Matrix Functions and Weighted Centers for Semidefinite Programming

Jan Brinkhuis* Zhi-Quan Luo† Shuzhong Zhang‡

August 2007

Abstract

In this paper, we develop various differentiation rules for general smooth matrix-valued functions, and for the class of matrix convex (or concave) functions first introduced by Löwner and Kraus in the 1930s. For a matrix monotone function, we present formulas for its derivatives of any order in an integral form. Moreover, for a general smooth primary matrix function, we derive a formula for all of its derivatives by means of the divided differences of the original function. As applications, we use these differentiation rules and the matrix concave function $\log X$ to study a new notion of weighted centers for Semidefinite Programming (SDP). We show that, with this definition, some known properties of weighted centers for linear programming can be extended to SDP. We also show how the derivative formulas can be used in the implementation of barrier methods for optimization problems involving nonlinear but convex matrix functions.

Keywords: matrix monotonicity, matrix convexity, semidefinite programming.

AMS subject classification: 15A42, 90C22.

*Econometric Institute, Erasmus University, Rotterdam, The Netherlands.

†Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455. Research supported by the National Science Foundation, grant No. DMS-0610037.

‡Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong. Research supported by Hong Kong RGC Earmarked Grant CUHK418406.

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 an LP

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

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

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

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 using either an integral representation or the 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; in particular, for an arbitrary matrix function, we offer a general formula for the derivatives of any order in terms of divided differences. This unifies and extends the known formulas for the first and second derivatives from [1] and [4]; 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) does 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 [2] 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 consider Hermitian matrices, and later restrict to the real case when we discuss 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

For a real function $f: J \mapsto \Re$, we can define the primary matrix function of $f(Z): \mathcal{H}^n(J) \mapsto \mathcal{H}^n$. Specifically, suppose $Z \in \mathcal{H}^n(J)$ is a Hermitian matrix with an eigen-decomposition $Z = Q^{\mathrm{H}} D Q$, where $Q^{\mathrm{H}} Q = I$ and D is a real-valued diagonal matrix with $D_{jj} \in J$, $j = 1, ..., n$. Then, the primary matrix function $f(Z)$ is defined as

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

where $f(D) = \mathrm{diag}\,(f(D_{11}), ..., f(D_{nn}))$. Although the eigen-decomposition of Z may not be unique, the matrix function $f(Z)$ 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 of a 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 $\alpha \in [0,1]$. If the above inequality holds strictly for all $X, Y \in \mathcal{H}_n(J)$ with $X - Y$ nonsingular and all $\alpha \in (0,1)$, then we say f is a *strictly matrix convex function on* $\mathcal{H}_n(J)$. Moreover, 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, including related results.

Theorem 2.3 *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 function $-1/x$ on $(0, \infty)$.

Lemma 2.4 *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 $0 \leq \alpha \leq 1$:

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

□

Lemma 2.5 *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 2.4 by a change of variable: $f_u(x) = -\tilde{x}^{-1}$ if we put $\tilde{x} = x - u$. □

Therefore, we can prove the following result:

Theorem 2.6 *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 2.3, using Lemma 2.5 and noting that the matrix concavity is preserved under summation and multiplication by a nonnegative number. Moreover, for the second statement one has to use that by Lemma 2.5 and the assumption that $d\mu$ has positive mass, the third term in the righthand side of (2) is a strictly matrix concave function and so it follows that f is a strictly matrix concave function. □

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 2.6 that the log 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 higher order derivatives of a matrix function

The logarithmic barrier plays an important role in the design and analysis of interior point methods for semidefinite programming (SDP); see Section 4. The need to calculate the derivatives of the log matrix function motivates us to study the calculus rules for a general matrix function. It turns out that there are two different ways to accomplish this goal. If the matrix function in question is monotone, such as the log function, then Löwner's result (Theorem 2.3) can be used to derive a closed-form formula for its derivatives of any order. This representation is discussed in Section 3.1 and later applied to study the weighted centers for SDP in Section 4. For a general matrix function, we present in Section 3.2 a formula for computing its derivatives of any order by means of divided differences. Special cases of the formula (first and second order derivatives) are known in the literature; see e.g. [10] and [1]. Our formula unifies and extends these existing results to the derivatives of any order.

