Modified log-Sobolev inequalities for strongly log-concave distributions

Heng Guo (University of Edinburgh)

Joint with Mary Cryan and Giorgos Mousa (Edinburgh)

Tsinghua University

Jun 25th, 2019

Strongly log-concave distributions

Discrete log-concave distribution

What is the correct definition of a log-concave distribution?

What about 1 dimension? For $\pi : [n] \to \mathbb{R}_{\geqslant 0}$, $\pi(i+1)\pi(i-1) \leqslant \pi(i)^2$?

Discrete log-concave distribution

What is the correct definition of a log-concave distribution?

What about 1 dimension? For $\pi : [n] \to \mathbb{R}_{\geqslant 0}$, $\pi(i+1)\pi(i-1) \leqslant \pi(i)^2$?

Consider $\pi(1) = 1/2, \pi(n) = 1/2$ and all other $\pi(i)$ are 0.

This distribution satisfies the condition, but it is not even unimodal.

What about high dimensions?

Log-concave polynomial

A polynomial $p \in \mathbb{R}_{\geqslant 0}[x_1, \ldots, x_n]$ is log-concave (at $\mathbf{x}$) if the Hessian $\nabla^2 \log p(\mathbf{x})$ is negative semi-definite.

$\Rightarrow \quad \nabla^2 p(\mathbf{x})$ has at most one positive eigenvalue.

Strongly log-concave polynomial

A polynomial $p \in \mathbb{R}_{\geqslant 0}[x_1, \ldots, x_n]$ is strongly log-concave if for any index set $I \subseteq [n]$, $\partial_I p$ is log-concave at $\mathbf{1}$.

Originally introduced by Gurvitz (2009), equivalent to:

- completely log-concave (Anari, Oveis Gharan, and Vinzant, 2018);
- Lorentzian polynomials (Brändén and Huh, 2019+).

Strongly log-concave distributions

A distribution $\pi : 2^{[n]} \to \mathbb{R}_{\geqslant 0}$ is strongly log-concave if so is its generating polynomial

$$g_\pi(\mathbf{x}) = \sum_{S \subseteq [n]} \pi(S) \prod_{i \in S} x_i.$$

An important example of homogeneous strongly log-concave distributions is the uniform distribution over bases of a matroid (Anari, Oveis Gharan, and Vinzant 2018; Brändén and Huh 2019+).

Matroid

A matroid $\mathcal{M} = (E, \mathcal{I})$ consists of a finite ground set E and a collection $\mathcal{I}$ of subsets of E (independent sets) such that:

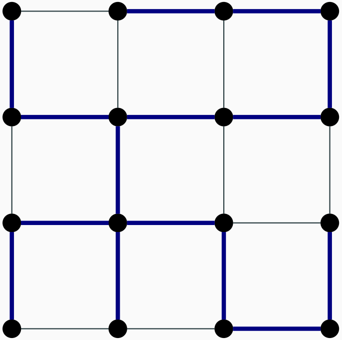
- $\emptyset \in \mathcal{I}$;
- if $S \in \mathcal{I}$, $T \subseteq S$, then $T \in \mathcal{I}$ (downward closed);
- if $S, T \in \mathcal{I}$ and $|S| > |T|$, then there exists an element $i \in S \setminus T$ such that $T \cup \{i\} \in \mathcal{I}$.

Maximum independent sets are the bases. For any two bases, there is a sequence of exchanges of ground set elements from one to the other.

Let $n = |E|$ and r be the rank, namely the size of any basis.

Example — graphic matroids

Spanning trees for graphs form the bases of graphic matroids.



Nelson (2018): Almost all matroids are non-representable!

Alternative characterisation for SLC

Brändén and Huh (2019+): An r-homogeneous multiaffine polynomial p with non-negative coefficients is strongly log-concave if and only if:

- the support of p is a matroid;
- after taking $r-2$ partial derivatives, the quadratic is real stable or $\mathbf{0}$.

Real stable: $p(\mathbf{x}) \neq 0$ if $\Im(x_i) > 0$ for all i.

Real stable polynomials (and strongly Rayleigh distributions) capture only "balanced" matroids, whereas SLC polynomials capture all matroids.

The following Markov chain $P_{BX,\pi}$ converges to a homogeneous SLC π:

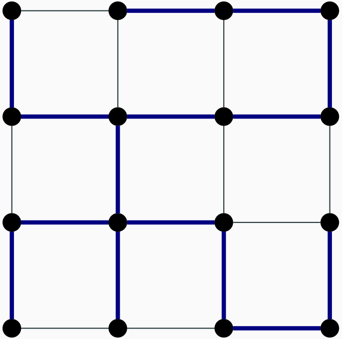
1. remove an element uniformly at random from the current basis (call the resulting set S);
2. add $i \notin S$ with probability proportional to $\pi(S \cup \{i\})$.

The implementation of the second step may be non-trivial.

The mixing time measures the convergence rate of a Markov chain:

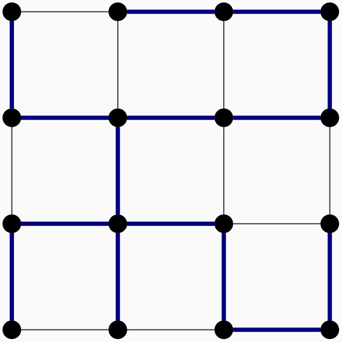
$$t_{mix}(P, \varepsilon) := \min_t \left\{ t \mid \|P^t(x_0, \cdot) - \pi\|_{TV} \leqslant \varepsilon \right\}.$$

Example — bases-exchange



1. Remove an edge uniformly at random;
2. Add back one of the available choices uniformly at random.

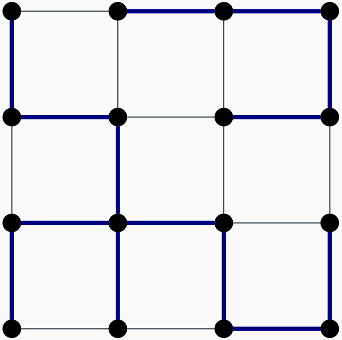
Example — bases-exchange



→ **1.** Remove an edge uniformly at random;

