Self-Concordance and Matrix Monotonicity with Applications to Quantum Entanglement Problems

Leonid Faybusovich*[1] and Cunlu Zhou†[2]

[1,2]Department of Mathematics, University of Notre Dame, IN

April 14, 2019

Abstract

Let V be an Euclidean Jordan algebra and Ω be a cone of invertible squares in V. Suppose that $g : \mathbb{R}_+ \to \mathbb{R}$ is a matrix monotone function on the positive semiaxis which naturally induces a function $\tilde{g} : \Omega \to V$. We show that $-\tilde{g}$ is compatible (in the sense of Nesterov-Nemirovski) with the standard self-concordant barrier $B(x) = -\ln\det(x)$ on Ω. As a consequence, we show that for any $c \in \Omega$, the functions of the form $-\operatorname{tr}(c\tilde{g}(x)) + B(x)$ are self-concordant on Ω. In particular, the function $x \mapsto -\operatorname{tr}(c \ln x)$ is a self-concordant barrier function on Ω. Using these results, we apply a long-step path-following algorithm developed in [L. Faybusovich and C. Zhou Long-step path-following algorithm for solving symmetric programming problems with nonlinear objective functions. *Comput Optim Appl*, 72(3):769-795, 2019] to a number of important optimization problems arising in quantum information theory. Results of numerical experiments and comparisons with existing methods are presented.

Keywords: Nonlinear symmetric programming, Self-concordant functions, Matrix monotone functions, Quantum entanglement, Quantum relative entropy, Interior-point method, Long-step path-following

1 Introduction

Many interesting and difficult convex optimization problems arise in quantum information theory [4, 11]. Their semidefinite relaxations are still difficult to solve mainly due to rather complicated nonlinear objective functions of (several) matrix arguments. First-order methods such as cutting plane and conditional gradient are typically used in practical computations in this area but they are usually quite slow and in general have no complexity estimates (see e.g. [6, 19]). An interesting approach was proposed in [7] where a class of nonlinear objective functions is approximated by rational functions with a subsequent

*lfaybuso@nd.edu

†zhou.43@nd.edu

reduction to semidefinite programming problems of a much larger size (with a linear objective function). In [10] we developed a long-step path-following algorithm for a class of symmetric programming problems with nonlinear objective functions of a certain type. Numerical experiments confirmed very good performance of our algorithm for a class of objective functions which include the so-called *quantum (von Neumann) entropy*.

In present paper we further extend the class of objective functions for which our method can be applied. In particular, we consider a quantum entanglement problem and compare our results with approach of [7]. Combining our results for the quantum entropy problem and the quantum entanglement problem, we show how to tackle the case of quantum relative entropy as an objective function. The latter class of optimization problems has numerous applications in quantum information theory and beyond [3].

The plan of the paper is as follows. In section 2 we briefly describe some basic Jordan algebraic concepts used in this paper. In section 3 we present our main theoretical results on a new class of functions that can be solved by our long-step path-following algorithm developed in [10]. Section 4 is devoted to applications and numerical experiments where we consider the relative entropy of entanglement problem and quantum relative entropy optimization problem. In section 5 we give some concluding remarks.

2 Jordan Algebraic Concepts

We adhere to the notation of an excellent book [5]. See also [1] for an introduction to applications of Jordan algebras in optimization.

Definition 2.1. Let $\mathbf{F}$ be the field $\mathbb{R}$ or $\mathbb{C}$. A vector space V over $\mathbf{F}$ is called an *algebra* over $\mathbf{F}$ if a bilinear mapping $(x, y) \to xy$ from $V \times V$ into V is defined.

For an element x in V, let $L(x) : V \to V$ be the linear map such that

$$L(x)y = xy.$$

Definition 2.2. An algebra V over $\mathbf{F}$ is a *Jordan algebra* if

$$xy = yx, \ x(x^2y) = x^2(xy), \ \forall x, y \in V.$$

In other words, Jordan algebra is always commutative but typically NOT associative. In an algebra V, one defines x^n recursively by $x^n = x \cdot x^{n-1}$. An algebra V is said to be *power associative* if $x^p \cdot x^q = x^{p+q}$ for any $x \in V$ and integers p, q.

Proposition 2.3. *A Jordan algebra is power associative. Besides,*

$$[L(x^p), L(x^q)] = L(x^p)L(x^q) - L(x^q)L(x^p) = 0,$$

for all $x \in V$ and any positive integers p and q. In other words, the corresponding linear operators commute.

This is Proposition II.1.2 in [5].

From now on we assume that the Jordan algebra has an identity element e, i.e.,

$$xe = ex = x, \ \forall x \in V.$$

Let V be a finite-dimensional power associative algebra over $\mathbf{F}$ with an identity element e, and let $\mathbf{F}[X]$ denote the algebra over $\mathbf{F}$ of polynomials in one variable with coefficients in $\mathbf{F}$. For $x \in V$ we define

$$\mathbf{F}[x] = \{p(x) : p \in \mathbf{F}[X]\}.$$

A nonzero polynomial $p \in \mathbf{F}[X]$ of minimal possible degree such that $p(x) = 0$ is called the *minimal polynomial* of x. Given $x \in V$, let $m(x)$ be the degree of the minimal polynomial of x. We define the rank of V as

$$r = \max\{m(x) : x \in V\}.$$

An element x is called *regular* if $m(x) = r$.

Proposition 2.4. *The set of regular elements is open and dense in V. There exist polynomials $a_1, \ldots, a_r$ on V such that the minimal polynomial of every regular element x is given by*

$$f(\lambda; x) = \lambda^r - a_1(x)\lambda^{r-1} + a_2(x)\lambda^{r-2} + \ldots + (-1)^r a_r(x).$$

The polynomials $a_1, \ldots, a_r$ are unique and a_j is homogeneous of degree j.

- This is Proposition II.2.1 in [5].
- The coefficient $a_1(x)$ is called the *trace* of x and is denoted by $\mathrm{tr}(x)$. In particular, trace is linear. The coefficient $a_r(x)$ is called the *determinant* of x and is denoted by $\det(x)$.

An element x is said to be *invertible* if there exists an element $y \in \mathbf{F}[x]$ such that $xy = e$. The set $\lambda \in \mathbf{F}$ such that $x - \lambda e$ is not invertible is called the *spectrum* of x, denoted by $\mathrm{spec}(x)$.

Given $x \in V$, we define

$$P(x) = 2L(x)^2 - L(x^2).$$

The map P is called the *quadratic representation* of V. Note that $P(x)e = x^2$. Let $DP(x)y$ be the Fréchet derivative of P at $x \in V$ evaluated on $y \in V$. Then we have

$$DP(x)y = 2P(x, y), \tag{1}$$

where

$$P(x, y) = L(x)L(y) + L(y)L(x) - L(xy),\ \ x, y \in V.$$

Proposition 2.5. *Let V be a finite-dimensional Jordan algebra over $\mathbf{F}$. An element $x \in V$ is invertible if and only if $P(x)$ is invertible. In this case*

$$P(x)x^{-1} = x, \quad P(x)^{-1} = P(x^{-1}).$$

This is Proposition II.3.1 in [5].

Proposition 2.6. *Let $\mathcal{J}$ be the (open) set of invertible elements in V. The map $x \to x^{-1} : \mathcal{J} \to \mathcal{J}$ is Fréchet differentiable and* *(i)* $D(x^{-1})u = -P(x^{-1})u, x \in \mathcal{J}, u \in V$.

(ii) *If* x *and* y *are invertible, then* $P(x)y$ *is invertible and*

$$(P(x)y)^{-1} = P(x^{-1})y^{-1}.$$

(iii) $P(P(x)y) = P(x)P(y)P(x),\ \forall x, y \in V$.

(iv) $P(P(x)y, P(x)z) = P(x)P(y,z)P(x),\ \forall x, y, z \in V$.

This is Proposition II.3.3 in [5].

Definition 2.7. A bilinear form β on V is called *associative* if

$$\beta(xy, z) = \beta(x, yz),\ \forall x, y, z \in V.$$

Proposition 2.8. *The symmetric bilinear forms* $\mathrm{Tr}(L(xy))$ *and* $\mathrm{tr}(xy)$ *are associative.*

This is Proposition II.4.3 in [5]. Here $\mathrm{Tr}(L(xy))$ is the usual trace of the linear operator $L(xy) : V \to V$.