3.1 Derivatives of monotone matrix functions: an integral representation

Before the computation of the derivatives of a matrix function, we recall the well-known linkage between the Fréchet derivatives and the Taylor expansion.

Proposition 3.1 *Let X be a normed vector space, let U be a neighborhood of a point $\widehat{x}$ in X, and let $f : U \to \Re$ be a function. If for each $x \in U$ and each $i \in \{1, \ldots, k\}$ there exists a continuous i-multilinear symmetric function $f^{((i))}(x) : X^i \to \Re$ that depends continuously on x and for which*

$$f(x+h) = f(x) + f^{((1))}(x)[h] + \frac{1}{2!} f^{((2))}(x)[h,h] + \cdots + \frac{1}{k!} f^{((k))}(x)[h, \ldots, h] + o(\|h\|^k),$$

then all (Fréchet) derivatives $f^{(i)}$, $1 \le i \le k$ of f up to order k exist and are continuous and given by $f^{(i)}(x) = f^{((i))}(x)$, $1 \le i \le k$, $x \in U$.

That is, for k-times continuously differentiable functions on an open set, the Fréchet derivatives can be defined implicitly by Taylor expansions at all points of this open set.

As an example, let us consider a very useful and concrete matrix function, $\log X$.

Theorem 3.2 *The following formula holds true:*

$$\begin{aligned} \log^{(i)}(X)[H, \ldots, H] &= \int_{-\infty}^{0} (uI - X)^{-1} (H(uI - X)^{-1})^i du \\ &= \int_{-\infty}^{0} ((uI - X)^{-1} H)^i (uI - X)^{-1} du, \end{aligned}$$

for all $H \in \mathcal{S}^n$, where $i \ge 1$.

Before we prove Theorem 3.2, we comment that the expression for the first order derivative for the primary matrix function of log (or the entropy function) 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 3.2, let us first introduce two lemmas.

Lemma 3.3 *The derivatives of the matrix function of $f : (0, +\infty) \mapsto \Re$ defined by $f(x) = x^{-1}$ are given by the following formulas*

$$f^{(i)}(X)[H, \ldots, H] = i! X^{-1} (-HX^{-1})^i,$$

for all $H \in \mathcal{S}^n$ and all $i \ge 1$.

Proof. Consider the Taylor expansion up to order k. We have, by Proposition 3.1,

$$(X+H)^{-1} = X^{-1} + \sum_{i=1}^{k} \frac{1}{i!} f^{(i)}(X)[H,\ldots,H] + o(\|H\|^k).$$

Multiplying both sides from the right with $(X+H)$, simplifying, and equating linear, quadratic and higher order functions of H respectively, gives the following equations:

$$\begin{aligned}
&X^{-1}H + f^{(1)}(X)[H]X = 0,\\
&\frac{1}{i!} f^{(i)}(X)[H,\ldots,H]H + \frac{1}{(i+1)!} f^{(i+1)}(X)[H,\ldots,H]X = 0,\ \ i = 1,...,k-1.
\end{aligned}$$

The first equation gives $f^{(1)}(X)[H] = -X^{-1}HX^{-1}$, and the second part of the equations can be used inductively to show that $f^{(i)}(X)[H,\ldots,H] = i!X^{-1}(-HX^{-1})^i$, for $i \geq 2$. □

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

$$f_u^{(i)}(X)[H,\ldots,H] = i!(uI-X)^{-1}(H(uI-X)^{-1})^i,$$

for all $H \in \mathcal{S}^n$ and all $i \geq 1$.

Proof of Theorem 3.2. We start with the identity (3). This gives the following formula:

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

Differentiating inside the integral and using Lemma 3.3 gives the required formulas. □

The range of integration can also be changed to $(0,+\infty)$ for convenience, as we shall do in the next section; that is,

$$\log^{(i)}(X)[H,\ldots,H] = -\int_{0}^{\infty} (uI+X)^{-1}(H(uI+X)^{-1})^i du. \tag{4}$$

By a similar argument and using Löwner's theorem (Theorem 2.3), we can extend the formulas for the derivatives of $\log X$ to the general matrix monotone functions.