2. Add back one of the available choices uniformly at random.

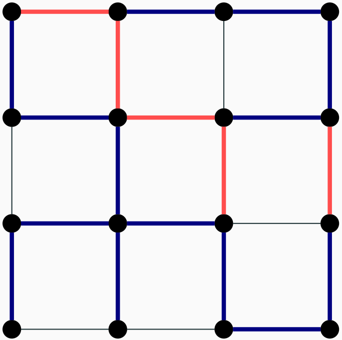
Example — bases-exchange



→ **1.** Remove an edge uniformly at random;

2. Add back one of the available choices uniformly at random.

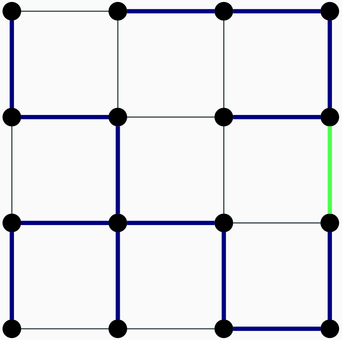
Example — bases-exchange



1. Remove an edge uniformly at random;

→ 2. Add back one of the available choices uniformly at random.

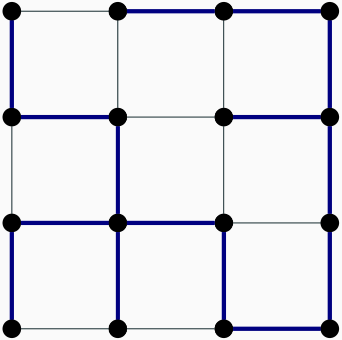
Example — bases-exchange



1. Remove an edge uniformly at random;

→ 2. Add back one of the available choices uniformly at random.

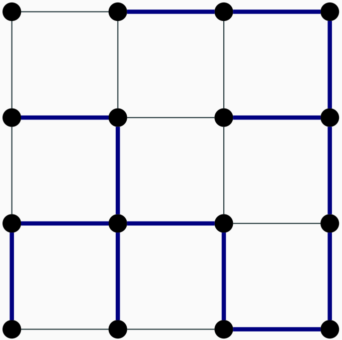
Example — bases-exchange



→ 1. Remove an edge uniformly at random;

2. Add back one of the available choices uniformly at random.

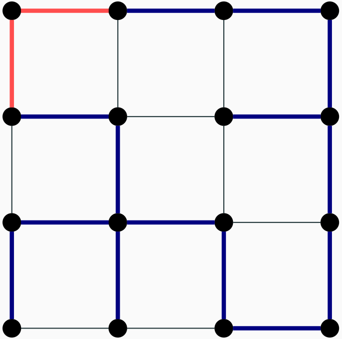
Example — bases-exchange



→ **1.** Remove an edge uniformly at random;

2. Add back one of the available choices uniformly at random.

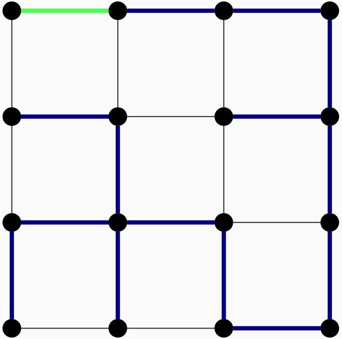
Example — bases-exchange



1. Remove an edge uniformly at random;

→ 2. Add back one of the available choices uniformly at random.

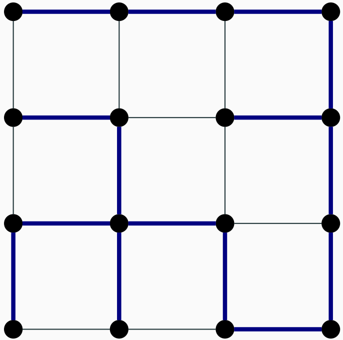
Example — bases-exchange



1. Remove an edge uniformly at random;

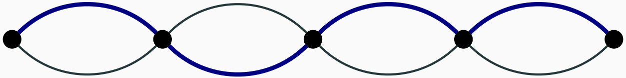
→ 2. Add back one of the available choices uniformly at random.

Example — bases-exchange



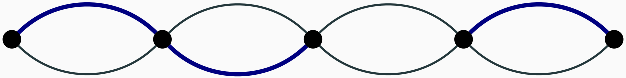
1. Remove an edge uniformly at random;
2. Add back one of the available choices uniformly at random.

Example — bases-exchange



1. Remove an edge uniformly at random;
2. Add back one of the two choices uniformly at random.

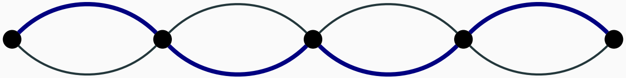
Example — bases-exchange



→ **1.** Remove an edge uniformly at random;

2. Add back one of the two choices uniformly at random.

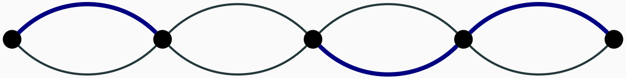
Example — bases-exchange



1. Remove an edge uniformly at random;

→ 2. Add back one of the two choices uniformly at random.

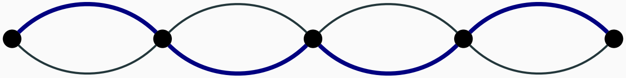
Example — bases-exchange



→ **1.** Remove an edge uniformly at random;

2. Add back one of the two choices uniformly at random.

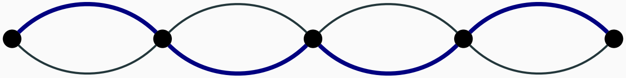
Example — bases-exchange



1. Remove an edge uniformly at random;

→ 2. Add back one of the two choices uniformly at random.

Example — bases-exchange



1. Remove an edge uniformly at random;
2. Add back one of the two choices uniformly at random.

If we encode the state as a binary string, then this is just the lazy random walk on the Boolean hypercube $\{0, 1\}^r$.

(The rank of this matroid is r and the ground set has size $n = 2r$.)

