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

James Saunderson

Electrical Engineering, University of Washington

Feb. 3, 2016

Joint work with Hamza Fawzi (MIT)

"How to optimize with quantum relative entropy?" (O. Fawzi)

Quantum relative entropy

$$S(A||B) = \mathrm{tr}[A \log(A) - A \log(B)] \quad \text{for } A, B \succ 0$$

Why?

- compute classical-quantum channel capacities (Sutter, Sutter, Esfahani, Renner 2014)
- test conjectures about properties of $S(\cdot||\cdot)$

"How to optimize with quantum relative entropy?" (O. Fawzi)

Quantum relative entropy is jointly convex in $A, B \succ 0$

$$S(A||B) = \mathrm{tr}[A\log(A) - A\log(B)] \quad \text{for } A, B \succ 0$$

Why?

- compute classical-quantum channel capacities (Sutter, Sutter, Esfahani, Renner 2014)
- test conjectures about properties of $S(\cdot||\cdot)$

"How to optimize with quantum relative entropy?" (O. Fawzi)

Quantum relative entropy is jointly convex in $A, B \succ 0$

$$S(A||B) = \mathrm{tr}[A\log(A) - A\log(B)] \quad \text{for } A, B \succ 0$$

Why?

- compute classical-quantum channel capacities (Sutter, Sutter, Esfahani, Renner 2014)
- test conjectures about properties of $S(\cdot||\cdot)$

“How to optimize with quantum relative entropy?” (O. Fawzi)

Quantum relative entropy is jointly convex in $A, B \succ 0$

$$S(A||B) = \mathrm{tr}[A\log(A) - A\log(B)] \quad \text{for } A, B \succ 0$$

Why?

- compute classical-quantum channel capacities (Sutter, Sutter, Esfahani, Renner 2014)
- test conjectures about properties of $S(\cdot||\cdot)$

Unsatisfactory answer: use (sub)gradient methods

"Can we use semidefinite optimization?" (H. Fawzi)

Does *epigraph*

$$\mathrm{epi}(S) = \{(A, B, t) : \mathrm{tr}[A \log(A) - A \log(B)] \leq t\}$$

have PSD lift?

“Can we use semidefinite optimization?” (H. Fawzi)

Does *epigraph*

$$\mathrm{epi}(S) = \{(A, B, t) : \mathrm{tr}[A \log(A) - A \log(B)] \leq t\}$$

have PSD lift?

Obvious answer: No!

"Can we use semidefinite optimization?" (H. Fawzi)

Does *epigraph*

$$\mathrm{epi}(S) = \{(A, B, t) : \mathrm{tr}[A \log(A) - A \log(B)] \leq t\}$$

have PSD lift?

Obvious answer: No!

Motivating questions:

- Does (compact base of) epi(S) have an approximate PSD lift?
- Is there a lift with size depending on $\log(1/\epsilon)$?

One approach: Lieb $\to$ Tsallis $\to$ $S(\cdot||\cdot)$

Theorem (Lieb 1973)

If $K \in \mathbb{C}^{m\times n}$ is fixed and $t \in [0,1]$ then

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

is jointly concave for positive definite A, B.

One approach: Lieb $\to$ Tsallis $\to$ $S(\cdot||\cdot)$

Theorem (Lieb 1973)

If $K \in \mathbb{C}^{m\times n}$ is fixed and $t \in [0,1]$ then

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

is jointly concave for positive definite A, B.

Convexity of $S(A||B) = \mathrm{tr}[A\log(A) - A\log(B)]$ follows from

$$S(A||B) = \lim_{t\to 0} \mathrm{tr}\left[A^{1-t}\left(\frac{A^t - I}{t} - \frac{B^t - I}{t}\right)\right]$$
$$= \lim_{t\to 0} \frac{1}{t}\mathrm{tr}\left[A - A^{1-t}B^t\right]$$

One approach: Lieb $\to$ Tsallis $\to$ $S(\cdot||\cdot)$

Theorem (Lieb 1973)

If $K \in \mathbb{C}^{m\times n}$ is fixed and $t \in [0,1]$ then

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

is jointly concave for positive definite A, B.

Central question:

Does epigraph of Lieb's function have small PSD lift (for $t \in \mathbb{Q}$)?

