Semidefinite approximations of matrix logarithm

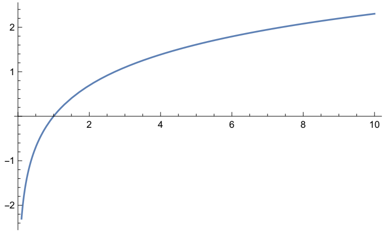
Hamza Fawzi
DAMTP, University of Cambridge

Joint work with
James Saunderson (Monash University) and Pablo Parrilo (MIT)

December 6, 2016

Logarithm

- Concave function



- Information theory:
  - Entropy $H(p) = -\sum_{i=1}^{n} p_i \log p_i$ (**Concave**).
  - Kullback-Leibler divergence (or relative entropy)

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

Convex jointly in (p, q).

Matrix logarithm function

- X symmetric matrix with positive eigenvalues (positive definite)

$$X = U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^* \quad \to \quad \log(X) = U \begin{pmatrix} \log(\lambda_1) & & \\ & \ddots & \\ & & \log(\lambda_n) \end{pmatrix} U^*$$

where U orthogonal.

Matrix logarithm function

- X symmetric matrix with positive eigenvalues (positive definite)

$$X = U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^* \quad \to \quad \log(X) = U \begin{pmatrix} \log(\lambda_1) & & \\ & \ddots & \\ & & \log(\lambda_n) \end{pmatrix} U^*$$

  where U orthogonal.

- von Neumann Entropy of X: $H(X) = -\operatorname{Tr}[X \log X]$. **Concave** in X.

- Quantum relative entropy:

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

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

Concavity of matrix logarithm

Striking property of the matrix logarithm (operator concavity):

$$\log(\lambda A + (1-\lambda)B) \quad \succeq \quad \lambda \log(A) \; + \; (1-\lambda)\log(B)$$

where

- $A, B \succ 0$ and $\lambda \in [0, 1]$
- "$X \succeq Y$" means $X - Y$ positive semidefinite (Löwner order)

Convex optimisation

- How can we solve convex optimisation problems involving matrix logarithm?
- Even for scalar logarithm, things are not so simple
  (solvers for exponential cone are not as well-developed as solvers for symmetric cones)
- CVX modeling tool developed by M. Grant and S. Boyd at Stanford

```
% Maximum entropy problem
cvx_begin
  variable p(n)
  maximize    sum(entr(p))
  subject to  p >= 0; sum(p) == 1;
              A*p == b;
cvx_end
```

- CVX uses a *successive approximation heuristic*. Works good in practice but:
  - sometimes fails (no guarantees)
  - slow for large problems
  - does not work for *matrix logarithm*.

Semidefinite programming

This talk:

- New method to treat matrix logarithm and derived functions using symmetric cone solvers (semidefinite programming)
- Based on accurate rational approximations of logarithm
- Much faster than successive approximation heuristic
- Works for matrix logarithm

Outline

- Semidefinite representations
- Approximating matrix logarithm
- Numerical examples, comparison with successive approximation (for scalars) and other matrix examples

Semidefinite programming

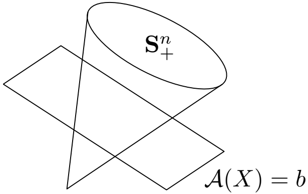
$$\underset{X \in \mathbf{S}^n}{\text{minimize}} \quad \langle C, X \rangle \quad \text{s.t.} \quad \mathcal{A}(X) = b, \; X \succeq 0$$

- Problem data: $C, \mathcal{A}, b$

- Available solvers: SeDuMi, SDPT3, Mosek, SDPA, etc. (e.g., `sedumi(A,b,C)`)

- Generalization of *linear programming* where

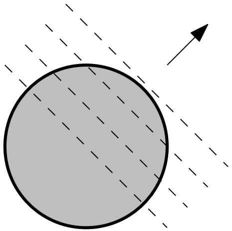
$$x \in \mathbb{R}^n \leftrightarrow X \in \mathbf{S}^n \qquad x \geq 0 \leftrightarrow X \succeq 0$$



Semidefinite formulation

- Not all optimisation problems are given in semidefinite form...
- Example:

