Sum-of-squares proofs for logarithmic Sobolev inequalities

Hamza Fawzi
Joint work with Oisín Faust

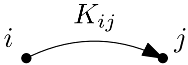
Department of Applied Mathematics and Theoretical Physics
University of Cambridge

Workshop on Real Algebraic Geometry and Algorithms
for Geometric Constraint Systems
The Fields Institute, June 2021

Markov chains

- $K : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ transition matrix

$$K_{ij} \geq 0, \qquad \sum_{j \in \mathcal{S}} K_{ij} = 1 \ \ \forall i \in \mathcal{S}$$

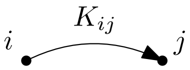


- Invariant distribution $\pi \in \mathbb{R}^{\mathcal{S}}$: $\sum_{i \in \mathcal{S}} K_{ij} \pi_i = \pi_j$ (i.e., $\pi K = \pi$).

Markov chains

- $K : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ transition matrix

$$K_{ij} \geq 0, \qquad \sum_{j \in \mathcal{S}} K_{ij} = 1 \ \ \forall i \in \mathcal{S}$$



- Invariant distribution $\pi \in \mathbb{R}^{\mathcal{S}}$: $\sum_{i \in \mathcal{S}} K_{ij} \pi_i = \pi_j$ (i.e., $\pi K = \pi$).

- Continuous-time Markov process ("heat equation")

$$\frac{d p(t)}{dt} = -p(t) L$$

where $L = I - K$ is *Laplacian*. $p(t) \in \mathbb{R}^{\mathcal{S}}$ distribution at time t

- Q: How fast does $p(t)$ converge to π?

Spectral theory / Poincaré inequality

- Let $x(t) = p(t)/\pi$ the density of $p(t)$ wrt π at time t

$$\forall t,\ \mathbf{E}_\pi[x(t)] = 1 \qquad \text{and} \qquad x(t) \to \mathbf{1} \text{ when } t \to \infty$$

- Define $\mathrm{Var}(x(t)) = \mathbf{E}_\pi[(x(t) - \mathbf{1})^2]$. Note $\mathrm{Var}(x(t)) \to 0$ as $t \to \infty$

- Evolution of $\mathrm{Var}(x(t))$:

$$\frac{d}{dt}\mathrm{Var}(x(t)) = -2\mathcal{E}(x(t), x(t)) \text{ where } \underbrace{\mathcal{E}(x, y) = \langle x, Ly\rangle_\pi}_{\text{Dirichlet form}}$$

Spectral theory / Poincaré inequality

- Let $x(t) = p(t)/\pi$ the density of $p(t)$ wrt π at time t

$$\forall t,\ \mathbf{E}_\pi[x(t)] = 1 \qquad \text{and} \qquad x(t) \to \mathbf{1} \text{ when } t \to \infty$$

- Define $\mathrm{Var}(x(t)) = \mathbf{E}_\pi[(x(t) - \mathbf{1})^2]$. Note $\mathrm{Var}(x(t)) \to 0$ as $t \to \infty$

- Evolution of $\mathrm{Var}(x(t))$:

$$\frac{d}{dt}\mathrm{Var}(x(t)) = -2\mathcal{E}(x(t), x(t)) \text{ where } \underbrace{\mathcal{E}(x, y) = \langle x, Ly \rangle_\pi}_{\text{Dirichlet form}}$$

- Poincaré inequality:

$$\boxed{\mathcal{E}(x, x) \geq \lambda\,\mathrm{Var}(x)} \implies \mathrm{Var}(x(t)) \leq \mathrm{Var}(x(0))e^{-2\lambda t}$$

λ is the second smallest eigenvalue of the Laplacian matrix L

Functional inequalities

- Logarithmic-Sobolev inequality:

$$\mathcal{E}(x,x) \;\geq\; \alpha \sum_i \pi_i x_i^2 \log(x_i^2) \qquad \forall x : \sum_i \pi_i x_i^2 = 1.$$

- Largest α for which this inequality holds is the logarithmic Sobolev constant

- Controls convergence of $p(t)$ to π in the *relative entropy* sense

$$D(p(t)\|\pi) \;\leq\; D(p(0)\|\pi)e^{-4\alpha t} \quad \text{where} \quad D(p\|q) := \sum_{i\in\mathcal{S}} p_i \log(p_i/q_i).$$

