The L^2-cutoff for reversible Markov chains

Guan-Yu Chen (Joint with J.-M. Hsu and Y.-C. Sheu)

Dept. of Applied Math, NCTU

July 14, 2016 at Jiangsu Normal University

1. The L^2-distances and L^2-mixing times
2. Cutoff of Laplace transforms
3. Product chains

Outline

The L^2-distance (I)

Consider an irreducible and reversible Markov chain $(X_m)_{m=0}^{\infty}$ on a finite set $\mathcal{S}$ with transition matrix K and stationary distribution π.

- If K has eigenvalues $\beta_1 = 1, \beta_2, ...$ and $L^2(\pi)$-orthonormal eigenvectors $\psi_1 = \mathbf{1}, \psi_2, ...$, then

$$K^m(x,y) = \pi(y) \sum_{i \geq 1} \beta_i^m \psi_i(x)\psi_i(y).$$

- The L^2-distance is defined by

$$d_2(x,m) := \sqrt{\sum_{y \in \mathcal{S}} \left(\frac{K^m(x,y)}{\pi(y)} - 1 \right)^2 \pi(y)} = \sqrt{\sum_{i>1} \psi_i^2(x)\beta_i^{2m}}.$$

- If $\beta_i \geq \beta_{i+1}$, then K is aperiodic if and only if $\beta_{|\mathcal{S}|} > -1$.

The L^2-distance (II)

Let N_t be a Poisson process with parameter 1 and independent of $(X_m)_{m=0}^\infty$. Set $Y_t = X_{N_t}$ and $H_t(x,\cdot)$ be the distribution of Y_t given $Y_0 = x$.

- The semigroup $(H_t)_{t\ge 0}$ has $K - I$ as the infinitesimal generator, where I is the identity matrix indexed by $\mathcal{S}$.
- The L^2-distance of Y_t is defined by

$$d_2(x,t) := \sqrt{\sum_{y\in\mathcal{S}} \left(\frac{H_t(x,y)}{\pi(y)} - 1\right)^2 \pi(y)} = \sqrt{\sum_{i>1} \psi_i^2(x) e^{-2t(1-\beta_i)}}.$$

For $(X_m)_{m=0}^\infty$ and $(Y_t)_{t\ge 0}$, the L^2-mixing time is defined by

$$T_2(x,\epsilon) = \min\{t \ge 0 | d_2(x,t) \le \epsilon\}.$$

The L^2-cutoff

Let $\mathcal{F} = (x_n, \mathcal{S}_n, H_{n,t}, \pi_n)_{n=1}^{\infty}$, where $(x_n, \mathcal{S}_n, H_{n,t}, \pi_n)$ is a continuous time Markov chain on the finite set $\mathcal{S}_n$ with stationary distribution π_n and initial state x_n. Let $d_{n,2}$ and $T_{n,2}$ be the L^2-distance and the L^2-mixing time of the nth chain.

Definition

$\mathcal{F}$ has a L^2-cutoff if there is a sequence $t_n > 0$ such that, for $c \in (0,1)$,

$$\lim_{n\to\infty} d_{n,2}(x_n, (1+c)t_n) = 0, \quad \lim_{n\to\infty} d_{n,2}(x_n, (1-c)t_n) = \infty,$$

or, equivalently,

$$T_{n,2}(x_n, \epsilon) \sim t_n, \quad \forall \epsilon > 0,$$

where $a_n \sim b_n$ means $a_n/b_n \to 1$. t_n is called a L^2-cutoff time.

Some history of cutoffs (I)

- In 1981, Diaconis & Shahshahani declared the total variation cutoff for random transpositions using the representation theory.
- In 1980's, Aldous & Diaconis introduced the coupling time and strong stationary time to bound the mixing times of classical examples in the total variation and separation.
- In 1990's, Bayer, Diaconis, Fill and Saloff-Coste etc. applied enumerative combinatorics and spectral theory to bound the total variation and L^2-distance for well-known shuffling schemes.
- In late 1990's, Diaconis and Saloff-Coste introduced the Poincaré inequality, the Log-Sobolev inequality and the Nash inequality to bound the L^p-mixing time for $1 < p < \infty$.

