Modified Logarithmic Sobolev Inequalities: Theory, Examples and Consequences

Prasad Tetali

Georgia Institute of Technology
and Indian Institute of Science

ICTS, Bengaluru, August 5-17, 2019

MLSI - (Log-Sobolev type) Entropy Ineq.

- Let (Ω, P, π): ergodic reversible Markov chain on finite state space Ω, with transition matrix P, and stationary distribn. π.

MLSI - (Log-Sobolev type) Entropy Ineq.

- Let (Ω, P, π): ergodic reversible Markov chain on finite state space Ω, with transition matrix P, and stationary distribn. π.
- Let $-L = I - P$. For $f, g : \Omega \to \mathbb{R}$, define

$$\mathcal{E}(f,g) := -\mathbb{E}_\pi(f\ Lg) := -\sum_{x\in\Omega} f(x)Lg(x)\pi(x)\,.$$

For $f > 0$, let

$$Ent_\pi f := E_\pi(f \log f) - (E_\pi f)\log(E_\pi f).$$

MLSI - (Log-Sobolev type) Entropy Ineq.

- Let (Ω, P, π): ergodic reversible Markov chain on finite state space Ω, with transition matrix P, and stationary distribn. π.
- Let $-L = I - P$. For $f, g : \Omega \to \mathbb{R}$, define

$$\mathcal{E}(f,g) := -\mathbb{E}_\pi(f \ Lg) := -\sum_{x\in\Omega} f(x)Lg(x)\pi(x)\,.$$

For $f > 0$, let

$$Ent_\pi f := E_\pi(f \log f) - (E_\pi f)\log(E_\pi f).$$

- Following [Dai Pra, Paganoni, Posta '02], [Gao-Quastel '03], [Bobkov-T. '03], [Goel, '04]: let $\alpha > 0$ be the opt. const. in

$$\alpha Ent_\pi(f) \le \frac{1}{2}\mathcal{E}(f, \log f),$$

over positive f on Ω.

Motivation: Rate of decay of entropy

For a probab. measure μ on Ω, absolut. cont. wrt π, recall

$$D(\mu\|\pi) = \sum_{x\in\Omega} \mu(x) \log \frac{\mu(x)}{\pi(x)},$$

the relative entropy (or informational divergence) of μ wrt π.
Letting $P_t = e^{tL}$; $\mu_t = \mu_0 P_t$, $f_t = \frac{\mu_t}{\pi}$, one gets, for all $t > 0$,

$$\frac{d}{dt} D(\mu_t\|\pi) = -\mathcal{E}(f_t, \log f_t) \le -2\alpha Ent_\pi(f_t) = -2\alpha D(\mu_t\|\pi)$$

$$\Rightarrow D(\mu_t\|\pi) \le D(\mu_0\|\pi) e^{-2\alpha t}, \quad t \ge 0,$$

where μ_0: initial distribution.

MODIFIED LOG-SOBOLEV INEQUALITIES

Recall (using reversibility)

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

In a graph setting one may define:

$$\mathcal{E}(f,g) = \sum_{x\in V}\sum_{y:(y,x)\in E}(f(x)-f(y))(g(x)-g(y))\pi(x)$$

where $G=(V,E)$ is the graph with π: probab. distribution on V.
Alternately, one may work with a discrete gradient:

$$\nabla f(x) = \left\{\frac{1}{\sqrt{2}}(f(x)-f(y))\sqrt{P(x,y)}\right\}_{y\in\Omega}$$

or

$$\nabla f(x) = \{f(x)-f(y)\}_{y:(y,x)\in E}$$

depending if one works with a Markov kernel or a graph.

Modified Log-Sobolev (Contd.)

Classical Log-Sob:

$$\rho Ent_\pi f^2 \le 2\mathcal{E}(f,f)$$

Entropy Ineq (MLSI):

$$\alpha Ent_\pi f \le \frac{1}{2}\mathcal{E}(f,\log f)$$

Modified L-S using gradients: (a la Bobkov-Ledoux)

$$\rho_1 Ent_\pi(e^f) \le \frac{1}{2}E_\pi(|\nabla f|^2 e^f)$$

$$\rho_2 Ent_\pi(e^f) \le \frac{1}{2}E_\pi(|\nabla e^f|^2 e^{-f})$$

or equivalently $\rho_2 Ent_\pi(f) \le \frac{1}{2}E_\pi\left(\frac{|\nabla f|^2}{f}\right)$, over $f > 0$.