The mixing time is $\Theta(r \log r)$.

Main result — mixing time

Theorem (mixing time)

For any r-homogeneous strongly log-concave distribution π,

$$t_{\mathrm{mix}}(P_{BX,\pi},\varepsilon) \leqslant r\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right),$$

where $\pi_{\min} = \min_{x\in\Omega}\pi(x)$.

Previously, Anari, Liu, Oveis Gharan, and Vinzant (2019):

$$t_{\mathrm{mix}}(P_{BX,\pi},\varepsilon) \leqslant r\left(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\varepsilon}\right)$$

E.g. for the uniform distribution over bases of matroids (with n elements and rank r), our bound is $O(r(\log r + \log\log n))$, whereas the previous bound is $O(r^2\log n)$.

The bound is asymptotically optimal, shown by the previous example.

Main result — concentration

Theorem (concentration bounds)

Let π and $P_{BX,\pi}$ be as before, and Ω be the support of π. For any observable function $f : \Omega \to \mathbb{R}$ and $a \geqslant 0$,

$$\Pr_{x\sim\pi}\left(|f(x) - \mathbb{E}_\pi f| \geqslant a\right) \leqslant 2\exp\left(-\frac{a^2}{2rv(f)}\right),$$

where $v(f)$ is the maximum of one-step variances

$$v(f) := \max_{x\in\Omega}\left\{\sum_{y\in\Omega} P_{BX,\pi}(x,y)(f(x)-f(y))^2\right\}.$$

For c-Lipschitz function f, $v(f) \leqslant c^2$.

Generalises concentration of Lipschitz functions in strongly Rayleigh distributions by Pemantle and Peres (2014); see also Hermon and Salez (2019+).

Dirichlet form

For a Markov chain P and two functions f and g over the state space Ω,

$$\mathcal{E}_P(f, g) := g^\top \operatorname{diag}(\pi)\mathcal{L}f.$$

(the Laplacian $\mathcal{L} := \mathbf{I} - P$)

For reversible Markov chains,

$$\mathcal{E}_P(f, g) = \frac{1}{2}\sum_{x,y\in\Omega} \pi(x)P(x, y)(f(x) - f(y))(g(x) - g(y)).$$

Modified log-Sobolev inequality

Theorem (modified log-Sobolev inequality)

For any $f : \Omega \to \mathbb{R}_{\geqslant 0}$,

$$\mathcal{E}_{P_{\mathrm{BX},\pi}}(f, \log f) \geqslant \frac{1}{r} \cdot \mathrm{Ent}_\pi(f),$$

Both main results are consequences of this.

$\mathrm{Ent}_\pi(f)$ is defined by

$$\mathrm{Ent}_\pi(f) := \mathbb{E}_\pi(f \circ \log f) - \mathbb{E}_\pi f \cdot \log \mathbb{E}_\pi f.$$

If we normalise $\mathbb{E}_\pi f = 1$, then $\mathrm{Ent}_\pi(f) = D(\pi \circ f \parallel \pi)$, the relative entropy (or Kullback–Leibler divergence) between $\pi \circ f$ and π.

Three "constants"

Poincare constant (spectral gap):

$$\lambda(P) := \inf_{\mathrm{Var}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\lambda(P)} \left(\log \frac{1}{\pi_{\min}} + \log \frac{1}{\varepsilon} \right)$$

log-Sobolev constant (Diaconis and Saloff-Coste, 1996):

$$\alpha(P) := \inf_{\mathrm{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(\sqrt{f},\sqrt{f})}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{4\alpha(P)} \left(\log\log \frac{1}{\pi_{\min}} + \log \frac{1}{2\varepsilon^2} \right)$$

modified log-Sobolev constant (Bobkov and Tetali, 2006):

$$\rho(P) := \inf_{\mathrm{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\rho(P)} \left(\log\log \frac{1}{\pi_{\min}} + \log \frac{1}{2\varepsilon^2} \right)$$

$$2\lambda(P) \geqslant \rho(P) \geqslant 4\alpha(P)$$ (Bobkov and Tetali, 2006)

$$\alpha(P) \leqslant \frac{1}{\log \pi_{\min}^{-1}}$$ (observed by Hermon and Salez, 2019+)

$$\rho(P_{\mathrm{BX},\pi}) \geqslant 1/r$$ (our result)

Three "constants"

Poincare constant (spectral gap):

$$\lambda(P) := \inf_{\mathrm{Var}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\lambda(P)}\left(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\varepsilon}\right)$$

log-Sobolev constant (Diaconis and Saloff-Coste, 1996):

$$\alpha(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(\sqrt{f},\sqrt{f})}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{4\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

modified log-Sobolev constant (Bobkov and Tetali, 2006):

$$\rho(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

$$2\lambda(P) \geqslant \rho(P) \geqslant 4\alpha(P)$$ (Bobkov and Tetali, 2006)

$$\alpha(P) \leqslant \frac{1}{\log \pi_{\min}^{-1}}$$ (observed by Hermon and Salez, 2019+)

$$\rho(P_{BX,\pi}) \geqslant 1/r$$ (our result)

Three "constants"

Poincare constant (spectral gap):

$$\lambda(P) := \inf_{\mathrm{Var}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\lambda(P)}\left(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\varepsilon}\right)$$

log-Sobolev constant (Diaconis and Saloff-Coste, 1996):

$$\alpha(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(\sqrt{f},\sqrt{f})}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{4\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

modified log-Sobolev constant (Bobkov and Tetali, 2006):

$$\rho(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

$$2\lambda(P) \geqslant \rho(P) \geqslant 4\alpha(P)$$ (Bobkov and Tetali, 2006)

$$\alpha(P) \leqslant \frac{1}{\log \pi_{\min}^{-1}}$$ (observed by Hermon and Salez, 2019+)

$$\rho(P_{\mathrm{BX},\pi}) \geqslant 1/\tau$$ (our result)

Three "constants"