Some history of cutoffs (II)

In 2004, Peres proposed the following conjecture.

$$\text{A cutoff exists} \quad \Leftrightarrow \quad \text{Mixing time} \times \text{Spectral gap} \to \infty. \qquad (1)$$

- In 2004, Aldous and Pak introduced counterexamples to (1) in the total variation.
- In 2006, Diaconis & Saloff-Coste proved (1) for birth and death chains in separation and provided a formula on the cutoff time.
- In 2008, Chen & Saloff-Coste proved (1) for reversible Markov chains and processes in the L^p-distance with $1 < p \le \infty$.
- In 2010, Peres et. al. proved (1) for birth and death chains in the total variation.

All above results focus on the maximum distance, e.g. $\max_x d_2(x, t)$.

Some history of cutoffs (III)

- In 2010, Chen & Saloff-Coste provided a similar version of (1) for reversible Markov chains and processes with prescribed initial states in the L^2-distance and provided a formula on the cutoff time.
- In 2015, Chen & Saloff-Coste introduce the hitting time to identify the cutoff for birth and death chains in the maximum total variation and separation and provided a formula on the cutoff time.
- In 2015, Peres & Sousi introduced (in a preprint) the hitting time to identify the maximum total variation cutoff for reversible Markov chains and confirmed (1) for random walks on trees.
- In 2016, Chen, Hsu & Sheu introduced (in a preprint) a different viewpoint to examine the L^2-cutoff from what was introduced by Chen & Saloff-Coste in 2010.

Outline

1. The L^2-distances and L^2-mixing times
2. Cutoff of Laplace transforms
3. Product chains

The Laplace transform

As before, let $(\beta_i, \psi_i)_{i=1}^{|\mathcal{S}|}$ be the spectral information of $(\mathcal{S}, K, \pi)$.

- In the continuous case, if $\beta_i \ge \beta_{i+1}$ and

$$V(\lambda) = \sum_{i=2}^{j-1} \psi_i^2(x), \quad \forall 2(1-\beta_{j-1}) \le \lambda < 2(1-\beta_j), \quad 2 \le j \le |\mathcal{S}|,$$

  where $\sum_{i=2}^{1} := 0$ and $1-\beta_{|\mathcal{S}|+1} := \infty$, then the L^2-distance can be rewritten as

$$d_2^2(x,t) = \int_{(0,\infty)} e^{-t\lambda} dV(\lambda). \tag{2}$$

- In the discrete time case, (2) remains true when $|\beta_i| \ge |\beta_{i+1}|$ and, in the definition of V, $1-\beta_i$ is replaced by $-\log|\beta_i|$.

Modified spectral gap

Let V be a non-decreasing, right-continuous function on $(0,\infty)$ satisfying $V(0^+)=0$ and $V(\infty)<\infty$. Define the Laplace transform of V by

$$\mathcal{L}_V(t) := \int_{(0,\infty)} e^{-t\lambda} dV(\lambda)$$

and the mixing time of $\mathcal{L}_V$ by

$$T_V(\epsilon) := \inf\{t : \mathcal{L}_V(t) \le \epsilon\}.$$

For $C>0$, consider the following modified spectral gap

$$\lambda_V(C) := \inf\{\lambda : V(\lambda) > C\}.$$

Conjectured mixing time

Observe that, for $C > 0$,

$$\mathcal{L}_V(t) \geq \int_{(0,\lambda_V(C)]} e^{-t\lambda} dV(\lambda) \geq Ce^{-t\lambda_V(C)}$$

and

$$\mathcal{L}_V(t) \leq C + \int_{(\lambda_V(C),\infty)} e^{-t\lambda} dV(\lambda).$$

The first inequality leads to

$$T_V\left(\frac{C}{1+C}\right) \geq \tau_V(C) := \sup_{\lambda \geq \lambda_V(C)} \left\{ \frac{\log(1+V(\lambda))}{\lambda} \right\}.$$

Cutoff criterion(Chen and Saloff-Coste))

A family of Laplace transforms $(\mathcal{L}_{V_n})_{n=1}^{\infty}$ is said to present a cutoff if there is a sequence $t_n > 0$ such that, for $c \in (0,1)$,

$$\lim_{n\to\infty} \mathcal{L}_{V_n}((1+c)t_n) = 0, \quad \lim_{n\to\infty} \mathcal{L}_{V_n}((1-c)t_n) = \infty.$$

Theorem

Let $\mathcal{F} = (\mathcal{L}_{V_n})_{n=1}^{\infty}$ be a family of Laplace transforms and, for $C > 0$, set

$$\lambda_n(C) = \lambda_{V_n}(C), \quad \tau_n(C) = \tau_{V_n}(C).$$

Then, $\mathcal{F}$ has cutoff if and only if, for some (all) $C > 0$ and $A > 0$,

(1) $\tau_n(C)\lambda_n(C) \to \infty$,

(2) $\int_{(0,\lambda_n(C))} e^{-A\tau_n(C)\lambda} dV_n(\lambda) \to 0$.