Theorem (B-T '03)

For reversible Markov kernels, $\rho \le \alpha \le \rho_1 \le \rho_2 \le \lambda$.

Log-Sob ρ versus Modified Log-Sob α

Proposition

If $f \geq 0$, then

$$2\mathcal{E}(\sqrt{f}, \sqrt{f}) \leq \mathcal{E}(f, \log f).$$

Proof. A bit of calculus:

$$a(\log a - \log b) = 2a \log \frac{\sqrt{a}}{\sqrt{b}} \geq 2a\left(1 - \frac{\sqrt{b}}{\sqrt{a}}\right) = 2\sqrt{a}(\sqrt{a} - \sqrt{b}),$$

the inequality from the relation $\log c \geq 1 - \frac{1}{c}$. Hence

$$\begin{aligned} \mathcal{E}(f, \log f) &= \sum_{x,y} f(x)(f(x) - f(y))P(x,y)\pi(x) \\ &\geq 2\sum_{x,y} \sqrt{f(x)}(\sqrt{f(x)} - \sqrt{f(y)})P(x,y)\pi(x) \\ &= 2\mathcal{E}(\sqrt{f}, \sqrt{f}). \end{aligned}$$

α versus λ

Proposition

$$\alpha \le 2\lambda\,.$$

Proof. More calculus: Apply the MLSI to functions $f = 1 + \varepsilon g$, with $g \in L^2(\pi)$, with $E_\pi g = 0$. Assume $\varepsilon \ll 1$ so that $f > 0$. Then using Taylor: $\log(1 + \varepsilon g) = \varepsilon g - (1/2)\varepsilon^2 g^2 + o(\varepsilon^2)$, we get to

$$Ent_\pi f = \frac{1}{2}\varepsilon^2 \pi(g^2) + o(\varepsilon^2), \qquad \text{and}$$

$$\mathcal{E}(f, \log f) = -\,\varepsilon E_\pi\Big((Lg)\log(1+\varepsilon g)\Big) = \varepsilon^2 \mathcal{E}(g,g) + o(\varepsilon^2)\,,$$

giving: $\varepsilon^2\mathcal{E}(g,g) \ge (\alpha/2)\varepsilon^2 E_\pi(g^2) + o(\varepsilon^2)$. Cancel ε^2 and let $\varepsilon \downarrow 0$. $\square$

The main "constants" and Mixing times

Poincare constant (spectral gap):

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

The main "constants" and Mixing times

Poincare constant (spectral gap):

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

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

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

The main "constants" and Mixing times

Poincare constant (spectral gap):

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

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

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

modified log-Sobolev constant (formalized by B-T., 2006):

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

The main "constants" and Mixing times

Poincare constant (spectral gap):

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

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

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

modified log-Sobolev constant (formalized by B-T., 2006):

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

$$2\lambda(P) \geq \alpha(P) \geq 4\rho(P)$$

(formalized B-T., 2006)

The main "constants" and Mixing times

Poincare constant (spectral gap):

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

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

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

modified log-Sobolev constant (formalized by B-T., 2006):

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

$$2\lambda(P) \ge \alpha(P) \ge 4\rho(P) \qquad \text{(formalized B-T., 2006)}$$

$$\rho(P) \le \frac{1}{\log \pi_{\min}^{-1}} \qquad \text{(folklore, D-SC)}$$

The main "constants" and Mixing times

Poincare constant (spectral gap):

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

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

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

modified log-Sobolev constant (formalized by B-T., 2006):

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

$$2\lambda(P) \geq \alpha(P) \geq 4\rho(P) \qquad \text{(formalized B-T., 2006)}$$

$$\rho(P) \leq \frac{1}{\log \pi_{\min}^{-1}} \qquad \text{(folklore, D-SC)}$$

K_n: the complete graph

For a simple r.w. on K_n,

$$\lambda = 1\,, \alpha \geq 1, \text{ but } \rho = O(1/\log n)\,.$$

Let π be arbitrary on the vertices of K_n, with $\pi_* := \min_x \pi(x) > 0$. Consider the kernel $P(x, y) = \pi(y)$. Then by Jensen's,

$$\mathrm{Ent}_\pi(f) \leq \mathbb{E}(f \log f) - \mathbb{E}f\mathbb{E}\log f = Cov_\pi(f, \log f) = \mathcal{E}(f, g)\,,$$

and so $\alpha \geq 1$.
And $\lambda = 1$, since $\mathcal{E}(f, f) = \mathrm{Var}_\pi f$. OTOH, tedious calculus gives:

$$\rho = \frac{(p - q)}{\log p - \log q},$$

where $p = \mu_*$ and $q = 1 - p$.