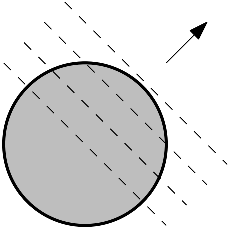
$$\underset{x,y\in\mathbb{R}}{\text{maximise}} \quad 2x + y \quad \text{s.t.} \quad x^2 + y^2 \leq 1$$



Semidefinite formulation

- Not all optimisation problems are given in semidefinite form...
- Example:

$$\underset{x,y\in\mathbb{R}}{\text{maximise}} \quad 2x + y \quad \text{s.t.} \quad x^2 + y^2 \leq 1$$

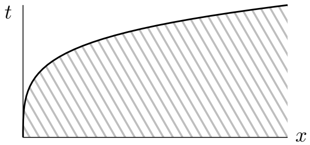


Formulate as *semidefinite optimisation* using the fact that:

$$x^2 + y^2 \leq 1 \quad \Leftrightarrow \quad \begin{bmatrix} 1 - x & y \\ y & 1 + x \end{bmatrix} \succeq 0$$

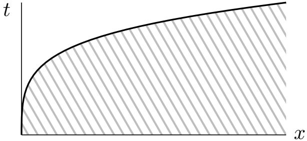
Examples of semidefinite formulation

$$\sqrt{x} \geq t \quad \Leftrightarrow \quad \begin{bmatrix} x & t \\ t & 1 \end{bmatrix} \succeq 0$$

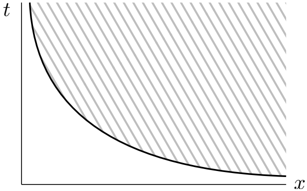


Examples of semidefinite formulation

$$\sqrt{x} \geq t \quad \Leftrightarrow \quad \begin{bmatrix} x & t \\ t & 1 \end{bmatrix} \succeq 0$$



$$\frac{1}{x} \leq t \quad \Leftrightarrow \quad \begin{bmatrix} x & 1 \\ 1 & t \end{bmatrix} \succeq 0$$



Semidefinite representations

- Concave function f has a *semidefinite representation* if:

$$f(x) \geq t \qquad \iff \qquad \mathcal{S}(x,t) \succeq 0$$

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

- Key fact: if f has a semidefinite representation then can solve optimisation problems involving f using semidefinite solvers.

Semidefinite representations

- Concave function f has a *semidefinite representation* if:

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

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

- Key fact: if f has a semidefinite representation then can solve optimisation problems involving f using semidefinite solvers.

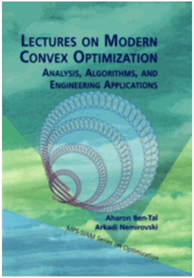
Semidefinite representations

- Concave function f has a *semidefinite representation* if:

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

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

- Key fact: if f has a semidefinite representation then can solve optimisation problems involving f using semidefinite solvers.



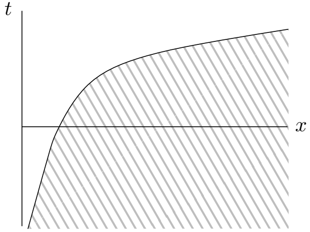
- Book by Ben-Tal and Nemirovski gives semidefinite representations of many convex/concave functions.

- Helton-Nie conjecture: "Any convex semialgebraic function has a semidefinite representation" (caveat: size of representation may be very large!)

Back to logarithm function

Goal: find a semidefinite representation of logarithm.

$$\log(x) \geq t$$



Logarithm is not semialgebraic! We have to resort to approximations.

Integral representation of log

Starting point of approximation is:

$$\log(x) = \int_0^1 \frac{x-1}{1+s(x-1)}ds$$

Integral representation of log

Starting point of approximation is:

$$\log(x) = \int_0^1 \frac{x-1}{1+s(x-1)} ds$$

- Key fact: integrand is concave and semidefinite rep. for any fixed s!

$$\frac{x-1}{1+s(x-1)} \geq t \quad \Leftrightarrow \quad \begin{bmatrix} 1+s(x-1) & 1 \\ 1 & 1-st \end{bmatrix} \succeq 0$$

