Analysis of non-reversible Markov chains

Michael Choi

Cornell University
School of ORIE
cc2373@cornell.edu

May 22, 2017

Introduction

- Consider a Markov chain P with time-reversal $\widehat{P}$ on state space $\mathcal{X}$

Introduction

- Consider a Markov chain P with time-reversal $\widehat{P}$ on state space $\mathcal{X}$
- Non-reversible Markov chains are of great theoretical and applied interest

Introduction

- Consider a Markov chain P with time-reversal $\widehat{P}$ on state space $\mathcal{X}$
- Non-reversible Markov chains are of great theoretical and applied interest
- The major theoretical challenge is to analyze non-self-adjoint operators

Introduction

- Consider a Markov chain P with time-reversal $\widehat{P}$ on state space $\mathcal{X}$
- Non-reversible Markov chains are of great theoretical and applied interest
- The major theoretical challenge is to analyze non-self-adjoint operators
- From a Markov chain Monte Carlo perspective, it has been demonstrated that non-reversibility can improve the rate of convergence:
  - Hwang et al. '93 and '05: acceleration by adding anti-symmetric drift
  - Diaconis et al. '00: proposes a non-reversible sampler
  - Sun et al. '10, Bierkens '16: non-reversible Metropolis-Hastings by vortices/perturbations
  - Duncan et al. '13: non-reversible Langevin samplers

Literature review

A historical account for analyzing non-reversible Markov chains:

- Kendall '59: dilation by Sz.-Nagy's dilation theorem
- Fill '91, Paulin '15: reversiblizations
- Kontoyiannis and Meyn '12: recasting to a weighted-L^∞ space
- Patie and Savov '15, Miclo '16, Choi and Patie '16: intertwining/similarity orbit

We will focus on the similarity orbit and the reversiblization approach today.

The $\mathcal{S}$ class

Definition (The $\mathcal{S}$ class: A class of Markov chain $\mathcal{S}$imilar to normal Markov chain)

We say that $P \in \mathcal{S}$ if

$$P\Lambda = \Lambda Q$$

where

- P: transition kernel of general chain on $\mathcal{X}$
- Λ: a bounded link kernel with bounded inverse
- Q: transition kernel of normal chain, i.e. $Q\widehat{Q} = \widehat{Q}Q$.

The $\mathcal{S}$ class

Definition (The $\mathcal{S}$ class: A class of Markov chain $\mathcal{S}$imilar to normal Markov chain)

We say that $P \in \mathcal{S}$ if

$$P\Lambda = \Lambda Q$$

where

- P: transition kernel of general chain on $\mathcal{X}$
- Λ: a bounded link kernel with bounded inverse
- Q: transition kernel of normal chain, i.e. $Q\widehat{Q} = \widehat{Q}Q$.

Remark: Recall that we analyzed the spectral theory where P is upward skip-free and Q is a birth-death chain

The $\mathcal{S}$ class

Definition (The $\mathcal{S}$ class: A class of Markov chain $\mathcal{S}$imilar to normal Markov chain)

We say that $P \in \mathcal{S}$ if

$$P\Lambda = \Lambda Q$$

where

- P: transition kernel of general chain on $\mathcal{X}$
- Λ: a bounded link kernel with bounded inverse
- Q: transition kernel of normal chain, i.e. $Q\widehat{Q} = \widehat{Q}Q$.

Remark: If Λ is a Markov kernel, then P and Q are said to be intertwined by the link Λ.

The $\mathcal{S}$ class: Spectral theory

Theorem

Assume that $P \in \mathcal{S}$ with $P \overset{\Lambda}{\sim} Q$. Then the following holds.