Theorem 3.4 *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

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

3.2 An eigenvalue representation

In this subsection we shall focus on an alternative approach to compute the derivatives of matrix functions, based on the divided differences of a function. (One is referred to [10] (Chapter 6) for a comprehensive introduction on the background of the topic.) This approach works exclusively under the assumption that X is a diagonal matrix. The following proposition shows that this assumption can be made without losing generality.

Proposition 3.5 *Let J be an open real interval and let $f \in C^k(J)$. Let $X \in \mathcal{H}^n(J)$. Choose a diagonal decomposition $X = Q^{\mathrm{H}} D Q$. Then the following formulas hold true for all $H \in \mathcal{S}^n$ (with $K = QHQ^{\mathrm{H}}$):*

$$f^{(i)}(X)[H, \ldots, H] = Q^{\mathrm{H}}(f^{(i)}(D)[K, \ldots, K])Q, \ 1 \leq i \leq k.$$

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 above of the derivatives of f at X and at D. □

Let J be an open real interval and let $f : J \to \Re$ be a k-times continuously differentiable function; that is, $f \in C^k(J)$. It can be proved — the idea of the proof will be given below — that there exists for each nonnegative number $r \leq k$ a unique continuous function $f^{[r]}$ on J^{k+1} such that the following recurrence relations hold:

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

The expression $f^{[r]}(\lambda)$ is called *the r-th divided difference* of f at λ.

The recurrence relations above allow to compute the divided differences $f^{[r]}(\lambda)$ recursively if all coordinates are distinct. The divided differences can be computed for other λ by using the continuity: this gives, 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 $f^{[r]}$ are symmetric, i.e., the value of the function is invariant with respect to permutation of its entries. For the power function $f(x) = x^m$, the following explicit formula holds true:

$$f^{[p]}(\lambda_1, \ldots, \lambda_{p+1}) = \sum_{\substack{l_1 + \cdots + l_{p+1} = m - p \\ l_1, \ldots, l_{p+1} \geq 0, \text{integers}}} \lambda_1^{l_1} \cdots \lambda_{p+1}^{l_{p+1}}, \text{ whenever } m \geq p. \tag{5}$$

We note that the following explicit formula holds for all λ (see e.g. formula (6.1.24) in Horn and Johnson [10]):

$$f^{[r]}(\lambda) = \int_{x \in \Delta_r} f^{(r)}(\lambda^T x) dx \tag{6}$$

where Δ_r is the standard r-dimensional simplex (the convex hull in $\Re^{r+1}$ of the unit vectors e_i, $i = 1, ..., r+1$) and $f^{(r)}$ denotes the r-th derivative of f. The righthand side of this formula is immediately seen to be a symmetric function and one can readily verify that the recurrence formulas above as well as the explicit formula for power functions given above hold true for the righthand side.

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. Now we present our main result of this section.

Theorem 3.6 *Let J be an open connected subset of the real line $\Re$ and let $f : J \to \Re$ be a C^k-function for some positive integer k. Let n be a positive integer. Then the n-th primary matrix function of f is k-times Fréchet differentiable and its k-th derivative is given by the following formula: for each diagonal matrix $X = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ whose spectrum is in J one has*

$$f^{(k)}(X)[H, \ldots, H] = k! \sum_{i_1, \ldots, i_{k+1}=1}^{n} f^{[k]}(\lambda_{i_1}, \ldots, \lambda_{i_{k+1}}) E_{i_1,i_1} H E_{i_2,i_2} H \cdots H E_{i_{k+1},i_{k+1}} \tag{7}$$

for all symmetric $n \times n$-matrices H.

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

$$f^{(1)}(X)[H] = f^{[1]}(X) \circ H, \tag{8}$$

which is a well known result (cf. [1, 10]).