Poincare constant (spectral gap):

$$\lambda(P) := \inf_{\mathrm{Var}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\lambda(P)}\left(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\varepsilon}\right)$$

log-Sobolev constant (Diaconis and Saloff-Coste, 1996):

$$\alpha(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(\sqrt{f},\sqrt{f})}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{4\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

modified log-Sobolev constant (Bobkov and Tetali, 2006):

$$\rho(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

$$2\lambda(P) \geqslant \rho(P) \geqslant 4\alpha(P) \qquad \text{(Bobkov and Tetali, 2006)}$$

$$\alpha(P) \leqslant \frac{1}{\log \pi_{\min}^{-1}} \qquad \text{(observed by Hermon and Salez, 2019+)}$$

$$\rho(P_{BX,\pi}) \geqslant 1/r \qquad \text{(our result)}$$

Three "constants"

Poincare constant (spectral gap):

$$\lambda(P) := \inf_{\mathrm{Var}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\lambda(P)}\left(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\varepsilon}\right)$$

log-Sobolev constant (Diaconis and Saloff-Coste, 1996):

$$\alpha(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(\sqrt{f},\sqrt{f})}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{4\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

modified log-Sobolev constant (Bobkov and Tetali, 2006):

$$\rho(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

$$2\lambda(P) \geqslant \rho(P) \geqslant 4\alpha(P)$$ (Bobkov and Tetali, 2006)

$$\alpha(P) \leqslant \frac{1}{\log \pi_{\min}^{-1}}$$ (observed by Hermon and Salez, 2019+)

$$\rho(P_{BX,\pi}) \geqslant 1/r$$ (our result)

Three "constants"

Poincare constant (spectral gap):

$$\lambda(P) := \inf_{\mathrm{Var}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\lambda(P)}\left(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\varepsilon}\right)$$

log-Sobolev constant (Diaconis and Saloff-Coste, 1996):

$$\alpha(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(\sqrt{f},\sqrt{f})}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{4\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

modified log-Sobolev constant (Bobkov and Tetali, 2006):

$$\rho(P) := \inf_{\mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)}, \qquad t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$$

$$2\lambda(P) \geqslant \rho(P) \geqslant 4\alpha(P)$$ (Bobkov and Tetali, 2006)

$$\alpha(P) \leqslant \frac{1}{\log \pi_{\min}^{-1}}$$ (observed by Hermon and Salez, 2019+)

$$\rho(P_{BX,\pi}) \geqslant 1/r$$ (our result)

Decay of relative entropy

Stratification

The set of all independent sets of a matroid $\mathcal{M}$ is downward closed.

Let $\mathcal{M}(k)$ be the set of independent sets of size k. Thus, $\mathcal{M}(r)$ is the set of all bases.

Let $\mathcal{M}_i$ denote the matroid $\mathcal{M}$ after contracting i, which is another matroid itself.

Weights for independent sets

We equip $\mathcal{M}$ with the following inductively defined weight function:

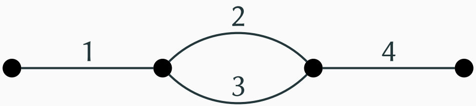
$$w(I) := \begin{cases} \pi(I) Z_r & \text{if } |I| = r, \\ \sum_{I' \supset I,\ |I'| = |I|+1} w(I') & \text{if } |I| < r, \end{cases}$$

for some normalisation constant $Z_r > 0$.

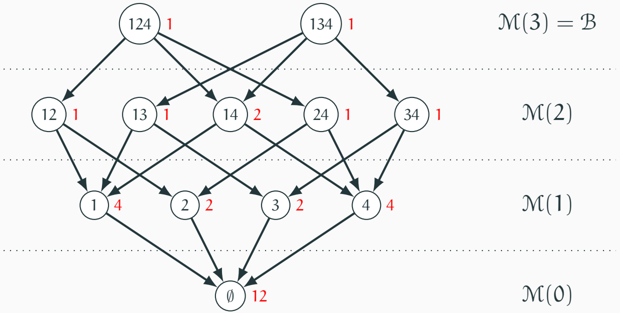
For example, we may choose $w(B) = 1$ for all $B \in \mathcal{B}$ and $Z_r = |\mathcal{B}|$, which corresponds to the uniform distribution over $\mathcal{B}$.

Let π_k be the distribution such that $\pi_k(I) \propto w(I)$, and Z_k be the corresponding normalising constant.

Example



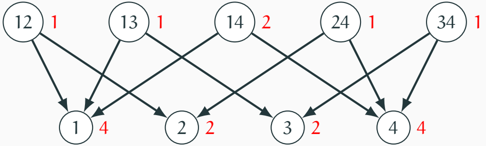
Independent sets of the matroid:



Three views

Polynomial	Matroid	Distribution
$\frac{\partial}{\partial x_i} p$	contraction over i	conditioning on having i
set $x_i = 0$	deletion of i	conditioning on not having i
$(r-k)! \cdot \partial_I p(\mathbf{1})$	$w(I)$	$\propto \pi_k(I)$
$p(\mathbf{1})$	$\lvert\mathcal{B}\rvert$	$\pi_0(\emptyset) = 1$

Random walk between levels



There are two natural random walks converging to π_k.

The "down-up" random walk $P_k^{\vee}$:

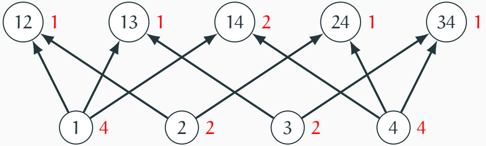
→ **1.** remove an element of $I \in \mathcal{M}(k)$ uniformly at random to get $I' \in \mathcal{M}(k-1)$;

2. move to J such that $J \in \mathcal{M}(k), J \supset I'$ with probability $\frac{w(J)}{w(I')}$.

The bases-exchange walk $P_{BX,\pi} = P_r^{\vee}$.

The "up-down" walk $P_k^{\wedge}$ is defined similarly.

Random walk between levels



There are two natural random walks converging to π_k.

The "down-up" random walk $P_k^{\vee}$:

1. remove an element of $I \in \mathcal{M}(k)$ uniformly at random to get $I' \in \mathcal{M}(k-1)$;
2. move to J such that $J \in \mathcal{M}(k), J \supset I'$ with probability $\frac{w(J)}{w(I')}$.