- *Denote the self-adjoint spectral measure of Q by $\mathcal{E} = \{E_B;\ B \in \mathcal{B}(\mathbb{C})\}$, then $\{F_B := \Lambda E_B \Lambda^{-1};\ B \in \mathcal{B}(\mathbb{C})\}$ defines a spectral measure and P is a spectral scalar-type operator with spectral resolution given by*

$$P = \int_{\sigma(P)} \lambda \, dF_\lambda,$$
$$\widehat{P} = \int_{\sigma(\widehat{P})} \lambda \, dF^*_\lambda.$$

Note that

$$\sigma(P) = \sigma(Q), \sigma(P) = \overline{\sigma(\widehat{P})}, \sigma_c(P) = \sigma_c(Q), \sigma_p(P) = \sigma_p(Q), \sigma_r(P) = \sigma_r(Q),$$

and the multiplicity of each eigenvalue in $\sigma_p(P)$ is the same as that of $\sigma_p(Q)$.

The $\mathcal{S}$ class: Spectral theory

Theorem

Assume that $P \in \mathcal{S}$ *with* $P \overset{\Lambda}{\sim} Q$*. Then the following holds.*

- *For analytic and single valued function* f *on* $\sigma(P)$*, we have*

$$f(P) = \int_{\sigma(P)} f(\lambda)\, dF_\lambda.$$

The $\mathcal{S}$ class: Spectral theory

Theorem

Assume that $P \in \mathcal{S}$ with $P \overset{\Lambda}{\sim} Q$. Then the following holds.

- *In particular, if P is compact on $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ with distinct eigenvalues then for any $f \in \ell^2(\pi)$ and $n \in \mathbb{N}$,*

$$P^n f = \sum_{k=0}^{\mathfrak{r}} \lambda_k^n \langle f, f_k^* \rangle_\pi f_k,$$

where the set $(f_k)_{k=0}^{\mathfrak{r}}$ are eigenfunctions of P associated to the eigenvalues $(\lambda_k)_{k=0}^{\mathfrak{r}}$ and form a Riesz basis of $\ell^2(\pi)$, and the set $(f_k^*)_{k=0}^{\mathfrak{r}}$ is the unique Riesz basis biorthogonal to $(f_k)_{k=0}^{\mathfrak{r}}$. For any $x, y \in \mathcal{X}$ and $n \in \mathbb{N}$, the spectral expansion of P is given by

$$P^n(x, y) = \sum_{k=0}^{\mathfrak{r}} \lambda_k^n f_k(x) f_k^*(y) \pi(y).$$

The $\mathcal{S}$ class: Spectral theory

Eigentime identity (Aldous and Fill '02, Cui and Mao '10, Miclo '15):

1. Sample two points x, y independently from π

The $\mathcal{S}$ class: Spectral theory

Eigentime identity (Aldous and Fill '02, Cui and Mao '10, Miclo '15):

1. Sample two points x, y independently from π
2. Calculate the expected hitting time from x to y: $\mathbb{E}_x(\tau_y)$

The $\mathcal{S}$ class: Spectral theory

Eigentime identity (Aldous and Fill '02, Cui and Mao '10, Miclo '15):

1. Sample two points x, y independently from π
2. Calculate the expected hitting time from x to y: $\mathbb{E}_x(\tau_y)$
3. Expected value of this procedure is the sum of the inverse of the non-zero (and negative of the) eigenvalues of the generator:

$$\sum_{x,y \in \mathcal{X}} \mathbb{E}_x(\tau_y)\pi(x)\pi(y) = \sum_{i=1,\lambda_i \neq 0}^{|\mathcal{X}|} \frac{1}{\lambda_i}$$

The $\mathcal{S}$ class: Spectral theory

Corollary (Eigentime identity)

Suppose that $\mathcal{X}$ is a finite state space. If $L \in \mathcal{S}(G)$ with eigenvalues $(-\lambda_i)_{i \in [\![|\mathcal{X}|]\!]}$, then $(P_t)_{t \geq 0}$ and $(Q_t)_{t \geq 0}$ share the same eigentime identity.

$$\sum_{x,y \in \mathcal{X}} \mathbb{E}_x(\tau_y^Q) \pi_Q(x) \pi_Q(y) = \sum_{x,y \in \mathcal{X}} \mathbb{E}_x(\tau_y^P) \pi(x) \pi(y) = \sum_{i=1, \lambda_i \neq 0}^{|\mathcal{X}|} \frac{1}{\lambda_i}.$$