General two-state Example

Proposition

For $0 \le a, b \le 1$, with $a + b > 0$, let

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \quad \textit{with } \pi = \left(\frac{b}{a+b}, \frac{a}{a+b}\right).$$

Then $\lambda(P) = a + b$, and

$$\alpha(P) \in [a+b, (a+b) + 2\sqrt{ab}], \quad \rho(P) = \begin{cases} \frac{a-b}{\log a - \log b} & \textit{if } a \neq b \\ a & \textit{if } a = b \end{cases}$$

Note. $\rho(P) \ge \min(a, b)$, since $|a - b| \ge \min(a, b)|\log a - \log b|$.

$\Omega_{n,k}$: k-slice of the n-cube

For $n > 1$ integer, consider slices of the discrete cube

$$\Omega(n,k) = \{x \in \{0,1\}^n : x_1 + \cdots + x_n = k\}, \quad 1 \leq k \leq n-1\,,$$

equipped with $\mu = \mu_{n,k}$ assigning mass $\mu(\{x\}) = \frac{1}{C_n^k} = \frac{k!(n-k)!}{n!}$, to each $x \in \Omega(n,k)$.
Every point $x \in \Omega(n,k)$ has $k(n-k)$ neighbors $\{s_{ij}x\}_{i \in I(x), j \in J(x)}$ parameterized by

$$I(x) = \{i \leq n : x_i = 1\}, \quad J(x) = \{j \leq n : x_j = 0\}.$$

Namely, $(s_{ij}x)_r = x_r$, for $r \neq i, j$, and $(s_{ij}x)_i = x_j = 0$, $(s_{ij}x)_j = x_i = 1$.

k-slice of the n-cube (contd.)

The canonical associated Dirichlet form (for the graph version) is given by

$$\mathbb{E}(f,g) = \frac{1}{\binom{n}{k}} \sum_{x \in \Omega(n,k)} \sum_{i \in I(x)} \sum_{j \in J(x)} (f(x) - f(s_{ij}x))(g(x) - g(s_{ij}x)),$$

where f, g are arbitrary functions on $\Omega(n,k)$.

Theorem

For every $f > 0$ on $\Omega(n,k)$, with respect to the uniform measure μ,

$$\mathrm{Ent}_\mu(f) \le \frac{1}{n+2} \mathbb{E}(f, \log f).$$

Markov chain version

The Dirichlet form for the Markov kernel version simply involves an extra factor of $1/[2k(n-k)]$ on the right hand side – $P(x,y) = 1/[k(n-k)]$ for neighbors x, y, and the extra half is due to the above definition of the Dirichlet form in the Markov case.

Thus the Markov chain version gives

$$\alpha \geq \frac{2(n+2)}{k(n-k)}.$$

Random Transpositions on S_n

Here the state space is S_n, the set of $n!$ permutations of $\{1, \ldots, n\}$ and in each step of the Markov chain, a transposition is uniformly randomly chosen (from the $\binom{n}{2}$ possible ones) and applied to the current state.

For $1 \leq i < j \leq n$, let $s_{ij} : S_n \to S_n$ denote the transpositions; i.e., for $x \in S_n$, we have $s_{ij}(x) = y$, where $y_k = x_k$, for $k \neq i, j$, and $y_j = x_i$ and $y_i = x_j$. Thus the associated Dirichlet form is given by

$$\begin{aligned}
\mathbb{E}(f, \log f) &= \frac{1}{2} \sum_{x \in S_n} \sum_{1 \leq i < j \leq n} R(f(x), f(s_{ij}(x))) P(x, s_{ij}(x)) \mu(x) \\
&= \frac{1}{n(n-1)} \sum_{x \in S_n} \sum_{1 \leq i < j \leq n} R(f(x), f(s_{ij}(x))) \mu(x),
\end{aligned}$$

where, $R(a, b) = (a - b)(\log a - \log b)$ and $\mu(x) \equiv 1/(n!)$.

