MLSI: Theory, Examples and Consequences
Lecture 3

Prasad Tetali

Georgia Institute of Technology
and Indian Institute of Science

ICTS, Bengaluru, August 5-17, 2019

Bochner-Bakry-Émery Criterion

The following 2nd derivative criterion is also useful, and inspires a notion of *discrete curvature*, mimicking the Ricci curvature.

Proposition (Bochner'46, Lichnérowicz'58)

$$\inf_f \frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi f} =: \lambda = \inf_f \frac{\mathcal{E}(-Lf,f)}{\mathcal{E}(f,f)} =: \mu_P =: \mu\,.$$

Proof. $\lambda \leq \mu$: Use Cauchy-Schwartz. Indeed, w/ $\mathbb{E}_\pi f = 0$,

$$\begin{aligned}
\mathcal{E}(f,f) = \mathbb{E}(f(-Lf)) &\leq \left(\mathrm{Var}_\pi f\right)^{1/2}\left(\mathbb{E}(-Lf)^2\right)^{1/2} \\
&\leq \frac{1}{\sqrt{\lambda}}(\mathcal{E}(f,f))^{1/2}\left(\mathbb{E}(-Lf)^2\right)^{1/2} \\
&= \frac{1}{\sqrt{\lambda}}(\mathcal{E}(f,f))^{1/2}\left(\mathcal{E}(-Lf,f)\right)^{1/2}.
\end{aligned}$$

Bochner-Bakry-Émery Criterion (contd.)

Other direction, $\lambda \geq \mu$: 2nd derivative and integration. Starting with

$$\frac{d}{dt}\mathrm{Var}(H_t f) = -2\mathcal{E}(H_t f, H_t f) \text{ and } \frac{d}{dt}\mathcal{E}(H_t f, H_t f) = -2\mathcal{E}(H_t f, -LH_t f)\,.$$

Observe that as $t \to \infty$, $H_t f \to \mathbb{E}_\pi f$, and so $\mathcal{E}(H_t f, H_t f) \to 0$. Hence,

$$\begin{aligned}
\mathcal{E}(f,f) &= -\int_0^\infty \frac{d}{dt}\mathcal{E}(H_t f, H_t f)dt = 2\int_0^\infty \mathcal{E}(H_t f, -L(H_t f))dt \\
&= 2\int_0^\infty \mathcal{E}(-L^*(H_t f), H_t f)dt \geq 2\mu_{P^*}\int_0^\infty \mathcal{E}(H_t f, H_t f)dt \\
&= -\mu_{P^*}\int_0^\infty \frac{d}{dt}\mathrm{Var}(H_t f)dt = \mu_{P^*}\mathrm{Var} f\,,
\end{aligned}$$

implying that $\lambda_P = \lambda_{P^*} \geq \mu_P$.

MLSI and Bakry-Émery Criterion

What about repeating the above for the entropy constant?

Proposition (Bakry-Émery'85)

$$\inf_{f>0} \frac{\mathcal{E}(f, \log f)}{\mathrm{Ent}_\pi f} =: \alpha \geq \inf_{f>0} \frac{u(f)}{\mathcal{E}(f, \log f)} ,$$

where $u(f) = \mathcal{E}(-Lf, \log f) + \mathcal{E}(f, (-Lf)/f)$.

Some relevant references include:

[Boudou, Caputo, Dai Pra, Posta, 2005]. "Spectral gap estimates for interacting particle systems via a Bochner type identity," J. Funct. Analysis.
[M. Erbar, J. Maas, P.T. 2015]. "Discrete Curvature Bounds for Bernoulli-Laplace and Random Transposition Models," Annales Fac. Sci. Toulouse.

Bochner formula (iterated gradient Γ_2 criterion) :

application: discrete Buser ...

Bakry-Émery, ... , Schmuckenschläger,...

$G = (V, E)$, and Graph Laplacian $\Delta = -(D - A)$

Bakry-Émery, ... , Schmuckenschläger,...