Functional inequalities

- Logarithmic-Sobolev inequality:

$$\mathcal{E}(x,x) \;\geq\; \alpha \sum_i \pi_i x_i^2 \log(x_i^2) \qquad \forall x : \sum_i \pi_i x_i^2 = 1.$$

- Largest α for which this inequality holds is the logarithmic Sobolev constant

- Controls convergence of $p(t)$ to π in the *relative entropy* sense

$$D(p(t)\|\pi) \;\leq\; D(p(0)\|\pi)e^{-4\alpha t} \quad \text{where} \quad D(p\|q) := \sum_{i\in\mathcal{S}} p_i \log(p_i/q_i).$$

- Advantage is that $D(p(0)\|\pi) \ll \mathrm{Var}(x(0))$

  Example: if $p(0) = \delta_i$ and $\pi = \mathbf{1}/|\mathcal{S}|$ (uniform) then $D(p(0)\|\pi) = \log(|\mathcal{S}|)$ and $\mathrm{Var}(x(0)) \approx |\mathcal{S}|$

- Compared to λ (Poincaré constant), α is much harder to compute

Lectures on finite Markov chains

Laurent Saloff-Coste
CNRS & Université Paul Sabatier, UMR 55830

École d'été de probabilités de St Flour 1996

⋮

This result shows that α is closely related to the quantity we want to bound, namely the "time to equilbrium" T_2 (more generally T_p) of the chain (K, π). The natural question now is:

can one compute or estimate the constant α?

Unfortunately, the present answer is that it seems to be a very difficult problem to estimate α. To illustrate this point we now present what, in some sense, is the only example of finite Markov chain for which α is known explicitely.

Computing α

Lectures on finite Markov chains

Laurent Saloff-Coste
CNRS & Université Paul Sabatier, UMR 55830

École d'été de probabilités de St Flour 1996

⋮

This result shows that α is closely related to the quantity we want to bound, namely the "time to equilbrium" T_2 (more generally T_p) of the chain (K, π). The natural question now is:

can one compute or estimate the constant α?

Unfortunately, the present answer is that it seems to be a very difficult problem to estimate α. To illustrate this point we now present what, in some sense, is the only example of finite Markov chain for which α is known explicitely.

This talk: Computational method to produce formal lower bounds on α

Sum-of-squares proofs

- Given $p, q \in \mathbb{R}[x_1, \ldots, x_n]$, decide:

$$\textit{is} \quad \boxed{p(x) \geq 0 \ \ \forall x \in \mathbb{R}^n \text{ s.t. } q(x) = 0} \quad ?$$

  Hard for general polynomials p, q.

Sum-of-squares proofs

- Given $p, q \in \mathbb{R}[x_1, \ldots, x_n]$, decide:

$$\textit{is} \quad p(x) \geq 0 \;\; \forall x \in \mathbb{R}^n \text{ s.t. } q(x) = 0 \quad ?$$

  Hard for general polynomials p, q.

- A sufficient condition:

$$p(x) = s(x) + h(x)q(x)$$

  where $h(x)$ is an arbitrary polynomial and $s(x)$ is a sum of squares of polynomials, i.e.,

$$s = \sum_k h_k^2$$

  where h_k are polynomials.

Sum-of-squares proofs

- Given $p, q \in \mathbb{R}[x_1, \ldots, x_n]$, decide:

$$\textit{is} \quad \boxed{p(x) \geq 0 \;\; \forall x \in \mathbb{R}^n \text{ s.t. } q(x) = 0} \quad ?$$

  Hard for general polynomials p, q.

- A sufficient condition:

$$\boxed{p(x) = s(x) + h(x)q(x)}$$

  where $h(x)$ is an arbitrary polynomial and $s(x)$ is a sum of squares of polynomials, i.e.,

$$s = \sum_k h_k^2$$

  where h_k are polynomials.

