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

Entropy Inequalities, Quantum Information and Quantum Physics
IPAM, February 2021

Markov chains

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

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

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

- Continuous-time Markov process

$$\frac{dp}{dt} = -p(I - K)$$

$L = I - K$ is Laplacian.

- Q: How fast does p converge to π?

Poincaré inequality / spectral theory

- Rate of convergence of $\mathrm{Var}(f(t))$, where $f(t) = p(t)/\pi$ is density

$$\mathrm{Var}(f(t)) = \sum_i \pi_i (p_i(t)/\pi_i - 1)^2$$

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

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

- Poincaré inequality:

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

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
- Implies $D(p(t)\|\pi) \;\leq\; D(p(0)\|\pi)e^{-4\alpha t}$
- Compared to λ (Poincaré constant), α is much harder to compute
- **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 $p, q \in \mathbb{R}[x_1, \dots, x_n]$, decide:

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

  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.

Sum-of-squares proofs and semidefinite programming

- $p(x, y) = 1 - x \geq 0$ whenever $q(x, y) = 1 - x^2 - y^2 = 0$

Sum-of-squares proofs and semidefinite programming

- $p(x, y) = 1 - x \geq 0$ whenever $q(x, y) = 1 - x^2 - y^2 = 0$
  Sum-of-squares proof: $p = s + hq$ where

$$s(x, y) = (1 - x)^2/2 + y^2/2 \quad \text{and} \quad h(x, y) = 1/2.$$

Sum-of-squares proofs and semidefinite programming

- $p(x, y) = 1 - x \geq 0$ whenever $q(x, y) = 1 - x^2 - y^2 = 0$
  Sum-of-squares proof: $p = s + hq$ where

$$s(x, y) = (1 - x)^2/2 + y^2/2 \quad \text{and} \quad h(x, y) = 1/2.$$

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

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

  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_{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

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

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:

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

- This is not a semidefinite program. Instead we impose

$$\hat{p} \geq t^2 R(t)$$

where $R(t) = \frac{P(t)}{Q(t)}$ is a rational upper bound on log (the $(k+1,k)$ Padé approximation at 1).

Implementation

- **Formal proofs from floating-point solutions:** Semidefinite programs are solved with floating-point arithmetic

  $\rightarrow$ To obtain formal proofs, we have to round the solution of the SDP to the rationals, while ensuring *exact* feasibility, and positivity of the Gram matrix [Peyrl-Parrilo]

- Solve slightly perturbed SDP, and round the solution of the perturbed SDP

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

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

Examples

- Simple walk on K_n = complete graph
- 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}

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**

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

- Relaxation based on the Taylor upper bound of degree 5 + symmetry reduction of the SDP + rounding in $\mathbb{Q}[\cos(2\pi/n)]$

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!