The $\mathcal{S}$ class: Convergence to equilibrium

Two notations:

- Second largest eigenvalue in modulus of P:

$$\lambda_* = \lambda_*(P) := \sup\{|\lambda| \in \sigma(P);\ \lambda \neq 1\}$$

- Second largest singular value of P:

$$\sigma_* = \sigma_*(P) := \sqrt{\lambda_*(P\widehat{P})}$$

- $\lambda_* \leq \sigma_*$ (with equality holds if P is normal)
- For a general P,

$$\lambda_*^n \leq \|P^n - \pi\|_{\ell^2(\pi)\to\ell^2(\pi)} \leq \sigma_*^n$$

- For a reversible P,

$$\|P^n - \pi\|_{\ell^2(\pi)\to\ell^2(\pi)} = \lambda_*^n$$

The $\mathcal{S}$ class: Convergence to equilibrium

Theorem ($\ell^2(\pi)$ distance)

Let $P \in \mathcal{S}$ with stationary distribution π. For $n \in \mathbb{N}$,

$$\lambda_*^n \leq \|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leq \sigma_*^n \mathbf{1}_{\{n < n^*\}} + \kappa_\Lambda \lambda_*^n \mathbf{1}_{\{n \geq n^*\}},$$

where $n^* = \lceil \frac{\ln \kappa_\Lambda}{\ln \sigma_* - \ln \lambda_*} \rceil$ and $\kappa_\Lambda = |||\Lambda|||\, |||\Lambda^{-1}||| \geq 1$ is the condition number of Λ.

The $\mathcal{S}$ class: Convergence to equilibrium

Theorem ($\ell^2(\pi)$ distance)

Let $P \in \mathcal{S}$ with stationary distribution π. For $n \in \mathbb{N}$,

$$\lambda_*^n \leq \|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leq \sigma_*^n \mathbf{1}_{\{n < n^*\}} + \kappa_\Lambda \lambda_*^n \mathbf{1}_{\{n \geq n^*\}},$$

where $n^* = \lceil \frac{\ln \kappa_\Lambda}{\ln \sigma_* - \ln \lambda_*} \rceil$ and $\kappa_\Lambda = |||\Lambda|||\,|||\Lambda^{-1}||| \geq 1$ is the condition number of Λ.

- A sufficient condition for which $\lambda_* < \sigma_*$ is given by $\max_i P(i,i) > \lambda_*$ using Sing-Thompson theorem.

The $\mathcal{S}$ class: Convergence to equilibrium

Theorem ($\ell^2(\pi)$ distance)

Let $P \in \mathcal{S}$ with stationary distribution π. For $n \in \mathbb{N}$,

$$\lambda_*^n \leq \|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leq \sigma_*^n \mathbf{1}_{\{n < n^*\}} + \kappa_\Lambda \lambda_*^n \mathbf{1}_{\{n \geq n^*\}},$$

where $n^* = \lceil \frac{\ln \kappa_\Lambda}{\ln \sigma_* - \ln \lambda_*} \rceil$ and $\kappa_\Lambda = |||\Lambda|||\,|||\Lambda^{-1}||| \geq 1$ is the condition number of Λ.

- A sufficient condition for which $\lambda_* < \sigma_*$ is given by $\max_i P(i,i) > \lambda_*$ using Sing-Thompson theorem.
- Recall the notion of hypocoercivity (Villani '06), i.e. there exists a constant $C < \infty$ and $\rho \in (0,1)$ such that

$$\|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leq C \rho^n .$$

  We provide an *explicit* spectral interpretation of the constant C by κ_Λ, and the convergence rate seems to be more involved.

