Semidefinite approximations of the matrix logarithm

Pablo A. Parrilo
EECS - MIT

Caltech CMS, November 2018

Joint work with:

Hamza Fawzi
(Cambridge)

James Saunderson
(Monash)

Original motivation: quantum information

How to solve convex optimization problems involving, e.g., quantum relative entropy?

$$D(X \| Y) := \operatorname{tr}\left[X(\log(X) - \log(Y))\right]$$

Original motivation: quantum information

How to solve convex optimization problems involving, e.g., quantum relative entropy?

$$D(X \| Y) := \operatorname{tr}\left[X(\log(X) - \log(Y))\right]$$

No existing off-the-shelf methods

Some bespoke algorithms for particular problems:

- Classical-to-quantum channel capacity [Sutter et al. 2016]
- Relative entropy of entanglement [Zinchenko et al. 2010]

Original motivation: quantum information

How to solve convex optimization problems involving, e.g., quantum relative entropy?

$$D(X\|Y) := \operatorname{tr}\left[X(\log(X) - \log(Y))\right]$$

No existing off-the-shelf methods

Some bespoke algorithms for particular problems:

- Classical-to-quantum channel capacity [Sutter et al. 2016]
- Relative entropy of entanglement [Zinchenko et al. 2010]

Can we exploit and leverage (or extend) successful existing technology (e.g., parsers/solvers for LP/SOCP/SDP, like CVX)?

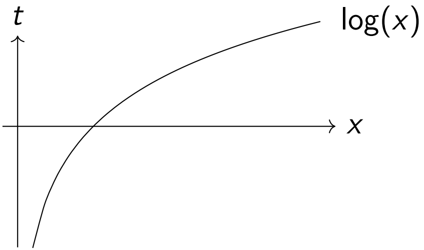
Fundamental issue:

- Semidefinite programming (SDP) can only solve *semialgebraic* problems
- Problems involving logarithms (or entropy) are *not* semialgebraic

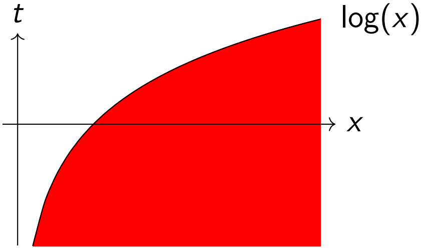
This talk:

- Principled approximations of logarithm that can be modeled using SDP
- Complexity of SDP approximation grows mildly with approximation quality
- Works for *matrix* logarithm and related functions (e.g., quantum entropy)
- Larger theme: what is the SDP complexity of sets and functions?

Logarithm



Logarithm



Properties

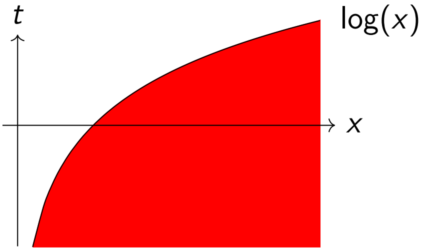
- Monotone:

$$x \geq y > 0 \quad \text{implies} \quad \log(x) \geq \log(y)$$

- Concave:

$$\{(x, \tau) : x > 0, \ \log(x) \geq \tau\} \qquad \text{is a convex set}$$

Logarithm



Related functions:

- Entropy: $H(p) = -\sum_{i=1}^{n} p_i \log(p_i)$ is concave
- Kullback-Leibler divergence (or relative entropy)

$$D(p\|q) = \sum_{i=1}^{n} p_i \log(p_i/q_i)$$

convex in (p, q)

Matrix logarithm

For positive definite X with eigendecomposition

$$X = U \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U^*$$

define

$$\log(X) = U \operatorname{diag}(\log(\lambda_1), \ldots, \log(\lambda_n)) U^*$$

Properties

- Operator monotone:

$$X \succeq Y \succ 0 \quad \text{implies} \quad \log(X) \succeq \log(Y)$$

- Operator concave:

$$\{(X, T) : X \succ 0,\ \log(X) \succeq T\} \qquad \text{is convex}$$

Matrix logarithm

For positive definite X with eigendecomposition

$$X = U \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U^*$$

define

$$\log(X) = U \operatorname{diag}(\log(\lambda_1), \ldots, \log(\lambda_n)) U^*$$

Related functions:

- Entropy $-\operatorname{tr}[X \log(X)]$ is concave in X
- Quantum relative entropy

$$D(X \| Y) = \operatorname{tr}[X(\log(X) - \log(Y))]$$

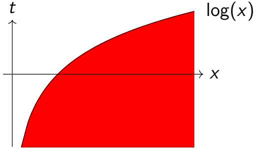
convex in (X, Y) [Lieb-Ruskai, 1973]

Semidefinite representations

Concave function f has a *semidefinite representation* of size d if:

$$f(x) \geq t \iff \exists u \in \mathbb{R}^m : \mathcal{S}(x, t, u) \succeq 0$$

for some affine function $\mathcal{S} : \mathbb{R}^{n+1+m} \to \mathbf{S}^d$.