Proof. We choose $[a,b]$ to be a closed interval in J such that the open interval (a,b) contains the spectrum of X, and write

$$f^{((p))}(X)[H,\ldots,H]=p!\sum_{i_1,\ldots,i_{p+1}=1}^{n} f^{[p]}(\lambda_{i_1},\ldots,\lambda_{i_{p+1}})E_{i_1,i_1}HE_{i_2,i_2}H\cdots HE_{i_{p+1},i_{p+1}}$$

for $p=0,1,\ldots,k$. In terms of this notation, we have to establish that

$$f^{(k)}(X)[H,\ldots,H]=f^{((k))}(X)[H,\ldots,H].$$

In light of (6), we have

$$|f^{[p]}(\lambda_{i_1},\ldots,\lambda_{i_{p+1}})|\le|\Delta_p|\max_{x\in[a,b]}|f^{(p)}(x)|,$$

for any $\lambda_{i_1},\ldots,\lambda_{i_{p+1}}\in J$, where $|\Delta_p|$ is the volume of the simplex Δ_p. Therefore,

$$\|f^{((p))}(X)[H,\ldots,H]\|\le C\max_{x\in[a,b]}|f^{(p)}(x)|\,\|H\|^p \tag{9}$$

for $p=0,1,...,k$, where C is a positive constant that depends only on n and k.

In the subsequent analysis, we treat several cases regarding the function f.

Case 1: f is a power function $f(x)=x^m$, with m a nonnegative integer.

In this case, $f(X+H)=(X+H)^m$. On writing $X=\sum_{i=1}^n\lambda_iE_{i,i}$ and $H=\sum_{i,j=1}^n E_{i,i}HE_{j,j}$ and expanding the brackets we get, using $E_{i,i}E_{j,j}=0$ whenever $i\neq j$, that

$$f(X+H)=\sum_{k=0}^{m}\sum_{i_1,\ldots,i_{k+1}=1}^{n}\sum_{l_{i_1}+\cdots+l_{i_{k+1}}=m-k}(\lambda_{i_1}E_{i_1,i_1})^{l_{i_1}}E_{i_1,i_1}HE_{i_2,i_2}(\lambda_{i_2}E_{i_2,i_2})^{l_{i_2}}\cdots$$
$$\cdots E_{i_k,i_k}HE_{i_{k+1},i_{k+1}}(\lambda_{i_{k+1}}E_{i_{k+1},i_{k+1}})^{l_{i_{k+1}}},$$

where $l_{i_1},\ldots,l_{i_{k+1}}$ are nonnegative integers. Simplifying the above expression yields

$$\sum_{k=0}^{m}\sum_{i_1,\ldots,i_{k+1}=1}^{n}\sum_{l_{i_1}+\cdots+l_{i_{k+1}}=m-k}\lambda_{i_1}^{l_{i_1}}\lambda_{i_2}^{l_{i_2}}\ldots\lambda_{i_{k+1}}^{l_{i_{k+1}}}(E_{i_1,i_1})^{l_{i_1}+1}H(E_{i_2,i_2})^{l_{i_2}+1}H\cdots H(E_{i_{k+1},i_{k+1}})^{l_{i_{k+1}}+1},$$

and since the $E_{i,i}$'s are idempotent, it follows further that the expression equals

$$\sum_{k=0}^{m}\sum_{i_1,\ldots,i_{k+1}=1}^{n}\sum_{l_{i_1}+\cdots+l_{i_{k+1}}=m-k}\lambda_{i_1}^{l_{i_1}}\lambda_{i_2}^{l_{i_2}}\ldots\lambda_{i_{k+1}}^{l_{i_{k+1}}}E_{i_1,i_1}HE_{i_2,i_2}H\cdots HE_{i_{k+1},i_{k+1}}.$$

By (5), the above explicit formula is equal to

$$\sum_{k=0}^{m}\sum_{i_1,\ldots,i_{k+1}=1}^{n}f^{[k]}(\lambda_{i_1},\ldots,\lambda_{i_{k+1}})E_{i_1,i_1}HE_{i_2,i_2}H\cdots HE_{i_{k+1},i_{k+1}}.$$

This proves the theorem in the case when f is a power function.

Case 2: f is a polynomial. This follows from the previous case by linearity.

Case 3: f is a general smooth function in $C^k(J)$.