In the case when $\mathbf{F} = \mathbb{R}$, we consider an important class of *Euclidean Jordan algebras*. A Jordan algebra V over $\mathbb{R}$ is called *Euclidean* if $\mathrm{tr}(x^2) > 0,\ \forall x \in V \setminus \{0\}$. A Euclidean Jordan algebra is *simple* if it is not the direct sum of two (nontrivial) Euclidean Jordan algebras.

An element $c \in V$ is called *idempotent* if $c^2 = c$. Two idempotents c and d are *orthogonal* if $cd = 0$. A system of idempotents $c_1, \ldots, c_k$ is a complete system of orthogonal idempotents if $c_i^2 = c_i,\ c_ic_j = 0,\ i \neq j$, and $c_1 + \ldots + c_k = e$.

Theorem 2.9. *Let* V *be a Euclidean Jordan algebra. Given* $x \in V$*, there exist unique real numbers* $\lambda_1, \ldots, \lambda_k$*, all distinct, and a unique complete system of orthogonal idempotents* $c_1, \ldots, c_k$ *such that*

$$x = \lambda_1 c_1 + \cdots + \lambda_k c_k.$$

In this case $\mathrm{spec}(x) = \{\lambda_1, \ldots, \lambda_k\},\ c_1, \ldots, c_k \in \mathbb{R}[x]$.

This is Theorem III.1.1 in [5].

An idempotent is *primitive* if it is non-zero and cannot be written as a sum of two non-zero idempotents. We say that $c_1, \ldots, c_m$ is a complete system of orthogonal primitive idempotents, or *Jordan frame*, if each c_j is primitive idempotent and if

$$c_jc_k = 0,\ j \neq k,\ c_1 + \ldots + c_m = e.$$

Note that in this case $m = r$ (rank of V).

Theorem 2.10. *Suppose* V *has rank* r*. Then for* $x \in V$ *there exists a Jordan frame* $c_1, \ldots, c_r$ *and real numbers* $\lambda_1, \ldots, \lambda_r$ *such that*

$$x = \sum_{j=1}^{r} \lambda_j c_j. \tag{2}$$

The numbers λ_j *(with multiplicities) are uniquely determined by* x*. Furthermore,*

$$\det(x) = \prod_{j=1}^{r} \lambda_j, \quad \mathrm{tr}(x) = \sum_{j=1}^{r} \lambda_j.$$

- This is Theorem III.1.2 in [5].
- Given a function f which is defined at least on $\text{spec}(x)$, we can define

$$f(x) = \sum_{i=1}^{r} f(\lambda_i) c_i,$$

if $x = \sum_{i=1}^{r} \lambda_i c_i$. In particular,

$$\exp(x) = \sum_{i=1}^{r} \exp(\lambda_i) c_i, \quad \ln x = \sum_{i=1}^{r} \ln(\lambda_i) c_i, \; \lambda_i > 0.$$

- Convexity and differentiability of such functions on Euclidean Jordan algebras have been studied in [8]. We will use these properties extensively.
- There exists a canonical scalar product on V defined as $\langle x, y\rangle = \text{tr}(xy)$, $x, y \in V$, and a norm $\|x\| = \langle x, x\rangle^{\frac{1}{2}}$, $x \in V$.

Let

$$\overline{\Omega} = \{x^2 : x \in V\}$$

be the so-called *cone of squares* of V.

Theorem 2.11. *Let V be a Euclidean Jordan algebra. The interior Ω of $\overline{\Omega}$ is a symmetric (i.e., self-dual, homogeneous) convex cone. Furthermore, Ω is the connected component of e in the set $\mathcal{J}$ of invertible elements, and also Ω is the set of elements x in V for which $L(x)$ is positive definite. In particular, the group of linear automorphisms $GL(\Omega)$ of Ω acts transitively on it. Moreover, $P(x) \in GL(\Omega)$ for any invertible x.*

This is Theorem III.2.1 and Proposition III.2.2 in [5].

Let $c_1, \ldots, c_k$ be a complete system of orthogonal idempotents. For each idempotent c, denote $V(c, 0)$, $V(c, 1)$, $V(c, 1/2)$ the eigenspaces of $L(c)$ corresponding to eigenvalues 0, 1, 1/2 respectively. Then $L(c_1), \ldots, L(c_k)$ pairwise commute and

$$V = \bigoplus_{1 \le i \le j} V_{ij},$$

where $V_{ii} = V(c_i, 1)$, $V_{ij} = V(c_i, 1/2) \cap V(c_j, 1/2)$. Such a decomposition of V corresponding to a complete system of orthogonal idempotents is called the Peirce decomposition. It is studied in detail in Section 1 of Chapter IV in [5].

Remark 2.12. In terms of spectral decomposition (2), the spectrum of linear operator $P(x)$ has the form $\lambda_i \lambda_j$, $i, j = 1, 2, \ldots, r$. This easily follows from the definition of $P(x)$, namely, the way it is expressed in $L(x)$ and its Peirce decomposition. It also follows that $P(x^\alpha) = P(x)^\alpha$, $\forall \alpha \in \mathbb{R}$, and $P(x) > 0$, $\forall x \in \Omega$.

Proposition 2.13. *Let $B(x) = -\ln\det(x)$, $x \in \Omega$., then*

$$DB(x)(u) = -\left\langle x^{-1}, u\right\rangle, \; x \in \Omega, \; u \in V, \tag{3}$$

and by proposition 2.6

$$D^2B(x)(u,u) = \left\langle P(x^{-1})u,\, u\right\rangle,\, x \in \Omega,\, u \in V, \tag{4}$$

where $D^2B(x)$ is the second Fréchet derivative of B at x.

- For a proof, see e.g. [5, Proposition III.4.2].
- $B(x) = -\ln\det(x)$ is called a *barrier* function on Ω: it is smooth and strongly convex (Hessian $H_B(x) > 0$, $x \in \Omega$), and

$$B(x) \to +\infty,\ x \to \partial\Omega.$$

Example 2.14. A typical example of a Jordan algebra over a field $\mathbf{F}$ is the vector space of symmetric matrices over $\mathbf{F}$ with multiplication operation

$$A \diamond B = \frac{AB + BA}{2},$$

where on the right we have a usual matrix multiplication.

- When $\mathbf{F} = \mathbb{R}$, we get an example of a (simple) Euclidean Jordan algebra: $\mathcal{S}^n$.
- Note that the $\diamond$ operation is commutative but not associative.
- For $L(X)$, the linear operator defined by $X \in \mathcal{S}^n$, we have

$$L(X)Y = \frac{XY + YX}{2}.$$

- The identity element $e = \mathrm{I}$, the $n \times n$ identity matrix.
- The rank of $\mathcal{S}^n$, $r(\mathcal{S}^n) = \mathrm{tr}(e) = r$.
- For the quadratic representation $P(X)$ of $\mathcal{S}^n$, we have for all $Y, Z \in \mathcal{S}^n$

$$\begin{aligned} P(X)Y &= (2L(X)^2 - L(X^2))Y \\ &= 2L(X)L(X)Y - L(X^2)Y \\ &= 2L(X)(X \diamond Y) - X^2 \diamond Y \\ &= XYX, \end{aligned}$$

$$\begin{aligned} P(X,Y)Z &= L(X)L(Y)Z + L(Y)L(X)Z - L(XY)Z \\ &= \frac{XZY + YZX}{2}, \end{aligned}$$

  and

$$\begin{aligned} P(X)^{-1}Y &= P(X^{-1})Y = X^{-1}YX^{-1}, \\ D(X^{-1})Y &= -X^{-1}YX^{-1} = -P(X^{-1})Y. \end{aligned}$$

- For each $X \in \mathcal{S}^n$, there exists a unique spectral decomposition:

$$X = \sum_{i=1}^{k} \lambda_i \Pi_i,$$

where λ_i's are distinct eigenvalues of X and Π_i's are corresponding orthogonal projectors.

- We can also have a spectral decomposition with a complete system of orthogonal primitive idempotents (or Jordan frame):

$$X = \sum_{i=1}^{n} \lambda_i \Pi_i = \sum_{i=1}^{n} \lambda_i v_i v_i^\intercal, \tag{5}$$

where $\{v_i\}_{i=1}^n$ constitute an orthonormal basis of $\mathbb{R}^n$.