Convexity of $R(\cdot,\cdot)$

It is now useful to notice that the function R is convex in the quadrant $a, b > 0$. Indeed,

$$\frac{\partial^2 R(a,b)}{\partial a^2} = \frac{a+b}{a^2}, \quad \frac{\partial^2 R(a,b)}{\partial b^2} = \frac{a+b}{b^2}, \quad \frac{\partial^2 R(a,b)}{\partial a\,\partial b} = -\frac{a+b}{ab}.$$

Consequently, by Jensen's inequality,

$$\begin{aligned} R(\varphi(i),\varphi(j)) &= R\Big(\int f(x)\,d\mu_i(x), \int f(s_{ij}x)\,d\mu_i(x)\Big) \\ &\leq \int R(f(x), f(s_{ij}x))\,d\mu_i(x)\,. \end{aligned}$$

Random Transpositions on S_n

Theorem

For every $f > 0$ on S_n, with respect to the uniform measure μ,

$$\mathrm{Ent}_\mu(f) \leq 2(n-1)\,\mathbb{E}(f, \log f).$$

Proof Sketch. Uses conditional relative entropy and the convexity of $R(a,b)$ a la [Lee-H.T.Yau], where for the continuous-time process the usual log-Sobolev constant ρ was estimated. Proof by induction on n in estimating the optimal constant C_n in

$$\mathrm{Ent}_\mu(f) \leq C_n\,\mathbb{E}(f, \log f)\,,$$

over all f on S_n. The relevant recurrence to target is

$$C_n \leq C_{n-1} + 2\frac{(n-1)}{n}\,, \tag{2.1}$$

which implies the theorem.

Random Transpositions on S_n

Proof contd. Observe that there are n ways to partition S_n into n classes, so that each class is isomorphic to S_{n-1}. For each $1 \le t \le n$, let

$$S_n = \cup_{k=1}^n S_k^{(t)}, \quad \text{where} \quad S_k^{(t)} = \{x \in S_n : x(t) = k\}, \quad \text{for } k = 1, \ldots, n.$$

Let $\mu_k^{(t)} \equiv 1/(n-1)!$ denote the (conditional) measure on $S_k^{(t)}$. Also, for each t, let $\mu^{(t)}$ denote the uniform measure on the n classes. That is,

$$\mu^{(t)}(k) = \sum_{x \in S_k^{(t)}} \mu(x) = 1/n, \quad \text{for } k = 1, \ldots, n.$$

Random Transpositions on S_n

Using the chain rule for relative entropy (or as can easily be verified directly), for every $f \geq 0$ with $E_\mu f = 1$, we have

$$\mathrm{Ent}_\mu f = \sum_{k=1}^{n} \mu^{(t)}(k) \mathrm{Ent}_{\mu_k^{(t)}}(f|_k) + \mathrm{Ent}_{\mu^{(t)}} \bar{f}^{(t)} , \tag{2.2}$$

where $f|_k = f|_k^{(t)} = f\mathbf{1}_{S_k^{(t)}}$, is f restricted to the class $S_k^{(t)}$, and $\bar{f}^{(t)} = \sum_{x \in S_k^{(t)}} f(x) \frac{\mu(x)}{\mu^{(t)}(k)}$, is simply the average (with respect to the conditional measure) of $f|_k$ over the class $S_k^{(t)}$.

Summing (2.2) over t, we get

$$n\mathrm{Ent}_\mu f = \sum_{t=1}^{n} \sum_{k=1}^{n} \mu^{(t)}(k) \mathrm{Ent}_{\mu_k^{(t)}}(f|_k) + \sum_{t} \mathrm{Ent}_{\mu^{(t)}} \bar{f}^{(t)}. \tag{2.3}$$

Random Transpositions on S_n

By induction, we may say, for each t and k,

$$\mathrm{Ent}_{\mu_k^{(t)}}(f|_k) \le \frac{C_{n-1}}{(n-1)(n-2)} \sum_{\substack{x\in S_n \\ x(t)=k}} \sum_{\substack{1\le i<j\le n \\ i,j\ne t}} R(x, s_{ij}(x)) \frac{\mu(x)}{\mu^{(t)}(k)}. \quad (2.4)$$

Multiplying the above by $\mu^{(t)}(k)$ and summing over k and t, we get

$$\begin{aligned} \sum_{t=1}^{n}\sum_{k=1}^{n} \mu^{(t)}(k)\mathrm{Ent}_{\mu_k^{(t)}}(f|_k) &\le \frac{C_{n-1}}{(n-1)} \sum_{x\in S_n} \sum_{1\le i<j\le n} R(f(x), f(s_{ij}(x)))\mu(x) \\ &\le nC_{n-1}\mathbb{E}(f, \log f), \quad (2.5) \end{aligned}$$

since when summed over t, each pair $(x, s_{ij}(x))$ is counted exactly $(n-2)$ times – in the partitions corresponding to $1 \le t \le n$, with $t \ne i, j$.