Moreover, $\tau_n(C)$ is a cutoff time.

Cutoff criterion(Chen, Hsu and Sheu)

As a result of the integration by parts, one may rewrite

$$\mathcal{L}_V(t) = t\int_0^\infty V(\lambda)e^{-t\lambda}d\lambda.$$

Theorem

Let $\mathcal{F}, \lambda_n(C), \tau_n(C)$ be constants as before. The following are equivalent.

(1) $\mathcal{F}$ has a cutoff.

(2) $\tau_n(C)\lambda_n(C) \to \infty$ for all $C > 0$.

Moreover, $\tau_n(C)$ is a cutoff time.

Remark

$\tau_n(C)\lambda_n(C) \to \infty$ for some $C > 0$ is not sufficient for a cutoff.

Outline

Product chains

Let $(\mathcal{S}_i, K_i, \pi_i)_{i=1}^n$ be irreducible finite Markov chains and $p = (p_1, ..., p_n)$ be a probability vector. Consider the following product chain.

$$\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_n, \quad \pi = \pi_1 \times \cdots \times \pi_n,$$

and

$$K = \sum_{i=1}^n p_i I_1 \otimes \cdots \otimes I_{i-1} \otimes K_i \otimes I_{i+1} \otimes \cdots \otimes I_n,$$

where I_i is the identity matrix indexed by $\mathcal{S}_i$ and $A \otimes B$ denotes the tensor product of matrices A, B. Concerning the continuous time case, we set $H_t = e^{-t(I-K)}$ and $H_{i,t} = e^{-t(I_i - K_i)}$. Then, one has

$$H_t = H_{1,p_1 t} \otimes \cdots \otimes H_{n,p_n t}.$$

Generally, the above identity fails in the discrete time case.

The L^2-distances of product chains

For the continuous time product chain, the L^2-distances, $d_{i,2}$ and d_2, of $(\mathcal{S}_i, K_i, \pi_i)$ and $(\mathcal{S}, K, \pi)$ satisfy, for $x = (x_1, .., x_n) \in \mathcal{S}$,

$$d_2^2(x, t) = \prod_{i=1}^{n} \left(d_{i,2}^2(x_i, p_i t) + 1 \right) - 1.$$

Note that the above equality holds without the assumption of reversibility. As a consequently, this leads to

$$\sum_{i=1}^{n} d_{i,2}^2(x_i, p_i t) \leq d_2^2(x, t) \leq \exp\left\{ \sum_{i=1}^{n} d_{i,2}^2(x_i, p_i t) \right\} - 1.$$

A practical example

Consider a machinery with a large number of components.

- Each component has two states and evolves independently in the way that the state changes to the other after an exponential waiting time.
- Concerning the effect of some external force, we assume that each component could speed up or slow down its evolution but still operates independently.
- The question is how (the existence of cutoffs) and when (the mixing time) this machinery gets close to its stability.

Mathematical setting

For simplicity, we quantize this problem as follows.

(1) Those components have $\mathcal{S} = \{0, 1\}$ as their state spaces and are indexed by positive integers.

(2) The nth component has unforced transition kernel $e^{-t(I-L_n)}$, where

$$L_n = \begin{pmatrix} 1-\alpha_n & \alpha_n \\ \beta_n & 1-\beta_n \end{pmatrix},$$

and $\alpha_n, \beta_n \in (0, 1)$.

(3) The accelerating rate of the nth component is $p_n > 0$ and, hence, the forced transition kernel is $e^{-p_n t(I-L_n)}$.

Products of two-state chains

Consider the set C_n of components indexed from m_n to $m_n+\ell_n-1$.

(1) Clearly, C_n has state space $\mathcal{S}_n=\{0,1\}^{\ell_n}$.

(2) As components operate independently, the transition kernel of C_n is $H_{n,t}=e^{-q_nt(I-K_n)}$, where $q_n=p_{m_n}+\cdots+p_{m_n+\ell_n-1}$, and

$$K_n=q_n^{-1}\sum_{i=1}^{\ell_n}p_{m_n+i-1}I_1\otimes\cdots\otimes I_{i-1}\otimes L_{m_n+i-1}\otimes I_{i+1}\otimes\cdots\otimes I_{\ell_n},$$

and I_i is the 2-by-2 identity matrix.