- With (5), the determinant and trace of X can be expressed as

$$\mathrm{Det}(X) = \prod_{i=1}^{n} \lambda_i, \quad \mathrm{Tr}(X) = \sum_{i=1}^{n} \lambda_i.$$

- The inner product $\langle A,\, B\rangle$ is defined as $\mathrm{Tr}(AB) = \sum_{i,j} A_{ij} B_{ij}$.

- With (5), we have a corresponding Peirce decomposition of $V = \mathcal{S}^n$:

$$V = \bigoplus_{1 \le i \le j \le n} V_{ij}, \tag{6}$$

where $V_{ii} = \{c \cdot v_i v_i^\intercal : c \in \mathbb{R}\}$, $i = 1, \ldots, n$, and

$$V_{ij} = \left\{ c \cdot \frac{v_i v_j^\intercal + v_j v_i^\intercal}{\sqrt{2}} : c \in \mathbb{R} \right\}, \; 1 \le i < j \le n.$$

- Recall that a symmetric matrix $A \in \mathcal{S}^n$ is positive semidefinite $(A \ge 0)$ if

$$x^T A x \ge 0, \; \forall x \in \mathbb{R}^n,$$

and positive definite $(A > 0)$ if

$$x^T A x > 0, \; \forall x \in \mathbb{R}^n \setminus \{0\}.$$

The cone of squares

$$\overline{\Omega} \triangleq \{X^2 : X \in \mathcal{S}^n\}$$

is the positive semidefinite cone, denoted by $\mathcal{S}_+^n$, and its interior Ω is the positive definite cone, denoted by $\mathcal{S}_{++}^n$.

3 Main Results

Proposition 3.1. *Let* $f : \Omega \to \Omega,\ f(x) = x^{-1}$*. For* $h \in V$*,* $\Gamma(x) = P(x)^{-\frac{1}{2}}h$*, we have*

$$\mathrm{D}\, f(x)(h) = -P(x)^{-1}h = -P(x^{-\frac{1}{2}})(\Gamma(x)), \tag{7}$$

$$\mathrm{D}^2\, f(x)(h,h) = 2P(x)^{-\frac{1}{2}}(\Gamma(x)^2) = 2P(x)^{-1}P(h)(x^{-1}), \tag{8}$$

$$\mathrm{D}^3\, f(x)(h,h,h) = -6P(x)^{-\frac{1}{2}}(\Gamma(x)^3). \tag{9}$$

Proof. Formula (7) follows from proposition 2.6 (i). We know that [9]

$$\mathrm{D}^2\, f(x)(h,h) = 2P(x)^{-\frac{1}{2}}(\Gamma(x)^2).$$

Now, using abbreviation $\Gamma = \Gamma(x)$, we have

$$\Gamma^2 = P(\Gamma)(e).$$

Let $\Delta = 2P(x)^{-\frac{1}{2}}(\Gamma^2)$. Then

$$\begin{aligned}\Delta &= 2P(x)^{-\frac{1}{2}}P(\Gamma)(e)\\ &= 2P(P(x)^{-\frac{1}{2}}(\Gamma))P(x^{\frac{1}{2}})(e)\\ &= 2P(P(x)^{-1}(h))(x)\\ &= 2P(x)^{-1}P(h)P(x)^{-1}(x)\\ &= 2P(x)^{-1}P(h)(x^{-1}),\end{aligned}$$

where in the second and the fourth inequality we used proposition 2.6 (iii).

Let

$$\psi(x) = \mathrm{D}^2\, f(x)(h,h) = 2P(x)^{-1}P(h)(x^{-1}).$$

Then

$$\begin{aligned}\mathrm{D}^3\, f(x)(h,h,h) &= \mathrm{D}\,\psi(x)(h)\\ &= 2\,\mathrm{D}\,\varphi(x)(h)P(h)(x^{-1}) + 2P(x)^{-1}P(h)(\mathrm{D}\, f(x)(h)),\end{aligned} \tag{10}$$

where $\varphi(x) = P(x)^{-1}$. Note that

$$\begin{aligned}\mathrm{D}\,\varphi(x)h &= -P(x)^{-1}\,\mathrm{D}\,P(x)(h)P(x)^{-1}\\ &= -2P(x)^{-1}P(x,h)P(x)^{-1}\\ &= -2P(x)^{-\frac{1}{2}}P(P(x)^{-\frac{1}{2}}(x), P(x)^{-\frac{1}{2}}(h))P(x)^{-\frac{1}{2}}\\ &= -2P(x)^{-\frac{1}{2}}P(e,\Gamma)P(x)^{-\frac{1}{2}}\\ &= -2P(x)^{-\frac{1}{2}}L(\Gamma)P(x)^{-\frac{1}{2}},\end{aligned}$$

where in the second equality we used (1) and in the third equality we used proposition 2.6 (iv). Substituting this into (10), we obtain

$$\mathrm{D}^3\, f(x)(h,h,h) = -4P(x)^{-\frac{1}{2}}L(\Gamma)P(x)^{-\frac{1}{2}}P(h)(x^{-1}) - 2P(x)^{-1}P(h)(P(x)^{-1}(h)),$$

where we used (7).

Now,

$$\begin{aligned} L(\Gamma)P(x)^{-\frac{1}{2}}P(h)(x^{-1}) &= L(\Gamma)P(x)^{-\frac{1}{2}}P(h)(P(x)^{-\frac{1}{2}}(e)) \\ &= L(\Gamma)P(P(x)^{-\frac{1}{2}}(h))(e) \\ &= L(\Gamma)P(\Gamma)(e) \\ &= \Gamma^3, \end{aligned}$$

where in the second equality we used proposition 2.6 (*iii*).

Furthermore,

$$\begin{aligned} P(x)^{-1}P(h)(P(x)^{-1}(h)) &= P(x)^{-\frac{1}{2}}P(x)^{-\frac{1}{2}}P(h)P(x)^{-\frac{1}{2}}(P(x)^{-\frac{1}{2}}(h)) \\ &= P(x)^{-\frac{1}{2}}P(P(x)^{-\frac{1}{2}}(h))(P(x)^{-\frac{1}{2}}(h)) \\ &= P(x)^{-\frac{1}{2}}(P(\Gamma)(\Gamma)) \\ &= P(x)^{-\frac{1}{2}}(\Gamma^3). \end{aligned}$$

Hence,

$$\mathrm{D}^3 f(x)(h,h,h) = -6P(x)^{-\frac{1}{2}}(\Gamma^3).$$

□

Let $x, y \in V$. Then

$$x \succeq y \Leftrightarrow x - y \in \overline{\Omega},$$

and

$$x \succ y \Leftrightarrow x - y \in \Omega.$$

Lemma 3.2. *Let $x \in V$ and $y \in \Omega$. Then*

$$\pm P(y)(x^3) \preceq \|x\|\, P(y)(x^2),$$

where

$$\|x\| = \sqrt{\langle x,\, x\rangle}.$$

Proof. Let

$$x = \sum_{i=1}^{r} \lambda_i e_i$$

be a spectral decomposition of x. Then

$$x^3 = \sum_{i=1}^{r} \lambda_i^3 e_i \text{ and } x^2 = \sum_{i=1}^{r} \lambda_i^2 e_i.$$

Denote by ρ the following expression

$$\rho \triangleq \max\{|\lambda_i| : i \in [1, r]\}.$$

We clearly have

$$\rho \le (\sum_{i=1}^{r} \lambda_i^2)^{\frac{1}{2}} = \|x\|,$$

and

$$\pm\, x^3 \preceq \rho x^2 \preceq \|x\|\, x^2. \tag{11}$$

Indeed, for (11), notice that

$$\pm\lambda_i^3 \le |\lambda_i|^3 \le \rho|\lambda_i|^2 = \rho\lambda_i^2 \le \|x\|\, \lambda_i^2,\ \ i = 1, \ldots, r.$$

Hence,

$$\|x\|\, x^2 - (\pm x^3) = \sum_{i=1}^{r} (\|x\|\, \lambda_i^2 \mp \lambda_i^3) e_i \in \overline{\Omega}.$$

It remains to notice that (see theorem 2.11)

$$P(y)\overline{\Omega} \subseteq \overline{\Omega}.$$

□

Theorem 3.3. *Let* $B(x) = -\ln\det(x),\ \ f(x) = x^{-1},\ \ x \in \Omega.$ *Then*

$$-\,\mathrm{D}^3 f(x)(h,h,h) \preceq 3\,\mathrm{D}^2 f(x)(h,h)\sqrt{\mathrm{D}^2 B(x)(h,h)},\ \ h \in V. \tag{12}$$

Remark 3.4. Compare (12) with [17, (5.4.21)].

