we can prove the latter inequality by contradiction. In particular, suppose $Z_w^d = Z_w^p$. Then the condition (15) would imply $X_w^p = F_W(Z_w^p)$. Notice that

$$Z_w^p = Q \begin{bmatrix} 1+\log 2 & 0 \\ 0 & 1-\log 2 \end{bmatrix} Q^T, \quad Q'WQ = \frac{1}{2}\begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}, \quad \text{where } Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Using the definition of F_W and simplifying the integral yields

$$\begin{aligned}
X_w^p &= \int_0^\infty (uI + Z_w^p)^{-1} W (uI + Z_w^p)^{-1} du \\
&= Q\left(\int_0^\infty (uI + \mathrm{Diag}\{1+\log 2, 1-\log 2\})^{-1} Q'WQ (uI + \mathrm{Diag}\{1+\log 2, 1-\log 2\})^{-1} du\right) Q' \\
&= Q\left(\log^{[1]}(1+\log 2, 1-\log 2) \circ (Q'WQ)\right) Q' \\
&= Q \begin{bmatrix} \frac{5}{2(1+\log 2)} & \frac{\log(1+\log 2)-\log(1-\log 2)}{4\log 2} \\ \frac{\log(1+\log 2)-\log(1-\log 2)}{4\log 2} & \frac{1}{2(1+\log 2)} \end{bmatrix} Q' \\
&= \begin{bmatrix} 2.1690 & -0.0765 \\ -0.0765 & 0.9370 \end{bmatrix},
\end{aligned}$$

contradicting the condition $X_w^p = \mathrm{Diag}\{2,1\}$. Therefore, we have established $Z_w^p \neq Z_w^d$.

The lack of commutability between X_w^p and Z_w^p (and similarly X_w^d, Z_w^d) further implies that the property $X_w^p Z_w^p = X_w^d Z_w^d = W$ cannot hold in the SDP case. Interestingly, a related property does hold as shown in the next result.

Theorem 4.4 *Given any $W \succ 0$, let $\{X_w^p, Z_w^p\}$, $\{X_w^d, Z_w^d\}$ be defined by* (13)-(14) *and* (15)-(16) *respectively. Then, there holds*

$$\mathrm{tr}\, X_w^p Z_w^p = \mathrm{tr}\, X_w^d Z_w^d = \mathrm{tr}\, W.$$

Proof. Since X_w^p and $(uI + X_w^p)^{-1}$ commute for any $u \geq 0$, it follows that

$$\begin{aligned}
\operatorname{tr} X_w^p Z_w^p &= \operatorname{tr} \int_0^\infty X_w^p (uI + X_w^p)^{-1} W (uI + X_w^p)^{-1} du \\
&= \operatorname{tr} \int_0^\infty (uI + X_w^p)^{-1} X_w^p W (uI + X_w^p)^{-1} du \\
&= \operatorname{tr} \int_0^\infty X_w^p W (uI + X_w^p)^{-2} du \\
&= \operatorname{tr} X_w^p W \, (X_w^p)^{-1} \\
&= \operatorname{tr} W,
\end{aligned}$$

where the third and the last steps are due to the identity $\operatorname{tr} AB = \operatorname{tr} BA$ for any matrices A and B. Similarly, we can show that $\operatorname{tr} X_w^d Z_w^d = \operatorname{tr} W$. **Q.E.D.**

Another property of weighted centers for linear programming is the fact that they fill up the entire primal and dual feasible region. Interestingly, this property no longer holds in the SDP case as is illustrated in the following example. Consider the primal SDP (P) with $m = 2n$ and

$$C = \text{Blockdiag}\left\{ \begin{bmatrix} 1 & 1-\epsilon \\ 1-\epsilon & 1 \end{bmatrix}, I_{n-2,n-2} \right\}, \quad A_i = E_{l,k}, \quad b_i = \delta_{l,k} + \delta_{l,1},$$

for $l = 1, 2$ and $k = 1, 2, ..., n$, or $k = 1, 2$ and $l = 1, 2, ..., n$, where $E_{l,k}$ denotes the $n \times n$ matrix whose entries are all zero except the (l, k)- and (k, l)-th entries which are 1; $\epsilon > 0$ is a constant to be chosen later; $\delta_{l,k}$ denotes the usual Kronecker function. In this case, the primal feasible set consists of all matrices of the form

$$X = \text{Blockdiag}\left\{ \begin{bmatrix} 1 & 1-\epsilon \\ 1-\epsilon & 1 \end{bmatrix}, M_{n-2,n-2} \right\}, \quad \text{with } M \succeq 0. \tag{17}$$