The bases-exchange walk $P_{BX,\pi} = P_r^{\vee}$.

The "up-down" walk $P_k^{\wedge}$ is defined similarly.

Decomposing the walks

Let A_k be the matrix whose rows are indexed by $\mathcal{M}(k)$ and columns by $\mathcal{M}(k+1)$ such that $A_k(I, J) = 1$ if and only if $I \subset J$.

Let $\mathbf{w}_k = \{w(I)\}_{I \in \mathcal{M}(k)}$, and

$$P^{\downarrow}_{k+1} := \frac{1}{k+1} \cdot A_k^{\top};$$
$$P^{\uparrow}_{k} := \operatorname{diag}(\mathbf{w}_k)^{-1} A_k \operatorname{diag}(\mathbf{w}_{k+1}).$$

We have

$$P^{\vee}_{k+1} = P^{\downarrow}_{k+1} P^{\uparrow}_{k};$$
$$P^{\wedge}_{k} = P^{\uparrow}_{k} P^{\downarrow}_{k+1}.$$

Decomposing the walks

Let A_k be the matrix whose rows are indexed by $\mathcal{M}(k)$ and columns by $\mathcal{M}(k+1)$ such that $A_k(I, J) = 1$ if and only if $I \subset J$.

Let $\mathbf{w}_k = \{w(I)\}_{I \in \mathcal{M}(k)}$, and

$$P^{\downarrow}_{k+1} := \frac{1}{k+1} \cdot A_k^{\top};$$
$$P^{\uparrow}_k := \mathrm{diag}(\mathbf{w}_k)^{-1} A_k \, \mathrm{diag}(\mathbf{w}_{k+1}).$$

We have

$$P^{\vee}_{k+1} = P^{\downarrow}_{k+1} P^{\uparrow}_k;$$
$$P^{\wedge}_k = P^{\uparrow}_k P^{\downarrow}_{k+1}.$$

Decomposing the walks

Let A_k be the matrix whose rows are indexed by $\mathcal{M}(k)$ and columns by $\mathcal{M}(k+1)$ such that $A_k(I, J) = 1$ if and only if $I \subset J$.

Let $\mathbf{w}_k = \{w(I)\}_{I \in \mathcal{M}(k)}$, and

$$\begin{aligned} P^{\downarrow}_{k+1} &:= \frac{1}{k+1} \cdot A_k^{\top}; \\ P^{\uparrow}_{k} &:= \mathrm{diag}(\mathbf{w}_k)^{-1} A_k \,\mathrm{diag}(\mathbf{w}_{k+1}). \end{aligned}$$

We have

$$\begin{aligned} P^{\vee}_{k+1} &= P^{\downarrow}_{k+1} P^{\uparrow}_{k}; \\ P^{\wedge}_{k} &= P^{\uparrow}_{k} P^{\downarrow}_{k+1}. \end{aligned}$$

Decomposing the walks

Let A_k be the matrix whose rows are indexed by $\mathcal{M}(k)$ and columns by $\mathcal{M}(k+1)$ such that $A_k(I, J) = 1$ if and only if $I \subset J$.

Let $\mathbf{w}_k = \{w(I)\}_{I \in \mathcal{M}(k)}$, and

$$P^{\downarrow}_{k+1} := \frac{1}{k+1} \cdot A_k^{\intercal};$$
$$P^{\uparrow}_{k} := \operatorname{diag}(\mathbf{w}_k)^{-1} A_k \operatorname{diag}(\mathbf{w}_{k+1}).$$

We have

$$P^{\vee}_{k+1} = {\color{red}P^{\downarrow}_{k+1}} P^{\uparrow}_{k};$$
$$P^{\wedge}_{k} = P^{\uparrow}_{k} {\color{red}P^{\downarrow}_{k+1}}.$$

Key lemma

Lemma

For any $k \geqslant 2$ *and* $f : \mathcal{M}(k) \to \mathbb{R}_{\geqslant 0}$,

$$\frac{\mathrm{Ent}_{\pi_k}(f)}{k} \geqslant \frac{\mathrm{Ent}_{\pi_{k-1}}(P^{\uparrow}_{k-1} f)}{k-1}.$$

- If $\mathbb{E}_{\pi_k} f = 1$, then $\pi_k \circ f$ is a distribution. View it as a row vector:

$$\pi_{k-1} \circ \left(P^{\uparrow}_{k-1} f\right) = (\pi_k \circ f) P^{\downarrow}_k.$$

  So applying $P^{\uparrow}_{k-1}$ to the left corresponds to the random walk $P^{\downarrow}_k$.
- Then the lemma is saying that $P^{\downarrow}_k$ contracts the relative entropy by at least $(1 - 1/k)$.

Key lemma

Lemma

For any $k \geqslant 2$ *and* $f : \mathcal{M}(k) \to \mathbb{R}_{\geqslant 0}$,

$$\frac{\mathrm{Ent}_{\pi_k}(f)}{k} \geqslant \frac{\mathrm{Ent}_{\pi_{k-1}}(P^{\uparrow}_{k-1} f)}{k-1}.$$

- If $\mathbb{E}_{\pi_k} f = 1$, then $\pi_k \circ f$ is a distribution. View it as a row vector:

$$\pi_{k-1} \circ \left(P^{\uparrow}_{k-1} f\right) = (\pi_k \circ f) P^{\downarrow}_k.$$

  So applying $P^{\uparrow}_{k-1}$ to the left corresponds to the random walk $P^{\downarrow}_k$.

- Then the lemma is saying that $P^{\downarrow}_k$ contracts the relative entropy by at least $(1 - 1/k)$.

Key lemma

Lemma

For any $k \geqslant 2$ *and* $f : \mathcal{M}(k) \to \mathbb{R}_{\geqslant 0}$,

$$\frac{\mathrm{Ent}_{\pi_k}(f)}{k} \geqslant \frac{\mathrm{Ent}_{\pi_{k-1}}(P_{k-1}^{\uparrow} f)}{k-1}.$$

- If $\mathbb{E}_{\pi_k} f = 1$, then $\pi_k \circ f$ is a distribution. View it as a row vector:

$$\pi_{k-1} \circ \left(P_{k-1}^{\uparrow} f\right) = (\pi_k \circ f) P_k^{\downarrow}.$$

  So applying $P_{k-1}^{\uparrow}$ to the left corresponds to the random walk $P_k^{\downarrow}$.

