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

What about 1 dimension? For $\pi : [n] \to \mathbb{R}_{\geq 0}$, $\pi(i + 1)\pi(i - 1) \leq \pi(i)^2$?
Discrete log-concave distribution

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

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

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?
Strongly log-concave polynomials

Log-concave polynomial

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

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

Strongly log-concave polynomial

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

Originally introduced by Gurvitz (2009), equivalent to:

- completely log-concave (Anari, Oveis Gharan, and Vinzant, 2018);
- Lorentzian polynomials (Brändén and Huh, 2019+).
A distribution $\pi : 2^{[n]} \to \mathbb{R}_{\geq 0}$ is strongly log-concave if so is its generating polynomial

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

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

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

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.
Spanning trees for graphs form the bases of graphic matroids.

Nelson (2018): Almost all matroids are non-representable!
Brändén and Huh (2019+): An $r$-homogeneous multi-affine 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 0.

Real stable: $p(x) \neq 0$ 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 $\pi$:

1. **remove** an element uniformly at random from the current basis (call the resulting set $S$);
2. **add** $i \not\in S$ 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:

$$t_{\text{mix}}(P, \varepsilon) := \min_{t} \left\{ t \mid \|P^t(x_0, \cdot) - \pi\|_{TV} \leq \varepsilon \right\}.$$
Example — bases-exchange

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

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


1. Remove an edge uniformly at random;

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

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

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

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

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

2. Add back one of the available choices uniformly at random.
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 \( \{0, 1\} \).

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

The mixing time is \( (r \log r) \).
1. Remove an edge uniformly at random;
2. Add back one of the two choices uniformly at random.
1. Remove an edge uniformly at random;

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

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

2. Add back one of the two choices uniformly at random.
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 \( \pi \),

\[
\tau_{\text{mix}}(P_{BX}, \pi, \varepsilon) \leq r \left( \log \log \frac{1}{\pi_{\min}} + \log \frac{1}{2\varepsilon^2} \right),
\]

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

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

\[
\tau_{\text{mix}}(P_{BX}, \pi, \varepsilon) \leq r \left( \log \frac{1}{\pi_{\min}} + \log \frac{1}{\varepsilon} \right)
\]

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

$$\Pr_{x \sim \pi} (|f(x) - \mathbb{E}_\pi f| \geq a) \leq 2 \exp \left( - \frac{a^2}{2r(v(f))} \right),$$

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 $f$, $v(f) \leq c^2$.

Generalises concentration of Lipschitz functions in strongly Rayleigh distributions by Pemantle and Peres (2014); see also Hermon and Salez (2019+).
For a Markov chain $P$ and two functions $f$ and $g$ over the state space $\Omega$,

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

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

For reversible Markov chains,

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

For any \( f : \Omega \to \mathbb{R}_{\geq 0} \),

\[
\mathcal{E}_{PBX,\pi}(f, \log f) \geq \frac{1}{r} \cdot \text{Ent}_\pi(f),
\]

Both main results are consequences of this.

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

\[
\text{Ent}_\pi(f) := \mathbb{E}_\pi(f \circ \log f) - \mathbb{E}_\pi f \cdot \log \mathbb{E}_\pi f.
\]

If we normalise \( \mathbb{E}_\pi f = 1 \), then \( \text{Ent}_\pi(f) = D(\pi \circ f \| \pi) \), the relative entropy (or Kullback–Leibler divergence) between \( \pi \circ f \) and \( \pi \).
Three “constants”

Poincare constant (spectral gap):

\[
\lambda(P) := \inf_{\text{Var}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f, f)}{\text{Var}_\pi(f)},
\]

\[
t_{\text{mix}}(P, \epsilon) \leq \frac{1}{\lambda(P)} \left( \log \frac{1}{\pi_{\text{min}}} + \log \frac{1}{\epsilon} \right)
\]

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

\[
\alpha(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(\sqrt{f}, \sqrt{f})}{\text{Ent}_\pi(f)},
\]

\[
t_{\text{mix}}(P, \epsilon) \leq \frac{1}{4\alpha(P)} \left( \log \log \frac{1}{\pi_{\text{min}}} + \log \frac{1}{2\epsilon^2} \right)
\]

