

Lieb's concavity theorem, matrix geometric means, and semidefinite optimization

Hamza Fawzi*

James Saunderson†

May 4, 2017

Abstract

A famous result of Lieb establishes that the map $(A, B) \mapsto \text{tr} [K^* A^{1-t} K B^t]$ is jointly concave in the pair (A, B) of positive definite matrices, where K is a fixed matrix and $t \in [0, 1]$. In this paper we show that Lieb's function admits an explicit semidefinite programming formulation for any rational $t \in [0, 1]$. Our construction makes use of a semidefinite formulation of weighted matrix geometric means. We provide an implementation of our constructions in Matlab.

Keywords: Matrix convexity; Semidefinite optimization; Linear matrix inequalities; Lieb's concavity theorem; Matrix geometric means

AMS Subject Classification: 90C22; 47A63; 81P45

1 Introduction

In 1973 Lieb [?] proved the following fundamental theorem.

Theorem 1 (Lieb). *Let K be a fixed matrix in $\mathbb{C}^{n \times m}$. Then for any $t \in [0, 1]$, the map*

$$(A, B) \mapsto \text{tr} [K^* A^{1-t} K B^t] \tag{1}$$

is jointly concave in (A, B) where A and B are respectively $n \times n$ and $m \times m$ Hermitian positive definite matrices.

This theorem plays a fundamental role in quantum information theory and was used for example to establish convexity of the quantum relative entropy as well as strong subadditivity [?]. In this paper we give an explicit representation of Lieb's function using semidefinite programming when t is a rational number. More precisely we prove:

Theorem 2. *Let K be a fixed matrix in $\mathbb{C}^{n \times m}$ and let $t = p/q$ be any rational number in $[0, 1]$. Then the convex set*

$$\{(A, B, \tau) : \text{tr} [K^* A^{1-t} K B^t] \geq \tau\}$$

has a semidefinite programming representation with at most $2 \lceil \log_2 q \rceil + 3$ linear matrix inequalities of size at most $2nm \times 2nm$.

*Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139. Email: hfawzi@mit.edu

†Department of Electrical and Computer Systems Engineering, Monash University, VIC 3800, Australia. Email: james.saunderson@monash.edu

Semidefinite programming is a class of convex optimization problems that can be solved in polynomial-time and that is supported by many existing numerical software packages. Having a semidefinite programming formulation of a function allows us to combine it with a wide family of other semidefinite representable functions and constraints, and solve the resulting problem to global optimality. In fact we have implemented our constructions in the Matlab-based modeling language CVX [?] and we are making them available online on the webpage

http://www.mit.edu/~hfawzi/lieb_cvx.html.

Matrix geometric means Our proof of Theorem ?? relies crucially on the notion of *matrix geometric mean*. Given $t \in [0, 1]$ and positive definite matrices A and B , the t -weighted matrix geometric mean of A and B denoted interchangeably by $G_t(A, B)$ or $A\#_t B$ is defined as:

$$G_t(A, B) = A\#_t B := A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^t A^{1/2}. \quad (2)$$

Note that when A and B are scalars (or commuting matrices) this formula reduces to the simpler expression $A^{1-t} B^t$. Equation (??) constitutes a generalization of the geometric mean to noncommuting matrices and satisfies many of the properties that are expected from a mean operation [?, ?]. One remarkable property of the matrix geometric mean is that it is *matrix concave*: if $t \in [0, 1]$, then for any pair $X = (A_1, B_1)$ and $Y = (A_2, B_2)$ we have:

$$G_t \left(\frac{X + Y}{2} \right) \succeq \frac{1}{2} (G_t(X) + G_t(Y))$$

where \succeq indicates the Löwner partial order on Hermitian matrices (i.e., $A \succeq B \Leftrightarrow A - B$ positive semidefinite). This fact can be used to give a simple proof of Lieb’s concavity theorem, see e.g., [?]. The matrix geometric mean was recently shown in [?] to have a semidefinite programming formulation. More precisely Sagnol showed that for any rational $t = p/q \in [0, 1]$ the convex set

$$\text{hyp}_t := \{ (A, B, T) \in \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \times \mathbf{H}^n : G_t(A, B) \succeq T \} \quad (3)$$

has a semidefinite programming representation with at most $O(\log_2(q))$ linear matrix inequalities of size $2n \times 2n$. In this paper we show how an SDP representation of the matrix geometric mean can be used to get an SDP representation of Lieb’s function as well as numerous other convex/concave functions. Table ?? summarizes the functions we consider in this paper, together with the size of the representations.

We only became aware of the result by Sagnol [?] after the first preprint of this paper appeared. As such, our alternative approach to constructing an SDP description of the matrix geometric mean, and a proof of its correctness, is included in Appendix ?? of this paper. Our construction has the same size (Theorem ??) as Sagnol’s, and extends to the regime $t \in [-1, 0] \cup [1, 2]$ for which G_t is matrix convex. Furthermore, our code, which is available online, is based on the construction in Appendix ??.

Implications for quantum relative entropy and related functions Our results can be used to solve, approximately, *quantum relative entropy programs* [?] using semidefinite programming. The quantum relative entropy function is defined as:

$$S(A\|B) = \text{tr}[A(\log A - \log B)]$$

Function	Properties	Size of SDP description ($t = p/q$)
Matrix geometric mean $(A, B) \mapsto A\#_t B$	matrix concave for $t \in [0, 1]$ matrix convex for $t \in [-1, 0] \cup [1, 2]$	$O(\log_2 q)$ LMIs of size $2n$ (Theorem ??) See also [?].
Lieb-Ando function $(A, B) \mapsto \text{tr} [K^* A^{1-t} K B^t]$ ($K \in \mathbb{C}^{n \times m}$ fixed)	concave for $t \in [0, 1]$ convex for $t \in [-1, 0] \cup [1, 2]$	$O(\log_2 q)$ LMIs of size $2nm$ (Theorem ??)
$A \mapsto \text{tr} [(K^* A^t K)^{1/t}]$ ($K \in \mathbb{C}^{n \times m}$ fixed)	concave for $t \in [-1, 1] \setminus \{0\}$ convex for $t \in [1, 2]$	$O(\log_2 q)$ LMIs of size $2nm$ (Theorem ??)
Tsallis entropy $A \mapsto \frac{1}{t} \text{tr} [A^{1-t} - A]$	concave for $t \in [0, 1]$ converges to von Neumann entropy $S(A)$ when $t \rightarrow 0$	$O(\log_2 q)$ LMIs of size $2n$ (Remark ??)
Tsallis relative entropy $(A, B) \mapsto \frac{1}{t} \text{tr} [A - A^{1-t} B^t]$	convex for $t \in [0, 1]$ converges to relative entropy $S(A\ B)$ when $t \rightarrow 0$	$O(\log_2 q)$ LMIs of size $2n^2$ (Remark ??)

Table 1: List of functions with SDP formulations considered in this paper.

where A and B are positive definite matrices. It is a simple corollary of Lieb's theorem that S is jointly convex in (A, B) . Indeed this follows from observing that:

$$S(A\|B) = \lim_{t \rightarrow 0^+} \frac{1}{t} \text{tr} [A - A^{1-t} B^t] \quad (4)$$

where we used the fact that for any matrix $X \succ 0$:

$$\log X = \lim_{t \rightarrow 0} \frac{1}{t} (X^t - I).$$

Identity (??) together with the semidefinite programming representation of Lieb's function can be used to get SDP approximations of the relative entropy function $S(A\|B)$ to arbitrary accuracy, by choosing t small enough. Unfortunately however, the convergence of $S_t(A\|B) := \frac{1}{t} \text{tr} [A - A^{1-t} B^t]$ to $S(A\|B)$ is slow (it is in $O(t)$) and obtaining decent approximations of $S(A\|B)$ thus requires to use very small values of t . While the size of the SDP descriptions of $S_t(A\|B)$ grows only like $\log(1/t)$, we observed that standard numerical algorithms to solve these SDPs become numerically ill-conditioned as t gets close to 0. There exist however other methods to obtain approximations of $S(A\|B)$ that converge much faster and are better behaved numerically and we discuss such methods in future work [?].