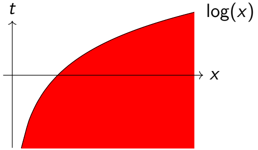
- Key fact: f has semidefinite representation $\implies$ can solve opt. problems involving f using semidefinite solvers

Semidefinite representations

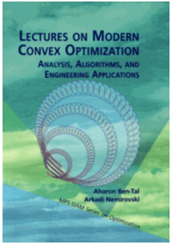
Concave function f has a *semidefinite representation* of size d if:

$$f(x) \geq t \iff \exists u \in \mathbb{R}^m : \mathcal{S}(x,t,u) \succeq 0$$

for some affine function $\mathcal{S} : \mathbb{R}^{n+1+m} \to \mathbf{S}^d$.



- Key fact: f has semidefinite representation $\implies$ can solve opt. problems involving f using semidefinite solvers
- Many convex/concave functions have SDP representations ("can solve using LMIs...")



Logarithm function

Goal: find a semidefinite representation of (matrix) logarithm.

$$\log X \succeq T \quad \Longleftrightarrow \quad ???$$

Problem: Logarithm not semialgebraic! We must approximate

Want: Size of representation to grow mildly with approximation quality

Logarithms and matrix friends

Many inter-related convex functions:

$$
\begin{array}{ccc}
\log(x) & \xrightarrow{\text{perspective}} & y\log(x/y) \\
\text{matrix arg} \downarrow & & \downarrow \text{bimatrix arg} \\
\log(X) & \xrightarrow{\text{NC perspective}} & Y^{\frac{1}{2}}\log(Y^{-\frac{1}{2}}XY^{-\frac{1}{2}})Y^{\frac{1}{2}}
\end{array}
$$

Logarithms and matrix friends

Many inter-related convex functions:

$$
\begin{array}{ccc}
\log(x) & \xrightarrow{\text{perspective}} & y\log(x/y) \\
\text{matrix arg}\downarrow & & \downarrow\text{bimatrix arg} \\
\log(X) & \xrightarrow{\text{NC perspective}} & Y^{\frac{1}{2}}\log(Y^{-\frac{1}{2}}XY^{-\frac{1}{2}})Y^{\frac{1}{2}}
\end{array}
$$

- For positive definite X with eigendecomposition:

$$X = U\Lambda U^* \qquad \to \qquad \log(X) := U\log(\Lambda)U^*$$

- Matrix log is *operator monotone* and *operator concave*

Starting point: Integral representation

$$\log(x) = \int_0^1 \frac{x-1}{1+\xi(x-1)}\, d\xi$$

Starting point: Integral representation

$$\log(x) = \int_0^1 \frac{x-1}{1+\xi(x-1)}\, d\xi$$

Integrand:

Rational, (operator) monotone and concave, has SDP rep. for fixed ξ:

$$\frac{x-1}{1+\xi(x-1)} \succeq \tau \quad \Longleftrightarrow \quad \begin{bmatrix} 1+\xi(x-1) & 1 \\ 1 & 1-\xi\tau \end{bmatrix} \succeq 0$$

In the background: Löwner's theorem on operator monotone functions

Idea 1: Approximate via quadrature

$$\log(x) = \int_0^1 \frac{x-1}{1+\xi(x-1)}\, d\xi$$
$$\approx \sum_{j=1}^{m} w_j \frac{x-1}{1+\xi_j(x-1)} =: r_m(x)$$

for quadrature nodes $\xi_j \in (0,1)$ and weights $w_j > 0$

- $r_m(x)$ is rational, operator monotone, operator concave
- $r_m(x)$ has semidefinite rep. with m LMIs of size 2×2

Idea 1: Approximate via quadrature

$$\log(x) = \int_0^1 \frac{x-1}{1+\xi(x-1)}\, d\xi$$

$$\approx \sum_{j=1}^{m} w_j \frac{x-1}{1+\xi_j(x-1)} =: r_m(x)$$

for quadrature nodes $\xi_j \in (0,1)$ and weights $w_j > 0$

- $r_m(x)$ is rational, operator monotone, operator concave
- $r_m(x)$ has semidefinite rep. with m LMIs of size 2×2

Which quadrature rule to use?