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

\[
\rho(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f, \log f)}{\text{Ent}_\pi(f)},
\]

\[
t_{\text{mix}}(P, \epsilon) \leq \frac{1}{\rho(P)} \left( \log \log \frac{1}{\pi_{\text{min}}} + \log \frac{1}{2\epsilon^2} \right)
\]

\[
2\lambda(P) \geq \rho(P) \geq 4\alpha(P)
\]  
(Bobkov and Tetali, 2006)

\[
\alpha(P) \leq \frac{1}{\log \pi_{\text{min}}}
\]  
(observed by Hermon and Salez, 2019+)

\[
\rho(P_{\text{BX}, \pi}) \geq 1/r
\]  
(our result)
Poincaré constant (spectral gap):

\[ \lambda(P) := \inf_{\text{Var}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f, f)}{\text{Var}_\pi(f)}, \quad t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\lambda(P)} \left( \log \frac{1}{\pi_{\min}} + \log \frac{1}{\varepsilon} \right) \]

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

\[ \alpha(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(\sqrt{f}, \sqrt{f})}{\text{Ent}_\pi(f)}, \quad t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{4\alpha(P)} \left( \log \log \frac{1}{\pi_{\min}} + \log \frac{1}{2\varepsilon^2} \right) \]

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

\[ \rho(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f, \log f)}{\text{Ent}_\pi(f)}, \quad t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\rho(P)} \left( \log \log \frac{1}{\pi_{\min}} + \log \frac{1}{2\varepsilon^2} \right) \]

\[ 2\lambda(P) \geq \rho(P) \geq 4\alpha(P) \quad (\text{Bobkov and Tetali, 2006}) \]

\[ \alpha(P) \leq \frac{1}{\log \pi_{\min}} \quad (\text{observed by Hermon and Salez, 2019+}) \]

\[ \rho(P_{\text{BX}, \pi}) \geq 1/r \quad (\text{our result}) \]
Three “constants”

Poincare constant (spectral gap):

\[
\lambda(P) := \inf_{\text{Var}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f, f)}{\text{Var}_\pi(f)},
\]

\[
t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\lambda(P)} \left( \log \frac{1}{\tau_{\text{min}}} + \log \frac{1}{\varepsilon} \right)
\]

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

\[
\alpha(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(\sqrt{f}, \sqrt{f})}{\text{Ent}_\pi(f)},
\]

\[
t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{4\alpha(P)} \left( \log \log \frac{1}{\tau_{\text{min}}} + \log \frac{1}{2\varepsilon^2} \right)
\]

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

\[
\rho(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f, \log f)}{\text{Ent}_\pi(f)},
\]

\[
t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\rho(P)} \left( \log \log \frac{1}{\tau_{\text{min}}} + \log \frac{1}{2\varepsilon^2} \right)
\]

\[
2\lambda(P) \geq \rho(P) \geq 4\alpha(P)
\]  

(Bobkov and Tetali, 2006)

\[
\alpha(P) \leq \frac{1}{\log \tau_{\text{min}}}
\]  

(Cheng et al., 2019)

\[
\rho(P_{\text{BS}, \pi}) \geq 1/r
\]  

(our result)

(observed by Hermon and Salez, 2019+).
Three “constants”

Poincaré constant (spectral gap):

$$\lambda(P) := \inf_{\text{Var}(f) \neq 0} \frac{E_P(f, f)}{\text{Var}(f)},$$

$$t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\lambda(P)} \left( \log \frac{1}{\tau_{\min}} + \log \frac{1}{\varepsilon} \right)$$

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

$$\alpha(P) := \inf_{\text{Ent}(f) \neq 0} \frac{E_P(\sqrt{f}, \sqrt{f})}{\text{Ent}(f)},$$

$$t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{4\alpha(P)} \left( \log \log \frac{1}{\tau_{\min}} + \log \frac{1}{2\varepsilon^2} \right)$$

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

$$\rho(P) := \inf_{\text{Ent}(f) \neq 0} \frac{E_P(f, \log f)}{\text{Ent}(f)},$$

$$t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\rho(P)} \left( \log \log \frac{1}{\tau_{\min}} + \log \frac{1}{2\varepsilon^2} \right)$$

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

$$\alpha(P) \leq \frac{1}{\log \tau_{\min}} \quad \text{(observed by Hermon and Salez, 2019+)}$$

$$\rho(P_{BX,\pi}) \geq 1/r \quad \text{(our result)}$$
Three “constants”