Related works It is well-known that the scalar functions $(x, y) \mapsto x^{1-t} y^t$ admit second-order cone representations when t is a rational number [?, Chapter 3]. The SDP representation of the matrix geometric mean can be seen as a matrix generalization of such results. The authors of [?] give a free semidefinite representations of the matrix power functions $X \mapsto X^t$ for rational $t \in [-1, 2]$,

however it seems that they were not aware of the paper by Sagnol [?] since such a representation already appears in this work. Furthermore the construction in [?] is in some cases smaller than [?]: for general rational $t = p/q \in [0, 1]$ the construction in [?] has size $O(\log_2(q))$ whereas in some cases the construction in [?] requires a number of LMIs that grows linearly with q . The authors of [?] also mentioned that certain multivariate versions of the matrix power function fail to have semidefinite representations. Working in the setting of geometric means, and then tensor products, seems to give one natural extension to the multivariate case (see Remark ??).

Outline In Section ?? we set up the basic notations and terminology for the paper and in Section ?? we prove the main results of the paper giving SDP representations of the functions given in Table ??.

2 Preliminaries

In this section we introduce basic notation and terminology used throughout the paper. Let \mathbf{H}^n be the space of $n \times n$ Hermitian matrices, $\mathbf{H}_+^n \subset \mathbf{H}^n$ the cone of $n \times n$ Hermitian positive semidefinite matrices and \mathbf{H}_{++}^n the cone of $n \times n$ strictly positive definite matrices. We use the notation $X \succeq Y$ if $X - Y$ is positive semidefinite, and $X \succ Y$ if $X - Y$ is positive definite. Suppose C is a convex set and $f : C \rightarrow \mathbf{H}^n$. We say that f is \mathbf{H}_+^n -convex if the \mathbf{H}_+^n -epigraph

$$\text{epi}_{\mathbf{H}_+^n}(f) := \{(X, T) \in C \times \mathbf{H}^n : f(X) \preceq T\}$$

is a convex set. Similarly f is \mathbf{H}_+^n -concave if the \mathbf{H}_+^n -hypograph

$$\text{hyp}_{\mathbf{H}_+^n}(f) := \{(X, T) \in C \times \mathbf{H}^n : f(X) \succeq T\}$$

is a convex set.

Semidefinite representations A semidefinite program is an optimization problem that takes the form

$$\begin{aligned} & \text{maximize} && \langle b, y \rangle \\ & \text{subject to} && A_0 + y_1 A_1 + \cdots + y_n A_n \succeq 0 \end{aligned}$$

where $y \in \mathbb{R}^n$ is the optimization variable, b is a fixed vector in \mathbb{R}^n and $A_0, A_1, \dots, A_n \in \mathbf{H}^m$ are fixed $m \times m$ Hermitian matrices. The condition

$$A_0 + y_1 A_1 + \cdots + y_n A_n \succeq 0$$

is known as a *linear matrix inequality* (LMI) of size m . We will say that a convex set C has a *SDP representation* if it can be expressed using LMIs (we allow for lifting variables). To evaluate the size of a semidefinite representation we record the number of LMIs of each size. For example consider the following convex set H :

$$H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1, x_2, x_3 \geq 0 \text{ and } x_1 x_2 x_3 \geq 1\}.$$

One can show that H admits the following SDP representation:

$$H = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 : \exists y, z \quad \text{s.t.} \quad \begin{bmatrix} x_1 & y \\ y & x_2 \end{bmatrix} \succeq 0, \begin{bmatrix} x_3 & z \\ z & 1 \end{bmatrix} \succeq 0, \begin{bmatrix} y & 1 \\ 1 & z \end{bmatrix} \succeq 0 \right\}. \quad (5)$$

This SDP representation consists of 3 LMIs of size 2 each.

Kronecker products and their properties If $A \in \mathbb{C}^{m \times n}$ we denote by $A^* \in \mathbb{C}^{n \times m}$ the conjugate transpose of A . The *Kronecker product* of $A \in \mathbb{C}^{m_1 \times n_1}$ and $B \in \mathbb{C}^{m_2 \times n_2}$ is the $\mathbb{C}^{m_1 m_2 \times n_1 n_2}$ matrix $A \otimes B$ with

$$[A \otimes B]_{(i,k)(j,\ell)} = A_{ij} B_{k\ell} \quad \text{for } 1 \leq i \leq n_1, 1 \leq j \leq n_2, 1 \leq k \leq m_1, 1 \leq \ell \leq m_2.$$

If A, B, C, D are matrices of compatible dimensions then $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ and $(A \otimes B)^* = A^* \otimes B^*$. Suppose $A \in \mathbf{H}^n$ and $B \in \mathbf{H}^m$ are Hermitian matrices with eigenvalue decompositions $A = U \Lambda_A U^*$ and $B = V \Lambda_B V^*$ where U, V are unitary matrices and Λ_A and Λ_B are diagonal. Then $U \otimes V$ is unitary and $\Lambda_A \otimes \Lambda_B$ is diagonal and so

$$A \otimes B = (U \otimes V)(\Lambda_A \otimes \Lambda_B)(U \otimes V)^*$$

is an eigenvalue decomposition of $A \otimes B$.

3 SDP representations

This is the main section of the paper where we describe the SDP representations of the various functions in Table ??.

3.1 Matrix geometric mean

We first consider the SDP representation of the matrix geometric mean. Recall that the *t-weighted geometric mean* $G_t : \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \rightarrow \mathbf{H}_{++}^n$ is defined by

$$G_t(A, B) = A \#_t B := A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^t A^{1/2}.$$

It is known [?] that G_t is matrix concave for $t \in [0, 1]$ and is matrix convex for $t \in [-1, 0] \cup [1, 2]$. We denote by hyp_t and epi_t the matrix hypograph and matrix epigraph of G_t respectively:

$$\text{hyp}_t = \{(A, B, T) \in \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \times \mathbf{H}^n : A \#_t B \succeq T\}$$

for $t \in [0, 1]$, and

$$\text{epi}_t = \{(A, B, T) \in \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \times \mathbf{H}^n : A \#_t B \preceq T\}$$

for $t \in [-1, 0] \cup [1, 2]$. These notations do not keep track of the dimension n explicitly but this omission should not cause any confusion.

The next theorem shows that the matrix geometric mean G_t for rational $t = p/q$ admits an SDP formulation involving $O(\log_2 q)$ LMIs of size at most $2n \times 2n$. The case $t \in [0, 1]$ was already obtained by Sagnol [?]. In Appendix ??, we explicitly describe our construction (on which our CVX code is based) and establish its correctness.

Theorem 3. *Let p, q be relatively prime integers with $p/q \in [-1, 2]$.*

- *If $p/q \in [0, 1]$ then $\text{hyp}_{p/q}$ has a SDP description with at most $2\lceil \log_2(q) \rceil + 1$ LMIs of size $2n \times 2n$ and one LMI of size $n \times n$.*
- *If $p/q \in [-1, 0] \cup [1, 2]$ then $\text{epi}_{p/q}$ has a SDP description with at most $2\lceil \log_2(q) \rceil + 2$ LMIs of size $2n \times 2n$ and one LMI of size $n \times n$.*

Proof. The construction is detailed in Appendix ??. □

We briefly discuss qualitative differences between our construction and that of Sagnol [?]. Our construction is recursive in nature, repeatedly expressing $\text{hyp}_{p/q}$ in terms of $\text{hyp}_{p'/q'}$ for a ‘simpler’ rational p'/q' , until reaching the base case of $\text{hyp}_{1/2}$. This recursive structure makes our construction particularly natural to implement in code. Sagnol’s construction is similar to what we would obtain if we explicitly unrolled our recursion, even though it is expressed quite differently. Indeed Sagnol’s approach assigns variables to the vertices of a binary tree and the LMIs describe relationships between each vertex and its children. To prove correctness of his SDP description, Sagnol requires a contraction argument with respect to the Riemannian metric on positive definite matrices. In contrast, the argument we give in Appendix ?? depends only on basic properties of the geometric mean and Schur complements.