Idea 1: Approximate via quadrature

$$\log(x) = \int_0^1 \frac{x-1}{1+\xi(x-1)}\, d\xi$$
$$\approx \sum_{j=1}^{m} w_j \frac{x-1}{1+\xi_j(x-1)} =: r_m(x)$$

for quadrature nodes $\xi_j \in (0,1)$ and weights $w_j > 0$

- $r_m(x)$ is rational, operator monotone, operator concave
- $r_m(x)$ has semidefinite rep. with m LMIs of size 2×2

Which quadrature rule to use?

Gaussian quadrature, with w_i given by Gauss-Legendre weights.

Idea 1: Approximate via quadrature

$$\begin{aligned}\log(x) &= \int_0^1 \frac{x-1}{1+\xi(x-1)}\, d\xi \\ &\approx \sum_{j=1}^{m} w_j \frac{x-1}{1+\xi_j(x-1)} =: r_m(x)\end{aligned}$$

for quadrature nodes $\xi_j \in (0,1)$ and weights $w_j > 0$

- $r_m(x)$ is rational, operator monotone, operator concave
- $r_m(x)$ has semidefinite rep. with m LMIs of size 2×2

Which quadrature rule to use?

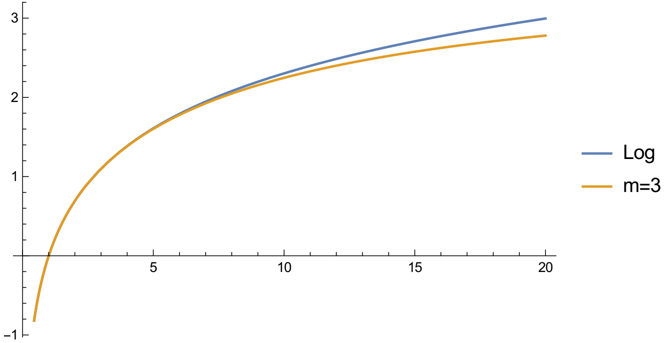
Gaussian quadrature, with w_i given by Gauss-Legendre weights.

Nice properties, e.g., gives Padé approximant at 1

Idea 2: Using the functional equation $\log(x^h) = h\log(x)$

Observations:

- $r_m(x)$ is very good approximation to $\log(x)$ when $x \approx 1$
- $x^{1/2^k} \approx 1$ (Briggs (1617) method for computing log)



Idea 2: Using the functional equation $\log(x^h) = h\log(x)$

Observations:

- $r_m(x)$ is very good approximation to $\log(x)$ when $x \approx 1$
- $x^{1/2^k} \approx 1$ (Briggs (1617) method for computing log)

Define: two-parameter family of approximations

$$r_{m,k}(x) := 2^k r_m(x^{1/2^k}) \approx 2^k \log(x^{1/2^k}) = \log(x)$$

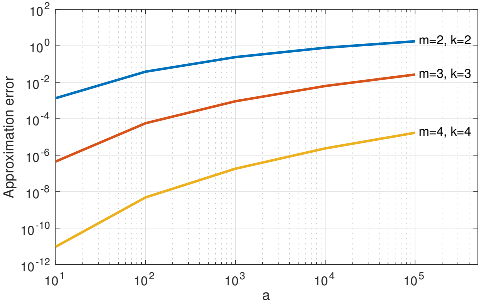
- operator monotone and operator concave
- has semidefinite rep. with $m + k$ LMIs of size 2×2

$$r_{m,k}(x) \geq \tau \quad \Longleftrightarrow \quad \exists u \;\; \text{s.t.} \;\; 2^k r_m(u) \geq \tau, \;\; x^{1/2^k} \geq u$$

Approximation error

Approximation error $\|r_{m,k} - \log\|_\infty$ on $[1/a, a]$

Optimal choice: $m \approx k$.