Poincare constant (spectral gap):

\[ \lambda(P) := \inf_{\text{Var}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f, f)}{\text{Var}_\pi(f)}, \]

\[ t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\lambda(P)} \left( \log \frac{1}{\pi_{\text{min}}} + \log \frac{1}{\varepsilon} \right) \]

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

\[ \alpha(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(\sqrt{f}, \sqrt{f})}{\text{Ent}_\pi(f)}, \]

\[ t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{4\alpha(P)} \left( \log \log \frac{1}{\pi_{\text{min}}} + \log \frac{1}{2\varepsilon^2} \right) \]

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

\[ \rho(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{\mathcal{E}_P(f, \log f)}{\text{Ent}_\pi(f)}, \]

\[ t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\rho(P)} \left( \log \log \frac{1}{\pi_{\text{min}}} + \log \frac{1}{2\varepsilon^2} \right) \]

\[ 2\lambda(P) \geq \rho(P) \geq 4\alpha(P) \] (Bobkov and Tetali, 2006)

\[ \alpha(P) \leq \frac{1}{\log \pi_{\text{min}}} \] (observed by Hermon and Salez, 2019+)

\[ \rho(P_{BX,\pi}) \geq 1/r \] (our result)
Three “constants”

Poincaré constant (spectral gap):

\[ \lambda(P) := \inf_{\text{Var}_\pi(f) \neq 0} \frac{E_P(f, f)}{\text{Var}_\pi(f)}, \]

\[ t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\lambda(P)} \left( \log \frac{1}{\tau_{\text{min}}} + \log \frac{1}{\varepsilon} \right) \]

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

\[ \alpha(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{E_P(\sqrt{f}, \sqrt{f})}{\text{Ent}_\pi(f)}, \]

\[ t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{4\alpha(P)} \left( \log \log \frac{1}{\tau_{\text{min}}} + \log \frac{1}{2\varepsilon^2} \right) \]

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

\[ \rho(P) := \inf_{\text{Ent}_\pi(f) \neq 0} \frac{E_P(f, \log f)}{\text{Ent}_\pi(f)}, \]

\[ t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\rho(P)} \left( \log \log \frac{1}{\tau_{\text{min}}} + \log \frac{1}{2\varepsilon^2} \right) \]

\[ 2\lambda(P) \geq \rho(P) \geq 4\alpha(P) \]

\[ \alpha(P) \leq \frac{1}{\log \tau_{\text{min}}} \]

\[ \rho(P_{BX, \pi}) \geq 1/\tau \]

(observed by Hermon and Salez, 2019+)

(our result)
Decay of relative entropy
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.
We equip $\mathcal{M}$ with the following inductively defined weight function:

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

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 $\pi_\kappa$ be the distribution such that $\pi_\kappa(I) \propto w(I)$, and $Z_\kappa$ be the corresponding normalising constant.
Independent sets of the matroid:

\[
\mathcal{M}(3) = \mathcal{B}
\]
\[
\mathcal{M}(2)
\]
\[
\mathcal{M}(1)
\]
\[
\mathcal{M}(0)
\]
<table>
<thead>
<tr>
<th>Polynomial</th>
<th>Matroid</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{\partial}{\partial x_i} p )</td>
<td>contraction over ( i )</td>
<td>conditioning on having ( i )</td>
</tr>
<tr>
<td>set ( x_i = 0 )</td>
<td>deletion of ( i )</td>
<td>conditioning on not having ( i )</td>
</tr>
<tr>
<td>((r - k)! \cdot \partial_1 p(1))</td>
<td>( w(I) )</td>
<td>( \propto \pi_k(I) )</td>
</tr>
<tr>
<td>( p(1) )</td>
<td>(</td>
<td>B</td>
</tr>
</tbody>
</table>
There are two natural random walks converging to $\pi_k$.

The “down-up” random walk $P_k^\vee$:

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

→ 2. move to $J$ such that $J \in M(k), J \supset I'$ 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.
There are two natural random walks converging to $\pi_k$.