Choose an infinite sequence of polynomials f_r, $r = 1, 2, ...$, such that $f_r^{(i)} \to f^{(i)}$ uniformly on $[a, b]$ for $i = 0, 1, \ldots, k$. We write

$$f^{((p))}(X)[H, \ldots, H] = p! \sum_{i_1,\ldots,i_{p+1}=1}^{n} f^{[p]}(\lambda_{i_1}, \ldots, \lambda_{i_{p+1}}) E_{i_1,i_1} H E_{i_2,i_2} H \cdots H E_{i_{p+1},i_{p+1}}$$

for $p = 0, 1, \ldots, k$. In terms of this notation, we have to prove that

$$f^{(k)}(X)[H, \ldots, H] = f^{((k))}(X)[H, \ldots, H].$$

We will do this by showing that for each $\epsilon > 0$, one has for $\|H\|$ sufficiently small that the inequality $R_f(H) \leq \epsilon \|H\|^k$ holds, where

$$R_f(H) := \left\| f(X + H) - f(X) - f^{((1))}(X)[H] - \frac{1}{2!} f^{((2))}(X)[H, H] - \cdots - \frac{1}{k!} f^{((k))}(X)[H, \ldots, H] \right\|.$$

By the triangle inequality one has $R_f(H) \leq R_1(H) + R_2(H) + R_3(H)$, where

$$\begin{aligned} R_1(H) &= \left\| (f - f_m)(X + H) - (f - f_m)(X) - \cdots - \frac{1}{(k-1)!} (f - f_m)^{((k-1))}(X)[H, \ldots, H] \right\|, \\ R_2(H) &= \left\| \frac{1}{k!} (f - f_m)^{((k))}(X)[H, \ldots, H] \right\|, \text{ and} \\ R_3(H) &= \left\| f_m(X + H) - f_m(X) - f_m^{((1))}(X)[H] - \cdots - \frac{1}{k!} f_m^{((k))}(X)[H, \ldots, H] \right\|. \end{aligned}$$

Therefore, to prove the theorem, it suffices to show that $R_i(H) \leq \frac{\epsilon}{3} \|H\|^k$, $i = 1, 2, 3$, if m is sufficiently large and $\|H\|$ is sufficiently small.

For the term $R_3(H)$ the estimation follows from **Case 2**, since the function in question is polynomial. Next, consider $R_2(H)$. By (9) we have $R_2(H) \leq C \max_{x \in [a,b]} |(f^{(k)} - f_m^{(k)})(x)| \, \|H\|^k$. Since $f_m^{(k)}(x) \to f^{(k)}(x)$ uniformly on $x \in [a, b]$ as $m \to \infty$, the desired relation holds. Finally, let us consider the first term $R_1(H)$. For any fixed integer q, using the usual estimation of the residue of the Taylor series, we have

$$\begin{aligned} & \left\| (f_q - f_m)(X + H) - (f_q - f_m)(X) - \cdots - \frac{1}{(k-1)!} (f_q - f_m)^{((k-1))}(X)[H, \ldots, H]) \right\| \\ \leq & \max_{Y \in [X, X+H]} \|(f_q - f_m)^{((k))}(Y)\| \frac{\|H\|^k}{k!} \\ \leq & \; C \max_{x \in [a,b]} |(f_q - f_m)^{((k))}(x)| \, \|H\|^k, \end{aligned}$$

where the second inequality is due to **Case 2** and (9). Letting $q \to +\infty$, we have

$$R_1(H) \le C \max_{x\in[a,b]} |(f-f_m)^{((k))}(x)| \, \|H\|^k.$$

This shows that $R_i(H) \le \frac{\epsilon}{3}\|H\|^k$, $i = 1, 2, 3$, and so $R_f(H) \le \epsilon\|H\|^k$, if m is sufficiently large and $\|H\|$ is sufficiently small. The theorem is proven. □

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 3.4 which are valid only for matrix monotone functions. As an exercise, let us apply Theorem 3.6 to some specific functions.