Integral representation of log

Starting point of approximation is:

$$\log(x) = \int_0^1 \frac{x-1}{1+s(x-1)} ds$$

- Key fact: integrand is concave and semidefinite rep. for any fixed s!

$$\frac{x-1}{1+s(x-1)} \geq t \quad \Leftrightarrow \quad \begin{bmatrix} 1+s(x-1) & 1 \\ 1 & 1-st \end{bmatrix} \succeq 0$$

- Get semidefinite approximation of log using quadrature:

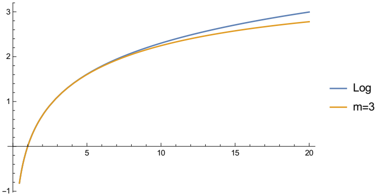
$$\log(x) \approx \sum_{j=1}^{m} w_j \frac{x-1}{1+s_j(x-1)}$$

  Right-hand side is semidefinite representable

Rational approximation

$$\log(x) \approx \underbrace{\sum_{j=1}^{m} w_j \frac{x-1}{1+s_j(x-1)}}_{r_m(x)}$$

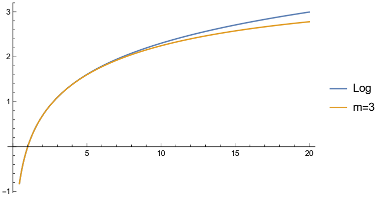
r_m = m'th diagonal Padé approximant of log at $x = 1$ (matches the first $2m$ Taylor coefficients).



Rational approximation

$$\log(x) \approx \underbrace{\sum_{j=1}^{m} w_j \frac{x-1}{1+s_j(x-1)}}_{r_m(x)}$$

r_m = m'th diagonal Padé approximant of log at $x = 1$ (matches the first $2m$ Taylor coefficients).



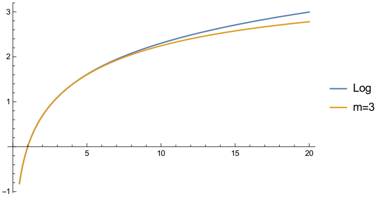
- Improve approximation by bringing x closer to 1 and using $\log(x) = \frac{1}{h}\log(x^h)$ $(0 < h < 1)$:

$$r_{m,h}(x) := \frac{1}{h} r_m(x^h)$$

Rational approximation

$$\log(x) \approx \underbrace{\sum_{j=1}^{m} w_j \frac{x-1}{1+s_j(x-1)}}_{r_m(x)}$$

$r_m = m$'th diagonal Padé approximant of log at $x = 1$ (matches the first $2m$ Taylor coefficients).



- Improve approximation by bringing x closer to 1 and using $\log(x) = \frac{1}{h}\log(x^h)$ $(0 < h < 1)$:

$$r_{m,h}(x) := \frac{1}{h} r_m(x^h)$$

- Remarkable fact: $r_{m,h}$ is still concave and semidefinite representable!

Quadrature + exponentiation

$$r_{m,h}(x) := \frac{1}{h} r_m(x^h)$$

- Semidefinite representation of $r_{m,h}$ (say $h = 1/2$ for concreteness):

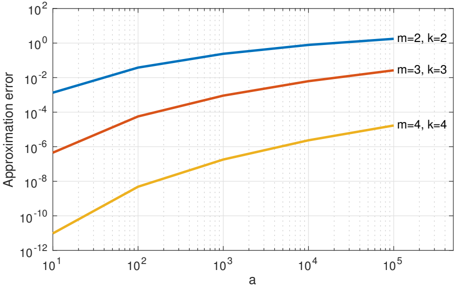
$$r_{m,1/2}(x) \geq t \quad \iff \quad \exists y \geq 0 \text{ s.t. } \begin{cases} x^{1/2} \geq y \\ r_m(y) \geq t/2 \end{cases}$$

- Uses fact that r_m is monotone and $x^{1/2}$ is concave and semidefinite rep.

- Can do the case $h = 1/2^k$ with iterative square-rooting.