- Key fact: **Can search for a sum-of-squares proof efficiently, using semidefinite programming**

Sum-of-squares proofs and semidefinite programming

- Let $\mathbb{R}[x]_{\leq d}$ = space of polynomials of degree $\leq d$, $N(n,d) = \dim \mathbb{R}[x]_{\leq d}$

- $s(x) \in \mathbb{R}[x]_{\leq d}$ is a sum of squares if, and only if, there exists a symmetric matrix Q of size $N(n, d/2)$ such that

$$Q \succeq 0 \text{ and } s_\gamma = \sum_{\alpha+\beta=\gamma} Q_{\alpha,\beta} \quad \forall |\gamma| \leq d$$

where $s(x) = \sum_{\gamma:|\gamma|\leq d} s_\gamma x^\gamma$
Rows/columns of Q indexed by monomials of degree $\leq d/2$

Log-Sobolev inequality

$$\mathcal{E}(x,x) - \alpha B(x) \geq 0 \qquad \forall x \in \mathbb{R}^n : S(x) = 0$$

where

- $\mathcal{E}(x,x) = \frac{1}{2}\sum_{ij} \pi_i K_{ij}(x_i - x_j)^2$
- $B(x) = \sum_i \pi_i x_i^2 \log(x_i^2)$
- $S(x) = \sum_i \pi_i x_i^2 - 1.$

Main problem: $B(x)$ is not a polynomial.

Log-Sobolev inequality

$$\mathcal{E}(x,x) - \alpha B(x) \geq 0 \qquad \forall x \in \mathbb{R}^n : S(x) = 0$$

where

- $\mathcal{E}(x,x) = \frac{1}{2}\sum_{ij} \pi_i K_{ij}(x_i - x_j)^2$
- $B(x) = \sum_i \pi_i x_i^2 \log(x_i^2)$
- $S(x) = \sum_i \pi_i x_i^2 - 1.$

Main problem: $B(x)$ is not a polynomial.

Approach: Find $\hat{B}(x)$ polynomial such that $B(x) \leq \hat{B}(x)$ and attempt to prove instead

$$\mathcal{E}(x,x) - \alpha \hat{B}(x) \geq 0 \qquad \forall x : S(x) = 0$$

using sums of squares. **How to choose $\hat{B}(x)$?**

Taylor bound

Simple fact: Let p^{Taylor}_{2d-1} be the degree $2d-1$ Taylor expansion of $t^2\log(t)$ at $t=1$. Then

$$p^{\text{Taylor}}(t) \geq t^2\log(t) \;\; \forall t \geq 0.$$

Consequence

$$\hat{B}(x) = 2\sum_i \pi_i p^{\text{Taylor}}(x_i) \geq B(x).$$

Taylor bound

Simple fact: Let p_{2d-1}^{Taylor} be the degree $2d-1$ Taylor expansion of $t^2\log(t)$ at $t=1$. Then

$$p^{\text{Taylor}}(t) \geq t^2\log(t) \ \ \forall t \geq 0.$$

Consequence

$$\hat{B}(x) = 2\sum_i \pi_i p^{\text{Taylor}}(x_i) \geq B(x).$$

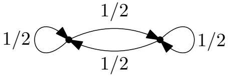
Semidefinite programming lower bound on α:

$$\begin{array}{ll} \max\limits_{\hat{\alpha}, s(x), h(x)} & \hat{\alpha} \\ \text{s.t.} & \mathcal{E}(x,x) - 2\hat{\alpha}\sum_i \pi_i p^{\text{Taylor}}(x_i) = s(x) + h(x)(\sum_i \pi_i x_i^2 - 1) \\ & s \text{ sum of squares}, \deg(s) = 2k \\ & h \text{ arbitrary polynomial}, \deg(h) = 2k-2. \end{array}$$

Solution of SDP gives formal lower bound on α

Example: two-point space

$$K = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

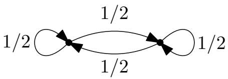


It is known that $\alpha = 1/2$. The inequality we have to prove is

$$\frac{1}{4}(x-y)^2 - \frac{1}{2}(x^2\log(x) + y^2\log(y)) \geq 0 \ \ \forall (x,y) \in \mathbb{R}_+^2 : x^2 + y^2 = 2.$$

Example: two-point space

$$K = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$



It is known that $\alpha = 1/2$. The inequality we have to prove is

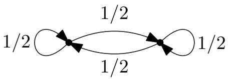
$$\frac{1}{4}(x-y)^2 - \frac{1}{2}(x^2\log(x) + y^2\log(y)) \geq 0 \;\; \forall (x,y) \in \mathbb{R}^2_+ : x^2 + y^2 = 2.$$