3.2 SDP description for functions in Table ??

In this section we show how the SDP description of the matrix geometric mean can be used to obtain an SDP description of the functions given in Table ??.

3.2.1 Lieb’s function

We first consider Lieb’s function. The following is a restatement of Theorem ?? from the introduction with the additional case $t \in [-1, 0] \cup [1, 2]$.

Theorem 4. *Let K be a fixed matrix in $\mathbb{C}^{n \times m}$ and let $t = p/q$ be any rational number in $[-1, 2]$. Let $F_t(A, B) = \text{tr}[K^* A^{1-t} K B^t]$.*

- *If $t = p/q \in [0, 1]$, then F_t is concave and its hypograph admits a semidefinite programming representation using at most $2\lceil \log_2 q \rceil + 1$ LMIs of size $2nm \times 2nm$, one LMI of size $nm \times nm$ and one scalar inequality.*
- *If $t = p/q \in [-1, 0] \cup [1, 2]$, then F_t is convex and its epigraph admits a semidefinite programming representation using at most $2\lceil \log_2 q \rceil + 2$ LMIs of size $2nm \times 2nm$, one LMI of size $nm \times nm$ and one scalar inequality.*

Proof. To prove this theorem we use the well-known relationship between F_t and the matrix-valued function $L_t(A, B) = A^{1-t} \otimes \bar{B}^t$ due to Ando. In fact it is not difficult to verify that we have the following identity:

$$\text{tr}[K^* A^{1-t} K B^t] = \text{vec}(K)^*(A^{1-t} \otimes \bar{B}^t) \text{vec}(K) \quad (6)$$

where $\text{vec}(K)$ is a column vector of size nm obtained by concatenating the rows of K and \bar{B} is the entrywise complex conjugate of B (see e.g., [?, Lemma 5.12]). Thus, if $t \in [0, 1]$ we have for any real number τ

$$\text{tr}[K^* A^{1-t} K B^t] \geq \tau \iff \exists T \in \mathbf{H}_{++}^{nm} \text{ s.t. } \begin{cases} A^{1-t} \otimes \bar{B}^t \succeq T \\ \text{vec}(K)^* T \text{vec}(K) \geq \tau. \end{cases} \quad (7)$$

We now show how to convert (7) into an SDP formulation. The key idea (see e.g., [?]) is to note that

$$A^{1-t} \otimes \bar{B}^t = (A \otimes I) \#_t (I \otimes \bar{B}) \quad (8)$$

where I denotes the identity matrix of appropriate size. To see why (8) holds, note that $A \otimes I$ and $I \otimes \bar{B}$ commute and so

$$(A \otimes I) \#_t (I \otimes \bar{B}) = (A \otimes I)^{1-t} (I \otimes \bar{B})^t \stackrel{(a)}{=} (A^{1-t} \otimes I) (I \otimes \bar{B}^t) \stackrel{(b)}{=} A^{1-t} \otimes \bar{B}^t$$

where (a) can be shown using the eigenvalue decompositions of $A \otimes I$ and $I \otimes \bar{B}$, and (b) follows from the properties of the Kronecker product. Using the SDP formulation of the matrix geometric mean (Theorem ??) we can thus formulate the constraint $A^{1-t} \otimes \bar{B}^t \succeq T$ using $2\lceil \log_2(q) \rceil + 1$ LMIs of size $2nm \times 2nm$ and one LMI of size $nm \times nm$ (where $t = p/q$). Plugging this in (??) gives us an SDP formulation of the hypograph of Lieb's function with the required size. The case $t \in [-1, 0] \cup [1, 2]$ is treated in the same way. \square

Remark 1. • *It is straightforward to extend Theorem ?? to get an SDP formulation of the functions $(A, B) \mapsto A^s \otimes B^t$ where s and t are nonnegative numbers such that $s + t \leq 1$. It suffices to observe that*

$$A^s \otimes B^t \succeq T \iff \exists S \in \mathbf{H}_+^{nm} \text{ s.t. } \begin{cases} A^{\frac{s}{s+t}} \otimes B^{\frac{t}{s+t}} \succeq S \\ S^{s+t} \succeq T. \end{cases}$$

- *Similarly one can also extend Theorem ?? to obtain an SDP formulation of a k -variate generalization of the Lieb function, namely*

$$(A_1, \dots, A_k) \mapsto A_1^{t_1} \otimes \dots \otimes A_k^{t_k}$$

where $t_1, \dots, t_k \geq 0$ are such that $t_1 + \dots + t_k = 1$. To do so we simply eliminate one matrix at a time. For example in the case $k = 3$ we use:

$$A_1^{t_1} \otimes A_2^{t_2} \otimes A_3^{t_3} \succeq T \iff \exists S \in \mathbf{H}_+^{n_1 n_2} \text{ s.t. } \begin{cases} A_1^{\frac{t_1}{t_1+t_2}} \otimes A_2^{\frac{t_2}{t_1+t_2}} \succeq S \\ S^{t_1+t_2} \otimes A_3^{t_3} \succeq T. \end{cases}$$

Remark 2 (Tsallis entropies). • *For $t \in [0, 1]$ the Tsallis entropy [?] is defined as*

$$S_t(A) := \frac{1}{t} \text{tr} [A^{1-t} - A].$$

It is easy to see that $S_t(A)$ converges (from above) to the von Neumann entropy $S(A) = -\text{tr}[A \log A]$ when $t \rightarrow 0$, i.e., $S_t(A) \geq S(A)$ for any $t \in [0, 1]$ and $\lim_{t \rightarrow 0} S_t(A) = S(A)$. Also note that S_t is concave for all $t \in [0, 1]$. Its hypograph, $\{(A, \tau) \in \mathbf{H}_{++}^n \times \mathbb{R} : S_t(A) \geq \tau\}$, can be expressed in terms of the matrix geometric mean as

$$\left\{ (A, \tau) \in \mathbf{H}_{++}^n \times \mathbb{R} : \exists T \in \mathbf{H}^n \text{ s.t. } A \#_t I \succeq T, \frac{1}{t} \text{tr} [T - A] \geq \tau \right\}.$$

By rewriting $A \#_t I \succeq T$ using the SDP description of the matrix geometric mean (with $B = I$), we obtain a SDP description of S_t (when $t = p/q$) having $O(\log_2 q)$ LMIs of size at most $2n$.

- *The Tsallis relative entropy is defined for $t \in [0, 1]$ as (see [?] and also [?])*

$$S_t(A||B) := \frac{1}{t} \text{tr} [A - A^{1-t} B^t].$$

As noted in (??) the Tsallis relative entropy $S_t(A||B)$ converges to the quantum relative entropy $S(A||B) = \text{tr}[A(\log A - \log B)]$ when $t \rightarrow 0$. It is also known that convergence is from below, i.e., $S_t(A||B) \leq S(A||B)$ for any $t \in [0, 1]$ (see e.g., [?, Proposition 2.1]). By

choosing $K = I$ in Lieb's theorem we see that $S_t(A\|B)$ is jointly convex in (A, B) . Indeed the epigraph of $S_t(\cdot\|\cdot)$ can be expressed as

$$\{(A, B, \tau) \in \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \times \mathbb{R} : S_t(A\|B) \leq \tau\} = \left\{ (A, B, \tau) \in \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \times \mathbb{R} : \exists s \in \mathbb{R} \text{ s.t. } \operatorname{tr}[A^{1-t}B^t] \geq s, \frac{1}{t}[\operatorname{tr}(A) - s] \leq \tau \right\}.$$

By rewriting $\operatorname{tr}[A^{1-t}B^t] \geq s$ using the SDP description from Theorem ?? (with $K = I$), we obtain a SDP description of $S_t(\cdot\|\cdot)$ (with $t = p/q$) having $O(\log_2 q)$ LMIs of size at most $2n^2$.