We claim that there *cannot* be any weight matrix $W \succ 0$ and any primal feasible matrix X which together with the dual feasible matrix $Z = C$ forms a pair of W-centers for this SDP (P), provided ϵ is small. Specifically, suppose there holds

$$C = \text{Blockdiag}\left\{ \begin{bmatrix} 1 & 1-\epsilon \\ 1-\epsilon & 1 \end{bmatrix}, I_{n-2,n-2} \right\} = \int_0^\infty (uI + X)^{-1} W (uI + X)^{-1} du$$

for some primal feasible matrix X of the form (17) and some symmetric weight matrix $W = [w_{ij}] \succ 0$. Since X has a block diagonal structure, the first principal 2×2 submatrix of the above right-hand-side integral can be easily calculated to be

$$\begin{bmatrix} \frac{1}{2} w_{11} & w_{12} \log 2 \\ w_{12} \log 2 & w_{22} \end{bmatrix}.$$

Equating this submatrix with that of C yields

$$w_{11} = 2, \quad w_{22} = 1, \quad w_{12} = \frac{1-\epsilon}{\log 2}.$$

This implies

$$w_{11}w_{22} - w_{12}^2 = 2 - \frac{(1-\epsilon)^2}{\log^2 2} < 0, \quad \text{for sufficiently small } \epsilon.$$

This contradicts the positive definiteness of W matrix. This shows that C cannot be a dual center Z_w^p for any choice of $W \succ 0$ and any primal feasible X_w^p.

5 Matrix convex programming

It is elementary to see that if f is matrix concave and g is matrix monotone, then the composite function $g \circ f$ is matrix concave. Also, the direct sum of matrix concave functions remain matrix concave.

Let us now consider the following matrix convex programming problem

$$\begin{array}{lll} (MCP) & \text{minimize} & \langle C, X \rangle \\ & \text{subject to} & f_j(X) \succeq B_j,\ j = 1, ..., m, \\ & & X \in \mathcal{S}^n, \end{array}$$

where f_j is matrix concave, $j = 1, ..., m$. This problem can be regarded as a kind of 'nonlinear' (but still convex) SDP. A different type of 'nonlinear' SDP model was studied in [21], with a provable polynomial-time computational complexity bound. The above model (MCP) is useful. For example, in many signal processing applications [14], we have $f_j(X) = C_j^T X + XC_j - X^2$ for some matrix C_j. A standard approach to handle the concave quadratic matrix inequality $f_j(X) \succeq B_j$ is to convert it to an equivalent linear matrix inequality by using Schur complement. However, such a conversion, while resulting in a polynomial time algorithm, will increase the problem dimension substantially, often leading to numerical difficulties in the solution of the resulting large scale SDP. A numerically more appealing approach is to treat the quadratic matrix inequality $f_j(X) \succeq B_j$ directly using a standard logarithmic barrier $-\text{tr} \log(f_j(X) - B_j)$. In this way, there is no increase in problem dimension nor the need to manage the sparse problem structure of an otherwise large SDP.

Let us consider a standard logarithmic barrier method for solving (MCP). Suppose that the Slater condition holds for (MCP), and then we introduce a barrier function for (MCP) as

$$g(X) := -\text{tr} \sum_{j=1}^{m} \log(f_j(X) - B_j).$$

The key step now is the ability to compute the Newton direction for the function

$$\langle C, X\rangle + \mu g(X),$$

at a given iterative point. Denote $g_j(X) := -\log(f_j(X) - B_j)$, $j = 1, ..., m$, which are all matrix concave functions.