The $\mathcal{S}$ class: Convergence to equilibrium

Theorem (total variation distance)

Let $P \in \mathcal{S}$ with stationary distribution π. For $n \in \mathbb{N}$,

$$||P^n(x,\cdot) - \pi||_{TV} \leq \frac{1}{2}\sqrt{\frac{1-\pi(x)}{\pi(x)}}\left(\sigma_*^n \mathbf{1}_{\{n<n^*\}} + \kappa_\Lambda \lambda_*^n \mathbf{1}_{\{n \geq n^*\}}\right)$$

The $\mathcal{S}$ class: Convergence to equilibrium

Theorem (total variation distance)

Let $P \in \mathcal{S}$ with stationary distribution π. For $n \in \mathbb{N}$,

$$||P^n(x,\cdot) - \pi||_{TV} \leq \frac{1}{2}\sqrt{\frac{1-\pi(x)}{\pi(x)}}\left(\sigma_*^n \mathbf{1}_{\{n<n^*\}} + \kappa_\Lambda \lambda_*^n \mathbf{1}_{\{n \geq n^*\}}\right)$$

Remark: Recall the notion of geometrically ergodicity (Kendall '59, Meyn and Tweedie '94, Baxendale '05), i.e. there exists a constant $C_x < \infty$ and $\rho \in (0,1)$ such that for $x \in E$ and $n \in \mathbb{N}$,

$$||P^n(x,\cdot) - \pi||_{TV} \leq C_x \rho^n$$

The constant κ_Λ provides an *explicit* and *computable* interpretation, and the rate of convergence seems to be more involved.

Example: Ehrenfest model

- Ehrenfest model: a birth-death process $(Q_t)_{t\geq 0}$ on $[\![0, N]\!]$ with parameter $0 < p < 1$ and
  - birth rate $\lambda_x = p(N - x)$
  - death rate $\mu_x = (1 - p)x$
  - stationary distribution: binomial distribution π_Q with parameters N, p
  - eigenfunctions: Krawtchouk polynomials

$$\phi_j(x) = {}_2F_1\left(\left.\begin{matrix}-j, -x\\ -N\end{matrix}\right| p^{-1}\right)$$

  - spectral representation:

$$Q_t(x, y) = \pi_Q(y) \sum_{j=0}^{N} e^{-jt} \phi_j(x) \phi_j(y) \frac{(-1)^{-j} p^j}{j!(1-p)^j} (-N)_j$$

Example: Ehrenfest model

- Three orbits:
  - Permutation orbit Λ_σ
  - (Zhou '08, Diaconis and Miclo '15) Random walk orbit Λ_{rw}
  - (In thesis) Pure birth orbit Λ_{pb}

Example: Ehrenfest model

- Three orbits:
  - Permutation orbit Λ_σ : $\Lambda_\sigma = (\mathbf{1}_{y=\sigma(x)})_{x,y \in \mathcal{X}}$, $\Lambda_\sigma^{-1} = \Lambda_\sigma^T$
  - (Zhou '08, Diaconis and Miclo '15) Random walk orbit Λ_{rw} : For $j = 1, 3, \ldots, 2N-1$, the eigenvalue β_j, right eigenfunction ψ_j and left eigenfunction Ψ_j are given by

$$\beta_j := \cos\left(\frac{j\pi}{2N+1}\right),$$

$$\psi_j(x) := \cos\left(\frac{(2x+1)j\pi}{2(2N+1)}\right), \quad x \in [\![0, N]\!],$$

$$\Psi_j(x) := \begin{cases} \psi_j(x), & \text{for } x \in [\![0, N-1]\!], \\ \dfrac{(-1)^{(j+1)/2}}{2} \cot\left(\dfrac{j\pi}{2(2N+1)}\right), & \text{for } x = N, \end{cases}$$

$$\Lambda_{rw} = \sum_{j \in \{0,1,3,\ldots,2N-1\}} \beta_j \psi_j \Psi_j^T,$$

  - (In thesis) Pure birth orbit Λ_{pb}

Example: Ehrenfest model

	Permutation orbit Λ_σ
Is Λ unitary?	✓
Is P_t reversible?	✓
Spectral representation of P_t	$P_t(x,y) = \sum_{j=0}^{N} e^{-jt} f_j(x) f_j^*(y) \dfrac{(-1)^{-j} p^j}{j!(1-p)^j}(-N)_j$
$f_j(x)$	$\phi_j(\sigma^{-1}(x))$
$f_j^*(y)$	$\phi_j(\sigma^{-1}(y))\pi_Q(\sigma^{-1}(y))$
$\\|P_t - \pi\\|_{\ell^2(\pi)\to\ell^2(\pi)}$	$e^{-t} = \kappa_{\Lambda_\sigma} e^{-t}$
Eigentime identity	$\sum_{j=1}^{N} \dfrac{1}{j}$