$G = (V, E)$, and Graph Laplacian $\Delta = -(D - A)$

Definition [Bochner w/ parameter K]

Curvature of G is at least K, if $\forall f : V \to \mathbb{R}$, and $\forall x \in V$,

$$\Delta\Gamma(f, f)(x) - 2\Gamma(f, \Delta f)(x) \geq K\, \Gamma(f, f)(x)\,,$$

Bakry-Émery, … , Schmuckenschläger,…

$G = (V, E)$, and Graph Laplacian $\Delta = -(D - A)$

Definition [Bochner w/ parameter K]

Curvature of G is at least K, if $\forall f : V \to \mathbb{R}$, and $\forall x \in V$,

$$\Delta\Gamma(f,f)(x) - 2\Gamma(f,\Delta f)(x) \geq K\, \Gamma(f,f)(x)\,,$$

where

$$\Delta f(x) = \sum_{y:(x,y)\in E} (f(y) - f(x))\,.$$

And given $f, g : V \to \mathbb{R}$, also define:

$$\Gamma(f,g)(x) = \frac{1}{2} \sum_{y:(x,y)\in E} (f(x) - f(y))(g(x) - g(y))\,.$$

Bakry-Émery, ... , Schmuckenschläger,...

$G = (V, E)$, and Graph Laplacian $\Delta = -(D - A)$

Definition [Bochner w/ parameter K]

Curvature of G is at least K, if $\forall f : V \to \mathbb{R}$, and $\forall x \in V$,

$$\Delta\Gamma(f, f)(x) - 2\Gamma(f, \Delta f)(x) \geq K\, \Gamma(f, f)(x)\,,$$

where

$$\Delta f(x) = \sum_{y:(x,y)\in E} (f(y) - f(x))\,.$$

And given $f, g : V \to \mathbb{R}$, also define:

$$\Gamma(f, g)(x) = \frac{1}{2} \sum_{y:(x,y)\in E} (f(x) - f(y))(g(x) - g(y))\,.$$

Note: $\Gamma(f)(x) := \Gamma(f, f)(x) = \frac{1}{2} \sum_{y:(x,y)\in E} (f(x) - f(y))^2 =: |\nabla f(x)|^2\,.$

Γ_2 Calculus

Basically, using the iterated gradient:

$$2\Gamma_2(f,g) = \Delta\Gamma(f,g) - \Gamma(f,\Delta g) - \Gamma(\Delta f,g)\,.$$

Γ_2 Calculus

Equivalently ...

Curvature of G is at least K, if $\forall f : V \to \mathbb{R}$, and $\forall x \in V$,

$$\Gamma_2(f)(x) \geq K\ \Gamma(f)(x)\,,$$

Basically, using the iterated gradient:

$$2\Gamma_2(f,g) = \Delta\Gamma(f,g) - \Gamma(f,\Delta g) - \Gamma(\Delta f, g)\,.$$

Γ_2 Calculus

Equivalently ...

Curvature of G is at least K, if $\forall f : V \to \mathbb{R}$, and $\forall x \in V$,

$$\Gamma_2(f)(x) \geq K\ \Gamma(f)(x)\,,$$

where

$$\Gamma_2(f) := \Gamma_2(f,f) = \frac{1}{2}\Delta\Gamma(f) - \Gamma(f,\Delta f)\,.$$

Basically, using the iterated gradient:

$$2\Gamma_2(f,g) = \Delta\Gamma(f,g) - \Gamma(f,\Delta g) - \Gamma(\Delta f, g)\,.$$

Γ_2 Calculus (contd.)

The iterated gamma can be written out:

Γ_2 Calculus (contd.)

The iterated gamma can be written out:

$$
\begin{aligned}
2\Gamma_2(f)(x) &= \Delta\Gamma(f)(x) - 2\Gamma(f,\Delta f)(x) \\
&= \sum_{v\sim x}\Gamma(f)(v) - d(x)\Gamma(f)(x) - \sum_{v\sim x} f(v)\big(\Delta f(v) - \Delta f(x)\big) \\
&= \Big(\sum_{v\sim x} f(v)\Big)^2 - \frac{d(x)}{2}\sum_{v\sim x} f^2(v) + \sum_{u\sim v\sim x}\frac{f^2(u) - 4f(u)f(v) + 3f^2(v)}{2} \\
&= \Big(\sum_{v\sim x} f(v)\Big)^2 - \sum_{v\sim x}\frac{d(x)+d(v)}{2} f^2(v) + \frac{1}{2}\sum_{u\sim v\sim x}\big((f(u) - 2f(v)\big)^2 .
\end{aligned}
\tag{2.1}
$$

Γ_2 Calculus (contd.)