Proof. By proposition 3.1,

$$\begin{aligned} -\,\mathrm{D}^3 f(x)(h,h,h) &= 6P(x)^{-\frac{1}{2}}(\Gamma(x)^3), \\ \mathrm{D}^2 f(x)(h,h) &= 2P(x)^{-\frac{1}{2}}(\Gamma(x)^2), \\ \Gamma(x) &= P(x)^{-\frac{1}{2}}(h). \end{aligned}$$

Hence, by lemma 3.2

$$-\,\mathrm{D}^3 f(x)(h,h,h) \preceq 3\,\mathrm{D}^2 f(x)(h,h)\, \|\Gamma(x)\|.$$

Now

$$\begin{aligned} \|\Gamma(x)\| &= \sqrt{\left\langle P(x)^{-\frac{1}{2}}(h),\, P(x)^{-\frac{1}{2}}(h)\right\rangle} \\ &= \sqrt{\langle P(x)^{-1}(h),\, h\rangle} \\ &= \sqrt{\mathrm{D}^2 B(x)(h,h)}, \end{aligned}$$

where in the last equality we used proposition 2.13. □

Corollary 3.5. *Consider the following set*

$$\text{epi} \triangleq \{(x, y) \in \Omega \times \Omega : x^{-1} \prec y\}.$$

Then the function

$$F(x, y) = -\ln \det(y - x^{-1}) - \ln \det(x)$$

is a self-concordant barrier for epi *with the barrier parameter* $2r$*.*

Proof. The proof is similar to one in [17, p. 413-414], where the case of the cone of positive semidefinite matrices was considered. $\square$

For real symmetric $n \times n$ matrices A and B, we will use notation

$$A \geq B \text{ if } A - B \geq 0,$$

and

$$A > B \text{ if } A - B > 0.$$

Let $g : [0, +\infty) \to \mathbb{R}$ be a real-valued function. We say that g is *matrix monotone* (*anti-monotone*) if for any real symmetric matrices of the same size such that $A \geq 0$, $B \geq 0$ and $A \geq B$, we have

$$g(A) \geq g(B) \quad (g(A) \leq g(B)).$$

Let $-g : [0, +\infty) \to \mathbb{R}$ be a matrix monotone function. Then g is anti-monotone and admits the following integral representation:

$$g(\lambda) = g(0) - \beta\lambda + \int_0^{+\infty} \left(\frac{\tau^2}{\lambda + \tau} - \tau \right) \mathrm{d}\nu(\tau),$$

where $\lambda \in [0, +\infty)$, $\beta \geq 0$, and ν is a positive measure on $(0, +\infty)$ such that

$$\int_0^{+\infty} \frac{\tau}{1 + \tau} \mathrm{d}\nu(\tau) < +\infty.$$

(See e.g. [13, Theorem 4.41].)

Hence, g induces a function mapping $\overline{\Omega}$ into V:

$$g(x) = g(0)e - \beta x + \int_0^{+\infty} \left(\tau^2 (x + \tau e)^{-1} - \tau e \right) \mathrm{d}\nu(\tau), \;\; x \in \overline{\Omega}. \tag{13}$$

(For details, see [9, p. 1520].)

Theorem 3.6. *Let* $B(x) = -\ln \det(x), \;\; x \in \Omega$*. Then*

$$-\mathrm{D}^3 g(x)(h, h, h) \leq 3 \, \mathrm{D}^2 g(x)(h, h) \sqrt{\mathrm{D}^2 B(x)(h, h)}, \;\; h \in V. \tag{14}$$

Corollary 3.7. *Consider the following set*

$$\mathrm{epi}_g = \{(x, y) \in \Omega \times \Omega : g(x) \prec y\}.$$

Then the function

$$F_g(x, y) = -\ln\det(y - g(x)) - \ln\det(x)$$

is a self-concordant barrier for epi_g *with barrier parameter* $2r$.

Proof of theorem 3.6. By proposition 3.1, using representation (13), we have

$$\mathrm{D}\, g(x)(h) = -\beta h - \int_0^{+\infty} P(x + \tau e)^{-1}(h)\tau^2 \,\mathrm{d}\nu(\tau),$$

$$\mathrm{D}^2\, g(x)(h, h) = 2\int_0^{+\infty} P(x + \tau e)^{\frac{1}{2}}\Gamma(\tau)^2\tau^2 \,\mathrm{d}\nu(\tau), \tag{15}$$

where

$$\Gamma(\tau) = P(x + \tau e)^{-\frac{1}{2}}(h),$$

and

$$\mathrm{D}^3\, g(x)(h, h, h) = -6\int_0^{+\infty} P(x + \tau e)^{\frac{1}{2}}\Gamma(\tau)^3\tau^3 \,\mathrm{d}\nu(\tau). \tag{16}$$

By lemma 3.2,

$$\pm P(x + \tau e)^{-\frac{1}{2}}\Gamma(\tau)^3 \preceq \|\Gamma(\tau)\|\, P(x + \tau e)^{-\frac{1}{2}}\Gamma(\tau)^2.$$

Now

$$\|\Gamma(\tau)\| = \sqrt{\langle h,\, P(x + \tau e)^{-1}h\rangle} \le \sqrt{\langle h,\, P(x)^{-1}h\rangle},\ \forall \tau \ge 0.$$

(See [9, p. 1527].) Since

$$\left\langle h,\, P(x)^{-1}h\right\rangle = \mathrm{D}^2\, B(x)(h, h),$$

we obtain

$$\pm P(x + \tau e)^{-\frac{1}{2}}\Gamma(\tau)^3 \preceq \sqrt{\mathrm{D}^2\, B(x)(h, h)}P(x + \tau e)^{-\frac{1}{2}}\Gamma(\tau)^2.$$

Using (15) and (16), we obtain (14). □

Corollary 3.8. *Let* $c \in \overline{\Omega}$. *Then under assumptions of theorem 3.6, we have*

$$|\mathrm{D}^3\, \varphi_c(x)(h, h, h)| \le 3\,\mathrm{D}^2\, \varphi_c(x)(h, h)\sqrt{\mathrm{D}^2\, B(x)(h, h)},\ x \in \Omega,\ h \in V,$$

where

$$\varphi_c(x) = \langle c,\, g(x)\rangle\,. \tag{17}$$

Proof. The result immediately follows from (14), taking into consideration of the self-duality of the cone $\overline{\Omega}$. □

Corollary 3.9. *For any* $\beta \ge 0$, *the function*

$$\beta\varphi_c(x) - \ln\det(x),\ x \in \Omega$$

is self-concordant on Ω.

For a proof, see e.g. [9, Theorem 3.5].

Remark 3.10. Corollary 3.9 allows one to apply a long-step path-following algorithm developed in [10] for optimization problems involving objective functions of the form φ_c, i.e.,

$$\varphi_c(x) = \mathrm{tr}(cg(x)),$$

where $-g : [0, +\infty) \to \mathbb{R}$ is matrix monotone.

Corollary 3.11. *Let $c \in \Omega$. Then the function*

$$\psi_c(x) = -\mathrm{tr}(c \ln(x))$$

is a self-concordant barrier function on Ω.