Random Transpositions on S_n

To bound the second term of the right hand side of (2.2), we will use Jensen's inequality twice. (Equivalently, we can think of using the MLSI for the r.w. on the complete graph!) Indeed, by Jensen's inequality,

$$\mathrm{Ent}_{\mu^{(t)}} \bar{f}^{(t)} \leq \frac{1}{n^2} \sum_{1 \leq k < l \leq n} R\left(\bar{f}^{(t)}(k), \bar{f}^{(t)}(l)\right). \tag{2.6}$$

To bound the term involving $R(.,.)$, observe that there is a natural bijection (namely, the restriction of s_{kl}) between $S_k^{(t)}$ and $S_l^{(t)}$, for $1 \leq k < l \leq n$, such that $P(x, y)$ for $x \in S_k^{(t)}$ and $y \in S_l^{(t)}$ is positive only when $y = s_{kl}(x)$.

Random Transpositions on S_n

This helps us in writing:

$$\begin{aligned} R\left(\bar{f}^{(t)}(k), \bar{f}^{(t)}(l)\right) &= R\Big(\sum_{x \in S_k^{(t)}} f(x)\mu_k^{(t)}(x), \sum_{x \in S_k^{(t)}} f(s_{kl}(x))\mu_k^{(t)}(s_{kl}(x))\Big) \\ &\leq \sum_{x \in S_k^{(t)}} R(f(x), f(s_{kl}(x)))\mu_k^{(t)}(x), \end{aligned}$$

where the inequality was using Jensen's, based on the convexity of $R(\cdot, \cdot)$. Plugging this into (2.6), we get

$$\mathrm{Ent}_{\mu^{(t)}} \bar{f}^{(t)} \leq \frac{1}{n} \sum_{1 \leq k < l \leq n} \sum_{x \in S_k^{(t)}} R(f(x), f(s_{kl}(x)))\mu(x) \tag{2.7}$$

Random Transpositions on S_n

Summing the above over t, we get

$$\sum_{t=1}^{n} \mathrm{Ent}_{\mu^{(t)}} \bar{f}^{(t)} \leq \frac{2}{n} \sum_{x \in S_n} \sum_{1 \leq k < l \leq n} R(f(x), f(s_{kl}(x))\mu(x)$$
$$= 2(n-1)\mathbb{E}(f, \log f), \qquad (2.8)$$

since when summed over t, each pair $(x, s_{kl}(x))$ was counted twice – once each for $t = k$ and $t = l$.
Finally, (2.3), (2.5), and (2.8) together imply the recurrence,

$$C_n \leq C_{n-1} + 2\frac{(n-1)}{n} \qquad (2.9)$$

which, together with the initial condition $C_2 \leq 2$ gives the bound $C_n \leq 2(n-1)$, completing the proof of the theorem. □

Remarks

- This implies an upper bound of $O(n \log n)$ on the total variation mixing time, which is known to be tight (see [D-Sh]). Note that the usual log-Sobolev constant only yields $O(n \log^2 n)$ (see [D-SC'96] and [L-Y'98]). Key difference is that recurrence for the inverse of the log-Sobolev constant has an extra factor of $\log n$ in the second term on the r.h.s. This is because ρ for the random walk on the complete graph is of the order of $1/(\log n)$, whereas $1/2 \leq \rho_0 \leq 1$.
- Theorem 2.2 implies a lower bound of $1/2(n-1)$ on the spectral gap. The truth is $2/(n-1)$ (see [Diaconis'88])

Random *Catalan* Transpositions

Problem. How about random transpositions of ('s and)' over the space of n pairs of *balanced* parentheses?

For $n = 5$:

((())) (()()) (())() ()(()) ()()()

More generally, $C_n = \frac{1}{(n+1)}\binom{2n}{n}$, the nth Catalan number many.

Random *Catalan* Transpositions

Problem. How about random transpositions of ('s and)' over the space of n pairs of *balanced* parentheses?

For $n = 5$:

$$((()))\quad (()())\quad (())()\quad ()(())\quad ()()()$$

More generally, $C_n = \frac{1}{(n+1)}\binom{2n}{n}$, the nth Catalan number many.