- Using Taylor bound of degree 3, we seek to prove the **stronger** polynomial inequality:

$$-1 + 3x + 3y - 3xy - x^3 - y^3 \geq 0 \;\; \forall (x,y) \in \mathbb{R}^2_+ : x^2 + y^2 = 2.$$

Example: two-point space

$$K = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$



It is known that $\alpha = 1/2$. The inequality we have to prove is

$$\frac{1}{4}(x-y)^2 - \frac{1}{2}(x^2\log(x) + y^2\log(y)) \geq 0 \ \ \forall (x,y) \in \mathbb{R}_+^2 : x^2 + y^2 = 2.$$

- Using Taylor bound of degree 3, we seek to prove the **stronger** polynomial inequality:

$$-1 + 3x + 3y - 3xy - x^3 - y^3 \geq 0 \ \ \forall (x,y) \in \mathbb{R}_+^2 : x^2 + y^2 = 2.$$

- Sum-of-squares proof:

$$-1 + 3x + 3y - 3xy - x^3 - y^3 = s(x,y)(1 + x + y) + h(x,y)(x^2 + y^2 - 2)$$

where $s(x,y) = 2(x/2 + y/2 - 1)^2$ and $h(x,y) = -3(x + y - 1)/2$.

Searching for the best polynomial bound

- We want the optimization program to *search for the best polynomial upper bound* on $B(x)$, i.e., we want to solve:

$$\begin{array}{ll} \max\limits_{\hat\alpha, s(x), h(x), \hat p} & \hat\alpha \\ \text{s.t.} & \mathcal{E}(x,x) - 2\hat\alpha \sum_i \pi_i \hat p(x_i) = s(x) + h(x)(\sum_i \pi_i x_i^2 - 1) \\ & s \text{ sum of squares}, \deg(s) = 2k \\ & h \text{ arbitrary polynomial}, \deg(h) = 2k-2 \\ & \hat p(t) \geq t^2 \log(t) \ \ \forall t \geq 0, \ \ \deg(\hat p) = \ell. \end{array}$$

- Need a tractable formulation of the convex set

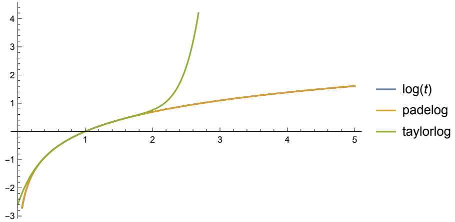
$$\left\{ \hat p \in \mathbb{R}[t], \deg(\hat p) = \ell \text{ s.t. } \hat p(t) \geq t^2 \log(t) \ \forall t > 0 \right\}$$

- We use rational approximations of log

Padé approximations

- The (m, n) Padé approximation of $f(t)$ at $t = t_0$ is a rational function P/Q with $\deg P = m, \deg Q = n$ so that around $t = t_0$

$$f(t) - P(t)/Q(t) = O((t - t_0)^{m+n+1})$$



Padé (4,3) vs Taylor of order 7 of log around $t = 1$

Padé upper bound on log

Proposition: For any integer m, the $(m+1, m)$ Padé approximant P_m/Q_m of log at $t = 1$ is an *upper bound* on log. Furthermore $Q_m(t) > 0$ for all $t > 0$

Thus a sufficient condition for $\hat{p}(t) \geq t^2 \log(t)$ is $\hat{p} \geq t^2 P_m/Q_m$, which we can impose via sum-of-squares as

$$Q_m \hat{p} - t^2 P_m \text{ is a sum-of-squares}$$

Padé upper bound on log

Proposition: For any integer m, the $(m+1, m)$ Padé approximant P_m/Q_m of log at $t = 1$ is an *upper bound* on log. Furthermore $Q_m(t) > 0$ for all $t > 0$

Thus a sufficient condition for $\hat{p}(t) \geq t^2 \log(t)$ is $\hat{p} \geq t^2 P_m/Q_m$, which we can impose via sum-of-squares as

$$Q_m \hat{p} - t^2 P_m \text{ is a sum-of-squares}$$

Theorem: The solution of the following sum-of-squares program is a lower bound on the log-Sobolev constant of (K, π):

$$\begin{array}{ll} \displaystyle\max_{\hat{\alpha}, s(x), h(x), \hat{p}} & \hat{\alpha} \\ \text{s.t.} & \mathcal{E}(x,x) - 2\hat{\alpha} \sum_i \pi_i \hat{p}(x_i) = s(x) + h(x)(\sum_i \pi_i x_i^2 - 1) \\ & s \text{ sum of squares}, \deg(s) = 2k \\ & h \text{ arbitrary polynomial}, \deg(h) = 2k-2 \\ & Q_m(t)\hat{p}(t) - t^2 P_m \text{ sum-of-squares}, \ \deg(\hat{p}) = \ell. \end{array}$$