Corollary 3.7 *Let $h(x) = -x^{-1}$ for $x \in (0,\infty)$. The following formulas hold true for the derivatives of the primary matrix function $h(X)$ at any positive diagonal matrix $X = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$:*

$$h^{(k)}(X)[H,\ldots,H] = k! \sum_{i_1,\ldots,i_{k+1}=1}^{n} \lambda_{i_1}^{-1}\lambda_{i_2}^{-1}\cdots\lambda_{i_{k+1}}^{-1} E_{i_1,i_1}HE_{i_2,i_2}H\cdots HE_{i_{k+1},i_{k+1}},$$

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

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

$$h^{[k]}(\lambda_1,\ldots,\lambda_{k+1}) = \prod_{i=1}^{k+1} \lambda_i^{-1}$$

for all $\lambda_1,\ldots,\lambda_{k+1} > 0$ for all $k \ge 1$. To see the formula for $k = 1$, we note $h^{[1]}(\lambda_1,\lambda_2) = \frac{-\lambda_1^{-1}+\lambda_2^{-1}}{\lambda_1-\lambda_2}$ $= (\lambda_1\lambda_2)^{-1}$, for $\lambda_1 \ne \lambda_2$, and so by continuity of $h^{[1]}$ for $\lambda_1 = \lambda_2$ as well, as desired. Continuing in the same way one can verify the formula recursively for all $k \ge 1$. □

As a second example, we apply Theorem 3.6 to the function $\log x$. Indeed, using (6) and observing that $\log^{(k)}(x) = -h^{(k-1)}(x)$ for $k \ge 1$, we have the following formula regarding the divided difference for log. Let λ_i $(i = 1, ..., k+1)$ be given. Then,

$$\log^{[k]}(\lambda) = \int_{x\in\Delta_{k+1}} \log^{(k)}(\lambda^T x)dx = -\int_{x\in\Delta_{k+1}} h^{(k-1)}(\lambda^T x)dx. \tag{10}$$

Introducing a variable transformation $x_1 = \tau$ and $x_{i+1} = (1-\tau)y_i$, $i = 1, ..., k$, the Jacobian of which is $(1-\tau)^k$, formula (10) further leads to

$$\begin{aligned}
\log^{[k]}(\lambda) &= -\int_0^1 (1-\tau)^k \int_{y\in\Delta_k} h^{(k-1)}\left(\sum_{j=1}^{k}(\lambda_1\tau + (1-\tau)\lambda_{j+1})y_j\right) dyd\tau \\
&= -\int_0^1 (1-\tau)^k h^{[k-1]}(\lambda_1\tau + (1-\tau)\lambda_2, \ldots, \lambda_1\tau + (1-\tau)\lambda_k)d\tau \\
&= -\int_0^1 (1-\tau)^k \prod_{j=1}^{k} \frac{1}{\lambda_1\tau + (1-\tau)\lambda_{j+1}}\, d\tau. \qquad (11)
\end{aligned}$$

Using the well-known algorithm for integrating a rational function with given factorization of the denominator into linear factors, one can achieve an explicit formula for $\log^{[k]}(\lambda)$, and so by Theorem 3.6 for $\log^{(k)}(X)[H,\ldots,H]$, in terms of elementary functions. We will see later that (11) yields an alternative proof for the concavity of the matrix $\log X$. Before proving this, we shall first show that we can use Theorem 3.6 to characterize the matrix convex functions.

Proposition 3.8 *Let $f \in C^2(J)$. The matrix function $f(X)$ is convex on $\mathcal{S}^n(J)$ if and only if*

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

for all $j = 1, \ldots, n$ and all $(\lambda_1, \ldots, \lambda_n)$ with $\lambda_l \in J$ for all $l = 1, \ldots, n$.

Proof. We need only to show that $f^{(2)}(X)[H, H] \succeq 0$ holds for all $X \in \mathcal{S}^n(J)$ and $H \in \mathcal{S}^n$ if and only if $M_j \succeq 0$ for all $j = 1, \ldots, n$ and all $(\lambda_1, \ldots, \lambda_n)$ with $\lambda_l \in J$ for all $l = 1, \ldots, n$.