Example: Ehrenfest model

	Random walk orbit Λ_{rw}
Is Λ unitary?	✗
Is P_t reversible?	✗
Spectral representation of P_t	$P_t(x,y) = \sum_{j=0}^{N} e^{-jt} f_j(x) f_j^*(y) \dfrac{(-1)^{-j} p^j}{j!(1-p)^j} (-N)_j$
$f_j(x)$	discrete cosine transform of $(\beta_k^{-1} \langle \Psi_k, \phi_j \rangle)_{\mathrm{odd}\, k}$
$f_j^*(y)$	$\sum_{k \in \{0,1,3,\ldots,2N-1\}} \beta_k \langle \psi_k, \phi_j \rangle_{\pi_Q} \Psi_k(y)$
$\Vert P_t - \pi \Vert_{\ell^2(\pi) \to \ell^2(\pi)}$	$\leq \kappa_{\Lambda_{rw}} e^{-t}$
Eigentime identity	$\sum_{j=1}^{N} \dfrac{1}{j}$

Example: Ehrenfest model

Remarks:

- More examples and their orbits in the thesis: $M/M/\infty$ queue, linear and quadratic birth-death processes.
- We can consider d-dimensional *non-reversible* chains via *tensorized* orbits, e.g.

$$P_t \Lambda_{rw}^{\otimes d} = \Lambda_{rw}^{\otimes d} Q_t,$$

where $(Q_t)_{t \geq 0}$ is a multivariate reversible Markov chain, e.g. generalized Bernoulli-Laplace (Khare and Zhou '09), Dirichlet-multinomial Gibbs sampler (Khare and Zhou '09), Griffths and Spano '13, Griffths '16...

Cutoff phenomena

- Separation cutoff: generalizes the “spectral gap times mixing time going to infinity" criterion to a subclass $\mathcal{GMC} \subset \mathcal{S}$.
- L^2-cutoff: generalizes the L^2-cutoff criteria for reversible chains to the class $\mathcal{S}$ using the Laplace transform cutoff criteria in Chen and Saloff-Coste '09.

The Metropolis reversiblization (Aldous and Fill '02)

9.4 Making reversible chains from irreversible chains

Let $\mathbf{P}$ be an irreducible transition matrix on I with stationary distribution π. The following straightforward lemma records several general ways in which to construct from $\mathbf{P}$ a transition matrix $\mathbf{Q}$ for which the associated chain still has stationary distribution π but is *reversible*. These methods all involve the time-reversed matrix $\mathbf{P}^*$

$$\pi_i p_{ij} = \pi_j p^*_{ji}$$

and so in practice can only be used when we know π explicitly (as we have observed several times previously, in general we cannot write down a useful explicit expression for π in the irreversible setting).

Lemma 9.20 *The following definitions each give a transition matrix* $\mathbf{Q}$ *which is reversible with respect to* π*.*

The additive reversiblization: $\mathbf{Q}^{(1)} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^*)$

The multiplicative reversiblization: $\mathbf{Q}^{(2)} = \mathbf{P}\mathbf{P}^*$

The Metropolis reversiblization; $\mathbf{Q}^{(3)}_{i,j} = \min(p_{i,j}, p^*_{j,i}),\ j \neq i.$

The Metropolis reversiblization

Definition (Classical MH kernel)

The first MH chain, with transition kernel denoted by $M_1 := M_1(P)$, is the MH kernel with proposal kernel P and target distribution π. That is, let

$$\alpha_1(x, y) = \begin{cases} \min\left(\frac{\pi(y)p(y, x)}{\pi(x)p(x, y)}, 1\right) & \text{if } \pi(x)p(x, y) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$m_1(x, y) = \alpha_1(x, y)p(x, y) = \min\{\widehat{p}(x, y), p(x, y)\},$$

$$r_1(x) = \int_{y \neq x} (1 - \alpha_1(x, y))p(x, y)\,\mu(dy),$$

then M_1 is given by

$$M_1(x, dy) = m_1(x, y)\mu(dy) + r_1(x)\delta_x(dy).$$

Metropolis-Hastings reversiblizations

Nothing new so far...