Convexity of $S(A||B) = \mathrm{tr}[A\log(A) - A\log(B)]$ follows from

$$S(A||B) = \lim_{t\to 0} \mathrm{tr}\left[A^{1-t}\left(\frac{A^t - I}{t} - \frac{B^t - I}{t}\right)\right]$$
$$= \lim_{t\to 0} \frac{1}{t}\mathrm{tr}\left[A - A^{1-t}B^t\right]$$

Outline

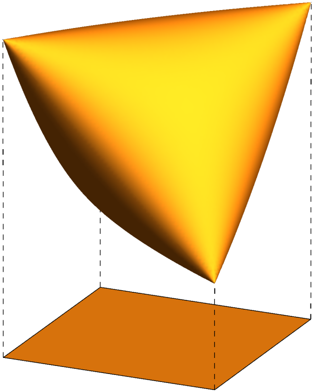
- Preliminaries
- Weighted matrix geometric mean
- Lieb-Ando function (and beyond)
- Back to approximating quantum relative entropy

PSD lifts

A convex set $C \subseteq \mathbb{R}^n$ has a PSD lift of size d if

$$C = \left\{ x \in \mathbb{R}^n : \quad \exists y \in \mathbb{R}^m \text{ s.t. } A_0 + \sum_{i=1}^n A_i x_i + \sum_{j=1}^m B_j y_j \succeq 0 \right\}$$

for symmetric $d \times d$ matrices $A_0, \ldots, A_n, B_1, \ldots, B_m$



Matrix concave/convex functions

Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex set.

Definition: A function $f : \mathcal{C} \to \mathcal{S}^k$ is *matrix convex* if for all $\lambda \in [0, 1]$ and all $A, B \in \mathcal{C}$

$$f(\lambda A + (1 - \lambda)B) \preceq \lambda f(A) + (1 - \lambda)f(B)$$

- For matrix convex functions $f : \mathcal{C} \to \mathcal{S}^k$ want PSD lift of $\mathcal{S}^k_+$-*epigraph*:

$$\text{epi}(f) = \{(x, T) \in \mathcal{C} \times \mathcal{S}^k : f(x) \preceq T\}$$

- For matrix concave functions $f : \mathcal{C} \to \mathcal{S}^k$ want PSD lift of $\mathcal{S}^k_+$-*hypograph*:

$$\text{hyp}(f) = \{(x, T) \in \mathcal{C} \times \mathcal{S}^k : f(x) \succeq T\}$$

Scalar weighted geometric mean

Weighted geometric mean:

$$(a, b) \mapsto a^{1-t} b^t$$

for $t \in \mathbb{R}$ and $a, b > 0$

Concavity/convexity properties:

- If $t \in [0, 1]$ jointly concave in $a, b > 0$
- If $t \in [-1, 0] \cup [1, 2]$ jointly convex in $a, b > 0$

Scalar weighted geometric mean

Weighted geometric mean:

$$(a, b) \mapsto a^{1-t} b^t$$

for $t \in \mathbb{R}$ and $a, b > 0$

Concavity/convexity properties:

- If $t \in [0, 1]$ jointly concave in $a, b > 0$
- If $t \in [-1, 0] \cup [1, 2]$ jointly convex in $a, b > 0$

PSD lift [Ben-Tal and Nemirovski (2001)]
If $p/q \in [0, 1]$ is rational,

$$\text{hyp}_t := \{(a, b, \tau) : a \geq 0,\ b \geq 0,\ a^{1-p/q} b^{p/q} \geq \tau\}$$

has PSD (actually SOCP) lift of size $O(\log(q))$

(Similar result for $t \in [-1, 0] \cup [1, 2]$ and epigraph)

Scalar weighted geometric mean

Weighted geometric mean:

$$(a, b) \mapsto a^{1-t} b^t = a(b/a)^t = a^{1-t} \otimes b^t$$

for $t \in \mathbb{R}$ and $a, b > 0$

Concavity/convexity properties:

- If $t \in [0, 1]$ jointly concave in $a, b > 0$
- If $t \in [-1, 0] \cup [1, 2]$ jointly convex in $a, b > 0$

PSD lift [Ben-Tal and Nemirovski (2001)]
If $p/q \in [0, 1]$ is rational,

$$\text{hyp}_t := \{(a, b, \tau) : a \geq 0,\ b \geq 0,\ a^{1-p/q} b^{p/q} \geq \tau\}$$

has PSD (actually SOCP) lift of size $O(\log(q))$

(Similar result for $t \in [-1, 0] \cup [1, 2]$ and epigraph)