Consider an iterative point $X^k \in \mathcal{S}^n$ with $f_j(X^k) \succ B_j$, $j = 1, ..., m$. Let $X^k = QD^kQ^T$ be an orthonormal decomposition of X^k, and $C^k := Q^TCQ$. Proposition 3.2 suggests that

$$\begin{aligned} g_j^{(1)}(X^k)[H] &= Q\left(g_j^{(1)}(D^k)[Q^THQ]\right)Q^T \\ g_j^{(2)}(X^k)[H,H] &= Q\left(g_j^{(2)}(D^k)[Q^THQ, Q^THQ]\right)Q^T \end{aligned}$$

for $j = 1, ..., m$. Hence, by letting $\bar{H} = Q^THQ$, and using Theorem 3.6 we have

$$\begin{aligned} g^{(1)}(X^k)[H] &= \sum_{j=1}^m \operatorname{tr} g_j^{(1)}(D^k)[\bar{H}] = \sum_{j=1}^m \sum_{p=1}^n g_j^{[1]}(d_p^k, d_p^k)\bar{h}_{pp} \\ g^{(2)}(X^k)[H,H] &= \sum_{j=1}^m \operatorname{tr} g_j^{(2)}(D^k)[\bar{H},\bar{H}] = \sum_{j=1}^m \sum_{p,q=1}^n g_j^{[2]}(d_p^k, d_q^k, d_p^k)\bar{h}_{pq}^2. \end{aligned}$$

Therefore, the Newton direction is given by $H = Q\bar{H}Q^T$, where $\bar{H} = (\bar{h}_{pq})_{n\times n} \in \mathcal{S}^n$ is the minimizer of the following separable convex quadratic function

$$\sum_{p,q=1}^n C_{pq}^k\bar{h}_{pq} + \mu\sum_{j=1}^m\sum_{p=1}^n g_j^{[1]}(d_p^k, d_p^k)\bar{h}_{pp} + \mu\sum_{j=1}^m\sum_{p,q=1}^n g_j^{[2]}(d_p^k, d_q^k, d_p^k)\bar{h}_{pq}^2.$$

In particular, we have

$$\bar{h}_{pq} = \begin{cases} -\dfrac{C_{pq}^k}{\mu\sum_{j=1}^m\left[g_j^{[2]}(d_p^k, d_q^k, d_p^k) + g_j^{[2]}(d_q^k, d_p^k, d_q^k)\right]}, & \text{for } p \neq q; \\[2ex] -\dfrac{C_{pp}^k + \mu\sum_{j=1}^m g_j^{[1]}(d_p^k, d_p^k)}{2\mu\sum_{j=1}^m g_j^{[2]}(d_p^k, d_p^k, d_p^k)}, & \text{for } p = q. \end{cases}$$

As a conclusion, we see that the total number of basic operations required to assemble such a Newton direction is $O(mn^3)$.

A Appendix: Proof of Theorem 3.6

Part 1 of Theorem 3.6 was established in [4]. We will only show part 2 here. The proof of part 3 is similar and therefore omitted. To establish the second order derivative formula, we consider a diagonal matrix $X = \mathrm{Diag}\{\lambda_1, ..., \lambda_n\}$ and suppose $f : J \to \Re$ is differentiable at $\lambda_1, ..., \lambda_n$. We can without loss of generality assume that the diagonal entries of X are distinct and ordered: $\lambda_1 < \cdots < \lambda_n$. [The case of equal diagonal entries can be handled using a simple continuity argument.] By Lemma 3 in reference [5], there exist scalars $\eta > 0$ and $\epsilon > 0$ such that for any $H \in \mathcal{S}^n$ with $\|H\| \le \epsilon$, there exists an orthnormal matrix P with the property that

$$X + H = P^T \mathrm{Diag}\{\mu_1, \mu_2, ..., \mu_n\} P, \quad \text{with } \mu_1 \le \cdots \le \mu_n, \ \text{ and } \|P - I\| \le \eta \|H\|. \tag{18}$$

This implies that the off-diagonal entries of P are of order $O(\|H\|)$. Moreover, according to a perturbation result of Weyl for eigenvalues of symmetric matrices (see [1, p. 63]),

$$|\lambda_i - \mu_i| \le \|H\|, \quad \forall\, i = 1, ..., n. \tag{19}$$

Notice that the orthonormality of P together with (18) imply

$$1 = P_{ii}^2 + \sum_{k \ne i} P_{ki}^2 = P_{ii}^2 + O(\|H\|^2), \tag{20}$$

$$0 = P_{ii}P_{ij} + P_{ji}P_{jj} + \sum_{k \ne i,j} P_{ki}P_{kj} = P_{ii}P_{ij} + P_{ji}P_{jj} + O(\|H\|^2), \quad i \ne j. \tag{21}$$