Metropolis-Hastings reversiblizations

Definition (The second MH kernel)

The second MH kernel M_2 and density m_2 are given by

$$m_2(x,y) = \max\{\widehat{p}(x,y), p(x,y)\},$$
$$M_2(x,dy) = m_2(x,y)\mu(dy) - r_1(x)\delta_x(dy).$$

Metropolis-Hastings reversiblizations

Definition (The second MH kernel)

The second MH kernel M_2 and density m_2 are given by

$$m_2(x,y) = \max\{\widehat{p}(x,y), p(x,y)\},$$
$$M_2(x,dy) = m_2(x,y)\mu(dy) - r_1(x)\delta_x(dy).$$

Lemma

1. $P + \widehat{P} = M_1 + M_2$.
2. M_1 *and* M_2 *are self-adjoint bounded operators on* $L^2(\pi)$.
3. $M_1 = M_2 = P$ *if and only if* P *is reversible with respect to* π.
4. $M_i(P) = M_i(\widehat{P})$ *for* $i = 1, 2$.

Remark: Note that M_2 may not even be a contraction kernel.

Weyl's inequality

Suppose that $\mathcal{X}$ is a finite state space with $n = |\mathcal{X}|$, and arrange eigenvalues of a symmetric matrix M in non-increasing order by $\lambda_1(M) \geq \ldots \geq \lambda_n(M)$.

Theorem (Weyl's inequality for additive reversiblization)

1. *For* $i, j, k \in [\![n]\!]$ *and* $i + 1 = j + k$,

$$\lambda_i(P + \widehat{P}) \leq \lambda_j(M_1) + \lambda_k(M_2).$$

2. *For* $i, l, m \in [\![n]\!]$ *and* $i + n = l + m$,

$$\lambda_i(P + \widehat{P}) \geq \lambda_l(M_1) + \lambda_m(M_2).$$

Weyl's inequality

Corollary (Spectral gap bound via Weyl's inequality)

Denote

$$L := \max_{l+m=2+n} \{\lambda_l(M_1) + \lambda_m(M_2)\},$$

$$U := \min_{j+k=3} \{\lambda_j(M_1) + \lambda_k(M_2)\}.$$

Then

$$1 - \frac{1}{2}U \leq \gamma(P) \leq 1 - \frac{1}{2}L.$$

Remark: This bound is tight for asymmetric (p, q) simple random walk on n-cycle, which gives

$$1-\frac{1}{2}U = \max\{p, q\}(1-\cos(2\pi/n)) \leq 1-\cos(2\pi/n) = \gamma(P) = 1-\frac{1}{2}L.$$

More examples can be found in the thesis.

Weyl's inequality

Corollary

If P is a lazy and ergodic Markov kernel on a finite state space, then M_2 is a contraction kernel.

Comparison theorem

Theorem

For real-valued function $f \in L^2(\pi)$*,*

$$\langle M_2 f, f\rangle_\pi \leq \langle Pf, f\rangle_\pi \leq \langle M_1 f, f\rangle_\pi.$$

In particular,

$$\lambda(M_2) \leq \lambda(M_1), \quad \Lambda(M_2) \leq \Lambda(M_1),$$

where for $i = 1, 2$*,*

$$\lambda(M_i) := \inf\{\alpha : \alpha \in \sigma(M_i), \alpha < 1\},$$
$$\Lambda(M_i) := \sup\{\alpha : \alpha \in \sigma(M_i), \alpha < 1\}.$$

Comparison theorem

Remarks:

- $\langle Pf, f \rangle_\pi \leq \langle M_1 f, f \rangle_\pi$ is known as Peskun ordering in MCMC.
- This kind of inequality is more generally known as comparison theorems of Markov chains - Diaconis and Saloff-Coste '93, Dyer et al. '06
- This allows us to state a variant of Cheeger's inequality for non-reversible chains.