Proof. Note that

$$\lambda \mapsto \ln(\lambda),\ \ \lambda > 0,$$

is a matrix monotone function. Hence, the function

$$\lambda \mapsto -\ln(\lambda),\ \ \lambda > 0,$$

is anti-monotone. By corollary 3.8,

$$|\mathrm{D}^3\,\psi_c(x)(h,h,h)| \leq 3\,\mathrm{D}^3\,\psi_c(x)(h,h)\sqrt{\mathrm{D}^2\,B(x)(h,h)}. \tag{18}$$

Note that

$$B(x) = \psi_e(x) = -\mathrm{tr}(\ln(x)),$$

but then

$$\mathrm{D}^2\,B(x)(h,h) = \mathrm{tr}\left(\mathrm{D}^2(-\ln(x))(h,h)\right), \tag{19}$$

$$\mathrm{D}^2\,\psi_c(x)(h,h) = \mathrm{tr}\left(c\,\mathrm{D}^2(-\ln(x))(h,h)\right). \tag{20}$$

Note that (see (15))

$$\mathrm{D}^2(-\ln(x))(h,h) \in \overline{\Omega}.$$

If $c_{\min} > 0$ is the minimal eigenvalue of c, then

$$c \succeq c_{\min} e.$$

Consequently, by (19) and (20), we have

$$\mathrm{D}^2\,B(x)(h,h) \leq \frac{D^2\psi_c(x)(h,h)}{c_{\min}}.$$

Combining this with (18), we obtain

$$|\mathrm{D}^3\,\psi_c(x)(h,h,h)| \leq \frac{1}{\sqrt{c_{\min}}}\left[\mathrm{D}^2\,\psi_c(x)(h,h)\right]^{\frac{3}{2}}.$$

□

Example 3.12. Consider the function

$$\lambda \mapsto \lambda^{-1},\ \lambda > 0.$$

Clearly, it is matrix anti-monotone on $(0, +\infty)$.

Let

$$\varphi_c(x) = \operatorname{tr}(cx^{-1}).$$

Using proposition 3.1,

$$\begin{aligned} \mathrm{D}\,\varphi_c(x)(h) &= -\operatorname{tr}\left(cP(x)^{-\frac{1}{2}}(\Gamma)\right), \\ \mathrm{D}^2\,\varphi_c(x)(h) &= 2\operatorname{tr}\left(c\left(P(x)^{-\frac{1}{2}}(\Gamma^2)\right)\right), \end{aligned}$$

where

$$\Gamma = P(x)^{-\frac{1}{2}}h.$$

Hence,

$$\begin{aligned} \mathrm{D}\,\varphi_c(x)(h) &= \langle \nabla\varphi_c(x),\, h\rangle \\ &= -\left\langle c,\, P(x)^{-\frac{1}{2}}\Gamma\right\rangle \\ &= -\left\langle P(x)^{-\frac{1}{2}}c,\, P(x)^{-\frac{1}{2}}h\right\rangle \\ &= -\left\langle P(x)^{-1}c,\, h\right\rangle. \end{aligned}$$

Hence,

$$\nabla\varphi_c(x) = -P(x)^{-1}c.$$

Furthermore,

$$\begin{aligned} \mathrm{D}^2\,\varphi_c(x)(h,h) &= 2\left\langle c,\, P(x)^{-\frac{1}{2}}(\Gamma^2)\right\rangle \\ &= 2\left\langle P(x)^{-\frac{1}{2}}c,\, L(\Gamma)\Gamma\right\rangle \\ &= 2\left\langle L(\Gamma)P(x)^{-\frac{1}{2}}c,\, P(x)^{-\frac{1}{2}}h\right\rangle \\ &= 2\left\langle L(P(x)^{-\frac{1}{2}}c)\Gamma,\, P(x)^{-\frac{1}{2}}h\right\rangle \\ &= 2\left\langle P(x)^{-\frac{1}{2}}L(P(x)^{-\frac{1}{2}}c)P(x)^{-\frac{1}{2}}h,\, h\right\rangle \\ &= \langle H_{\varphi_c}(x)(h),\, h\rangle. \end{aligned}$$

Hence,

$$H_{\varphi_c}(x) = 2P(x)^{-\frac{1}{2}}L(P(x)^{-\frac{1}{2}}c)P(x)^{-\frac{1}{2}}.$$

Correspondingly, inverting the Hessian is equivalent to solving the *"generalized Lyapunov"* equation.

4 Applications and Numerical Experiments

4.1 Relative Entropy of Entanglement

In quantum information theory, the so-called *relative entropy of entanglement* (REE) is an important measure of entanglement for a given quantum state. For a detailed description of the concept, see [19].

We are interested in solving the following optimization problem which gives a lower bound for the REE [19] [1]:

$$\begin{aligned} f(X) &= -\operatorname{Tr}(C\ln(X)) \to \min, \\ \operatorname{Tr}(X) &= 1, \\ \mathcal{L}(X) &\geq 0, \\ X &\geq 0, \end{aligned} \tag{21}$$

where C is a density matrix, i.e., $C \geq 0$ and $\operatorname{Tr}(C) = 1$. For the entanglement problem, we assume that the matrix size is $n \times n$ and $n = kl$. We further assume that the feasible set has a nonempty interior. Without loss of generality, here we only consider the real symmetric matrices (the case of Hermitian matrices fits naturally into the general Jordan algebraic framework).

$\mathcal{L}(\cdot)$ in (21) is the so-called *partial transpose* operator which is given as (in block-matrix form)

$$\mathcal{L}(\rho) = \begin{bmatrix} A_{11}^T & A_{12}^T & \cdots & A_{1k}^T \\ A_{21}^T & A_{22}^T & \cdots & A_{2k}^T \\ \vdots & \vdots & \ddots & \vdots \\ A_{k1}^T & A_{k2}^T & \cdots & A_{kk}^T \end{bmatrix}, \tag{22}$$

where

$$\rho = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & \cdots & A_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{kk} \end{bmatrix}, \tag{23}$$

and A_{ij}'s are $l \times l$ sub-block matrices. It is easy to verify that $\mathcal{L}(\cdot)$ is a linear symmetric operator for $(kl) \times (kl)$ symmetric matrices and $\mathcal{L}^{-1} = \mathcal{L}$.

By corollary 3.11,

$$F_\beta(X) = \beta f(X) - \ln\det(X)$$

is self-concordant on S^n_{++} (the cone of positive definite matrices). Hence, we can apply the long-step path-following algorithm developed in [10] to solve (21). More specifically, we

[1] More precisely, the lower bound for the REE of a quantum state C is given by $\operatorname{Tr}(C\ln(C)) - \operatorname{Tr}(C\ln(X^*))$, where X^* is the optimal solution of (21).

solve the following barrier family of auxiliary problems:

$$\begin{aligned} F_\beta(X) &= \beta f(X) + B(X) + \zeta(X) \to \min, \\ \mathrm{Tr}(X) &= 1, \\ X &> 0, \\ \mathcal{L}(X) &> 0, \\ \beta &\geq 0, \end{aligned} \tag{24}$$

where

$$B(X) = -\ln\det(X) \text{ and } \zeta(X) = -\ln\det(\mathcal{L}(X)).$$

For implementation of the long-step path-following algorithm, we need to obtain expressions for $\nabla F_\beta(X)$ and $\mathrm{H}_{F_\beta}(X)$.

4.1.1 Calculations of Gradient and Hessian

Let I be the identity matrix of the same size as X. For $X > 0$, consider the integral representation of $\ln(X)$:

$$\ln(X) = \int_0^{+\infty} \left[\frac{1}{1+t}\mathrm{I} - (X + t\mathrm{I})^{-1}\right] \mathrm{d}t, \tag{25}$$

with which we can calculate its first and second Fréchet derivatives:

$$\mathrm{D}\ln(X)(\xi) = \int_0^{+\infty} (X + t\mathrm{I})^{-1}\xi(X + t\mathrm{I})^{-1}\,\mathrm{d}t, \tag{26}$$

$$\mathrm{D}^2\ln(X)(\xi,\xi) = -2\int_0^{+\infty} (X + t\mathrm{I})^{-1}\xi(X + t\mathrm{I})^{-1}\xi(X + t\mathrm{I})^{-1}\,\mathrm{d}t. \tag{27}$$