The iterated gamma can be written out:

$$
\begin{aligned}
2\Gamma_2(f)(x) &= \Delta\Gamma(f)(x) - 2\Gamma(f,\Delta f)(x)\\
&= \sum_{v\sim x}\Gamma(f)(v) - d(x)\Gamma(f)(x) - \sum_{v\sim x} f(v)\big(\Delta f(v) - \Delta f(x)\big)\\
&= \Big(\sum_{v\sim x} f(v)\Big)^2 - \frac{d(x)}{2}\sum_{v\sim x} f^2(v) + \sum_{u\sim v\sim x}\frac{f^2(u) - 4f(u)f(v) + 3f^2(v)}{2}\\
&= \Big(\sum_{v\sim x} f(v)\Big)^2 - \sum_{v\sim x}\frac{d(x)+d(v)}{2} f^2(v) + \frac{1}{2}\sum_{u\sim v\sim x}((f(u) - 2f(v))^2 .
\end{aligned}
\tag{2.1}
$$

For d-regular, further simplifies to:

$$
2\Gamma_2(f)(x) = d\sum_{v\sim x} f^2(v) + \Big(\sum_{v\sim x} f(v)\Big)^2 + \sum_{v\sim x}\sum_{u\sim v}\Big(\frac{f^2(u)}{2} - 2f(u)f(v)\Big).
\tag{2.2}
$$

Cheeger is tight if curvature ≥ 0

$h := h(G) := \min_{A \subset V} |\partial A|/|A|$, the Cheeger const, of G; $\lambda > 0$: gap(G).

Theorem (A-M, L-S, J-S 80's)

$$\frac{h^2}{c_1 \,\mathrm{maxdeg}(G)} \leq \lambda \leq c_2 h\,.$$

Theorem (Klartag-Kozma-Ralli-T. '14)

1. *Bochner w/ parameter K implies: for any subset $A \subset V$,*

$$|\partial A| \geq \frac{1}{2} \min\left\{\sqrt{\lambda}, \frac{\lambda}{\sqrt{2|K|}}\right\} |A|\left(1 - \frac{|A|}{|V|}\right).$$

2. *Bochner w/ parameter $K \geq 0$ implies $\lambda \geq K$, and hence:*

$$\lambda \leq 16 h^2\,.$$

3. *Abelian Cayley graphs satisfy Bochner w/ parameter $K \geq 0$.*

CD(K, ∞) with $K \geq 0$: Examples

- 1. Complete graph K_n: Curvature $= 1 + n/2$.
- 2. Discrete n-cube Q_n: Curvature $= 2$.
- 3. Symmetric group $S(n)$ with the transposition metric: Curvature $= 2$
- 4. **General Proposition**. If T is the maximum number of triangles containing any edge, then $K \leq 2 + T/2$.
- 5. **Abelian necessary!** On S_n, the (left) Cayley graph generated by $(12), (12...n) \pm 1$, the Cheeger constant is $c_1 n^{-2}$, while the spectral gap is $c_2 n^{-3}$, with $c_1, c_2 > 0$, independent of n.

Cheeger, CD(K, ∞) with $K \geq 0$

- Qn. 1 Can we characterize the graphs or Markov kernels which satisfy non-negative curvature?
- Qn. 2 Can we characterize the graphs for which the Cheeger inequality is tight?
- Qn. 3 More examples? Non-crossing partition lattice $NC(n)$?

Ollivier-Villani

" In positive curvature, balls are closer than their centers."

Coarse Ricci: Take two small balls and compute the transportation distance between them. If it is smaller than the distance between the centers of the balls, then coarse Ricci is *positive*.

$$W_1(\mu_x, \mu_y) =: (1 - \kappa(x, y))d(x, y)\,,$$

Examples.
1. n-cube: μ_x uniform on the n+1 neighbors of x (including itself). For x, y, neighbors, $\kappa(x, y) = 2/(n + 1)$.

2. S_n with transpositions: For σ, τ differing in a transposition, $\kappa(x, y) = 1/\binom{n}{2}$.

(Coarse) Ricci of Hypercube

Proposition (folklore)

The n-cube has coarse Ricci $= 2/(n+1)$.

Proof sketch.
(i) Consider the *lazy* random walk on the n-cube.

(ii) Use a simple (path) coupling argument to show that two copies of the chain started at *neighboring vertices* $x, y \in \{0,1\}^n$ can be "coupled" with probability at least $2/(n+1)$ in one step.

(Coarse) Ricci of Transposition Graph

Proposition

S_n with transpositions has coarse Ricci $= 1/\binom{n}{2}$.

Proof sketch.

Lower bound:

(i) Consider the *lazy* random transposition chain on S_n.