In light of Proposition 3.5, 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

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

'$\Longrightarrow$': Take any $x, y \in \Re^n$ with $y_l \neq 0$, $l = 1, ..., n$. Let $w = x \circ y^{-1}$ and $H = yy^T$. We have

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

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 $j = 1, ..., n$.

'$\Longleftarrow$': Since $M_j \in \mathcal{S}^n_+$, $j = 1, ..., n$, we have

$$f^{(2)}(X)[H, H] = 2\sum_{j=1}^{n} M_j \circ (H(:, j)H(:, j)^T) \in \mathcal{S}^n_+,$$

where '$\circ$' is the Hadamard product and $H(:, j)$ is the j-th column of H. $\square$

Using Proposition 3.8 and formula (11), it is now obvious that $\log X$ is matrix concave. Observe

$$\begin{aligned} & -\left(\log^{[2]}(\lambda_i,\lambda_j,\lambda_k)\right)_{(i,k)} \\ = & \left(\int_0^1 \frac{(1-\tau)^2 d\tau}{(\lambda_j\tau+(1-\tau)\lambda_i)(\lambda_j\tau+(1-\tau)\lambda_k)}\right)_{(i,k)} \\ = & \int_0^1 \left(\frac{(1-\tau)^2}{(\lambda_j\tau+(1-\tau)\lambda_i)(\lambda_j\tau+(1-\tau)\lambda_k)}\right)_{(i,k)} d\tau, \end{aligned} \tag{12}$$

and since

$$\begin{aligned} & \left(\frac{(1-\tau)^2}{(\lambda_j\tau+(1-\tau)\lambda_i)(\lambda_j\tau+(1-\tau)\lambda_k)}\right)_{(i,k)} \\ = & (1-\tau)^2 \begin{pmatrix} \frac{1}{\lambda_j\tau+(1-\tau)\lambda_1} \\ \vdots \\ \frac{1}{\lambda_j\tau+(1-\tau)\lambda_n} \end{pmatrix} \left(\frac{1}{\lambda_j\tau+(1-\tau)\lambda_1},\ldots,\frac{1}{\lambda_j\tau+(1-\tau)\lambda_n}\right) \succeq 0 \end{aligned}$$

for each fixed $\tau \in [0,1]$ and $\lambda \in \Re^n_{++}$, it follows from (12) that $-\left(\log^{[2]}(\lambda_i,\lambda_j,\lambda_k)\right)_{(i,k)} \succeq 0$. Therefore, $\log X$ is a matrix concave function as Proposition 3.8 asserts.

Let us now specialize Theorem 3.6 to the matrix exponential function e^X (which is known to be not matrix monotone so Theorem 3.4 does not apply).

Proposition 3.9 *For any symmetric X and H, there holds*

$$(e^X)^{(1)}[H] = \int_0^1 e^{(1-u)X} H e^{uX} du \tag{13}$$

$$(e^X)^{(2)}[H,H] = 2\int_0^1\int_0^1 (1-u)e^{uX} H e^{v(1-u)X} H e^{(1-v)(1-u)X} du dv \tag{14}$$

Proof. We only need to prove the proposition for diagonal matrix $X = \mathrm{Diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$. Assume that λ_i's are all distinct. In light of (8), we can compute the (i,j)-th entry $(i \neq j)$ of the matrix derivative $(e^X)^{(1)}[H]$:

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

where the third equality follows from the identity

$$\frac{e^{\lambda_i} - e^{\lambda_j}}{\lambda_i - \lambda_j} = \int_0^1 e^{\lambda_i u} e^{(1-u)\lambda_j} du. \tag{15}$$

This proves (13) for the case $i \neq j$. The case $i = j$ can be considered in a similar fashion.