Consider a spectral decomposition of X,

$$X = U\Lambda U^T,$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $UU^T = \mathrm{I}$. Then

$$(X + t\mathrm{I})^{-1} = UDU^T,$$

where $D = (\Lambda + t\mathrm{I})^{-1}$. Let

$$\tilde{C} = U^T CU,\ \tilde{\xi} = U^T \xi U,$$

and

$$d_i = (\lambda_i + t)^{-1},\ i = 1, \ldots, n.$$

Furthermore, let $h(\lambda) = \ln(\lambda)$, $\lambda > 0$, and denote the first divided difference matrix by $h^{[1]}(\Lambda)$, where

$$\begin{aligned} [h^{[1]}(\Lambda)]_{ij} &= h^{[1]}(\lambda_i, \lambda_j) \\ &= \begin{cases} \dfrac{h(\lambda_j) - h(\lambda_i)}{\lambda_j - \lambda_i}, & \lambda_j \neq \lambda_i, \\ h'(\lambda_i), & \lambda_j = \lambda_i. \end{cases} \end{aligned}$$

We have

$$\int_0^{+\infty} D\tilde{\xi}D\,\mathrm{d}t = h^{[1]}(\Lambda)\circ\tilde{\xi}, \tag{28}$$

where $\circ$ is the Schur product: for $m\times n$ matrices A and B,

$$[A\circ B]_{ij} = A_{ij}B_{ij}.$$

Indeed,

$$\begin{aligned}
\int_0^{+\infty} [D\tilde{\xi}D]_{ij}\,\mathrm{d}t &= \int_0^{+\infty} d_i\tilde{\xi}_{ij}d_j\,\mathrm{d}t \\
&= \tilde{\xi}_{ij}\int_0^{+\infty} d_id_j\,\mathrm{d}t \\
&= \tilde{\xi}_{ij}\int_0^{+\infty} (\lambda_i+t)^{-1}(\lambda_j+t)^{-1}\,\mathrm{d}t \\
&= \tilde{\xi}_{ij}\cdot\begin{cases} \dfrac{\ln(\lambda_j)-\ln(\lambda_i)}{\lambda_j-\lambda_i}, & \lambda_j\neq\lambda_i \\ \dfrac{1}{\lambda_j}, & \lambda_j=\lambda_i. \end{cases} \\
&= \tilde{\xi}_{ij}\cdot h^{[1]}(\lambda_i,\lambda_j).
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{D}\,f(X)(\xi) &\overset{(26)}{=} -\operatorname{Tr}\left(\int_0^{+\infty} CUDU^T\xi UDU^T\,\mathrm{d}t\right) \\
&= -\int_0^{+\infty}\operatorname{Tr}(U^TCUDU^T\xi UD)\,\mathrm{d}t \\
&= -\int_0^{+\infty}\operatorname{Tr}(\tilde{C}D\tilde{\xi}D)\,\mathrm{d}t \\
&= -\operatorname{Tr}\left(\tilde{C}\int_0^{+\infty} D\tilde{\xi}D\,\mathrm{d}t\right) \\
&\overset{(28)}{=} -\operatorname{Tr}\left(\tilde{C}\left(h^{[1]}(\Lambda)\circ\tilde{\xi}\right)\right) \\
&= -\operatorname{Tr}\left(\left(\tilde{C}\circ h^{[1]}(\Lambda)\right)\tilde{\xi}\right) \\
&= \left\langle U\left(\tilde{C}\circ h^{[1]}(\Lambda)\right)U^T,\,\xi\right\rangle,
\end{aligned}$$

where in the second last equality we used properties of Schur product ([15, p. 306]):

$$\operatorname{Tr}((A\circ B)C^T) = \operatorname{Tr}((A\circ C)B^T).$$

Hence,

$$\nabla f(X) = -U\left((U^TCU)\circ h^{[1]}(\Lambda)\right)U^T. \tag{29}$$

Furthermore, we have (by proposition 2.13)

$$\nabla B(X) = -X^{-1}, \tag{30}$$

and

$$\begin{aligned}
\mathrm{D}\,\zeta(X)(\xi) &= -\left\langle \mathcal{L}(X)^{-1},\, \mathcal{L}(\xi)\right\rangle \\
&= -\left\langle \mathcal{L}^T(\mathcal{L}(X)^{-1}),\, \xi\right\rangle,
\end{aligned}$$

which implies that

$$\nabla\zeta(X) = -\mathcal{L}^T(\mathcal{L}(X)^{-1}) = -\mathcal{L}(\mathcal{L}(X)^{-1}), \tag{31}$$

where in the last equality we used the fact that $\mathcal{L}$ is a symmetric operator.

With (29), (30) and (31), we obtain

$$\nabla F_\beta(X) = \beta\nabla f(X) - X^{-1} - \mathcal{L}(\mathcal{L}(X)^{-1}). \tag{32}$$

Next we calculate the Hessian of $f(X)$. We have

$$\begin{aligned}
\mathrm{D}^2 f(X)(\xi,\xi) &\overset{(27)}{=} -\operatorname{Tr}\left(C\left(-2\int_0^{+\infty} UDU^T\xi UDU^T\xi UDU^T\,\mathrm{d}t\right)\right) \\
&= 2\int_0^{+\infty}\operatorname{Tr}(\tilde{C}D\tilde{\xi}D\tilde{\xi}D)\,\mathrm{d}t \\
&= 2\int_0^{+\infty}\operatorname{Tr}(D\tilde{C}D\tilde{\xi}D\tilde{\xi})\,\mathrm{d}t \\
&= 2\int_0^{+\infty}\operatorname{Tr}\left(\frac{D\tilde{C}D\tilde{\xi}D + D\tilde{\xi}D\tilde{C}D}{2}\cdot\tilde{\xi}\right)\mathrm{d}t \\
&= \operatorname{Tr}\left(\left(\int_0^{+\infty}(D\tilde{C}D\tilde{\xi}D + D\tilde{\xi}D\tilde{C}D)\,\mathrm{d}t\right)\tilde{\xi}\right) \\
&= \left\langle U\left(\int_0^{+\infty}(D\tilde{C}D\tilde{\xi}D + D\tilde{\xi}D\tilde{C}D)\,\mathrm{d}t\right)U^T,\, \xi\right\rangle \\
&= \left\langle \mathrm{H}_f(X)(\xi),\, \xi\right\rangle.
\end{aligned}$$

Hence,

$$\mathrm{H}_f(X)(\xi) = U\left(\int_0^{+\infty}(D\tilde{C}D\tilde{\xi}D + D\tilde{\xi}D\tilde{C}D)\,\mathrm{d}t\right)U^T. \tag{33}$$

Recall the vectorization operator $\operatorname{vec}(\cdot)$ for an $n\times m$ matrix $A = [a_{ij}]$:

$$\operatorname{vec}(A) = [a_{11},\ldots,a_{n1},a_{12},\ldots,a_{n2},\ldots,a_{1m},\ldots,a_{nm}]^T.$$

Using vectorization, we obtain

$$\mathrm{H}_f(X) = (U\otimes U)\left(\int_0^{+\infty}((D\tilde{C}D)\otimes D + D\otimes(D\tilde{C}D))\,\mathrm{d}t\right)(U\otimes U)^T. \tag{34}$$

Let

$$S = \int_0^{+\infty}((D\tilde{C}D)\otimes D + D\otimes(D\tilde{C}D))\,\mathrm{d}t.$$

Note that S is a sparse block matrix. Indeed, recalling that $d_i = (\lambda_i + t)^{-1}$, we have the (ij, kl)-th entry of S:

$$\begin{aligned} S_{ij,kl} &= \int_0^{+\infty} (d_i \tilde{C}_{ij} d_j \delta_{kl} d_l + \delta_{ij} d_j d_k \tilde{C}_{kl} d_l) \, \mathrm{d}t \\ &= \delta_{kl} \tilde{C}_{ij} \Gamma_{ijl} + \delta_{ij} \tilde{C}_{kl} \Gamma_{jkl}, \end{aligned} \tag{35}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \text{ and } \Gamma_{ijk} = \int_0^{+\infty} d_i d_j d_k \, \mathrm{d}t,$$

from which we notice that the ij-th sub-block matrix is diagonal if $i \neq j$. Such sparsity can be utilized when performing inversion (see e.g. [16]).