- Then the lemma is saying that $P_k^{\downarrow}$ contracts the relative entropy by at least $(1 - 1/k)$.

Base case

For the base case, we want to show that

$$\mathrm{Ent}_{\pi_2}(f) - 2\mathrm{Ent}_{\pi_1}(P_1^{\uparrow} f) \geqslant 0.$$

Using $a \log \frac{a}{b} \geqslant a - b$ for $a, b > 0$, we can get

$$\mathrm{Ent}_{\pi_2}(f) - 2\mathrm{Ent}_{\pi_1}(P_1^{\uparrow} f) \geqslant 1 - \frac{1}{2Z_2} \cdot h^{\mathsf{T}} W h,$$

where $W_{ij} = w(\{i, j\})$ and $h = P_1^{\uparrow} f$.

Since $W = (r-2)! Z_r \nabla^2 g_\pi(\mathbf{1})$, it has at most one positive eigenvalue. The quadratic form is maximised at $h = P_1^{\uparrow} f = \mathbf{1}$, which proves the base case.

Base case

For the base case, we want to show that

$$\mathrm{Ent}_{\pi_2}(f) - 2\mathrm{Ent}_{\pi_1}(P_1^{\uparrow} f) \geqslant 0.$$

Using $a \log \frac{a}{b} \geqslant a - b$ for $a, b > 0$, we can get

$$\mathrm{Ent}_{\pi_2}(f) - 2\mathrm{Ent}_{\pi_1}(P_1^{\uparrow} f) \geqslant 1 - \frac{1}{2Z_2} \cdot h^{\mathsf{T}} W h,$$

where $W_{ij} = w(\{i, j\})$ and $h = P_1^{\uparrow} f$.

Since $W = (r-2)! Z_r \nabla^2 g_\pi(\mathbf{1})$, it has at most one positive eigenvalue. The quadratic form is maximised at $h = P_1^{\uparrow} f = \mathbf{1}$, which proves the base case.

Base case

For the base case, we want to show that

$$\mathrm{Ent}_{\pi_2}(f) - 2\mathrm{Ent}_{\pi_1}(P_1^{\uparrow} f) \geqslant 0.$$

Using $a \log \frac{a}{b} \geqslant a - b$ for $a, b > 0$, we can get

$$\mathrm{Ent}_{\pi_2}(f) - 2\mathrm{Ent}_{\pi_1}(P_1^{\uparrow} f) \geqslant 1 - \frac{1}{2Z_2} \cdot h^{\mathsf{T}} W h,$$

where $W_{ij} = w(\{i, j\})$ and $h = P_1^{\uparrow} f$.

Since $W = (r-2)! Z_r \nabla^2 g_{\pi}(\mathbf{1})$, it has at most one positive eigenvalue. The quadratic form is maximised at $h = P_1^{\uparrow} f = \mathbf{1}$, which proves the base case.

Decomposing π_k

Consider the following process:

1. draws a basis $B \sim \pi$;
2. repeatedly removes an element from the current set uniformly at random for at most r repetitions.

The outcome X_k after removing $r-k$ elements follows exactly π_k.

By the Law of Total Probability,

$$\Pr(X_k = I) = \sum_{i \in \mathcal{M}(1)} \Pr(X_k = I \mid X_1 = \{i\}) \cdot \Pr(X_1 = \{i\}).$$

Noticing that $\Pr(X_k = I \mid X_1 = \{i\}) = \pi_{i,k-1}(I)$ and $\Pr(X_1 = \{i\}) = \pi_1(i)$,

$$\pi_k = \sum_{i \in \mathcal{M}(1)} \pi_{i,k-1} \cdot \pi_1(i).$$

Decomposing π_k

Consider the following process:

1. draws a basis $B \sim \pi$;
2. repeatedly removes an element from the current set uniformly at random for at most r repetitions.

The outcome X_k after removing $r - k$ elements follows exactly π_k.

By the Law of Total Probability,

$$\Pr(X_k = I) = \sum_{i \in \mathcal{M}(1)} \Pr(X_k = I \mid X_1 = \{i\}) \cdot \Pr(X_1 = \{i\}).$$

Noticing that $\Pr(X_k = I \mid X_1 = \{i\}) = \pi_{i,k-1}(I)$ and $\Pr(X_1 = \{i\}) = \pi_1(i)$,

$$\pi_k = \sum_{i \in \mathcal{M}(1)} \pi_{i,k-1} \cdot \pi_1(i).$$

Induction step

The distribution π_k has the decomposition:

$$\pi_k = \sum_{i \in \mathcal{M}(1)} \pi_1(i) \cdot \pi_{i,k-1}.$$

This leads to a decomposition of relative entropy:

$$\mathrm{Ent}_{\pi_k}(f) = \sum_{i \in \mathcal{M}(1)} \pi_1(i) \mathrm{Ent}_{\pi_{i,k-1}}(f) + \mathrm{Ent}_{\pi_1}(f^{(1)}).$$

where $f^{(1)}(i) := \mathbb{E}_{\pi_{i,k-1}} f$. In fact, $f^{(1)} = \prod_{j=1}^{k-1} P_j^{\uparrow} f$.

Induction step (cont.)