We will show that, for any $H \in \mathcal{S}^n$ with $\|H\| \le \epsilon$, such that

$$f(X + H) - f(X) - f^{(1)}(X)[H] - f^{(2)}(X)[H, H] = o(\|H\|^2), \tag{22}$$

where the constants in $o(\cdot)$ depend on f and X only, and $f^{(1)}(X)[H]$, $f^{(2)}(X)[H, H]$ are given by

$$\begin{aligned} f^{(1)}(X)[H] &= \sum_{i,j=1}^{n} f^{[1]}(\lambda_i, \lambda_j) E_{i,i} H E_{j,j} \\ f^{(2)}(X)[H, H] &= \sum_{i,j,k=1}^{n} f^{[2]}(\lambda_i, \lambda_j, \lambda_k) E_{i,i} H E_{j,j} H E_{k,k}. \end{aligned}$$

This would show that $f(X)$ is twice differentiable at the diagonal matrix X with the first and second order directional derivatives given by $f^{(1)}(X)[H]$ and $f^{(2)}(X)[H, H]$ respectively. Substituting the definitions of $f^{(1)}(X)[H]$ and $f^{(2)}(X)[H, H]$ into the left side of (22) yields

$$\begin{aligned} & f(X + H) - f(X) - f^{(1)}(X)[H] - f^{(2)}(X)[H, H] \\ & \quad = P^T \mathrm{Diag}\{f(\mu_1), \ldots, f(\mu_n)\} P - \mathrm{Diag}\{\lambda_1, \ldots, \lambda_n\} \\ & \qquad - \sum_{i,j=1}^{n} f^{[1]}(\lambda_i, \lambda_j) E_{i,i} H E_{j,j} - \sum_{i,j,k=1}^{n} f^{[2]}(\lambda_i, \lambda_k, \lambda_j) E_{i,i} H E_{k,k} H E_{j,j}. \end{aligned} \tag{23}$$

We need to show that each entry of the above matrix equation is $o(\|H\|^2)$. We separate two cases: diagonal entries and off-diagonal entries.

Let us first consider the (i,i)-th diagonal entry of the above matrix equation (23). Notice that from the relation $X+H = P^T \mathrm{Diag}\{\mu_1,\dots,\mu_n\}P$ (cf. (18)) we have

$$\lambda_i + H_{ii} = \sum_{k=1}^{n} P_{ki}^2 \mu_k. \tag{24}$$

Substituting this relation into the (i,i)-th entry of (23) and simplifying yields

$$\begin{aligned}\text{the } (i,i)\text{-th entry} &= \sum_{k=1}^{n} P_{ki}^2 f(\mu_k) - f(\lambda_i) - f'(\lambda_i)H_{ii} - \sum_{k=1}^{n} f^{[2]}(\lambda_i,\lambda_k,\lambda_i)H_{ki}^2 \\ &= f(\mu_i) - f(\lambda_i) - f'(\lambda_i)(\mu_i-\lambda_i) + \sum_{k\neq i}\left(f(\mu_k)-f(\mu_i)-f'(\lambda_i)(\mu_k-\mu_i)\right)P_{ki}^2 \\ &\quad -\sum_{k=1}^{n} f^{[2]}(\lambda_i,\lambda_k,\lambda_i)H_{ki}^2. \end{aligned} \tag{25}$$

We need to bound the last term of (25) which can be written as

$$\sum_{k=1}^{n} f^{[2]}(\lambda_i,\lambda_k,\lambda_i)H_{ki}^2 = \frac{1}{2}f''(\lambda_i)H_{ii}^2 + \sum_{k\neq i} f^{(2)}(\lambda_i,\lambda_k,\lambda_i)H_{ki}^2, \tag{26}$$

where we have used the fact $f^{[2]}(\lambda_i,\lambda_i,\lambda_i) = \frac{1}{2}f''(\lambda_i)$. Therefore, we need to bound the two terms of the above expression separately. The first term can be estimated as follows:

$$\begin{aligned} f''(\lambda_i)H_{ii}^2 &\overset{\text{(a)}}{=} f''(\lambda_i)\left(\lambda_i - \sum_{k=1}^{n} P_{ki}^2\mu_k\right)^2 \\ &= f''(\lambda_i)\left(\lambda_i - \mu_i + \sum_{k=1}^{n} P_{ki}^2(\mu_i-\mu_k)\right)^2 \\ &\overset{\text{(b)}}{=} f''(\lambda_i)\left(\lambda_i-\mu_i\right)^2 + 2f''(\lambda_i)\left(\lambda_i-\mu_i\right)\sum_{k\neq i} P_{ki}^2(\mu_i-\mu_k) + O(\|H\|^4) \\ &\overset{\text{(c)}}{=} f''(\lambda_i)\left(\lambda_i-\mu_i\right)^2 + O(\|H\|^3), \end{aligned} \tag{27}$$