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.
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 $w_k = \{w(I)\}_{I \in \mathcal{M}(k)}$, and

\[
P_{k+1}^\uparrow := \frac{1}{k+1} A_k^T;
\]

\[
P_k^\uparrow := \text{diag}(w_k)^{-1} A_k \text{diag}(w_{k+1}).
\]

We have

\[
p_{k+1}^\lor = p_{k+1}^\downarrow p_k^\uparrow;
\]

\[
p_k^\land = p_k^\uparrow p_{k+1}^\downarrow.
\]
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 $w_k = \{w(I)\}_{I \in \mathcal{M}(k)}$, and

$$P_{k+1}^\uparrow := \frac{1}{k+1} \cdot A_k^T;$$
$$P_k^\uparrow := \text{diag}(w_k)^{-1} A_k \text{diag}(w_{k+1}).$$

We have

$$p_{k+1}^\lor = p_{k+1}^\downarrow p_k^\uparrow;$$
$$p_k^\land = p_k^\uparrow p_{k+1}^\downarrow.$$
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 $I \subseteq J$.

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

$$P_{k+1}^\downarrow := \frac{1}{k+1} \cdot A_k^\top;$$

$$P_k^\uparrow := \text{diag}(w_k)^{-1} A_k \text{diag}(w_{k+1}).$$

We have

$$P_{k+1}^\vee = P_{k+1}^\downarrow P_k^\uparrow,$$

$$P_k^\wedge = P_k^\uparrow P_{k+1}^\downarrow.$$
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 $I \subseteq J$.

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

$$P_{k+1}^\downarrow := \frac{1}{k+1} \cdot A_k^\top;$$

$$P_k^\uparrow := \text{diag}(w_k)^{-1} A_k \text{diag}(w_{k+1}).$$

We have

$$P_{k+1}^\lor = P_{k+1}^\downarrow P_k^\uparrow,$$

$$P_k^\land = P_k^\uparrow P_{k+1}^\downarrow.$$
Key lemma

**Lemma**

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

\[
\frac{\text{Ent}_{\pi_k}(f)}{k} \geq \frac{\text{Ent}_{\pi_{k-1}}(P_{k-1}^\uparrow f)}{k-1}.
\]

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

  \[
  \pi_{k-1} \circ \left( P_{k-1}^\uparrow f \right) = (\pi_k \circ f)P_{k}^\uparrow.
  \]

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

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

Lemma

For any \( k \geq 2 \) and \( f : M(k) \to \mathbb{R}_{\geq 0} \),

\[
\frac{\text{Ent}_{\pi_k}(f)}{k} \geq \frac{\text{Ent}_{\pi_{k-1}}(P_{k-1}^\uparrow f)}{k-1}.
\]

- 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 \((1 - 1/k)\).
Key lemma

Lemma

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

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

- If $E_{\pi_k} f = 1$, 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 $(1 - 1/k)$. 
For the base case, we want to show that

$$\text{Ent}_{\pi_2}(f) - 2\text{Ent}_{\pi_1}(P_{\uparrow}^1 f) \geq 0.$$  

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

$$\text{Ent}_{\pi_2}(f) - 2\text{Ent}_{\pi_1}(P_{\uparrow}^1 f) \geq 1 - \frac{1}{2Z_2} \cdot h^T Wh,$$

where $W_{ij} = w(i,j)$ and $h = P_{\uparrow}^1 f$.

Since $W = (\tau - 2)!Z_\tau \nabla^2 g_{\pi}(1)$, it has at most one positive eigenvalue. The quadratic form is maximised at $h = P_{\uparrow}^1 f = 1$, which proves the base case.
For the base case, we want to show that
\[
\text{Ent}_{\pi_2}(f) - 2\text{Ent}_{\pi_1}(P_1^\uparrow f) \geq 0.
\]

Using \(a \log \frac{a}{b} \geq a - b\) for \(a, b > 0\), we can get
\[
\text{Ent}_{\pi_2}(f) - 2\text{Ent}_{\pi_1}(P_1^\uparrow f) \geq 1 - \frac{1}{2Z_2} \cdot h^\top W h,
\]
where \(W_{ij} = w(\{i, j\})\) and \(h = P_1^\uparrow f\).