3.2.2 The map $A \mapsto \operatorname{tr}[(K^*A^tK)^{1/t}]$

Let K be a fixed $n \times m$ matrix and consider the function $\Upsilon_t : \mathbf{H}_{++}^n \rightarrow \mathbb{R}$ defined by

$$\Upsilon_t(A) = \operatorname{tr}[(K^*A^tK)^{1/t}].$$

The following result is due to Carlen and Lieb [?] where they established the case $t \in [0, 2]$ (the same arguments were used to prove the case $t \in [-1, 0]$ in [?]; the case $t \in (0, 1]$ was first established by Epstein [?]).

Theorem 5. *If $t \in [1, 2]$ then Υ_t is convex on \mathbf{H}_{++}^n . If $t \in [-1, 1] \setminus \{0\}$ then Υ_t is concave on \mathbf{H}_{++}^n .*

Rather than working with Υ_t , it is slightly more natural to focus on the function $t\Upsilon_t$. Since t is a fixed parameter, this simply changes the signs of some expressions for the cases $t \in [-1, 0]$. It follows directly from Theorem ?? that $t\Upsilon_t$ is convex on \mathbf{H}_{++}^n for $t \in [-1, 0) \cup [1, 2]$, and concave on \mathbf{H}_{++}^n for $t \in (0, 1]$.

In this section we show how to give SDP formulations of $\operatorname{hyp}(t\Upsilon_t)$ for $t \in (0, 1]$ and $\operatorname{epi}(t\Upsilon_t)$ for $t \in [-1, 0) \cup [1, 2]$ by using our SDP formulations of Lieb's function in different regimes of the parameters. Our SDP formulations rely on variational expressions for $t\Upsilon_t$ (equations (??) and (??) to follow) established in [?] (see also [?]). We include a proof of these variational descriptions, for completeness, en route to our expressions for the hypograph/epigraph of $t\Upsilon_t$ in terms of Lieb's function (equations (??) and (??) to follow).

Lemma 1. *Let $A \in \mathbf{H}_{++}^n$ and $t \in [-1, 2] \setminus \{0\}$.*

- *If $t \in (0, 1]$ then*

$$t\Upsilon_t(A) = \max_{X \in \mathbf{H}_{++}^m} \operatorname{tr}[K^*A^tKX^{1-t}] - (1-t)\operatorname{tr}[X]. \quad (9)$$

Hence

$$\operatorname{hyp}(t\Upsilon_t) = \{(A, \tau) \in \mathbf{H}_{++}^n \times \mathbb{R} : \exists X \in \mathbf{H}_{++}^m \text{ s.t. } \operatorname{tr}[K^*A^tKX^{1-t}] - (1-t)\operatorname{tr}[X] \geq \tau\}. \quad (10)$$

- *If $t \in [-1, 0) \cup [1, 2]$ then*

$$t\Upsilon_t(A) = \min_{X \in \mathbf{H}_{++}^m} \operatorname{tr}[K^*A^tKX^{1-t}] - (1-t)\operatorname{tr}[X]. \quad (11)$$

Hence

$$\operatorname{epi}(t\Upsilon_t) = \{(A, \tau) \in \mathbf{H}_{++}^n \times \mathbb{R} : \exists X \in \mathbf{H}_{++}^m \text{ s.t. } \operatorname{tr}[K^*A^tKX^{1-t}] - (1-t)\operatorname{tr}[X] \leq \tau\}. \quad (12)$$

Proof. First observe that if $t \in [0, 1]$ and $y, z > 0$ then the arithmetic-mean geometric-mean inequality gives

$$ty \geq y^t z^{1-t} - (1-t)z \quad (13)$$

for all $y, z > 0$. If $t \in [1, 2]$ and $y, z > 0$ then the arithmetic-mean geometric-mean inequality gives $\frac{1}{t}y^t + \frac{t-1}{t}z^t \geq yz^{t-1}$. Rearranging gives

$$ty \leq y^t z^{1-t} - (1-t)z \quad (14)$$

for all $y, z > 0$. If $t \in [-1, 0]$ and $a, b > 0$ then $s = 1 - t \in [1, 2]$. Hence $sa + (1-s)b \leq a^s b^{1-s}$. Putting $y = b$ and $z = a$ we obtain that (??) also holds when $t \in [-1, 0]$ for all $y, z > 0$.

We can apply inequalities (??) and (??) to the eigenvalues of the positive definite commuting matrices $Y = (K^* A^t K)^{1/t} \otimes I$ and $Z = I \otimes \bar{X}$ (with $A \in \mathbf{H}_{++}^n$ and $X \in \mathbf{H}_{++}^m$). Doing so we see that if $t \in (0, 1]$ then

$$t(K^* A^t K)^{1/t} \otimes I \succeq K^* A^t K \otimes \bar{X}^{1-t} - (1-t)(I \otimes \bar{X})$$

for all $A \in \mathbf{H}_{++}^n$ and all $X \in \mathbf{H}_{++}^m$. Similarly if $t \in [-1, 0) \cup [1, 2]$ then

$$t(K^* A^t K)^{1/t} \otimes I \preceq K^* A^t K \otimes \bar{X}^{1-t} - (1-t)(I \otimes \bar{X})$$

for all $A \in \mathbf{H}_{++}^n$ and all $X \in \mathbf{H}_{++}^m$. If we apply the map $\mathbf{H}^{m^2} \ni M \mapsto \text{vec}(I)^* M \text{vec}(I)$ to both sides of these matrix inequalities and use identity (??) we get that for $t \in [0, 1]$

$$t\Upsilon_t(A) \geq \text{tr} [K^* A^t K X^{1-t}] - (1-t)\text{tr}[X]$$

for all $A \in \mathbf{H}_{++}^n$ and all $X \in \mathbf{H}_{++}^m$, and for $t \in [-1, 0) \cup [1, 2]$

$$t\Upsilon_t(A) \leq \text{tr} [K^* A^t K X^{1-t}] - (1-t)\text{tr}[X]$$

for all $A \in \mathbf{H}_{++}^n$ and all $X \in \mathbf{H}_{++}^m$. To ensure that the variational formulas (??) and (??) hold, one simply checks that putting $X = (K^* A K)^{1/t}$ gives equality in both cases. The descriptions of $\text{hyp}(t\Upsilon_t)$ for $t \in (0, 1]$ and $\text{epi}(t\Upsilon_t)$ for $t \in [-1, 0) \cup [1, 2]$ are direct consequences of (??) and (??) respectively. \square

When t is rational, each of the convex sets (??) and (??) can be expressed explicitly in terms of LMIs by using the SDP description of Lieb's function from Theorem ???. The following summarizes the size of these descriptions.

Theorem 6. *Let p, q be relatively prime integers such that $p/q \in [-1, 2] \setminus \{0\}$.*

- *If $t = p/q \in (0, 1]$ then $\text{hyp}(t\Upsilon_t)$ has a SDP description with at most $2\lceil \log_2(q) \rceil + 1$ LMIs of size $2mn \times 2mn$, one LMI of size $mn \times mn$, and one scalar inequality.*
- *If $t = p/q \in [-1, 0) \cup [1, 2]$ then $\text{epi}(t\Upsilon_t)$ has a SDP description with at most $2\lceil \log_2(q) \rceil + 2$ LMIs of size $2mn \times 2mn$, one LMI of size $mn \times mn$, and one scalar inequality.*

4 Numerical experiments

In this section we present some numerical results for the semidefinite programming representations given in this paper.

4.1 Maximum entropy problem

We consider maximum entropy optimization problems of the form

$$\begin{aligned} & \text{maximize} && S_t(\sum_{i=1}^s p_i M_i) \\ & \text{subject to} && p \geq 0, \sum_{i=1}^s p_i = 1 \end{aligned} \tag{15}$$

where M_1, \dots, M_s are fixed $n \times n$ positive semidefinite matrices of trace one, and $S_t(A) = \frac{1}{t} \text{tr}[A^{1-t} - A]$ is the Tsallis entropy considered in Remark ???. We used CVX [?, ?] to formulate the problem with the new function `tsallis_entr` available in our package [?]. The listing below shows the Matlab code used to solve (??) in the case $s = 3$:

```
% M1,M2,M3 are three positive semidefinite matrices
cvx_begin
    variable p(3);
    maximize (tsallis_entr(p(1)*M1+p(2)*M2+p(3)*M3,1/8));
    subject to
        p >= 0;
        sum(p) == 1;
cvx_end
```

Note that `tsallis_entr` is automatically recognized by CVX as being a concave function of its first argument (the matrix). The second argument to `tsallis_entr` is the parameter t which we set here to be $1/8$.