where (a) follows from (24), (b) is due to $P_{ki} = O(\|H\|)$ for $k \neq i$ (cf. (18)), and (c) follows from (19). Next we estimate the second term of (26). Since $X + H = P^T\mathrm{Diag}\{\mu_1, ...\mu_n\}P$, it follows that $H_{ki} = \sum_j P_{ji}P_{jk}\mu_j$ for $i \neq k$. Thus, we have

$$\begin{aligned} H_{ki}^2 &= \left(\sum_{j=1}^{n} P_{ji}P_{jk}\mu_j\right)^2 \\ &= \left(P_{ii}P_{ik}\mu_i + P_{ki}P_{kk}\mu_k + \sum_{j\neq i,k} P_{ji}P_{jk}\mu_j\right)^2 \end{aligned}$$

$$
\begin{aligned}
&\overset{\text{(i)}}{=} \left(P_{ki}P_{kk}(\mu_k - \mu_i) + \sum_{j\neq i,k} P_{ji}P_{jk}(\mu_j - \mu_i) \right)^2 \\
&\overset{\text{(ii)}}{=} P_{ki}^2 P_{kk}^2 (\mu_k - \mu_i)^2 + O(\|H\|^3) \\
&\overset{\text{(iii)}}{=} P_{ki}^2 (\lambda_k - \lambda_i)^2 + O(\|H\|^3),
\end{aligned}
$$

where step (i) follows from the orthonormality condition $\sum_j P_{ji}P_{jk} = 0$ when $i \neq k$, step (ii) is due to the fact all the off-diagonal entries of P are of order $O(\|H\|)$, and step (iii) follows from the fact $P_{kk}^2 = 1 + O(\|H\|^2)$ (see (20)) and the fact $|\lambda_k - \lambda_i| = |\mu_k - \mu_i| + O(\|H\|)$ (see (19)). The above estimate implies

$$
\begin{aligned}
\sum_{k\neq i} f^{[2]}(\lambda_i, \lambda_k, \lambda_i) H_{ki}^2 &= \sum_{k\neq i} f^{[2]}(\lambda_i, \lambda_k, \lambda_i) P_{ki}^2 (\lambda_k - \lambda_i)^2 + O(\|H\|^3) \\
&= \sum_{k\neq i} \left(f(\lambda_k) - f(\lambda_i) - f'(\lambda_i)(\lambda_k - \lambda_i) \right) P_{ki}^2 + O(\|H\|^3),
\end{aligned}
$$

where the last step follows from the definition of $f^{[2]}(\lambda_i, \lambda_k, \lambda_i)$. Combining this with (25), (26) and (27), we obtain

$$
\begin{aligned}
\text{the } (i,i)\text{-th entry} &= f(\mu_i) - f(\lambda_i) - f'(\lambda_i)(\mu_i - \lambda_i) + \sum_{k\neq i} \left(f(\mu_k) - f(\mu_i) - f'(\lambda_i)(\mu_k - \mu_i) \right) P_{ki}^2 \\
&\quad - \sum_{k=1}^{n} f^{[2]}(\lambda_i, \lambda_k, \lambda_i) H_{ki}^2 \\
&= f(\mu_i) - f(\lambda_i) - f'(\lambda_i)(\mu_i - \lambda_i) + \sum_{k\neq i} \left(f(\mu_k) - f(\mu_i) - f'(\lambda_i)(\mu_k - \mu_i) \right) P_{ki}^2 \\
&\quad - \frac{1}{2} f''(\lambda_i)(\lambda_i - \mu_i)^2 - \sum_{k\neq i} \left(f(\lambda_k) - f(\lambda_i) - f'(\lambda_i)(\lambda_k - \lambda_i) \right) P_{ki}^2 + O(\|H\|^3) \\
&\overset{\text{(i)}}{=} f(\mu_i) - f(\lambda_i) - f'(\lambda_i)(\mu_i - \lambda_i) - \frac{1}{2} f''(\lambda_i)(\lambda_i - \mu_i)^2 + O(\|H\|^3) \\
&\overset{\text{(ii)}}{=} o(\|H\|^2),
\end{aligned}
$$

where step (i) follows from $P_{ki} = O(\|H\|)$ and $|\lambda_i - \mu_i| \leq \|H\|$ for all i, and step (ii) is due to the second order differentiability of f at λ_i.