Many challenges - 214 Catalan structures! Few tight results :-/

Random *Catalan* Transpositions

Problem. How about random transpositions of ('s and)' over the space of n pairs of *balanced* parentheses?

For $n = 5$:

$$((()))\quad (()())\quad (())()\quad ()(())\quad ()()()$$

More generally, $C_n = \frac{1}{(n+1)}\binom{2n}{n}$, the nth Catalan number many.

Many challenges - 214 Catalan structures! Few tight results :-/

[Cohen-T.-Yeliussizov'15]. $\rho, \lambda \geq 1/(2n^2)$, hence $O(n^2 \log n)$ mixing time.

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.*

General Decomposition Theorems

Let (Ω, P, π) be the usual triple, with P being reversible with respect to π on Ω.

Consider a partition $\Omega = \cup_{i \in \mathcal{I}} \Omega_i$. We may introduce a *projection* chain $\hat{P}$ on $\mathcal{I}$ and *restriction* chains P_i on Ω_i.

$$\hat{P}(i,j) := \frac{1}{\hat{\pi}(i)} \sum_{x \in \Omega_i} \sum_{y \in \Omega_j} \pi(x) P(x,y), \quad \text{and } \hat{\pi}(i) := \sum_{x \in \Omega_i} \pi(x) \,.$$

Clearly, $\hat{P}$ is reversible w.r.t. $\hat{\pi}$ on the space $\mathcal{I}$:

$$\hat{\pi}(i)\hat{P}(i,j) = \hat{\pi}(j)\hat{P}(j,i) \,,$$

for all $i, j \in \mathcal{I}$.

Decomposition (contd.)

Restriction P_i is defined so that it is a Markov generator within Ω_i, for each $i \in \mathcal{I}$:

$$P_i(x,y) = P(x,y)\,,$$

for $x, y \in \Omega_i$, with diagonal entries suitably adjusted. Easy to see that

$$\pi_i(x) := \frac{\pi(x)}{\hat{\pi}(i)}$$

is reversible under P_i.

Decomposition (contd.)

Suppose for each i, j with $\hat{P}(i,j) > 0$, we have couplings $\kappa_{ij} : \Omega_i \times \Omega_j \to [0,1]$ of probab. distributions π_i and π_j. Let the "quality" of the coupling be:

$$\chi := \min\left\{ \frac{\pi(x)P(x,y)}{\hat{\pi}(i)\hat{P}(i,j)\kappa_{ij}(x,y)} \right\},$$

where the min. runs over tuples s.t. the denom. is positive. Then

Lemma (Hermon-Salez, 2019)

$$\alpha(P) \geq \min\left\{ \chi\alpha(\hat{P}), \min_i \alpha(P_i) \right\}.$$

Note: Also holds for ρ and λ, implicit in [Jerrum-Son-T.-Vigoda'04].

Mass Transport (Monge-Kantorvich)

Let $G = (V, E)$ be a graph, with distance d. Given μ, ν: probab. measures on V, we have

$$W_1(\mu, \nu) = \inf_{\Gamma \to (\mu,\nu)} \sum_{x,y} d(x, y) \Gamma(x, y)$$

(Γ: coupling of μ and ν)
Dual Version:

$$\sup_{f \in \mathrm{Lip}(G)} \sum_{x \in V} f(x)(\mu(x) - \nu(x)),$$

where $f \in \mathrm{Lip}(G) \Leftrightarrow |f(x) - f(y)| \leq d(x, y), \forall\, x, y \in V$.

Subgaussian and Transport-Entropy

For a probab. measure μ on Ω, absolut. cont. wrt π, recall

$$D(\mu\|\pi) = \sum_{x\in\Omega} \mu(x) \log \frac{\mu(x)}{\pi(x)},$$

the relative entropy (or informational divergence) of μ wrt π.

Proposition (Bobkov-Götze '99)

Let $c > 0$. Then TFAE:

1. $E_\pi[e^{t(f-E_\pi f)}] \leq e^{ct^2/2}$,
   for all $f \in Lip(G)$, $t \in \mathbb{R}$.

(Follow up by Otto-Villani '00, Bobkov-Gentil-Ledoux '01, ...)

Subgaussian and Transport-Entropy

For a probab. measure μ on Ω, absolut. cont. wrt π, recall

$$D(\mu \| \pi) = \sum_{x \in \Omega} \mu(x) \log \frac{\mu(x)}{\pi(x)},$$

the relative entropy (or informational divergence) of μ wrt π.