In Table ?? we present numerical results obtained with different values of n (matrix size) while fixing $s = 10$. The results were obtained with the solver SeDuMi [?], which is currently packaged with CVX.

n	optimal value	time (s)	n	optimal value	time (s)
10	2.6027	0.638	30	4.1619	2.048
	2.6051	0.501		4.1689	3.336
	2.6280	0.292		4.1767	2.038
	2.6092	0.250		4.1638	2.261
	2.6262	0.260		4.1784	2.040
20	3.5662	0.521	40	4.6191	8.763
	3.5706	0.519		4.6144	9.051
	3.5625	0.523		4.6172	9.272
	3.5665	0.516		4.6125	10.131
	3.5717	0.516		4.6131	10.498

Table 2: Results of numerical experiments for the maximum entropy problem (??) with $t = 1/8$. The matrices M_1, \dots, M_s were generated at random. Each cell (labeled by a value of n) corresponds to five different random trials. Note that the SDP representation of (??) with $t = 1/8$ consists of 3 semidefinite constraints of size $2n$ each.

4.2 Relative entropy of entanglement

We now consider another numerical illustration of our results to compute lower bounds on the so-called *relative entropy of entanglement* in quantum information theory which is used to measure

the distance of a given bipartite state ρ to the set of separable states. This quantity is in general intractable to compute [?] and we consider here a popular relaxation using the *positive partial transpose* (PPT) criterion [?]. This relaxation is defined in terms of the following optimization problem, where ρ is a fixed positive semidefinite matrix of trace one and τ is the optimization variable:

$$\begin{aligned} & \underset{\tau}{\text{minimize}} && S(\rho||\tau) \\ & \text{subject to} && \tau \succeq 0, \text{tr}[\tau] = 1 \\ & && \tau \in \text{PPT}. \end{aligned} \tag{16}$$

The constraint $\tau \in \text{PPT}$ is a linear matrix inequality constraint. We omit its precise meaning here. Note that the variable τ of the optimization problem (??) enters the second argument of the relative entropy S in the cost function. As such the cost function is *not* a matrix trace function of the form considered e.g., in [?, Theorem 3.1].

If we replace the cost function $S(\rho||\tau)$ in (??) by the Tsallis relative entropy $S_t(\rho||\tau)$ (for t rational) the resulting optimization problem can be expressed as an SDP using the formulation given in Remark ???. Furthermore since $S_t(\rho||\tau) \leq S(\rho||\tau)$ the optimal value we get is always a lower bound to (??). The function `tsallis_rel_entr` available in our package [?] can be used to formulate the resulting problem using CVX on Matlab. The code is shown below. (Note that the code uses the function `Tx` from the `quantinf` package [?] to implement the PPT constraint on τ .)

```
na = 2;
nb = 2;
% Generate a random positive semidefinite matrix rho of size na*nb of trace one
rho = randn(na*nb,na*nb); rho = rho*rho';
rho = rho/trace(rho);
cvx_begin
    variable tau(na*nb,na*nb) hermitian;
    minimize (tsallis_rel_entr(rho,tau,2^(-8)));
    subject to
        tau == hermitian_semidefinite(na*nb);
        trace(tau) == 1;
        % Positive partial transpose constraint
        Tx(tau,2,[na nb]) == hermitian_semidefinite(na*nb);
cvx_end
```

Note that CVX automatically recognizes `tsallis_rel_entr` as a convex function of its arguments ρ and τ . The third argument of `tsallis_rel_entr` specifies the value of t to use in the definition of Tsallis relative entropy. Here we use $t = 2^{-8}$.

We now present numerical experiments where we solve the optimization problem for random bipartite states ρ . We use the solver SeDuMi and compare our results with the tailored algorithm developed in [?, ?] based on a cutting-plane approach. Table ?? shows the results for different matrix sizes $n = n_A \times n_B$ (where n_A and n_B are the sizes of the subsystems). Note that the cutting-plane approach of [?, ?] returns an interval of length ϵ that is guaranteed to contain the optimal value of (??) (we chose $\epsilon = 10^{-3}$ in the experiments). We see that our method consistently gives lower bounds that are better than the cutting-plane approach of [?, ?] in a fraction of the time it takes.

$n = n_A \times n_B$	Our approach		Cutting-plane approach of [?, ?]	
	value	time (s)	value	time (s)
$4 = 2 \times 2$	0.0670	0.68 s	[0.0669,0.0678]	6.38 s
	0.0002	0.54 s	[0.0000,0.0010]	4.33 s
	0.0157	0.52 s	[0.0150,0.0160]	6.58 s
	0.0478	0.52 s	[0.0473,0.0480]	7.45 s
	0.0027	0.60 s	[0.0020,0.0030]	6.70 s
$6 = 3 \times 2$	0.0088	0.63 s	[0.0083,0.0093]	14.21 s
	0.0052	0.69 s	[0.0047,0.0057]	17.28 s
	0.0476	0.63 s	[0.0473,0.0483]	17.40 s
	0.0133	0.64 s	[0.0130,0.0140]	14.69 s
	0.0169	0.62 s	[0.0166,0.0174]	26.28 s
$9 = 3 \times 3$	0.0109	1.04 s	[0.0105,0.0115]	44.65 s
	0.0342	1.02 s	[0.0339,0.0349]	39.47 s
	0.0062	1.01 s	[0.0056,0.0066]	52.37 s
	0.0278	1.05 s	[0.0276,0.0286]	40.20 s
	0.0249	1.01 s	[0.0247,0.0257]	26.35 s

Table 3: Solving (??) for random choices of bipartite states ρ of size $n = n_A \times n_B$. In our approach we replace the cost function $S(\rho||\tau)$ by the Tsallis relative entropy $S_t(\rho||\tau)$ with $t = 2^{-8}$ and use the SDP formulations given in this paper. Note that $S_t(\rho||\tau) \leq S(\rho||\tau)$ for any ρ, τ and as such our approach always returns a lower bound on (??). We see that on all the matrices tested, our method is much faster than the cutting-plane approach of [?, ?] and gives better lower bounds. Note that the approach of [?, ?] returns an interval of length ϵ guaranteed to contain the optimal value of (??) (we set $\epsilon = 10^{-3}$ in the experiments).

5 Conclusion

We conclude by discussing the possibility of a SDP representation for a related jointly convex/concave function.

Sandwiched Rényi divergence The *sandwiched Rényi divergence* introduced in [?, ?] is defined as

$$(A, B) \mapsto \text{tr} \left[\left(A^{\frac{1-t}{2t}} B A^{\frac{1-t}{2t}} \right)^t \right]. \quad (17)$$

In [?] Frank and Lieb proved that (??) is jointly concave for $t \in [1/2, 1]$ and jointly convex for $t \geq 1$. Note that if A and B commute then (??) reduces to $\text{tr} [A^{1-t} B^t]$; however these two expressions are different for general noncommuting matrices A and B . The quantity (??) has found applications in quantum information theory, see e.g., [?]. In the case $t = 1/2$, the expression (??) is called the *fidelity* of A and B and is known to have the following semidefinite programming formulation [?, Section 3.2]:

$$\text{tr} \left[\left(A^{1/2} B A^{1/2} \right)^{1/2} \right] = \max_{Z \in \mathbb{C}^{n \times n}} \frac{1}{2} (\text{tr}[Z] + \text{tr}[Z^*]) : \begin{bmatrix} A & Z \\ Z^* & B \end{bmatrix} \succeq 0.$$

A natural question is:

Problem 1. *Find a semidefinite programming formulation for (??) for any $t \geq 1/2$ rational.*

Acknowledgments

Hamza Fawzi was supported in part by AFOSR FA9550-11-1-0305. James Saunderson was supported by NSF grant CCF-1409836.

Hamza Fawzi would like to thank Omar Fawzi for fruitful discussions that provided the motivation for this paper, and for useful comments on a previous version of this manuscript that helped improve the presentation.