It remains to show that the off-diagonal entries of (23) are of order $o(\|H\|^2)$. To this end, consider the (i,j)-th entry of (23), $i \neq j$:

$$
\begin{aligned}
\text{the } (i,j)\text{-th entry} &= \sum_{k=1}^{n} P_{ki}P_{kj} f(\mu_k) - f^{[1]}(\lambda_i, \lambda_j) H_{ij} - \sum_{k=1}^{n} f^{[2]}(\lambda_i, \lambda_k, \lambda_j) H_{ik} H_{kj} \\
&= \underbrace{\sum_{k=1}^{n} P_{ki}P_{kj} f(\mu_k) - f^{[1]}(\lambda_i, \lambda_j) H_{ij}}_{\text{term I}} - \underbrace{\sum_{k\neq i,j} f^{[2]}(\lambda_i, \lambda_k, \lambda_j) H_{ik} H_{kj}}_{\text{term II}}
\end{aligned}
$$

$$-\underbrace{(f^{[2]}(\lambda_i,\lambda_i,\lambda_j)H_{ii}+f^{[2]}(\lambda_i,\lambda_j,\lambda_j)H_{jj})H_{ij}}_{\text{term III}}. \tag{28}$$

Since

$$H = P^T\text{Diag}\{\mu_1,...,\mu_n\}P - X = P^T\text{Diag}\{\mu_1,...,\mu_n\}P - \text{Diag}\{\lambda_1,...,\lambda_n\},$$

it follows that

$$H_{\ell,m} = \begin{cases} \sum_{k=1}^{n} P_{k\ell}P_{km}\mu_k - \lambda_\ell, & \text{if } \ell = m, \\ \sum_{k=1}^{n} P_{k\ell}P_{km}\mu_k, & \text{else.} \end{cases}$$

Substituting this into term I and using (18)–(21), we can obtain the following alternative expression for term I:

$$\begin{aligned} \text{term I} &= P_{ji}P_{jj}(f(\mu_j)-f(\mu_i)-f^{[1]}(\lambda_i,\lambda_j)(\mu_j-\mu_i)) + (P_{ii}P_{ij}+P_{ji}P_{jj})(f(\mu_i)-f^{[1]}(\lambda_i,\lambda_j)\mu_i) \\ &\quad + \sum_{k\neq i,j} P_{ki}P_{kj}(f(\mu_k)-f^{[1]}(\lambda_i,\lambda_j)\mu_k). \end{aligned} \tag{29}$$

To estimate term II, we first notice that for $k \neq i$:

$$\begin{aligned} H_{ik} &= \sum_{m=1}^{n} P_{mi}P_{mk}\mu_m \\ &= P_{ii}P_{ik}\mu_i + P_{ki}P_{kk}\mu_k + O(\|H\|^2) \\ &= P_{kk}P_{ki}(\mu_k-\mu_i) + (P_{ii}P_{ik}+P_{ki}P_{kk})\mu_i + O(\|H\|^2) \\ &= P_{kk}P_{ki}(\mu_k-\mu_i) + O(\|H\|^2) \\ &= P_{ki}(\mu_k-\mu_i) + O(\|H\|^2), \end{aligned} \tag{30}$$

where the second, fourth and fifth steps follow from (18) and (21). Similarly, we have

$$H_{kj} = P_{kj}(\mu_k-\mu_j) + O(\|H\|^2). \tag{31}$$

Since, for $k \neq i,j$, both P_{ki} and P_{kj} are of order $O(\|H\|)$, we can use (30) and (31) to estimate term II as follows:

$$\begin{aligned} \text{term II} &= \sum_{k\neq i,j} f^{[2]}(\lambda_i,\lambda_k,\lambda_j)H_{ik}H_{kj} \\ &= \sum_{k\neq i,j} f^{[2]}(\lambda_i,\lambda_k,\lambda_j)(\mu_k-\mu_i)(\mu_k-\mu_j)P_{ki}P_{kj} + O(\|H\|^3) \\ &= \sum_{k\neq i,j} f^{[2]}(\lambda_i,\lambda_k,\lambda_j)(\lambda_k-\lambda_i)(\lambda_k-\lambda_j)P_{ki}P_{kj} + O(\|H\|^3), \end{aligned}$$