Proposition (Bobkov-Götze '99)

Let $c > 0$. Then TFAE:

1. $E_\pi[e^{t(f - E_\pi f)}] \le e^{ct^2/2}$, *for all $f \in Lip(G)$, $t \in \mathbb{R}$.*
2. $W_1^2(\mu, \pi) \le 2cD(\mu \| \pi)$, *for every μ absolut. cont. wrt π.*

(Follow up by Otto-Villani '00, Bobkov-Gentil-Ledoux '01, ...)

Entropy Inequality implies Transport Ineq.

Theorem (Sammer-T '05, '09)

Given $G = (V, E)$ and a prob. meas. π on V. Let $L = P - I$, with supp(P) on E. Let $d : V \times V \to \mathbb{R}^+$ be the graph distance. Suppose L and d satisfy:

$$\sum_y d^2(x,y)L(x,y) \leq 1, \quad \forall\, x \in V.$$

Then

$$2\alpha Ent_\pi(f) \leq \mathcal{E}(f, \log f),$$

for all densities wrt π

$$\Rightarrow W_1^2(\mu,\pi) \leq 2\left(\frac{1}{2\alpha}\right) D(\mu\|\pi),$$

for all μ: absolut. cont. wrt π.

Proof (a la Otto-Villani)

- Let $\nu_t = \nu P_t$ where $P_t = e^{tL}$, and let $f_t = \nu_t / \pi$. Let g_t be a solution to Kantorovich's problem wrt ν_t and π for each $t \geq 0$.

Proof (a la Otto-Villani)

- Let $\nu_t = \nu P_t$ where $P_t = e^{tL}$, and let $f_t = \nu_t/\pi$. Let g_t be a solution to Kantorovich's problem wrt ν_t and π for each $t \geq 0$.
- First show that

$$\frac{d^+}{dt} W(\nu_t, \pi) \geq \sum_{x \in V} g_t(x) L f_t(x) \pi(x) = -\mathcal{E}(g_t, f_t).$$

Proof (a la Otto-Villani)

- Let $\nu_t = \nu P_t$ where $P_t = e^{tL}$, and let $f_t = \nu_t/\pi$. Let g_t be a solution to Kantorovich's problem wrt ν_t and π for each $t \geq 0$.
- First show that

$$\frac{d^+}{dt} W(\nu_t, \pi) \geq \sum_{x \in V} g_t(x) L f_t(x) \pi(x) = -\mathcal{E}(g_t, f_t).$$

- And then that $-\mathcal{E}(g_t, f_t) \geq -\frac{1}{\sqrt{2}} \sqrt{\mathcal{E}(f_t, \log f_t)}$.

Proof (contd.)

- To prove $-\mathcal{E}(g_t, f_t) \geq -\frac{1}{\sqrt{2}}\sqrt{\mathcal{E}(f_t, \log f_t)}$ requires:

Proof (contd.)

- To prove $-\mathcal{E}(g_t, f_t) \geq -\frac{1}{\sqrt{2}}\sqrt{\mathcal{E}(f_t, \log f_t)}$ requires:
- (i) Lipschitz property of g_t;

  (ii) $\mathcal{E}(e^{f/2}, e^{f/2}) \leq \frac{1}{4}\mathcal{E}(e^f, f)$; and

  (iii) $\sum_x \sum_y 2(f_t(x) + f_t(y))d^2(x,y)L(x,y)\pi(x) \leq 4$,

  (which in turn uses the hypo. that $\sum_y d^2(x,y)L(x,y) \leq 1$).

Proof (contd.)

Then one may proceed as in [O-V]:

$$\frac{d}{dt}\sqrt{\frac{D(\nu_t\|\pi)}{\alpha}} = -\frac{\frac{1}{2}\mathcal{E}(f_t,\log f_t)}{\sqrt{\alpha D(\nu_t\|\pi)}}$$

$$\leq -\frac{1}{\sqrt{2}}\sqrt{\mathcal{E}(f_t,\log f_t)} \leq \frac{d^+}{dt}W(\nu_t,\pi).$$

Letting $\phi(t) = \sqrt{\frac{D(\nu_t\|\pi)}{\alpha}} - W(\nu_t,\pi)$, the above asserts that $\frac{d^+}{dt}\phi(t) \leq 0$. Hence

$$0 = \lim_{t\to\infty}\phi(t) \leq \phi(0) = -W(\nu_0,\pi) + \sqrt{\frac{D(\nu_0\|\pi)}{\alpha}}$$

$$\Rightarrow W^2(\nu_0,\pi) \leq \frac{1}{\alpha}D(\nu_0\|\pi).$$

LSI versus MLSI

Note: Log-Sobolev implies (sub)gaussian tails, but

$$\rho(P) \leq \min_{x \in \Omega} \left\{ \log \frac{1}{\pi(x)} \right\},$$

and RHS could be arbitrarily small without some assumption on π.