(3) The stationary distribution for C_n is

$$\pi_n(x)=\prod_{i=1}^{\ell_n}\frac{\alpha_{m_n+i-1}^{x_i}\beta_{m_n+i-1}^{1-x_i}}{\alpha_{m_n+i-1}+\beta_{m_n+i-1}},\quad \forall x=(x_1,...,x_{\ell_n})\in\mathcal{S}_n.$$

Cutoffs for products of two-state chains

Theorem

Let $\mathcal{F} = (\mathbf{0}, \mathcal{S}_n, H_{n,t}, \pi_n)$, where $\mathbf{0}$ is the zero vector, and assume

$$\alpha_n + \beta_n = A > 0, \quad 0 < \inf_{n \geq 1} \frac{\alpha_n}{\beta_n} \leq \sup_{n \geq 1} \frac{\alpha_n}{\beta_n} < \infty.$$

(1) If $p_n = e^{an}$, then $\mathcal{F}$ has no L^2-cutoff.

(2) If $p_n = \exp\{a[\log(n+1)]^b\}$ with $a > 0, b > 0$, then $\mathcal{F}$ has a L^2-cutoff if and only if $m_n \to \infty$ and $\ell_n \to \infty$.

(3) If $p_n = [\log(n+1)]^a$, then $\mathcal{F}$ has a L^2-cutoff if and only if

$$\begin{cases} m_n \to \infty, \ell_n \to \infty & \text{for } a \geq 1, \\ \ell_n \to \infty & \text{for } 0 < a < 1. \end{cases}$$

Cutoff times for products of two-state chains

Theorem

Let t_n be a L^2-cutoff time for cases (2) and (3).

- In the case that $p_n = \exp\{a[\log(n+1)]^b\}$ or $p_n = [\log(n+1)]^a$ with $a \ge 1$, if $m_n \to \infty$ and $\ell_n \to \infty$, then

$$t_n \sim \frac{\log(m_n \wedge \ell_n)}{2Ap_{m_n}}.$$

- In the case of $p_n = [\log(n+1)]^a$ with $0 < a < 1$, if $\ell_n \to \infty$, then

$$t_n \sim \frac{[\log(1 + m_n \wedge \ell_n)]^a (\log \ell_n)^{1-a}}{2Ap_{m_n}}.$$

Remark

Recently, Chen and Kumagai proved the above theorems in total variation.

A concrete example

Consider the following concrete setting.

- $m_n = \lfloor n^\alpha \rfloor$ and $\ell_n = n - \lfloor n^\alpha \rfloor + 1$: This means that C_n is the set of components indexed from $\lfloor n^\alpha \rfloor$ to n.
- $p_n = n + 1$: The accelerating rates are of case (2) with $a = b = 1$.
- $\alpha_n + \beta_n = 1$ and $\inf_n(\alpha_n \wedge \beta_n) > 0$: For each component, the transition rates between 0 and 1 are comparable with each other and sum up to 1.

The conclusion says:

(1) For $\alpha = 0$, $\mathcal{F}$ has no L^2-cutoff and the L^2-mixing times are bounded above and below by positive constants.

(2) For $0 < \alpha < 1$, $\mathcal{F}$ has a L^2-cutoff with cutoff time $\alpha(\log n)/(2n^\alpha)$.

Reference

Books:

1. Diaconis, P. *Group representations in probability and statistics*, 1988.
2. Aldous, D. and Fill, J. *An unpublished book on finite Markov chians.*
3. Saloff-Coste, L. *Lectures on finite Markov chains*, 1996.
4. Saloff-Coste, L. *Random walks on finite groups*, 2004.
5. Levin, D. and Peres, Y. and Wilmer, E. *Markov Chains and Mixing Times*, 2009.

Articles: (To capture this talk)

1. Diaconis, P. *The cutoff phenomenon in finite Markov chains*, 1996.
2. Chen, G.-Y. and Saloff-Coste, L. *The L2-cutoff for reversible Markov processes*, 2010.
3. Chen, G.-Y., Hsu, J.-M. and Sheu, Y.-C. *The L2-cutoff for reversible Markov chains*, 2016.

Thank you for your attention!