(Higher-order) Cheeger's inequality

Theorem (Lee et al. '12, Wang '14, Miclo '15)

Suppose that P is the transition kernel of a reversible finite Markov chain with eigenvalues $1 = \lambda_1 \geq \ldots \geq \lambda_n$. For $k \in [\![n]\!]$,

$$\frac{1-\lambda_k}{2} \leq \Phi_*(k) \leq O(k^4)\sqrt{1-\lambda_k},$$

where $\Phi_*(k)$ is the k-way expansion defined to be

$$\Phi_*(k) := \min_{(A_1,\ldots,A_k)\in\mathcal{D}_k} \max_{i\in[\![k]\!]} \frac{\langle P\mathbf{1}_{A_i}, \mathbf{1}_{A_i^c}\rangle_\pi}{\langle \mathbf{1}_{A_i}, \mathbf{1}_{A_i}\rangle_\pi},$$

and $\mathcal{D}_k$ is the set of k-uples of disjoint and π-non-negligible subsets of $\mathcal{X}$.

(Higher-order) Cheeger's inequality

Corollary

For non-reversible P and $k \in [\![n]\!]$,

$$\frac{1 - \lambda_k(M_1)}{2} \leq \Phi_*(k) \leq O(k^4)\sqrt{1 - \lambda_k(M_2)}\,.$$

MH-spectral gap

Suppose that P is the transition kernel of a non-reversible chain with time-reversal $\widehat{P}$ on a finite state space with stationary distribution π and spectral gap $\gamma(P)$.

$$||P^n(x, \cdot) - \pi||_{TV} \leq C_x \beta^n \,,$$

where β can be (possibly with a different C_x):

- $\sigma_*(P)$ (Fill '91)
- $1 - \gamma((P + \widehat{P})/2)$ for lazy chain (Fill '91)
- $\sqrt{1 - \gamma_{ps}}$ with $\gamma_{ps} = \max_{k \geq 1} \gamma(\widehat{P}^k P^k)/k$ ("*Pseudo*" spectral gap, Paulin '15)
- We now propose a gap based on M_1 and M_2.

MH-spectral gap

Definition (MH-spectral gap)

Denote

$$\mathcal{C} := \{n \in \mathbb{N} : |\lambda(M_2(P^n))| < 1, \Lambda(M_1(P^n)) < 1\},$$
$$\beta^{MH} := \sup_{n \in \mathcal{C}} \{|\lambda(M_2(P^n))|^{1/n}, \Lambda(M_1(P^n))^{1/n}\}.$$

The MH-spectral gap $\gamma^{MH} = \gamma^{MH}(P)$ is given by

$$\gamma^{MH} := 1 - \beta^{MH}.$$

Remark: For reversible P, $\gamma^{MH} = \gamma$, the classical L^2-spectral gap.

MH-spectral gap

Theorem

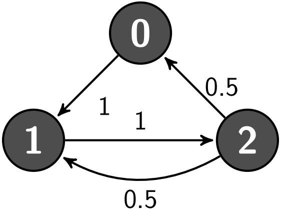
If $|\mathcal{C}^c| < \infty$, *then for all* $n \in \mathbb{N}$ *and* $x \in \mathcal{X}$,

$$||P^n(x,\cdot) - \pi||_{TV} \leq C_x(\beta^{MH})^n.$$

Remark: In some (but not all) numerical examples, the convergence rate β^{MH} outperform the rate of Fill '91 and Paulin '15. See the thesis.