As for Γ_{ijk}, recall that $h(\lambda) = \ln(\lambda)$, $\lambda > 0$, and notice that

$$\Gamma_{ijk} = -h^{[2]}(\lambda_i, \lambda_j, \lambda_k),$$

where for distinct λ_i, λ_j, and λ_k,

$$h^{[2]}(\lambda_i, \lambda_j, \lambda_k) = \frac{h^{[1]}(\lambda_i, \lambda_j) - h^{[1]}(\lambda_i, \lambda_k)}{\lambda_j - \lambda_k}, \tag{36}$$

is the second divided difference function [2, p. 128]; for cases when λ_i, λ_j, and λ_k are not distinct, the function is defined by taking limits in (36), e.g.,

$$h^{[2]}(\lambda, \lambda, \lambda) = \frac{1}{2} h''(\lambda).$$

For the remaining parts of $F_\beta(X)$, by proposition 2.13 we have

$$\mathrm{H}_B(X)(\xi) = P(X^{-1})(\xi) = X^{-1} \xi X^{-1}. \tag{37}$$

Furthermore, we have

$$\mathrm{H}_\zeta(X)(\xi) = \mathcal{L}\left(\mathcal{L}(X)^{-1} \mathcal{L}(\xi) \mathcal{L}(X)^{-1} \right), \tag{38}$$

and

$$\mathrm{H}_\zeta(X)^{-1}(\eta) = \mathcal{L}\left(\mathcal{L}(X) \mathcal{L}(\eta) \mathcal{L}(X) \right), \tag{39}$$

where we used the fact that $\mathcal{L}^T = \mathcal{L}$ and $\mathcal{L}^{-1} = \mathcal{L}$.

Remark 4.1. Note that our calculations of the gradient and Hessian are applicable to all matrix monotone functions since the first and second divided differences appeared here can be computed in a similar fashion for all matrix monotone functions. Hence the whole numerical scheme works for all matrix monotone functions.

4.1.2 Numerical Results

In this section, we present some numerical results for the relative entropy of entanglement (REE) problem (21). We compare our results to the ones obtained by using the `cvxquad` package described in [6, 7] which uses a rational approximation of the natural logarithm function and then employs available SDP (semidefinite programming) solvers to solve the approximation SDP problems of a much larger size. Here we use MOSEK as their solver (one of the most powerful SDP solvers available) and call their method "cvxquad + mosek" or `cvxquad` in the following discussions.

Our algorithm is implemented in MATLAB, and all the numerical experiments are performed on a personal 15-in Macbook Pro with Intel core i7 and 16 GB memory. Data are randomly generated with an initial interior feasible point. Table 1 shows the numerical results for the REE problem (21), where N is the number of total Newton steps performed in our path-following algorithm [10].

Table 1: Numerical Results for Relative Entropy of Entanglement

	Long-Step Path-Following			cvxquad + mosek	
$n = k \times l$	Time(s)	f_{min}	N	Time(s)	f_{min}
$4 = 2 \times 2$	0.0245	0.0045	8	0.2372	0.0044
$6 = 3 \times 2$	0.0548	0.0468	11	0.2473	0.0468
$8 = 4 \times 2$	0.1176	0.0090	18	0.2611	0.0090
$9 = 3 \times 3$	0.1141	0.0883	13	0.2719	0.0883
$12 = 4 \times 3$	0.2291	0.0153	14	0.3127	0.0153
$16 = 4 \times 4$	0.4703	0.0138	13	0.4473	0.0138

- Note that the problem sizes we used are up to the ones used in [6]. In [19], a first-order cutting plane method was developed for the same REE optimization problem (21), but their performance is not as good as `cvxquad` so we only compare to `cvxquad` here. To the best of our knowledge, these are the only known alternative options for the REE optimization problem.
- Compared to the other two options, our long-step path-following is the only one with complexity estimates (in terms of Newton steps): $\mathcal{O}(n \ln(\frac{n}{\epsilon}))$. For details, see [10].
- Such numerical experiments are mainly for showing the viability and potential of our approach. The performance of our long-step path-following algorithm can be greatly improved by more efficient implementations, e.g., implement the whole or part of the program in **C** instead of MATLAB and use sparsity for the Hessian calculation (see (35)). Potentially more efficient inexact Newton's method can be explored as well.
- For a more robust implementation of the first divided difference (and hence the second

divided difference), we used techniques described in [14, p. 279-280], namely,

$$h^{[1]}(\lambda_i,\lambda_j) = \begin{cases} \dfrac{1}{\lambda_i}, & \lambda_i = \lambda_j, \\ \dfrac{\ln(\lambda_i)-\ln(\lambda_j)}{\lambda_i-\lambda_j}, & |\lambda_i| < \dfrac{|\lambda_j|}{2} \text{ or } |\lambda_j| < \dfrac{|\lambda_i|}{2}, \\ \dfrac{2\,\mathrm{atanh}(z)}{\lambda_i-\lambda_j}, & \text{otherwise}\ , \end{cases}$$

where $z = (\lambda_i - \lambda_j)/(\lambda_i + \lambda_j)$, and

$$\mathrm{atanh}(z) = \frac{1}{2}\ln\left(\frac{1+z}{1-z}\right).$$

By corollary 3.11, we know that for $C > 0$, $f(X) = -\operatorname{Tr}(C\ln(X))$ is self-concordant on S^n_{++} (the cone of $n \times n$ positive definite matrices). Hence, we can apply Newton's method directly to the following optimization problem:

$$\begin{aligned} f(X) = -\operatorname{Tr}(C\ln(X)) &\to \min, \\ \operatorname{Tr}(A_iX) = b_i,\ i &= 1,\dots,m, \\ X &\geq 0. \end{aligned} \tag{40}$$

Table 2 shows numerical results for (40) compared to the ones obtained by using the *trace_logm* function in the `cvxquad` package.

Table 2: Numerical Results for (40) with Newton's Method

		Newton's Method			cvxquad + mosek	
n	m	Time(s)	f_{min}	N	Time(s)	f_{min}
4	2	0.0086	1.0452	12	0.2185	1.0452
6	3	0.0199	1.7242	14	0.2364	1.7242
8	4	0.0321	2.7675	13	0.2824	2.7675
9	4	0.0532	1.6394	17	0.2750	1.6394
12	6	0.1123	2.7024	18	0.3152	2.7023
16	8	0.2781	2.4900	19	0.4443	2.4900
20	10	0.4820	2.8335	15	0.6899	2.8334

4.2 Quantum Relative Entropy Optimization

Quantum relative entropy is another important concept in quantum information theory (see e.g. [18]), which is defined as

$$\mathcal{D}(\rho \parallel \sigma) = \operatorname{Tr}(\rho\ln(\rho)) - \operatorname{Tr}(\rho\ln(\sigma)), \tag{41}$$

where ρ and σ are $n \times n$ density matrices, i.e., $\rho, \sigma \geq 0$ and $\operatorname{Tr}(\rho) = \operatorname{Tr}(\sigma) = 1$.

Quantum relative entropy optimization has been shown to play an important role in various applications [3]. We consider the following optimization problem involving quantum relative entropy:

$$\begin{aligned} f(X,Y) &= \mathrm{Tr}(X\ln(X)) - \mathrm{Tr}(X\ln(Y)) \to \min, \\ \mathrm{Tr}(A_iX) &= b_i,\ i=1,\dots,m, \\ \mathrm{Tr}(B_iY) &= c_i,\ i=1,\dots,l, \\ X,Y &\geq 0. \end{aligned} \tag{42}$$

Again, we assume the feasible set is bounded and has a nonempty interior.

To solve (42), we propose an alternating minimization procedure as follows: starting from a feasible $Y_0 > 0$, we solve for X_0:

$$\begin{aligned} f(X,Y_0) &= \mathrm{Tr}(X\ln(X)) - \mathrm{Tr}(X\ln(Y_0)) \to \min, \\ \mathrm{Tr}(A_iX) &= b_i,\ i=1,\dots,m, \\ X &\geq 0, \end{aligned} \tag{43}$$

and then with X_0, we solve for Y_1:

$$\begin{aligned} f(X_0,Y) &= \mathrm{Tr}(X_0\ln(X_0)) - \mathrm{Tr}(X_0\ln(Y)) \to \min, \\ \mathrm{Tr}(B_iY) &= c_i,\ i=1,\dots,l, \\ Y &\geq 0. \end{aligned} \tag{44}$$

Such an alternating procedure is repeated until a stopping criterion is met.