A Construction for the matrix geometric mean

In this section we give an SDP description of the matrix geometric mean. Our construction heavily relies on the properties of the geometric mean which we review below.

A.1 Properties of the matrix geometric mean

For convenience, we first recall the definition of the t -weighted geometric mean $G_t : \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \rightarrow \mathbf{H}_{++}^n$:

$$G_t(A, B) = A\#_t B := A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^t A^{1/2}.$$

The following lemma summarizes important and well-known properties of the weighted geometric mean used in our construction.

Lemma 2. *Suppose $A, B \in \mathbf{H}_{++}^n$.*

- (i) *If X is an $n \times n$ invertible matrix and $t \in [0, 1]$ then $X(A\#_t B)X^* = (XAX^*)\#_t(XBX^*)$.*
- (ii) *(Monotonicity) If $A \succeq B \succeq 0$ and $C \succeq D \succeq 0$ and $t \in [0, 1]$ then $A\#_t C \succeq B\#_t D$.*
- (iii) *For any $s, t \in \mathbb{R}$*

$$A\#_t B = B\#_{1-t} A \tag{18}$$

$$A\#_s(A\#_t B) = A\#_{st} B \quad \text{and} \tag{19}$$

$$(A\#_t B)\#_s B = A\#_{s+t-st} B. \tag{20}$$

- (iv) *For any $s, t \in \mathbb{R}$, and any $X \in \mathbf{H}_{++}^n$,*

$$X\#_s A \succeq X\#_t B \iff X\#_{-s} A \preceq X\#_{-t} B \iff A\#_{s+1} X \preceq B\#_{t+1} X. \tag{21}$$

Proof. Properties (i)-(iii) are well-known, see e.g., [?, Lemma 2.1]. We only include a proof of (iv). By first multiplying on the left and right by $X^{-1/2}$, then inverting both sides, then multiplying on the left and right by $X^{1/2}$ we have that

$$\begin{aligned} X\#_s A \succeq X\#_t B &\iff (X^{-1/2} A X^{-1/2})^s \succeq (X^{-1/2} B X^{-1/2})^t \\ &\iff (X^{-1/2} B X^{-1/2})^{-t} \succeq (X^{-1/2} A X^{-1/2})^{-s} \\ &\iff X\#_{-t} B \succeq X\#_{-s} A. \end{aligned}$$

Finally it follows from (21) that $X\#_{-t} B \succeq X\#_{-s} A$ is equivalent to $B\#_{t+1} X \succeq A\#_{s+1} X$. \square

The properties given in Lemma ?? can be directly translated to relationships between the hypographs/epigraphs of the matrix geometric mean. Recall that hyp_t and epi_t are defined as:

$$\text{hyp}_t := \{(A, B, T) \in \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \times \mathbf{H}^n : A\#_t B \succeq T\}$$

for $t \in [0, 1]$, and

$$\text{epi}_t := \{(A, B, T) \in \mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \times \mathbf{H}^n : A\#_t B \preceq T\}$$

for $t \in [-1, 0] \cup [1, 2]$.

Lemma 3. *The following holds:*

(i) *If $t \in [0, 1]$ then*

$$\text{hyp}_{1-t} = \{(A, B, T) : (B, A, T) \in \text{hyp}_t\}. \quad (22)$$

(ii) *If $t \in [-1, 0] \cup [1, 2]$ then*

$$\text{epi}_{1-t} = \{(A, B, T) : (B, A, T) \in \text{epi}_t\}. \quad (23)$$

(iii) *For any $s, t \in [0, 1]$ we have*

$$\text{hyp}_{st} = \{(A, B, T) : \exists Z \text{ s.t. } (A, B, Z) \in \text{hyp}_t, (A, Z, T) \in \text{hyp}_s\}. \quad (24)$$

(iv) *For any $t \in [0, 1]$,*

$$\text{epi}_{-t} = \left\{ (A, B, T) : \exists S \text{ s.t. } (A, B, S) \in \text{hyp}_t, \begin{bmatrix} T & A \\ A & S \end{bmatrix} \succeq 0 \right\}. \quad (25)$$

Proof. The proof of this lemma is a direct consequence of the properties of the matrix geometric mean stated in Lemma ?. We include a proof of (iv), the other items can be proved in a similar way. First observe that for any A, S positive definite we have $A\#_{-1}S = AS^{-1}A$ thus by the Schur complement lemma we have

$$\begin{bmatrix} T & A \\ A & S \end{bmatrix} \succeq 0 \iff A\#_{-1}S \preceq T. \quad (26)$$

To prove (??), suppose $(A, B, T) \in \text{epi}_{-t}$, i.e., $A\#_{-t}B \preceq T$. Let $S = A\#_tB$. Then $A\#_{-1}S = A\#_{-1}(A\#_tB) = A\#_{-t}B \preceq T$. So, by (??) we have

$$\begin{bmatrix} T & A \\ A & S \end{bmatrix} \succeq 0$$

as desired. For the reverse inclusion, suppose there exists $S \in \mathbf{H}_{++}^n$ such that $A\#_tB \succeq S$ and $S \succeq A\#_{-1}T$. Then by (??) of Lemma ?? we have that

$$A\#_tB \succeq A\#_{-1}T \implies A\#_{-t}B \preceq A\#_1T = T$$

Hence $(A, B, T) \in \text{epi}_{-t}$ as required. \square

A.2 Semidefinite representation of the matrix geometric mean

The remainder of this section is devoted to getting an SDP description of the matrix geometric mean (Theorem ??). Our construction is recursive in nature and heavily relies on the properties of the matrix geometric mean given above. Section ?? treats the base case $t = 1/2$, Section ?? treats the case where $t = p/q \in [0, 1]$ and q is a power of two and Section ?? treats the case where $t = p/q \in [1/2, 1]$ and p is a power of two. In Section ?? we combine these cases together and complete the proof of Theorem ??.

A.2.1 Base case $t = 1/2$

The following well-known lemma gives an SDP description of $\text{hyp}_{1/2}$. It forms the base case of our construction and we thus include a proof for completeness.

Lemma 4. *We have*

$$\text{hyp}_{1/2} = \left\{ (A, B, T) : \begin{bmatrix} A & T \\ T & B \end{bmatrix} \succeq 0 \right\}. \quad (27)$$

Proof. Using the Schur complement lemma and monotonicity of the matrix square root we have:

$$\begin{aligned} \begin{bmatrix} A & T \\ T & B \end{bmatrix} \succeq 0 &\iff B \succeq TA^{-1}T \iff A^{-1/2}BA^{-1/2} \succeq (A^{-1/2}TA^{-1/2})^2 \\ &\iff (A^{-1/2}BA^{-1/2})^{1/2} \succeq A^{-1/2}TA^{-1/2} \\ &\iff A\#_{1/2}B \succeq T. \end{aligned}$$

□

Remark 3. *In the description (??) (A, B, T) is understood to be restricted to $\mathbf{H}_{++}^n \times \mathbf{H}_{++}^n \times \mathbf{H}^n$. This will be implicit in our subsequent SDP descriptions of hyp_t and epi_t .*

A.2.2 Denominator is a power of two

In this section we give an SDP description of $\text{hyp}_{p/2^\ell}$ when p is odd and ℓ is a positive integer such that $p < 2^\ell$. Observe that $p/2^\ell$ has a binary expansion of length ℓ as $(0.m_\ell m_{\ell-1} \cdots m_1)_2$ where $m_1 = 1$ (because p is odd) and $m_i \in \{0, 1\}$ for $i = 2, \dots, \ell$. The construction can be expressed explicitly in terms of this binary expansion (see Proposition ?? to follow) by repeatedly applying the following result.