Formal proofs from approximate SDP solutions

- Sum-of-squares programs are transformed into standard form semidefinite programs

$$\max_{X \in \mathbf{S}^n} \langle C, X \rangle \quad \text{s.t.} \quad X \succeq 0 \text{ and } \langle A_i, X \rangle = b_i \ (i = 1, \ldots, m)$$

  Numerical solvers yield approximate (floating-point) solutions. Need to extract formal lower bounds on α

- We use a perturb-and-project approach [Peyrl-Parrilo]. We first perturb the SDP to

$$\max_{X \in \mathbf{S}^n} \langle C, X \rangle \quad \text{s.t.} \quad X \succeq \varepsilon I \text{ and } \langle A_i, X \rangle = b_i \ (i = 1, \ldots, m)$$

  and project the returned $\hat{X}$ (using rational arithmetic) on the subspace $\mathcal{A}(X) = b$.

- All of this implemented in the Julia language, available at

```
https://github.com/oisinfaust/LogSobolevRelaxations
```

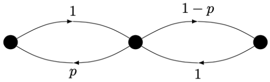
Examples

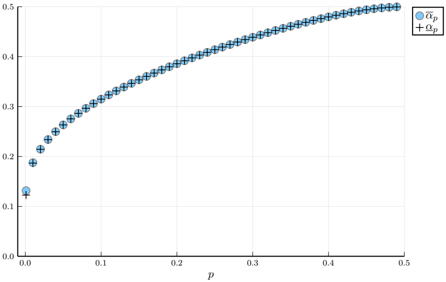
- Simple walk on the complete graph K_n
- Exact value known $\alpha = \frac{n-2}{(n-1)\log(n-1)}$ [Diaconis-Saloff-Coste]

n	$\hat{\alpha}$	ϵ_{rel}
3	0.72134751987	7.96×10^{-10}
4	0.6068261485	4.25×10^{-9}
5	0.541010629	2.16×10^{-8}
6	0.497067908	7.95×10^{-8}
7	0.46509209	2.22×10^{-7}
8	0.44048407	5.06×10^{-7}
9	0.4207856	1.02×10^{-6}
10	0.4045500	1.85×10^{-6}
11	0.3908638	3.13×10^{-6}
12	0.3791184	5.06×10^{-6}
13	0.3688909	7.81×10^{-6}

Using Padé approach with $m = 5$

3-point stick





The cycle

- Simple walk on $\mathbb{Z}_n$: $K_{i,i\pm 1} = 1/2$ for $i \in \mathbb{Z}_n$.
- It is known that $\alpha = \frac{\lambda}{2} = \frac{1}{2}(1 - \cos(2\pi/n))$ for all even n and $n = 5$.
  [Chen-Sheu],[Chen-Liu-Saloff-Coste]
- Open question: is $\alpha = \lambda/2$ for all odd $n \geq 5$?
- **We give formal proofs that**

$$\alpha = \frac{1}{2}(1 - \cos(2\pi/n)) \;\; \forall n \in \{5, 7, 9, \ldots, 21\}$$

Several ingredients:

- Relaxation based on the Taylor upper bound of degree 5
- Symmetry reduction reduces SDP from a large block of size $\sim 3n^2/2$ to smaller blocks of size $\sim 3n/2$
- Rounding in $\mathbb{Q}[\cos(2\pi/n)]$ (instead of just $\mathbb{Q}$)

Conclusion

Paper at `arXiv:2101.04988`

Open directions

- Fastest Mixing Markov Chain: can use the relaxation to search for a Markov chain with the largest log-Sobolev constant. Compare with Markov chains with largest Poincaré constant [Boyd-Diaconis-Xiao].

- Modified log-Sobolev constant

- Quantum (modified) log-Sobolev constant?

Thank you!