Matrix weighted geometric mean

If A and B positive definite and $t \in \mathbb{R}$ define

$$A \#_t B := A^{1/2}(A^{-1/2} B A^{-1/2})^t A^{1/2}$$

If A and B commute then $A \#_t B = A^{1-t} B^t$

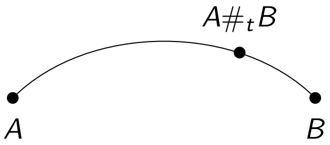
Matrix weighted geometric mean

If A and B positive definite and $t \in \mathbb{R}$ define

$$A \#_t B := A^{1/2}(A^{-1/2} B A^{-1/2})^t A^{1/2}$$

If A and B commute then $A \#_t B = A^{1-t} B^t$

Geometric interpretation



If $s, t \in \mathbb{R}$

- $t \mapsto A \#_t B$ parameterizes geodesic through A and B
- $A \#_t B = B \#_{1-t} A$
- $A \#_s (A \#_t B) = A \#_{st} B$

Concavity properties

If $t \in [0,1]$ then $(A,B) \mapsto A\#_t B$ is matrix concave

Proof idea (Bhatia) For $t \in (0,1)$

$$A\#_t B = \int_0^\infty (\lambda A^{-1} + B^{-1})^{-1} \lambda^{t-1} \frac{\sin(\pi t)}{\pi}\, d\lambda$$

If $t \in [-1,0] \cup [1,2]$ then $(A,B) \mapsto A\#_t B$ is matrix convex

Semidefinite descriptions of $A\#_t B$

Theorem

- If $t = p/q \in [0,1]$ is rational then

$$\mathrm{hyp}_{p/q} = \{(A, B, T) : A \succeq 0,\ B \succeq 0,\ A\#_{p/q}B \succeq T\}$$

has a PSD lift of size at most $(4\lfloor \log_2(q) \rfloor + 3)n$.

Semidefinite descriptions of $A\#_t B$

Theorem

- If $t = p/q \in [0,1]$ is rational then

$$\mathrm{hyp}_{p/q} = \{(A,B,T) : A \succeq 0,\ B \succeq 0,\ A\#_{p/q}B \succeq T\}$$

has a PSD lift of size at most $(4\lfloor \log_2(q) \rfloor + 3)n$.

- If $t = p/q \in [-1,0] \cup [1,2]$ is rational then

$$\mathrm{epi}_{p/q} = \{(A,B,T) : A \succeq 0,\ B \succeq 0,\ A\#_{p/q}B \preceq T\}$$

has a PSD lift of size at most $(4\lfloor \log_2(q) \rfloor + 5)n$.

Semidefinite descriptions of $A\#_t B$

Theorem

- If $t = p/q \in [0, 1]$ is rational then

$$\text{hyp}_{p/q} = \{(A, B, T) : A \succeq 0,\ B \succeq 0,\ A\#_{p/q}B \succeq T\}$$

  has a PSD lift of size at most $(4\lfloor \log_2(q) \rfloor + 3)n$.

- If $t = p/q \in [-1, 0] \cup [1, 2]$ is rational then

$$\text{epi}_{p/q} = \{(A, B, T) : A \succeq 0,\ B \succeq 0,\ A\#_{p/q}B \preceq T\}$$

  has a PSD lift of size at most $(4\lfloor \log_2(q) \rfloor + 5)n$.

- Concave case due to Sagnol (2013), not very explicit
- We give explicit lifts—CVX code is available

Flavor of construction

Dyadic rational case ($0 \leq p \leq q = 2^{\ell}$): Construction comes from

- PSD lift for $p/q = 1/2$ (base case)
- Digits of binary expansion of $p/2^{\ell}$

Flavor of construction

Dyadic rational case ($0 \le p \le q = 2^\ell$): Construction comes from

- PSD lift for $p/q = 1/2$ (base case)
- Digits of binary expansion of $p/2^\ell$

Example: $5/8$ in binary is $(0.101)_2$

$$A\#_{5/8}B = A\#_{(0.101)_2}B = B\#_{1/2}(A\#_{1/2}(B\#_{1/2}A))$$

Leads to (free) PSD lift:

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

Flavor of construction

Dyadic rational case ($0 \leq p \leq q = 2^\ell$): Construction comes from

- PSD lift for $p/q = 1/2$ (base case)
- Digits of binary expansion of $p/2^\ell$

Example: $5/8$ in binary is $(0.101)_2$

$$A\#_{5/8}B = A\#_{(0.101)_2}B = B\#_{1/2}(A\#_{1/2}(B\#_{1/2}A))$$

Leads to (free) PSD lift:

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

- Proof requires monotonicity of matrix geometric mean
- Other cases obtained by transformations that reduce to dyadic rational case

Lieb-Ando theorem

Theorem (Lieb 1973, Ando 1979)

Fix $K \in \mathbb{C}^{m \times n}$.

$$\mathcal{S}_+^n \times \mathcal{S}_+^m \ni (A, B) \mapsto \mathrm{tr}[KA^{1-t}K^*B^t]$$

- *jointly concave for* $t \in [0, 1]$
- *jointly convex for* $t \in [-1, 0] \cup [1, 2]$