Since \(W = (\tau - 2)!Z_\tau \nabla^2 g_\pi(1)\), it has at most one positive eigenvalue. The quadratic form is maximised at \(h = P_1^\uparrow f = 1\), which proves the base case.
For the base case, we want to show that

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

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

$$\text{Ent}_{\pi_2}(f) - 2\text{Ent}_{\pi_1}(P_1^\uparrow f) \geq 1 - \frac{1}{2Z_2} \cdot h^T Wh,$$

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

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

By the Law of Total Probability,

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

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

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

By the Law of Total Probability,

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

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

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

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

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


As \( f^{(1)} = \prod_{j=1}^{k-1} p_j f \),

\[
\begin{align*}
\text{Ent}_{\pi_k}(f) &= \sum_{i \in \mathcal{M}(1)} \pi_1(i) \text{Ent}_{\pi_i, k-1}(f) + \text{Ent}_{\pi_1}(f^{(1)}) \\
\text{Ent}_{\pi_{k-1}}(P_{k-1} f) &= \sum_{i \in \mathcal{M}(1)} \pi_1(i) \text{Ent}_{\pi_i, k-2}(P_{k-1} f) + \text{Ent}_{\pi_1}(f^{(1)})
\end{align*}
\]

Induction hypothesis on \( \mathcal{M}_i \) implies that

\[
\text{Ent}_{\pi_i, k-1}(f) \geq \frac{k - 1}{k - 2} \cdot \text{Ent}_{\pi_i, k-2}(P_{k-1} f).
\]

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

\[
\sum_{i \in \mathcal{M}(1)} \pi_1(i) \text{Ent}_{\pi_i, k-2}(P_{k-1} f) \geq (k - 2) \text{Ent}_{\pi_1}(f^{(1)}).
\]

Finally, notice that

\[
\frac{k - 1}{k - 2} = \frac{k}{k - 1} + \frac{1}{(k - 1)(k - 2)}.
\]
We have shown entropy contraction from level $k$ to level $k - 1$:

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

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

\[
D \left( (P_k^\vee)^T \tau \| \pi_k \right) = \operatorname{Ent}_{\pi_k}(D_k^{-1}(P_k^\vee)^T \tau)
\]
\[
= \operatorname{Ent}_{\pi_k}(P_k^\vee D_k^{-1} \tau)
\]
\[
= \operatorname{Ent}_{\pi_k}(P_k^\downarrow P_k^\uparrow D_k^{-1} \tau)
\]
\[
\leq \operatorname{Ent}_{\pi_{k-1}}(P_k^\uparrow D_k^{-1} \tau) \quad \text{(Jensen's inequality)}
\]
\[
\leq \left(1 - \frac{1}{k}\right) \operatorname{Ent}_{\pi_k}(D_k^{-1} \tau) \quad \text{(entropy contraction)}
\]
\[
= \left(1 - \frac{1}{k}\right) D(\tau \| \pi_k).
\]

The mixing time bound follows from Pinsker’s inequality

\[
2 \|\tau - \sigma\|_{TV}^2 \leq D(\tau \| \sigma).
\]
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 \( c = \frac{\nu(f)}{\rho(P)} \),

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

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

Note that

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

The following inequalities thus finish the argument

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

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

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

Note that

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

The following inequalities thus finish the argument

$$\text{Ent}_\pi(F_t) \leq \frac{1}{\rho(P)} \mathcal{E}_P(F_t, \log F_t) \leq \frac{t^2 v(f)}{2 \rho(P)} \mathbb{E}[F_t].$$
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 $c = \frac{v(f)}{\rho(P)}$,

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

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

Note that

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

The following inequalities thus finish the argument

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

$$\mathbb{E}[e^{tf}] \leq e^{tf + ct^2}.$$ 

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

Note that

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

The following inequalities thus finish the argument

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

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

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

Note that

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

The following inequalities thus finish the argument

\[
\text{Ent}_\pi(F_t) \leq \frac{1}{\rho(P)} \mathcal{E}_P(F_t, \log F_t) \leq \frac{t^2 \nu(f)}{2\rho(P)} \mathbb{E}[F_t].
\]
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.
• The decomposition of $\text{Ent}_{\pi_k}(f)$ 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?
Recall

\[ P_{k+1}^\vee = P_{k+1}^\downarrow P_k^\uparrow; \]
\[ P_k^\wedge = P_k^\uparrow P_{k+1}^\downarrow. \]

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

For modified log-Sobolev constants, we showed

\[ \rho(P_{k+1}^\vee) \geq \frac{1}{k+1}; \]
\[ \rho(P_k^\wedge) \geq \frac{1}{k+1}, \]

but

\[ \rho(P_{k+1}^\vee) = \rho(P_k^\wedge)? \]
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