Concentration a la Herbst

Theorem (J. Hermon - J. Salez'19)

Let P be a reversible Markov generator with respect to π on Ω. Then

$$\pi\left(f \geq \mathbb{E}_\pi f + a\right) \leq \exp\left(-\frac{\alpha(P)a^2}{4v(f)}\right),$$

for all $f : \Omega \to \mathbb{R}$ and all $a \geq 0$, where

$$v(f) := \max_{x\in\Omega}\left\{\sum_{y\in\Omega} P(x,y)[f(y) - f(x)]_+^2\right\}.$$

Proof

- (Standard) Recall that $\mathbb{E}\left[e^{t(f-\mathbb{E}f)}\right] \leq e^{ct^2}$ implies (by Chebyshev/Chernoff?) for all $a \geq 0$,

  $$\pi\left(f \geq \mathbb{E}_\pi f + a\right) \leq e^{ct^2 - at}.$$

  So if we prove the hypothesis above for all $t \geq 0$ and choose $t = a/(2c)$, we get

  $$\pi\left(f \geq \mathbb{E}_\pi f + a\right) \leq \exp\left(-\frac{ca^2}{4}\right).$$

Proof

- (Standard) Recall that $\mathbb{E}\left[e^{t(f-\mathbb{E}f)}\right] \leq e^{ct^2}$ implies (by Chebyshev/Chernoff?) for all $a \geq 0$,

$$\pi\left(f \geq \mathbb{E}_\pi f + a\right) \leq e^{ct^2 - at}.$$

So if we prove the hypothesis above for all $t \geq 0$ and choose $t = a/(2c)$, we get

$$\pi\left(f \geq \mathbb{E}_\pi f + a\right) \leq \exp\left(-\frac{ca^2}{4}\right).$$

- So we will derive, for all $t \geq 0$,

$$\mathbb{E}\left[e^{t(f-\mathbb{E}f)}\right] \leq e^{ct^2},$$

with $c := v(f)/\alpha(P)$ to prove the theorem.

Proof (contd)

- (Herbst, ..., Hermon-Salez) For $t \in (0, \infty)$ and $x \in \Omega$, let

$$F_t(x) := e^{tf(x)-ct^2} ,$$

where the choice of $c > 0$ is no longer a secret! By reversibility, $\mathcal{E}(F_t, \log F_t)$

$$
\begin{aligned}
&= \frac{t}{2} \sum_{x,y \in \Omega} \pi(x)P(x,y)\big(F_t(x) - F_t(y)\big)(f(x) - f(y)) \\
&= t \sum_{x,y \in \Omega} \pi(x)P(x,y)F_t(x)\left(1 - e^{-t(f(x)-f(y))}\right)[f(x) - f(y)]_+ \\
&\leq t^2 \sum_{x,y \in \Omega} \pi(x)P(x,y)F_t(x)[f(x) - f(y)]_+^2 \quad (\text{using } 1 - e^{-u} \leq u) \\
&\leq t^2 v(f)\mathbb{E}(F_t) .
\end{aligned}
$$

Proof (contd)

- By the definition of $\alpha(P)$, we deduce:

$$Ent(F_t) \leq \frac{v(f)}{\alpha(P)} t^2 \mathbb{E}[F_t] \,.$$

OTOH, for $t > 0$, by computation,

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

Proof (contd)

- By the definition of $\alpha(P)$, we deduce:

$$Ent(F_t) \leq \frac{v(f)}{\alpha(P)} t^2 \mathbb{E}[F_t] \, .$$

OTOH, for $t > 0$, by computation,

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

- So the choice of $c := v(f)/\alpha(P)$, ensures that $t \to \frac{\log \mathbb{E}[F_t]}{t}$ is non-increasing on $(0, \infty)$, yielding:

$$\frac{\log \mathbb{E}[F_t]}{t} \leq \lim_{h \to 0} \frac{\log \mathbb{E}[F_h]}{h} = \mathbb{E}[f] \, ,$$

by taking the limit appropriately. This establishes as desired:

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