where the last step is due to (19). Since

$$f^{[2]}(\lambda_i,\lambda_k,\lambda_j)(\lambda_k-\lambda_i)(\lambda_k-\lambda_j) = (f(\lambda_k)-f^{[1]}(\lambda_i,\lambda_j)\lambda_k) - (f(\lambda_i)-f^{[1]}(\lambda_i,\lambda_j)\lambda_i)$$

and

$$\sum_{k\neq i,j} P_{ki}P_{kj} = -(P_{ii}P_{ij} + P_{ji}P_{jj}),$$

it follows that

$$\begin{aligned}
\text{term II} &= \sum_{k\neq i,j} P_{ki}P_{kj}(f(\lambda_k) - f^{[1]}(\lambda_i,\lambda_j)\lambda_k) - (f(\lambda_i) - f^{[1]}(\lambda_i,\lambda_j)\lambda_i)\sum_{k\neq i,j} P_{ki}P_{kj} + O(\|H\|^3) \\
&= \sum_{k\neq i,j} P_{ki}P_{kj}(f(\lambda_k) - f^{[1]}(\lambda_i,\lambda_j)\lambda_k) + (P_{ii}P_{ij} + P_{ji}P_{jj})(f(\lambda_i) - f^{[1]}(\lambda_i,\lambda_j)\lambda_i) + O(\|H\|^3).
\end{aligned}$$

By (19), we have $|\mu_i - \lambda_i| \leq \|H\|$, we further obtain

$$\text{term II} = \sum_{k\neq i,j} P_{ki}P_{kj}(f(\lambda_k) - f^{[1]}(\lambda_i,\lambda_j)\lambda_k) + (P_{ii}P_{ij} + P_{ji}P_{jj})(f(\mu_i) - f^{[1]}(\lambda_i,\lambda_j)\mu_i) + O(\|H\|^3). \quad (32)$$

To estimate term III, we first notice that

$$H_{ii} = -\lambda_i + \sum_{k=1}^{n} P_{ki}^2\mu_k = P_{ii}^2\mu_i - \lambda_i + O(\|H\|^2) = \mu_i - \lambda_i + O(\|H\|^2)$$

and similarly

$$H_{jj} = \mu_j - \lambda_j + O(\|H\|^2).$$

Since $H_{ij} = O(\|H\|)$, it follows that

$$\begin{aligned}
\text{term III} &= (f^{[2]}(\lambda_i,\lambda_i,\lambda_j)H_{ii} + f^{[2]}(\lambda_i,\lambda_j,\lambda_j)H_{jj})H_{ij} \\
&= (f^{[2]}(\lambda_i,\lambda_i,\lambda_j)(\mu_i - \lambda_i) + f^{[2]}(\lambda_i,\lambda_j,\lambda_j)(\mu_j - \lambda_j))H_{ij} + O(\|H\|^3)
\end{aligned}$$

By an argument similar to (30), we have

$$H_{ij} = P_{jj}P_{ji}(\mu_j - \mu_i) + O(\|H\|^2).$$

Thus, we have

$$\text{term III} = P_{jj}P_{ji}(f^{[2]}(\lambda_i,\lambda_i,\lambda_j)(\mu_i - \lambda_i) + f^{[2]}(\lambda_i,\lambda_j,\lambda_j)(\mu_j - \lambda_j))(\mu_j - \mu_i) + O(\|H\|^3), \quad (33)$$

where we have used the fact that $|\mu_i - \lambda_i| \leq \|H\|$ and $|\mu_j - \lambda_j| \leq \|H\|$. It can be checked from the definition of second order divided difference $f^{[2]}$ that

$$\begin{aligned}
&(\lambda_i - \lambda_j)(f^{[2]}(\lambda_i,\lambda_i,\lambda_j)(\mu_i - \lambda_i) + f^{[2]}(\lambda_i,\lambda_j,\lambda_j)(\mu_j - \lambda_j)) \\
&\quad = f^{[1]}(\lambda_i,\lambda_j)(\mu_j - \mu_i) + (f(\lambda_i) + (\mu_i - \lambda_i)f'(\lambda_i)) - (f(\lambda_j) + (\mu_j - \lambda_j)f'(\lambda_j)) \\
&\quad = f^{[1]}(\lambda_i,\lambda_j)(\mu_j - \mu_i) + f(\mu_i) - f(\mu_j) + o(\|H\|),
\end{aligned}$$