Lemma 5. *If $t \in [0, 1/2]$ then*

$$\text{hyp}_t = \left\{ (A, B, T) : \exists Z \in \mathbf{H}^n \text{ s.t. } (A, B, Z) \in \text{hyp}_{2t}, \begin{bmatrix} A & T \\ T & Z \end{bmatrix} \succeq 0 \right\}. \quad (28)$$

If $t \in [1/2, 1]$ then

$$\text{hyp}_t = \left\{ (A, B, T) : \exists Z \in \mathbf{H}^n \text{ s.t. } (A, B, Z) \in \text{hyp}_{2t-1}, \begin{bmatrix} B & T \\ T & Z \end{bmatrix} \succeq 0 \right\}. \quad (29)$$

Proof. If $t \in [0, 1/2]$ then $2t \in [0, 1]$. Hence $\text{hyp}_t = \text{hyp}_{(2t)(1/2)}$ can be expressed in terms of hyp_{2t} and $\text{hyp}_{1/2}$ using (??). The condition $(A, Z, T) \in \text{hyp}_{1/2}$ can then be rewritten as an LMI using Lemma ?? and this yields (??).

If $t \in [1/2, 1]$ then $1 - t \in [0, 1/2]$, so we use (??) together with the relationship between hyp_t and hyp_{1-t} . Applying (??) we have $\text{hyp}_t = \{(A, B, T) : (B, A, T) \in \text{hyp}_{1-t}\}$. Then, because $1 - t \in [0, 1/2]$, we can apply (??) giving

$$\text{hyp}_t = \left\{ (A, B, T) : \exists Z \in \mathbf{H}^n \text{ s.t. } (B, A, Z) \in \text{hyp}_{2(1-t)}, \begin{bmatrix} B & T \\ T & Z \end{bmatrix} \succeq 0 \right\}.$$

Finally using (??) again we see that $(B, A, Z) \in \text{hyp}_{2(1-t)}$ if and only if $(A, B, Z) \in \text{hyp}_{1-2(1-t)} = \text{hyp}_{2t-1}$, completing the proof. \square

Suppose p is odd and ℓ is a positive integer with $p < 2^\ell$. Then by repeatedly applying Lemma ??, and using the expression for $\text{hyp}_{1/2}$ in Lemma ?? as a base case, we can describe $\text{hyp}_{p/2^\ell}$ in terms of ℓ LMIs, each of size $2n \times 2n$. Proposition ??, to follow, explicitly gives this semidefinite formulation of $\text{hyp}_{p/2^\ell}$. It is most naturally expressed in terms of the binary expansion of $p/2^\ell$. Note that if $m = 0$ then $A \#_m B = A$ and if $m = 1$ then $A \#_m B = B$. In particular, in each case the expression is actually linear in A and B .

Proposition 1. *Suppose p is an odd positive integer and ℓ is a positive integer such that $p < 2^\ell$. Let $p/2^\ell = (0.m_\ell m_{\ell-1} \cdots m_1)_2$ be the binary expansion of $p/2^\ell$ where $m_1 = 1$ and $m_i \in \{0, 1\}$ for $i = 2, \dots, \ell$. Then*

$$\text{hyp}_{p/2^\ell} = \left\{ (A, B, Z_\ell) : \exists Z_1, \dots, Z_{\ell-1} \in \mathbf{H}^n \text{ s.t. } \begin{bmatrix} A \#_{m_i} B & Z_i \\ Z_i & Z_{i-1} \end{bmatrix} \succeq 0 \text{ for } i = 2, 3, \dots, \ell, \right. \\ \left. \begin{bmatrix} A & Z_1 \\ Z_1 & B \end{bmatrix} \succeq 0 \right\}. \quad (30)$$

Hence $\text{hyp}_{p/2^\ell}$ has an SDP description with ℓ LMIs, each of size $2n \times 2n$.

Proof. The proof of correctness is by induction on ℓ . If $\ell = 1$ then necessarily $p = 1$. In this case the description of $\text{hyp}_{1/2}$ in (??) matches that from Lemma ?? and so is correct. Now suppose $\ell > 1$, and let $p/2^\ell = (0.m_\ell m_{\ell-1} \cdots m_1)_2$ be the binary expansion of $p/2^\ell$. Then $p/2^\ell > 1/2$ if and only if $m_\ell = 1$ and $p/2^\ell < 1/2$ if and only if $m_\ell = 0$.

Hence if $m_\ell = 0$ then we apply (??) to express $\text{hyp}_{p/2^\ell}$ in terms of $\text{hyp}_{p/2^{\ell-1}}$ as

$$\text{hyp}_{p/2^\ell} = \left\{ (A, B, Z_\ell) : \exists Z_{\ell-1} \in \mathbf{H}^n \text{ s.t. } (A, B, Z_{\ell-1}) \in \text{hyp}_{p/2^{\ell-1}}, \begin{bmatrix} A \#_{m_\ell} B & Z_\ell \\ Z_\ell & Z_{\ell-1} \end{bmatrix} \succeq 0 \right\}.$$

Note that in this case the binary expansion of $p/2^{\ell-1}$ is $(0.m_{\ell-1} \cdots m_1)_2$, so we can apply the induction hypothesis to see that (??) is correct.

Similarly if $m_\ell = 1$ then we apply (??) to express $\text{hyp}_{p/2^\ell}$ in terms of $\text{hyp}_{p/2^{\ell-1}-1}$ as

$$\text{hyp}_{p/2^\ell} = \left\{ (A, B, Z_\ell) : \exists Z_{\ell-1} \in \mathbf{H}^n \text{ s.t. } (A, B, Z_{\ell-1}) \in \text{hyp}_{p/2^{\ell-1}-1}, \begin{bmatrix} A \#_{m_\ell} B & Z_\ell \\ Z_\ell & Z_{\ell-1} \end{bmatrix} \succeq 0 \right\}.$$

Note that in this case the binary expansion of $p/2^{\ell-1} - 1$ is $(0.m_{\ell-1} \cdots m_1)_2$, so we can apply the induction hypothesis to see that (??) is correct. \square

We conclude with an example in which the denominator is a power of two.

Example 1 (SDP representation of $\text{hyp}_{5/8}$). Let $p = 5$ and $\ell = 3$ so that $p/2^\ell = 5/8 = (0.101)_2$. Consider constructing a SDP representation of $\text{hyp}_{5/8}$. We have that $m_1 = m_3 = 1$ and $m_2 = 0$ so that $A\#_{m_1}B = B$ and $A\#_{m_2}B = A$. Applying Proposition ?? gives

$$\text{hyp}_{5/8} = \left\{ (A, B, Z_3) : \exists Z_1, Z_2 \text{ s.t. } \begin{bmatrix} B & Z_3 \\ Z_3 & Z_2 \end{bmatrix} \succeq 0, \begin{bmatrix} A & Z_2 \\ Z_2 & Z_1 \end{bmatrix} \succeq 0, \begin{bmatrix} A & Z_1 \\ Z_1 & B \end{bmatrix} \succeq 0 \right\}$$

using $\ell = 3$ LMIs of size $2n \times 2n$.

A.2.3 Numerator is a power of two

In this section we show how to construct an SDP representation of hyp_t when t has a numerator that is a power of two and $t \in [1/2, 1]$. We do this by relating hyp_t and $\text{hyp}_{\frac{2t-1}{t}}$ (see Lemma ?? to follow). This is useful because if $t = 2^\ell/q$ with $t \in [1/2, 1]$, then $\frac{2t-1}{t} = \frac{2^{\ell+1}-q}{2^\ell}$ has a denominator that is a power of two. Hence we can relate $\text{hyp}_{2^\ell/q}$ with $\text{hyp}_{\frac{2^{\ell+1}-q}{2^\ell}}$, an SDP description of which we can obtain from Proposition ??.

Lemma 6. *If $t \in [1/2, 1]$ then*

$$\text{hyp}_t = \left\{ (A, B, T) : \exists Z, W \in \mathbf{H}_{++}^n \text{ s.t. } (A, W, Z) \in \text{hyp}_{\frac{2t-1}{t}}, \begin{bmatrix} Z & W \\ W & B \end{bmatrix} \succeq 0, W \succeq T \right\}. \quad (31)$$

Proof. We first prove \subseteq . Suppose $A\#_t B \succeq T$. Then let $Z = A\#_{2t-1} B$ and $W = A\#_t B$. It is easy to see that the conditions on the right-hand side of (??) are satisfied. Indeed first we have

$$A\#_{\frac{2t-1}{t}} W = A\#_{\frac{2t-1}{t}} (A\#_t B) = A\#_{2t-1} B = Z$$

and this shows that $(A, W, Z) \in \text{hyp}_{\frac{2t-1}{t}}$. Second, using Property (??) and Lemma ?? we have,

$$Z\#_{1/2} B = (A\#_{2t-1} B)\#_{1/2} B = A\#_t B = W \quad \text{which implies that } \begin{bmatrix} Z & W \\ W & B \end{bmatrix} \succeq 0.$$

Finally we have that $W = A\#_t B \succeq T$ by assumption.