Equivalently:

$$\mathcal{S}_+^n \otimes \mathcal{S}_+^m \ni (A, B) \mapsto A^{1-t} \otimes B^t$$

- jointly matrix concave for $t \in [0, 1]$
- jointly matrix convex for $t \in [-1, 0] \cup [1, 2]$

Lieb-Ando function

If A and B are positive definite and $t \in \mathbb{R}$ consider

$$(A, B) \mapsto A^{1-t} \otimes B^t$$

Lieb-Ando vs weighted matrix geometric mean

Nikoufar, Ebadian, Gordji (2013): Since $A \otimes I$ and $I \otimes B$ commute

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

Lieb-Ando function

If A and B are positive definite and $t \in \mathbb{R}$ consider

$$(A, B) \mapsto A^{1-t} \otimes B^t$$

Lieb-Ando vs weighted matrix geometric mean

Nikoufar, Ebadian, Gordji (2013): Since $A \otimes I$ and $I \otimes B$ commute

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

Corollaries:

- For $t \in [0, 1]$, $A^{1-t} \otimes B^t$ jointly matrix concave (Lieb)
- For $p/q \in [0, 1]$ rational, matrix hypograph has PSD lift of size $O(mn \log(q))$
- For $t \in [-1, 0] \cup [1, 2]$ $A^{1-t} \otimes B^t$ jointly matrix convex (Ando)
- For $p/q \in [0, 1]$ rational, matrix epigraph has PSD lift of size $O(mn \log(q))$

Lieb-Ando function

$$(A, B) \mapsto A^{1-t} \otimes B^t$$

Concavity/convexity properties

- used in many proofs in quantum information
- extends to $(A_1, \ldots, A_k) \mapsto A^{t_1} \otimes \cdots \otimes A^{t_k}$ where $t_1, \ldots, t_k \geq 0$ and $\sum_i t_i = 1$
- also hold when restrict to symmetry classes of tensors, Hadamard products, etc.

Straightforward to obtain PSD lifts for

- multivariate case (but they get much bigger...)
- matrix hypo/epigraphs when restricted to symmetry classes of tensors, Hadamard products, etc.

Related functions

Function	Properties	PSD lift size ($t = p/q$)
Matrix geometric mean $(A, B) \mapsto A \#_t B$	$\mathcal{S}_+^n$-concave if $t \in [0, 1]$ $\mathcal{S}_+^n$-convex if $t \in [-1, 0] \cup [1, 2]$	$O(n \log(q))$
Lieb-Ando function $(A, B) \mapsto A^{1-t} \otimes B^t$	$\mathcal{S}_+^{mn}$-concave if $t \in [0, 1]$ $\mathcal{S}_+^{mn}$-convex if $t \in [-1, 0] \cup [1, 2]$	$O(mn \log(q))$
$A \mapsto t\,\mathrm{tr}[(K^* A^t K)^{1/t}]$ (fixed $n \times m$ matrix K)	concave if $t \in (0, 1]$ convex if $t \in [-1, 0) \cup [1, 2]$	$O(mn \log(q))$
Tsallis relative entropy $(A, B) \mapsto \frac{1}{t}\mathrm{tr}[A - A^{1-t} B^t]$	convex if $t \in [0, 1]$ $\to S(A\|\|B)$ as $t \to 0$	$O(n^2 \log(q))$

Motivation: towards quantum relative entropy

Quantum relative entropy

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

- Limit of Tsallis relative entropy

$$S(A||B) = \frac{1}{\epsilon}\mathrm{tr}[A - A^{1-\epsilon}B^{\epsilon}] + O(\epsilon)$$

- Get good approximation by choosing $\epsilon = 2^{-k}$
- PSD lift has size $O(n^2 k)$.
- But standard numerical algorithms become ill-conditioned as k grows

Conclusion

This talk

- Matrix hypograph of $A\#_{p/q}B$ has PSD lift of size $O(n\log(q))$
- Gives PSD lift of hypograph/epigraph of Lieb-Ando function (and others) for rational parameters
- Gives approximations of quantum relative entropy

Problems

- Improved approximate PSD lifts for quantum relative entropy
- PSD lifts for epigraph of sandwiched Renyi entropy

$$(A, B) \mapsto \operatorname{tr}\left[\left(A^{\frac{1-t}{2t}} B A^{\frac{1-t}{2t}}\right)^{t}\right] \qquad \text{for rational } t > 1$$

For more information

Preprint

- Fawzi, Saunderson, "Lieb's concavity theorem, matrix geometric means, and semidefinite optimization" arXiv:1512.03401

Code

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

Thank you!