Note that (43) is the entropy minimization problem solved in [10], which guarantees its optimization solution attained in the interior at each iteration, and (44) is the relative entropy of entanglement problem without the partial transpose operator (40), which also guarantees its optimization solution stays in the interior for each $X > 0$. For the general results on convergence of the alternating method, we refer to [12].

4.2.1 Numerical Results

In this section, we present some numerical results for the previous discussed alternating minimization procedure. Again the data are randomly generated with an initial interior feasible point for (44) and the analytic center calculated for (43). Note that the analytic center for (43) needs only to be calculated once and for small dimensions (e.g. $n \leq 20$) the time is about $0.3 - 0.4$ seconds (see [10]). Without loss of generality, we assume that $\mathrm{Tr}(X) = \mathrm{Tr}(Y) = 1$. Table 3 shows numerical results by using the alternating method combined with our long-step path-following algorithm. We compare to the results obtained by using the *quantum_rel_entr* function in `cvxquad` [6].

- Note that m and l can be different values. We choose the same values for convenience.
- "failed" means that the program runs more than one hour and stops responding.
- For instances when $n = 4$ and 9, we have $f_{min} \approx 0$. However, since the quantum relative entropy function is nonnegative when $\mathrm{Tr}(X) = \mathrm{Tr}(Y)$ (see [13, p. 117]) and we assume $\mathrm{Tr}(X) = \mathrm{Tr}(Y) = 1$, our results are more accurate compared to the ones obtained by `cvxquad`.

Table 3: Numerical Results for Quantum Relative Entropy

			Long-Step Path-Following (alternating)			cvxquad + mosek (quantum_rel_entr)	
n	m	l	Time(s)	f_{min}	Iters	Time(s)	f_{min}
4	2	2	0.06	2.078e-06	8	0.37	-1.78e-07
6	3	3	0.55	0.0226	31	3.86	0.0225
8	4	4	1.86	0.0429	70	59.96	0.0425
9	4	4	0.11	1.258e-08	4	110.48	-3.016e-06
12	6	6	0.56	5.642e-06	8	N/A	failed
16	8	8	0.88	9.540e-07	6	N/A	failed
20	10	10	2.17	3.856e-06	7	N/A	failed

- As shown in table 3, using the *quantum_rel_entr* function in `cvxquad` directly is not practical (see instances when $n \geq 8$), compared to which, our long-step path-following alternating scheme has a clear advantage.

Table 4: Numerical Results for Quantum Relative Entropy

			Long-Step Path-Following (alternating)			cvxquad + mosek (alternating)		
n	m	l	Time(s)	f_{min}	Iters	Time(s)	f_{min}	Iters
4	2	2	0.06	2.078e-06	8	3.64	-1.192e-07	8
6	3	3	0.55	0.0226	31	15.87	0.0226	31
8	4	4	1.86	0.0429	70	36.63	0.0428	70
9	4	4	0.11	1.258e-08	4	2.11	-6.398e-07	4
12	6	6	0.56	5.642e-06	8	5.34	8.156e-07	8
16	8	8	0.88	9.540e-07	6	5.93	-5.443e-06	6
20	10	10	2.17	3.856e-06	7	11.90	-7.529e-06	7

In table 4, we run both methods with the alternating scheme and our long-step path-following algorithm still performs better. Note that at each iteration in (44), with our long-step path-following algorithm, we can use the previous optimization solution Y as a starting point (a "warm" start). In contrast, there are no such options for `cvxquad`. Again, for instances when $n = 4, 9, 16$ and 20, $f_{min} \approx 0$, and our results are more accurate as explained above.

Remark 4.2. All the data used in our numerical experiments can be accessed from here: `https://doi.org/10.13140/RG.2.2.10779.31521`.

5 Concluding Remarks

In the present paper we considered a class of important objective functions in quantum information theory which involve functions of matrix arguments. We showed that a long-step path-following algorithm developed (along with complexity estimates) in [10] can be applied to optimization problems with such objective functions. We performed numerical experiments for optimization problems involving quantum entanglement and quantum relative entropy which are two critical concepts in quantum information theory and quantum computing. We compared our results with the ones obtained by existing methods and showed that our approach is more efficient for most of the cases. In particular, for the quantum relative entropy optimization problem we considered, our results are significantly better. That being said, such numerical experiments are mainly for showing the viability and potential of our approach. As mentioned early, various aspects of the implementation can be improved. In particular, for the objective functions arising in quantum information theory, their Hessians are quite complicated which put a limit on the size of the problems that can be solved by using second-order methods. However, we noticed a certain sparsity pattern for the Hessians of the objective functions we considered which potentially can be exploited to mitigate the "curse of dimensionality." Other options such as quasi-Newton methods (e.g., a "modern" version called Newton sketch) can be explored as well.

Lastly, notice that our numerical scheme is universal in the sense that the natural logarithm function can be substituted by arbitrary matrix monotone function on the positive semiline. The only difference will be the calculations of the first and second divided differences.

Acknowledgement

This research is supported in part by Simmons Foundation Grant 275013.

References

[1] F. Alizadeh. *An Introduction to Formally Real Jordan Algebras and Their Applications in Optimization*, pages 297–337. Springer US, Boston, MA, 2012. doi: 10.1007/978-1-4614-0769-0_11.

[2] R. Bhatia. *Matrix Analysis*. Springer-Verlag, New York, 1997.

[3] V. Chandrasecaran and P. Shah. Relative entropy optimization and its applications. *Math. Program. Ser. A*, 161(1-2):1–32, 2017. doi: 10.1007/s10107-016-0998-2.

[4] B. Coutts, M. Girard, and J. Watrous. Certifying optimality for convex quantum channel optimization problems, 2018. arXiv:1810.13295.

[5] J. Faraut and A. Korányi. *Analysis on Symmetric Cones*. Cambridge University Press, Cambridge, 1994.

[6] H. Fawzi and O. Fawzi. Efficient optimization of the quantum relative entropy. *J Phys A Math Theor*, 51(15):154003, 2018. doi: 10.1088/1751-8121/aab285.

[7] H. Fawzi, J. Saunderson, and P. A. Parrilo. Semidefinite approximations of the matrix logarithm. *Found. Comput. Math.*, 2018. doi: 10.1007/s10208-018-9385-0.

[8] L. Faybusovich. Several Jordan-algebraic aspects of optimization. *Optimization*, 57 (3):379–393, 2008. doi: 10.1080/02331930701523510.

[9] L. Faybusovich and T. Tsuchiya. Matrix monotonicity and self-concordance: how to handle quantum entropy in optimization problems. *Optim. Lett.*, 11(8):1513–1526, 2017. doi: 10.1007/s11590-017-1145-6.

[10] L. Faybusovich and C. Zhou. Long-step path-following algorithm for solving symmetric programming problems with nonlinear objective functions. *Comput Optim Appl*, 72 (3):769–795, Apr 2019. doi: 10.1007/s10589-018-0054-7.

[11] M. W. Girard, G. Gour, and S. Friedland. On convex optimization problems in quantum information theory. *J Phys A Math Theor*, 47(50):505302, Nov 2014. doi: 10.1088/1751-8113/47/50/505302.

[12] L. Grippo and M. Sciandrone. On the convergence of the block nonlinear gauss-seidel method under convex constraints. *Oper. Res. Lett.*, 26(3):127–136, 2000. doi: 10.1016/S0167-6377(99)00074-7.

[13] F. Hiai and D. Petz. *Introduction to Matrix Analysis and Applications.* Springer International Publishing, 2014.

[14] N. J. Higham. *Functions of Matrices: Theory and Computation.* Society for Industrial and Applied Mathematics, Philadelphia, USA, 2008.

[15] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[16] M. Kojima and M. Muramatsu. A note on sparse SOS and SDP relaxations for polynomial optimization problems over symmetric cones. *Comput Optim Appl*, 42(1): 31–41, Jan 2009. doi: 10.1007/s10589-007-9112-2.

[17] Y. Nesterov. *Lectures on Convex Optimization.* Springer International Publishing, 2018.

[18] J. Watrous. *The Theory of Quantumm Information.* Cambridge University Press, 2018.

[19] Y. Zinchenko, S. Friedland, and G. Gour. Numerical estimation of the relative entropy of entanglement. *Phys. Rev. A*, 82:052336, Nov 2010. doi: 10.1103/PhysRevA.82.052336.