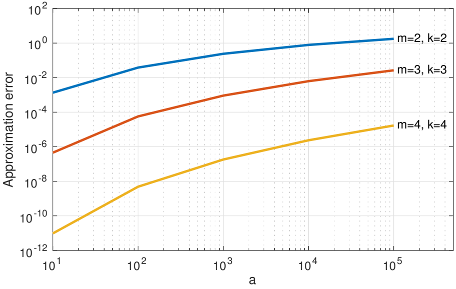
Approximation error

Approximation error $\|r_{m,h} - \log\|_\infty$ on $[1/a, a]$ ($h = 1/2^k$):



Approximation error

Approximation error $\|r_{m,h} - \log\|_\infty$ on $[1/a, a]$ ($h = 1/2^k$):



Recap: Two ingredients

- Rational approximation via quadrature
- Use $\log(x) = \frac{1}{h}\log(x^h)$ with small h to bring x closer to 1.

Key fact: resulting approximation is concave and semidefinite representable.

Matrix logarithm

What about matrix logarithm?

- Integral representation is valid for matrix log as well:

$$\log(X) = \int_0^1 (X - I)(I + s(X - I))^{-1} ds$$

Matrix logarithm

What about matrix logarithm?

- Integral representation is valid for matrix log as well:

$$\log(X) = \int_0^1 (X - I)(I + s(X - I))^{-1} ds$$

- Key fact: integrand is **operator concave** and semidefinite rep. for any fixed s (use Schur complements)

$$(X - I)(I + s(X - I))^{-1} \succeq T \quad \Leftrightarrow \quad \begin{bmatrix} I + s(X - I) & I \\ I & I - sT \end{bmatrix} \succeq 0$$

Matrix logarithm

What about matrix logarithm?

- Integral representation is valid for matrix log as well:

$$\log(X) = \int_0^1 (X - I)(I + s(X - I))^{-1} ds$$

- Key fact: integrand is **operator concave** and semidefinite rep. for any fixed s (use Schur complements)

$$(X - I)(I + s(X - I))^{-1} \succeq T \quad \Leftrightarrow \quad \begin{bmatrix} I + s(X - I) & I \\ I & I - sT \end{bmatrix} \succeq 0$$

- Get semidefinite approximation of matrix log using quadrature:

$$\log(X) \approx \sum_{j=1}^{m} w_j \frac{X - 1}{1 + s_j(X - 1)}$$

  Right-hand side is semidefinite representable

Exponentiation

- Exponentiation idea also works for matrices:

$$r_{m,h}(X) := \frac{1}{h} r_m(X^h) \qquad (0 < h < 1)$$

  - r_m is not only monotone concave but *operator monotone* and *operator concave*. Also $X \mapsto X^h$ is *operator concave* and semidefinite rep.

$$X^{1/2} \succeq T \quad \Leftrightarrow \quad \begin{bmatrix} X & T \\ T & I \end{bmatrix} \succeq 0$$

Exponentiation

- Exponentiation idea also works for matrices:

$$r_{m,h}(X) := \frac{1}{h} r_m(X^h) \qquad (0 < h < 1)$$

  - r_m is not only monotone concave but *operator monotone* and *operator concave*. Also $X \mapsto X^h$ is *operator concave* and semidefinite rep.

$$X^{1/2} \succeq T \quad \Leftrightarrow \quad \begin{bmatrix} X & T \\ T & I \end{bmatrix} \succeq 0$$

- Approximation $\log(X) \approx r_{m,h}(X)$ called *inverse scaling and squaring method* by Kenney-Laub, widely used in numerical computations.

- Remarkable that it "preserves" concavity and can be implemented in semidefinite programming.

From (matrix) logarithm to (matrix) relative entropy

$$\log(x) \approx r_{m,h}(x)$$

- Perspective transform (homogenization):

$$f : \mathbb{R} \to \mathbb{R} \text{ concave} \quad \Rightarrow \quad g(x,y) := yf(x/y) \text{ also concave on } \mathbb{R} \times \mathbb{R}_{++}$$

From (matrix) logarithm to (matrix) relative entropy

$$\log(x) \approx r_{m,h}(x)$$

- Perspective transform (homogenization):

$$f : \mathbb{R} \to \mathbb{R} \text{ concave} \quad \Rightarrow \quad g(x,y) := yf(x/y) \text{ also concave on } \mathbb{R} \times \mathbb{R}_{++}$$

- Perspective of log is $(x,y) \mapsto y \log(x/y)$ related to *relative entropy*. Can simply approximate with the perspective of $r_{m,h}$:

$$y \log(x/y) \approx y r_{m,h}(x/y)$$

Semidefinite representation is obtained by homogenization (replace 1 by y).