Theorem

There exists a semidefinite representable function r such that

$$|r(x) - \log(x)| \leq \epsilon \quad \text{for all } x \in [1/a, a]$$

and r has semidefinite rep. of size $O(\sqrt{\log(1/\epsilon)} + \log\log(a))$

Logarithm and matrix friends (SDP version)

What about matrix logarithm?

$$\log(X) \approx r_{m,k}(X) \succeq T,$$

$$\begin{array}{ccc} \log(x) & \xrightarrow{\text{perspective}} & y\log(x/y) \\ \text{matrix arg}\downarrow & & \downarrow\text{bimatrix arg} \\ \log(X) & \xrightarrow{\text{NC perspective}} & Y^{\frac{1}{2}}\log(Y^{-\frac{1}{2}}XY^{-\frac{1}{2}})Y^{\frac{1}{2}} \end{array}$$

- 2×2 linear matrix inequalities become $2n \times 2n$

$$\begin{bmatrix} 1+\xi(x-1) & 1 \\ 1 & 1-\xi\tau \end{bmatrix} \succeq 0 \quad \rightarrow \quad \begin{bmatrix} I+\xi(X-I) & I \\ I & I-\xi T \end{bmatrix} \succeq 0$$

Logarithm and matrix friends (SDP version)

What about matrix logarithm?

$$\log(X) \approx r_{m,k}(X) \succeq T,$$

$$\begin{array}{ccc} \log(x) & \xrightarrow{\text{perspective}} & y\log(x/y) \\ \downarrow{\scriptstyle \text{matrix arg}} & & \downarrow{\scriptstyle \text{bimatrix arg}} \\ \log(X) & \xrightarrow{\text{NC perspective}} & Y^{\frac{1}{2}}\log(Y^{-\frac{1}{2}}XY^{-\frac{1}{2}})Y^{\frac{1}{2}} \end{array}$$

- 2×2 linear matrix inequalities become $2n\times 2n$

$$\begin{bmatrix} 1+\xi(x-1) & 1 \\ 1 & 1-\xi\tau \end{bmatrix} \succeq 0 \quad \rightarrow \quad \begin{bmatrix} I+\xi(X-I) & I \\ I & I-\xi T \end{bmatrix} \succeq 0$$

- Related to *inverse scaling and squaring method* or *Briggs-Padé method* in numerical analysis
- Preserves operator concavity, via SDP.
- Links to "free spectrahedra" (Helton et al.)

Relative entropy cone

$$K_{\mathrm{re}} = \{(x, y, \tau) : x, y > 0,\ -y \log(y/x) \leq \tau\}$$

- Can approximate by homogenizing LMIs in our approximation for logarithm $(1 \leftrightarrow y)$
- Can model, e.g., geometric programs in conic form w.r.t. products of K_{re}
- Can then approximate with second-order cone programs

What about matrices?

Operator relative entropy cone

Theorem [Effros, Ebadian et al.]
If f operator concave then matrix perspective of f, i.e.,

$$g(X, Y) = Y^{1/2} f(Y^{-1/2} X Y^{-1/2}) Y^{1/2}$$

is jointly matrix concave in (X, Y).

Operator relative entropy cone

$$K_{\mathrm{re}}^{n} = \{(X, Y, T) : X, Y \succ 0, \\ - Y^{1/2} \log(Y^{-1/2} X Y^{-1/2}) Y^{1/2} \preceq T\}$$

Operator relative entropy cone

Theorem [Effros, Ebadian et al.]
If f operator concave then matrix perspective of f, i.e.,

$$g(X,Y) = Y^{1/2} f(Y^{-1/2} X Y^{-1/2}) Y^{1/2}$$

is jointly matrix concave in (X, Y).

Operator relative entropy cone

$$K_{\text{re}}^n = \{(X, Y, T) : X, Y \succ 0, \\ -Y^{1/2} \log(Y^{-1/2} X Y^{-1/2}) Y^{1/2} \preceq T\}$$

- Can approximate by 'homogenizing' LMIs in approximation for matrix logarithm ($I \leftrightarrow Y$)

Approximating quantum relative entropy

Quantum relative entropy

$$D(Y\|X) = \mathrm{tr}\left[Y\log(Y) - Y\log(X)\right]$$

(Effros 2009, Tropp 2015) $D(Y\|X)$ can be written as

$$-\phi\left[(I\otimes Y)^{1/2}\log((I\otimes Y)^{-1/2}(X\otimes I)(I\otimes Y)^{-1/2})(I\otimes Y)^{1/2}\right]$$

where ϕ is the positive linear map s.t. $\phi(X\otimes Y) = \mathrm{tr}(XY)$.