As $f^{(1)} = \prod_{j=1}^{k-1} P_j^{\uparrow} f$,

$$\mathrm{Ent}_{\pi_k}(f) = \sum_{i \in \mathcal{M}(1)} \pi_1(i) \mathrm{Ent}_{\pi_{i,k-1}}(f) + \mathrm{Ent}_{\pi_1}(f^{(1)})$$

$$\mathrm{Ent}_{\pi_{k-1}}(P_{k-1}^{\uparrow} f) = \sum_{i \in \mathcal{M}(1)} \pi_1(i) \mathrm{Ent}_{\pi_{i,k-2}}(P_{k-1}^{\uparrow} f) + \mathrm{Ent}_{\pi_1}(f^{(1)})$$

Induction hypothesis on $\mathcal{M}_i$ implies that

$$\mathrm{Ent}_{\pi_{i,k-1}}(f) \geqslant \frac{k-1}{k-2} \cdot \mathrm{Ent}_{\pi_{i,k-2}}(P_{k-1}^{\uparrow} f).$$

Induction hypothesis from $\mathcal{M}(k-1)$ to $\mathcal{M}(1)$ implies that

$$\sum_{i \in \mathcal{M}(1)} \pi_1(i) \mathrm{Ent}_{\pi_{i,k-2}}(P_{k-1}^{\uparrow} f) \geqslant (k-2) \mathrm{Ent}_{\pi_1}(f^{(1)}).$$

Finally, notice that

$$\frac{k-1}{k-2} = \frac{k}{k-1} + \frac{1}{(k-1)(k-2)}.$$

Recap

We have shown entropy contraction from level k to level $k-1$:

$$\frac{\mathrm{Ent}_{\pi_k}(f)}{k} \geqslant \frac{\mathrm{Ent}_{\pi_{k-1}}(P^{\uparrow}_{k-1} f)}{k-1}.$$

It is straightforward from this to derive the modified log-Sobolev inequality, with the help of Jensen's inequality.

Bound the mixing time directly

For a distribution τ on $\mathcal{M}(k)$, the relative entropy $D(\tau \parallel \pi_k) = \mathrm{Ent}_{\pi_k}(D_k^{-1}\tau)$ where $D_k = \mathrm{diag}(\pi_k)$. Moreover, after one step of $P_k^{\vee}$, the distribution is $(\tau^{\mathsf{T}} P_k^{\vee})^{\mathsf{T}} = (P_k^{\vee})^{\mathsf{T}}\tau$. Since $P_k^{\vee}$ is reversible, $D_k^{-1}(P_k^{\vee})^{\mathsf{T}} = P_k^{\vee} D_k^{-1}$.

$$\begin{aligned}
D\left((P_k^{\vee})^{\mathsf{T}}\tau \parallel \pi_k\right) &= \mathrm{Ent}_{\pi_k}(D_k^{-1}(P_k^{\vee})^{\mathsf{T}}\tau) \\
&= \mathrm{Ent}_{\pi_k}(P_k^{\vee} D_k^{-1}\tau) \\
&= \mathrm{Ent}_{\pi_k}(P_k^{\downarrow} P_{k-1}^{\uparrow} D_k^{-1}\tau) \\
&\leqslant \mathrm{Ent}_{\pi_{k-1}}(P_{k-1}^{\uparrow} D_k^{-1}\tau) && \text{(Jensen's inequality)} \\
&\leqslant \left(1 - \frac{1}{k}\right) \mathrm{Ent}_{\pi_k}(D_k^{-1}\tau) && \text{(entropy contraction)} \\
&= \left(1 - \frac{1}{k}\right) D(\tau \parallel \pi_k).
\end{aligned}$$

The mixing time bound follows from Pinsker's inequality

$$2\,\|\tau - \sigma\|_{\mathrm{TV}}^2 \leqslant D(\tau \parallel \sigma).$$

Herbst argument

The Herbst argument is a standard trick to get sub-Gaussian concentration bounds from log-Sobolev inequalities.

The key is to show, for $t > 0$ and $c = \frac{\nu(f)}{\rho(P)}$,

$$\mathbb{E}[e^{tf}] \leqslant e^{t\,\mathbb{E}\,f + ct^2}.$$

Let $F_t := e^{tf - ct^2}$. Then we just need to show $\frac{\log \mathbb{E}[F_t]}{t} \leqslant \mathbb{E}\,f$. This, in turn, follows from the claim that $t \mapsto \frac{\log \mathbb{E}[F_t]}{t}$ is non-increasing.

Note that

$$\frac{d}{dt}\left(\frac{\log \mathbb{E}[F_t]}{t}\right) = \frac{\mathrm{Ent}_\pi(F_t) - ct^2\,\mathbb{E}[F_t]}{t^2\,\mathbb{E}[F_t]}.$$

The following inequalities thus finish the argument

$$\mathrm{Ent}_\pi(F_t) \leqslant \frac{1}{\rho(P)}\mathcal{E}_P(F_t, \log F_t) \leqslant \frac{t^2\nu(f)}{2\rho(P)}\,\mathbb{E}[F_t].$$

Herbst argument

The Herbst argument is a standard trick to get sub-Gaussian concentration bounds from log-Sobolev inequalities.

The key is to show, for $t > 0$ and $c = \frac{\nu(f)}{\rho(P)}$,

$$\mathbb{E}[e^{tf}] \leqslant e^{t\,\mathbb{E}\,f + ct^2}.$$

Let $F_t := e^{tf - ct^2}$. Then we just need to show $\frac{\log \mathbb{E}[F_t]}{t} \leqslant \mathbb{E}\,f$. This, in turn, follows from the claim that $t \mapsto \frac{\log \mathbb{E}[F_t]}{t}$ is non-increasing.

Note that

$$\frac{d}{dt}\left(\frac{\log \mathbb{E}[F_t]}{t}\right) = \frac{\mathrm{Ent}_\pi(F_t) - ct^2\,\mathbb{E}[F_t]}{t^2\,\mathbb{E}[F_t]}.$$

The following inequalities thus finish the argument

$$\mathrm{Ent}_\pi(F_t) \leqslant \frac{1}{\rho(P)}\mathcal{E}_P(F_t, \log F_t) \leqslant \frac{t^2\nu(f)}{2\rho(P)}\,\mathbb{E}[F_t].$$

Herbst argument

The Herbst argument is a standard trick to get sub-Gaussian concentration bounds from log-Sobolev inequalities.