From (matrix) logarithm to (matrix) relative entropy

$$\log(x) \approx r_{m,h}(x)$$

- Perspective transform (homogenization):

$$f : \mathbb{R} \to \mathbb{R} \text{ concave} \quad \Rightarrow \quad g(x,y) := yf(x/y) \text{ also concave on } \mathbb{R} \times \mathbb{R}_{++}$$

- Perspective of log is $(x,y) \mapsto y\log(x/y)$ related to *relative entropy*. Can simply approximate with the perspective of $r_{m,h}$:

$$y\log(x/y) \approx y r_{m,h}(x/y)$$

  Semidefinite representation is obtained by homogenization (replace 1 by y).

- What about for matrices? What is the perspective transform?

Matrix perspective

- Matrix perspective of f:

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

- **Theorem** [Effros, Ebadian et al.]: If f operator concave then matrix perspective of f is jointly operator concave in (X, Y).

Matrix perspective

- Matrix perspective of f:

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

- **Theorem** [Effros, Ebadian et al.]: If f operator concave then matrix perspective of f is jointly operator concave in (X, Y).
- For $f = \log$ matrix perspective is related to *operator relative entropy*

$$D_{\mathrm{op}}(X \| Y) = -Y^{1/2} \log(Y^{-1/2} X Y^{-1/2}) Y^{1/2}$$

- Approximate with the matrix perspective of $r_{m,h}$:

$$D_{\mathrm{op}}(X \| Y) \approx -Y^{1/2} r_{m,h}(Y^{-1/2} X Y^{-1/2}) Y^{1/2}$$

- Semidefinite representation obtained by homogenization

Numerical experiments: maximum entropy problem

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

Numerical experiments: maximum entropy problem

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

		CVX's successive 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 of log instead of Padé approx + successively refine linearization point

Geometric programming

- Geometric program:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_j(x) \leq 1, \quad j = 1, \ldots, \ell \\ & x > 0 \end{array}$$

where $f_0, \ldots, f_\ell$ are *posy*nomials (polynomials with nonnegative coeffs)

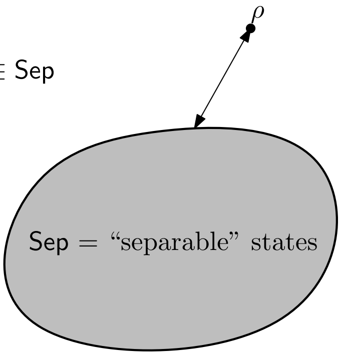
- Important class of convex optimization problems (applications in circuit design, communications, etc.)

		CVX's successive approx.		**Our approach** $m = 3, h = 1/8$	
n	ℓ	time (s)	accuracy	time (s)	accuracy
100	200	7.60 s	1.853e-06	2.69 s	3.769e-06
200	200	7.47 s	2.441e-07	3.72 s	7.505e-07
200	400	42.71 s	3.666e-06	14.36 s	2.855e-06
200	600	184.33 s	7.899e-06	35.45 s	4.480e-06

Application in quantum information theory: 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;
              Tx(tau,2,[na nb]) >= 0;  % Positive partial transpose constraint
cvx_end
```

Conclusion

- Approximation theory with convexity
- Approach extends to other operator concave functions via their integral representation (Löwner theorem)
- Our approximation for scalar log has size (second-order cone rep.) $\sqrt{\log(1/\epsilon)}$ where ϵ error on $[e^{-1}, e]$. Is this best possible?
- Paper soon to be posted on arXiV with Matlab code

Conclusion

- Approximation theory with convexity
- Approach extends to other operator concave functions via their integral representation (Löwner theorem)
- Our approximation for scalar log has size (second-order cone rep.) $\sqrt{\log(1/\epsilon)}$ where ϵ error on $[e^{-1}, e]$. Is this best possible?
- Paper soon to be posted on arXiV with Matlab code

Thank you!