Now we prove the second order differential formula (14). 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 (15) to obtain

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

which establishes (14). □

Notice that the first order derivative formula (13) 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 X, Z. However, such a pair may be not 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 natural to define the weighted centers here 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 $X = P^T DP$ 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

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

□

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 an optimal solution is unique. $\square$

Let X_w^p be the optimal solution for (P_w). Using Theorem 3.2 we obtain the following Karush-Kuhn-Tucker optimality condition for X_w^p: $\exists\, y^p \in \Re^m$ such that

$$C - \sum_{i=1}^m y_i^p A_i - \int_0^\infty (uI + X_w^p)^{-1} W (uI + X_w^p)^{-1} du = 0. \tag{16}$$

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, (16) induces a dual solution

$$Z_w^p = C - \sum_{i=1}^m y_i^p A_i = F_W(X_w^p). \tag{17}$$

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 that (D_w) has a unique optimal solution, which we denote by y^d. Again, by Theorem 3.2, the KKT optimality condition for (D_w) reduces to

$$b_i - \left\langle A_i, \int_0^\infty \left(vI + C - \sum_{i=1}^m y_i^d A_i \right)^{-1} W \left(vI + C - \sum_{i=1}^m y_i^d A_i \right)^{-1} dv \right\rangle = 0,\ i = 1, 2, ..., m. \tag{18}$$

The condition (18) induces a primal solution

$$X_w^d = \int_0^\infty \left(vI + C - \sum_{i=1}^m y_i^d A_i \right)^{-1} W \left(vI + C - \sum_{i=1}^m y_i^d A_i \right)^{-1} dv = F_W \left(C - \sum_{i=1}^m y_i^d A_i \right). \tag{19}$$

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 commutativity fails to hold in general. This can be seen from the following simple 2×2 example: let

$$A_1 = \frac{1}{2}E_{1,1}, \quad A_2 = E_{2,2}, \quad A_3 = E_{1,2}, \quad b_1 = b_2 = 1, \; b_3 = 0, \quad C = E_{1,1} + E_{2,2} + E_{1,2}, \quad W = C + E_{1,1},$$

where $E_{i,j}$ denotes the symmetric matrix with all entries zero except the (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 (18) would imply $X_w^p = F_W(Z_w^p)$. Notice that

$$Z_w^p = Q \begin{bmatrix} 1+\log 2 & 0 \\ 0 & 1-\log 2 \end{bmatrix} Q^T, \quad Q^T W Q = \frac{1}{2}\begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}, \quad \text{where } Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Using the definition of F_W and simplifying the integral yields

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

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

The lack of commutativity 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 below.

Theorem 4.4 *Given any $W \succ 0$, let $\{X_w^p, Z_w^p\}$, $\{X_w^d, Z_w^d\}$ be defined by* (16)-(17) *and* (18)-(19) *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

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

where the third and the last steps are due to the identity $\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)$ for any matrices A and B. Similarly, we can show that $\mathrm{tr}\,(X_w^d Z_w^d) = \mathrm{tr}\,(W)$. □

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

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

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

$$X = \mathrm{Blockdiag}\left\{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, M_{n-2,n-2}\right\}, \quad \text{with } M \succeq 0. \tag{20}$$

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

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

for some primal feasible matrix X of the form (20) 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 $f_j(X) = C_j^T X + XC_j - XQ_jX$ for some matrices C_j and $Q_j \succeq 0$, $j = 1, ..., m$. 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 a 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

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

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 3.5 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

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

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

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

As a conclusion, we see that the total number of basic operations required to assemble such a Newton direction is $O(mn^2 + n^3)$.

In the case that $f_j(X) = C_j^TX + XC_j - XQ_jX$, $j = 1, ..., m$, a direct transformation using the standard Schur complement would turn the constraints $f_j(X) \succeq B_j$ into the Linear Matrix

Inequalities

$$\left[\begin{array}{cc} C_j^T X + X C_j - B_j & Q_j^{1/2} X \\ X Q_j^{1/2} & I \end{array}\right] \succeq 0, \; j = 1, ..., m.$$

Including the slack variables the dimension of the variable will be $O(mn^2)$. Hence, finding the Newton direction in the most straightforward way will involve $O((mn^2)^3) = O(m^3n^6)$ operations. In view of this, working with the quadratic matrix function directly is without doubt an attractive option.

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," *Mathematical Programming* 95 (2003) 431-474.

[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*, Cambridge 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 SEEM 2003-09, Department of Systems Engineering & Engineering Management, The Chinese University of Hong Kong (2003).