(ii) Use a simple (path) coupling argument to show that two copies of the chain started at $\sigma, \tau \in S_n$ with $d(\sigma, \tau) = 1$, can be coupled with probability at least $1/\binom{n}{2}$ in one step.

Upper bound: Follows by showing that there is no better coupling – use (Kantorovich's) dual formulation of W_1:

$$W_1(\nu, \mu) = \sup_{f:\ 1-\mathrm{Lip}} \Big(\mathrm{E}_\nu f - \mathrm{E}_\mu f \Big) .$$

Peres-Tetali "Conjecture"!

Does *coarse Ricci* at least $\kappa > 0$ imply the MLSI constant $\alpha \geq \kappa$?

Something weaker is known: namely that a W_1 transport-Entropy inequality follows, thanks to (Marton??, Eldan, Lehec, Lee'16)

MLSI for Bases Exchange Walk

I am thankful to Heng Guo for letting me borrow some of his slides in this section.

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 $T \cup \{i\} \in \mathcal{I}$ (augment axiom).

Maximum independent sets are the bases.

For any two bases, there is a sequence of exchanges of ground set elements that take one basis to the other.

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

Bases-exchange walk

The following Markov chain $P_{\mathrm{BX},\pi}$ 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.

MLSI for Matroid Basis Exchange

Theorem (Mary Cryan-Heng Guo-Giorgos Mousa)

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

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

where r *is the rank of the matroid.*

Many open problems remain!

- Sampling spanning trees of G with no broken circuit - the same as evaluation $T_G(1,0)$ of the Tutte polynomial, leading coefficient $|a_1[\chi_G(\lambda)]|$ of the chromatic polynomial (up to sign), number of maximum G-parking functions, sampling acyclic orientations *with a unique sink* ...
- Sampling all acyclic orientations of G
- Negative correlation for random forests of G
- (More generally) characterization of matroids with negative correlation for random bases.

Strongly log-concave polynomials

Log-concave polynomial

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

Originally introduced by Gurvits (2009), equivalent to:

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

Strongly log-concave distributions

A distribution $\pi : 2^{[n]} \to \mathbb{R}_{\geq 0}$ 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+).

Alternative characterization 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 the bases of some 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.

Bases-exchange walk

The following Markov chain $P_{\mathrm{BX},\pi}$ 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.

Recall the mixing time defn.

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

Theorem (Cryan-Guo-Mousa 2019)

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

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

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

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

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

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

Levels of independent sets

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

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 normalization 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 normalizing constant.

Different views

- Polynomial: $\frac{\partial \mathbf{p}}{\partial x_i}$; setting $x_i = 0$; $(r-k)!\frac{\partial}{\partial x_I}\mathbf{p}(\mathbf{1})$
- Matroid: contraction over i; deletion of i; w(I)
- Distribution: conditioning on having i; conditioning on *not* having i; proportional to $\pi_k(I)$

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 $P_{\mathrm{BX},\pi} = P_r^\vee$.

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 $I' \in \mathcal{M}(k-1)$;

$\rightarrow$ 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 $P_{\mathrm{BX},\pi} = P_r^\vee$.

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 $I \subset J$.

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

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

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

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

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

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

We have

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

Key lemma

Lemma

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

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

Key lemma

Lemma

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

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

- Applying $P_{k-1}^{\uparrow}$ to the left corresponds to the random walk $P_k^{\downarrow}$.
- The lemma asserts that $P_k^{\downarrow}$ contracts the relative entropy by at least $(1 - 1/k)$:

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

Base case

For the base case, we want to show that

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

Base case

For the base case, we want to show that

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

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

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

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

Base case

For the base case, we want to show that

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

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

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

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 maximized at $h = P_1^{\uparrow} f = \mathbf{1}$, which helps prove the base case...

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, $D_k^{-1}(P_k^{\vee})^{\mathsf{T}} = P_k^{\vee} D_k^{-1}$.

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

The mixing time bound follows from Pinsker's inequality

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

Many open problems remain!

- Sampling spanning trees of G with no broken circuit - the same as evaluation $T_G(1,0)$ of the Tutte polynomial, leading coefficient $|a_1[\chi_G(\lambda)]|$ of the chromatic polynomial (up to sign), number of maximum G-parking functions, sampling acyclic orientations *with a unique sink* ...
- Sampling all acyclic orientations of G
- Negative correlation for random forests of G
- (More generally) characterization of matroids with negative correlation for random bases.

Fin

Thank you!