We now prove \supseteq . Suppose there exist $Z, W \in \mathbf{H}_{++}^n$ such that $A\#_{\frac{2t-1}{t}} W \succeq Z$ and $W\#_{-1} B \preceq Z$ and $W \succeq T$. Then since $1 - \frac{2t-1}{t} = \frac{1}{t} - 1$ we have that $W\#_{1/t-1} A \succeq Z$. Then

$$W\#_{1/t-1} A \succeq Z \succeq W\#_{-1} B.$$

Applying (??) from Lemma ?? it follows that

$$B = B\#_{-1+1} W \succeq A\#_{1/t-1+1} W = A\#_{1/t} W.$$

Then since $t \in [1/2, 1]$ and G_t is monotone for $t \in [0, 1]$, applying $G_t(A, \cdot)$ to both sides gives

$$A\#_t B \succeq A\#_t (A\#_{1/t} W) = A\#_1 W = W \succeq T$$

as required. \square

Note that if $t = 2^\ell/q$ then $\frac{2t-1}{t} = \frac{2^{\ell+1}-q}{2^\ell}$ is a dyadic number and so $\text{hyp}_{\frac{2t-1}{t}}$ has a SDP description from the previous section (Proposition ??).

Proposition 2. Assume ℓ, q are integers such that $\frac{2^\ell}{q} \in [1/2, 1]$. Then

$$\text{hyp}_{2^\ell/q} = \left\{ (A, B, T) : \exists Z, W \text{ s.t. } (A, W, Z) \in \text{hyp}_{\frac{2^{\ell+1}-q}{2^\ell}}, \begin{bmatrix} Z & W \\ W & B \end{bmatrix} \succeq 0, W \succeq T \right\}. \quad (32)$$

Hence $\text{hyp}_{2^\ell/q}$ has a SDP representation using $\ell + 1$ LMIs of size $2n \times 2n$ and one LMI of size $n \times n$.

Proof. The SDP description follows directly from Lemma ?? with $t = \frac{2^\ell}{q}$. Since $\text{hyp}_{\frac{2^{\ell+1}-q}{2^\ell}}$ has a SDP description with ℓ LMIs of size $2n \times 2n$ (cf. Proposition ??) the conclusion about the size of the description (??) holds. \square

We conclude with an example in which the numerator is a power of two.

Example 2 (SDP representation of $\text{hyp}_{8/13}$). Let $q = 13$ and $\ell = 3$ so that $2^\ell/q = 8/13$. Note that $8/13 \in [1/2, 1]$. Consider constructing an SDP description of $\text{hyp}_{8/13}$. We have that $(2^{\ell+1}-q)/2^\ell = 3/8 = (0.011)_2$. Hence, by Proposition ??,

$$\text{hyp}_{8/13} = \left\{ (A, B, T) : \exists Z_3, W \text{ s.t. } (A, W, Z_3) \in \text{hyp}_{3/8}, \begin{bmatrix} Z_3 & W \\ W & B \end{bmatrix} \succeq 0, W \succeq T \right\}.$$

Using Proposition ?? to obtain a semidefinite description of $\text{hyp}_{3/8}$ gives

$$\text{hyp}_{8/13} = \left\{ (A, B, T) : \exists Z_3, W, Z_1, Z_2 \text{ s.t. } \begin{bmatrix} A & Z_3 \\ Z_3 & Z_2 \end{bmatrix} \succeq 0, \begin{bmatrix} W & Z_2 \\ Z_2 & Z_1 \end{bmatrix} \succeq 0, \begin{bmatrix} W & Z_1 \\ Z_1 & A \end{bmatrix} \succeq 0, \begin{bmatrix} Z_3 & W \\ W & B \end{bmatrix} \succeq 0, W \succeq T \right\},$$

a SDP representation of $\text{hyp}_{8/13}$ using $\ell + 1 = 4$ LMIs of size $2n \times 2n$ and one LMI of size $n \times n$.

A.2.4 Putting everything together and summary of construction

We now complete the proof of Theorem ??.

Proof of Theorem ??. First observe that, using relations established in Lemma ??, we only need to consider the case $p/q \in [0, 1/2]$: indeed if we have an SDP representation of hyp_t for $t \in [0, 1/2]$ then we can use the relationship between hyp_{1-t} and hyp_t in (??) to get an SDP representation for hyp_t in the range $t \in [1/2, 1]$ with no additional LMIs. Then using the relationship (??) between epi_{-t} and hyp_t we can get an SDP representation of epi_t for $t \in [-1, 0]$ with the addition of a single $2n \times 2n$ LMI. Finally using again the relationship (??) between epi_{1-t} and epi_t we get an SDP representation for epi_t where $t \in [1, 2]$.

It thus remains to prove the case where t is an arbitrary rational in $[0, 1/2]$. We show how to do this using the results from the two previous sections. If $t = p/q \in [0, 1/2]$ we decompose t as $t = (p/2^\ell) \cdot (2^\ell/q)$ where $\ell = \lfloor \log_2(q) \rfloor$. By applying Propositions ?? and ?? to construct respectively $\text{hyp}_{p/2^\ell}$ and $\text{hyp}_{2^\ell/q}$ and appealing to (??) we get an SDP description of hyp_t (note that $2^\ell/q \in [1/2, 1]$ since $\ell = \lfloor \log_2(q) \rfloor$ and so Proposition ?? applies to get an SDP description of $\text{hyp}_{2^\ell/q}$).

To see that our SDP representation has the right size, the SDP representation of $\text{hyp}_{p/2^\ell}$ uses at most ℓ LMIs of size $2n \times 2n$ and the SDP representation of $\text{hyp}_{2^\ell/q}$ uses at most $\ell + 1$ LMIs of size $2n \times 2n$ and one LMI of size $n \times n$. Hence our description has at most $2\ell + 1 = 2\lfloor \log_2(q) \rfloor + 1$ LMIs of size $2n \times 2n$ and one LMI of size $n \times n$. The size of the SDP representation for the epigraph case $t \in [-1, 0] \cup [1, 2]$ requires an additional $2n \times 2n$ LMI which comes from identity (??). \square

Table ?? summarizes our SDP construction of the hypograph/epigraph of the matrix geometric mean for arbitrary rationals $t = p/q \in [-1, 2]$.

Semidefinite representation of hyp_t for $t = p/q \in [0, 1]$

- (i) If q is a power of two
Use construction in Proposition ??.
- (ii) If $t \in [1/2, 1]$ and p is a power of two
Use Proposition ?? which expresses hyp_t in terms of the hypograph of a dyadic number, then use (i).
- (iii) If t is any rational in $[0, 1/2]$
Express t as $t = (p/2^\ell) \cdot (2^\ell/q)$ where $q = \lfloor \log_2(q) \rfloor$. Use (i) and (ii) to construct $\text{hyp}_{p/2^\ell}$ and $\text{hyp}_{2^\ell/q}$ and combine them using (??) to get hyp_t .
- (iv) If t is any rational in $[1/2, 1]$
Use relationship (??) between hyp_t and hyp_{1-t} then apply (iii).

Semidefinite representation of epi_t for $t = p/q \in [-1, 0] \cup [1, 2]$

- (i) If $t \in [-1, 0]$
Use (??) to express epi_t in terms of hyp_{-t} and apply box above.
- (ii) If $t \in [1, 2]$
Use relationship (??) between epi_t and epi_{1-t} then apply (i).

Table 4: Semidefinite representation of the matrix geometric mean (Theorem ??).