where the last step is due to the second order differentiability of f at λ_i and λ_j. Substituting this bound into (33) and noting $P_{ji} = O(\|H\|)$, we obtain

$$\begin{aligned}\text{term III} &= P_{jj}P_{ji}(f^{[1]}(\lambda_i,\lambda_j)(\mu_j-\mu_i)+f(\mu_i)-f(\mu_j))\frac{(\mu_j-\mu_i)}{\lambda_i-\lambda_j}+o(\|H\|^2)\\ &= P_{jj}P_{ji}(f(\mu_j)-f(\mu_i)-f^{[1]}(\lambda_i,\lambda_j)(\mu_j-\mu_i))+o(\|H\|^2), \qquad (34)\end{aligned}$$

where the last step follows from the fact (cf. (19))

$$\frac{(\mu_j-\mu_i)}{\lambda_i-\lambda_j} = -1 + O(\|H\|).$$

Combining the estimates (29), (32), (34) with (28), we immediately obtain

$$\text{the } (i,j)\text{-th entry} = o(\|H\|^2),$$

as desired. This completes the proof of part 2 of Theorem 3.6.

References

[1] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1997.

[2] J.S. Chen, “The Convex and Monotone Functions Associated with Second-Order Cone,” Working Paper, 2004.

[3] J.S. Chen, X. Chen, P. Tseng, “Analysis of Nonsmooth Vector-Valued Functions Associated with Second-Order Cones,” *Mathematical Programming* 101 (2004) 95–117.

[4] X. Chen, H.-D. Qi, and P. Tseng, “Analysis of Nonsmooth Symmetric Matrix Functions with Applications to Semidefinite Complementarity Problems,” *SIAM Journal Optimization* 13 (2003) 960–985.

[5] X. Chen and P. Tseng, “Non-interior Continuation Methods for Solving Semidefinite Complementarity Problems,” *Math. Programming,* to appear.

[6] C.B. Chua, “A New Notion of Weighted Centers for Semidefinite Programming,” Working Paper, Department of IEOR, Cornell University (2004).

[7] T. Georgiou, “Relative Entropy and the Multi-variable Multi-Dimensional Moment Problems,” Submitted to *IEEE Transactions on Information Theory* (2004).

[8] R.P. Feynman, “An Operator Calculus Having Applications in Quantum Electrodynamics,” *Physical Review* 84 (1951) 108–128.

[9] S.P. Heims and E.T. Jaynes, "Theory of Gyromagnetic Effects and Some Related Magnetic Phenomena," *Reviews of Modern Physics* 34 (1962) 143–165.

[10] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Campbridge University Press (1991).

[11] F. Kraus, "Über Konvexe Matrixfunctionen," *Math. Zeit.* 41 (1936) 18–42.

[12] A.S. Lewis and H.S. Sendov, "Twice Differentiable Spectral Functions," *SIAM Journal Matrix Analysis and Applications* 23 (2001) 368–386.

[13] K. Löwner, "Über Monotone Matrixfunctionen," *Math. Zeit.* 38 (1934) 177–216.

[14] Z.-Q. Luo, T.N. Davidson, G.B. Giannakis, and K.M. Wong, "Transceiver Optimization for Multiple Access through ISI Channels", *IEEE Transactions on Signal Processing*, 52 (2004) 1037–1052.

[15] I. Najfeld and T.F. Havel, "Derivatives of the Matrix Exponential and Their Computation," *Advances in Appllied Mathematics* 16 (1985) 321–375.

[16] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000).

[17] J.F. Sturm, *Theory and Algorithms for Semidefinite Programming.* In J.B.G. Frenk, C. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization*, Kluwer Academic Publishers. (2000) 3–196.

[18] J.F. Sturm and S. Zhang, "On Weighted Centers for Semidefinite Programming," *European Journal of Operational Research* 126 (2000) 391–407.

[19] D. Sun and J. Sun, "Semismooth Matrix Valued Functions," *Mathematics of Operations Research* 27 (2002) 150–169.

[20] H. Wolkowicz, R. Saigal, L. Vandenberghe, eds., *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, Kluwer Academic Publishers, Dordrecht, 2000.

[21] S. Zhang, "On Conically Ordered Convex Programs," Technical Report SEEM2003-09, Department of Systems Engineering & Engineering Management, The Chinese University of Hong Kong (2003).