The key is to show, for $t > 0$ and $c = \frac{\nu(f)}{\rho(P)}$,

$$\mathbb{E}[e^{tf}] \leqslant e^{t\,\mathbb{E}\,f + ct^2}.$$

Let $F_t := e^{tf - ct^2}$. Then we just need to show $\frac{\log \mathbb{E}[F_t]}{t} \leqslant \mathbb{E}\,f$. This, in turn, follows from the claim that $t \mapsto \frac{\log \mathbb{E}[F_t]}{t}$ is non-increasing.

Note that

$$\frac{d}{dt}\left(\frac{\log \mathbb{E}[F_t]}{t}\right) = \frac{\mathrm{Ent}_\pi(F_t) - ct^2\,\mathbb{E}[F_t]}{t^2\,\mathbb{E}[F_t]}.$$

The following inequalities thus finish the argument

$$\mathrm{Ent}_\pi(F_t) \leqslant \frac{1}{\rho(P)}\mathcal{E}_P(F_t, \log F_t) \leqslant \frac{t^2\nu(f)}{2\rho(P)}\,\mathbb{E}[F_t].$$

Herbst argument

The Herbst argument is a standard trick to get sub-Gaussian concentration bounds from log-Sobolev inequalities.

The key is to show, for $t > 0$ and $c = \frac{\nu(f)}{\rho(P)}$,

$$\mathbb{E}[e^{tf}] \leqslant e^{t\,\mathbb{E}\,f + ct^2}.$$

Let $F_t := e^{tf - ct^2}$. Then we just need to show $\frac{\log \mathbb{E}[F_t]}{t} \leqslant \mathbb{E}\,f$. This, in turn, follows from the claim that $t \mapsto \frac{\log \mathbb{E}[F_t]}{t}$ is non-increasing.

Note that

$$\frac{d}{dt}\left(\frac{\log \mathbb{E}[F_t]}{t}\right) = \frac{\mathrm{Ent}_\pi(F_t) - ct^2\,\mathbb{E}[F_t]}{t^2\,\mathbb{E}[F_t]}.$$

The following inequalities thus finish the argument

$$\mathrm{Ent}_\pi(F_t) \leqslant \frac{1}{\rho(P)}\mathcal{E}_P(F_t, \log F_t) \leqslant \frac{t^2\nu(f)}{2\rho(P)}\,\mathbb{E}[F_t].$$

Herbst argument

The Herbst argument is a standard trick to get sub-Gaussian concentration bounds from log-Sobolev inequalities.

The key is to show, for $t > 0$ and $c = \frac{\nu(f)}{\rho(P)}$,

$$\mathbb{E}[e^{tf}] \leqslant e^{t\,\mathbb{E}\,f + ct^2}.$$

Let $F_t := e^{tf - ct^2}$. Then we just need to show $\frac{\log \mathbb{E}[F_t]}{t} \leqslant \mathbb{E}\,f$. This, in turn, follows from the claim that $t \mapsto \frac{\log \mathbb{E}[F_t]}{t}$ is non-increasing.

Note that

$$\frac{d}{dt}\left(\frac{\log \mathbb{E}[F_t]}{t}\right) = \frac{\mathrm{Ent}_\pi(F_t) - ct^2\,\mathbb{E}[F_t]}{t^2\,\mathbb{E}[F_t]}.$$

The following inequalities thus finish the argument

$$\mathrm{Ent}_\pi(F_t) \leqslant \frac{1}{\rho(P)}\mathcal{E}_P(F_t, \log F_t) \leqslant \frac{t^2\nu(f)}{2\rho(P)}\,\mathbb{E}[F_t].$$

Concluding remarks

Why strongly log-concave?

Apparently, strong log-concavity was used in two places:

- Base case: log-concavity;
- Inductive step: closure property under contractions.

The approach should still work with some distribution property that is closed under contractions (namely conditioning) but has perhaps a "weaker" base case.

Entropy decomposition

- The decomposition of $\mathrm{Ent}_{\pi_k}(f)$ seems to be the key to our argument. This differs from the traditional Markov chain decomposition techniques, where the state space is partitioned.

- Is there a more general technique?

An oddity

Recall

$$P^{\vee}_{k+1} = P^{\downarrow}_{k+1} P^{\uparrow}_{k};$$
$$P^{\wedge}_{k} = P^{\uparrow}_{k} P^{\downarrow}_{k+1}.$$

Their spectral gaps are the same: $\lambda(P^{\vee}_{k+1}) = \lambda(P^{\wedge}_{k})$.

For modified log-Sobolev constants, we showed

$$\rho(P^{\vee}_{k+1}) \geqslant \frac{1}{k+1}, \qquad \rho(P^{\wedge}_{k}) \geqslant \frac{1}{k+1},$$

but

$$\rho(P^{\vee}_{k+1}) = \rho(P^{\wedge}_{k})?$$

Open problems

- Fast implementation of the (modified) bases-exchange?
- An $\Omega(r \log r)$ lower bound of the mixing time?
- Deterministic counting algorithms?
  - What can we say about the zeros of (inhomogeneous) SLC polynomials? E.g. the reliability polynomial?
- Common bases / independent sets of matroids?

Open problems

- Fast implementation of the (modified) bases-exchange?
- An $\Omega(r \log r)$ lower bound of the mixing time?
- Deterministic counting algorithms?
    - What can we say about the zeros of (inhomogeneous) SLC polynomials? E.g. the reliability polynomial?
- Common bases / independent sets of matroids?

Open problems

- Fast implementation of the (modified) bases-exchange?
- An $\Omega(r \log r)$ lower bound of the mixing time?
- Deterministic counting algorithms?
    - What can we say about the zeros of (inhomogeneous) SLC polynomials? E.g. the reliability polynomial?
- Common bases / independent sets of matroids?

Open problems

- Fast implementation of the (modified) bases-exchange?
- An $\Omega(r \log r)$ lower bound of the mixing time?
- Deterministic counting algorithms?
    - What can we say about the zeros of (inhomogeneous) SLC polynomials? E.g. the reliability polynomial?
- Common bases / independent sets of matroids?

A professor is one who can speak on any subject for precisely fifty minutes.

— Norbert Wiener

Thank you!

arXiv:1903.06081