Approximating quantum relative entropy

Quantum relative entropy

$$D(Y\|X) = \mathrm{tr}\left[Y\log(Y) - Y\log(X)\right]$$

(Effros 2009, Tropp 2015) $D(Y\|X)$ can be written as

$$-\phi\left[(I\otimes Y)^{1/2}\log((I\otimes Y)^{-1/2}(X\otimes I)(I\otimes Y)^{-1/2})(I\otimes Y)^{1/2}\right]$$

where ϕ is the positive linear map s.t. $\phi(X\otimes Y) = \mathrm{tr}(XY)$.

Representation with operator relative entropy cone

$$K_{\mathrm{re}}^{n} = \{(X,Y,T) : X,Y\succ 0, -Y^{1/2}\log(Y^{-1/2}XY^{-1/2})Y^{1/2}\preceq T\}$$

$$D(Y\|X)\le\tau \iff \exists T \text{ s.t. } (X\otimes I, I\otimes Y, T)\in K_{\mathrm{re}}^{n^2}, \phi(T)\le\tau\}.$$

Maximum entropy problems

$$\begin{array}{ll} \text{maximize} & -\sum_{i=1}^{n} x_i \log(x_i) \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \qquad (A \in \mathbb{R}^{\ell \times n}, b \in \mathbb{R}^{\ell})$$

		CVX's succ. approx.		**Our approach** $m = 3, h = 1/8$	
n	ℓ	time (s)	accuracy*	time (s)	accuracy*
200	100	1.10 s	6.635e-06	0.88 s	2.767e-06
400	200	3.38 s	2.662e-05	0.72 s	1.164e-05
600	300	9.14 s	2.927e-05	1.84 s	2.743e-05
1000	500	52.40 s	1.067e-05	3.91 s	1.469e-04

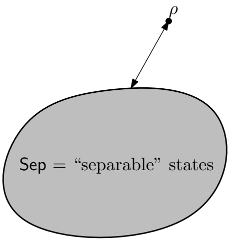
*accuracy measured wrt specialized MOSEK routine

- CVX's successive approx.: Uses Taylor expansion instead of Padé approx + successively refine linearization point

Relative entropy of entanglement

Quantify *entanglement* of a bipartite state ρ:

$$\min \ D(\rho \| \tau) \ \ \text{s.t.} \ \ \tau \in \text{Sep}$$



n	**Cutting-plane** [Zinchenko et al.]	**Our approach** $m = 3, h = 1/8$
4	6.13 s	0.55 s
6	12.30 s	0.51 s
8	29.44 s	0.69 s
9	37.56 s	0.82 s
12	50.50 s	1.74 s
16	100.70 s	5.55 s

```
cvx_begin sdp
  variable tau(na*nb,na*nb) hermitian;
  minimize    (quantum_rel_entr(rho,tau));
  subject to  tau >= 0; trace(tau) == 1;
              % PPT constraint
              Tx(tau,2,[na nb]) >= 0;
cvx_end
```

Beyond logarithm (and friends)

Recall two parts to the approximation:

1. Integral representation (with positive measure μ)

$$f(x) = \int F(x,\xi)\, d\mu(\xi)$$

   where $x \mapsto F(x,\xi)$ has semidefinite rep. for fixed ξ.
2. Functional equation $h \log(x) = \log(x^h)$

First idea generalizes to other classes of functions:

- hypergeometric functions (for certain parameter ranges)
- operator monotone and concave functions on $(0,\infty)$

Sometimes second idea generalizes: AGM, logarithmic mean, ...

Conclusion

Broad issues:

- What can we describe with small SDPs (or SOCPs)?
- What can we approximate with small SDPs (or SOCPs)?
- How to approximate and preserve structural properties?

This talk:

- Matrix logarithm has ϵ-approximate semidefinite description with $O(\sqrt{\log(1/\epsilon)})$, $2n \times 2n$ LMIs
- Gives approximate semidefinite description for quantum relative entropy, operator relative entropy
- Gives new SOCP approx. for relative entropy cone

More information

Paper: H. Fawzi, J. Saunderson, P. Parrilo, 'Semidefinite approximations of the matrix logarithm' `arXiv:1705.00812`. *Foundations of Computational Mathematics*, 2018.

Accompanying paper: H. Fawzi, O. Fawzi, 'Relative entropy optimization in quantum information theory via semidefinite programming approximations.' `arXiv:1705.06671`, *Journal of Physics A: Mathematical and Theoretical*, 2018.

Code: `www.github.com/hfawzi/cvxquad`