MH-spectral gap

Non-reversible Markov chain on triangle (Montenegro and Tetali '06)



$$||P^n(x,\cdot) - \pi||_{TV} \leq C_x \beta^n$$

where β can be

- $\sigma_*(P) = 1$ (Fill '91)
- ~~$1 - \gamma((P + \widehat{P})/2)$ for lazy chain (Fill '91)~~
- $\sqrt{1 - \gamma_{ps}} = 0.866$ (Paulin '15)
- $\beta^{MH} = 0.849$ ☺

Variance bounds

Theorem (Variance bounds for reversible chains)

For reversible P and $f \in L^2(\pi)$,

$$V_f := \mathrm{Var}_\pi(f),$$
$$\sigma^2_{as} := \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right).$$

Then,

$$\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right) \leq nV_f\left(\frac{2}{\gamma}\right),$$
$$\left|\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right) - n\sigma^2_{as}\right| \leq V_f\left(\frac{4}{\gamma^2}\right).$$

Variance bounds

Theorem (Variance bounds for non-reversible chains, Paulin '15)

For non-reversible P and $f \in L^2(\pi)$,

$$V_f := \mathrm{Var}_\pi(f),$$
$$\sigma_{as}^2 := \lim_{n\to\infty} \frac{1}{n} \mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right).$$

Then,

$$\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right) \leq n V_f \left(\frac{4}{\gamma_{ps}}\right),$$
$$\left|\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right) - n\sigma_{as}^2\right| \leq V_f \left(\frac{16}{\gamma_{ps}^2}\right).$$

Variance bounds

Theorem (Variance bounds for non-reversible chains)

For non-reversible P and $f \in L^2(\pi)$,

$$V_f := \mathrm{Var}_\pi(f),$$
$$\sigma^2_{as} := \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right).$$

Then,

$$\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right) \leq nV_f\left(|\mathcal{C}^c| + \frac{4}{\gamma^{MH}}\right),$$
$$\left|\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right) - n\sigma^2_{as}\right| \leq V_f 4\left(1 + |\mathcal{C}^c| + \frac{4(\beta^{MH})^{|\mathcal{C}^c|+1}}{\gamma^{MH}}\right)^2.$$

Summary

1. Similarity orbit of normal Markov chains
    - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
    - Eigentime identity
    - Convergence to equilibrium
    - Separation and L^2-cutoff
    - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
    - Weyl's inequality
    - Comparison theorem
    - MH-spectral gap
    - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
   - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
   - Eigentime identity
   - Convergence to equilibrium
   - Separation and L^2-cutoff
   - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
   - Weyl's inequality
   - Comparison theorem
   - MH-spectral gap
   - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
    - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
    - Eigentime identity
    - Convergence to equilibrium
    - Separation and L^2-cutoff
    - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
    - Weyl's inequality
    - Comparison theorem
    - MH-spectral gap
    - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
   - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
   - Eigentime identity
   - Convergence to equilibrium
   - Separation and L^2-cutoff
   - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
   - Weyl's inequality
   - Comparison theorem
   - MH-spectral gap
   - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
   - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
   - Eigentime identity
   - Convergence to equilibrium
   - Separation and L^2-cutoff
   - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
   - Weyl's inequality
   - Comparison theorem
   - MH-spectral gap
   - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
   - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
   - Eigentime identity
   - Convergence to equilibrium
   - Separation and L^2-cutoff
   - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
   - Weyl's inequality
   - Comparison theorem
   - MH-spectral gap
   - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
    - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
    - Eigentime identity
    - Convergence to equilibrium
    - Separation and L^2-cutoff
    - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
    - Weyl's inequality
    - Comparison theorem
    - MH-spectral gap
    - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
   - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
   - Eigentime identity
   - Convergence to equilibrium
   - Separation and L^2-cutoff
   - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
   - Weyl's inequality
   - Comparison theorem
   - MH-spectral gap
   - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
   - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
   - Eigentime identity
   - Convergence to equilibrium
   - Separation and L^2-cutoff
   - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
   - Weyl's inequality
   - Comparison theorem
   - MH-spectral gap
   - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
   - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
   - Eigentime identity
   - Convergence to equilibrium
   - Separation and L^2-cutoff
   - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
   - Weyl's inequality
   - Comparison theorem
   - MH-spectral gap
   - Variance bounds

Summary

1. Similarity orbit of normal Markov chains
   - Spectral theory and functional calculus for Markov chains in this $\mathcal{S}$ class
   - Eigentime identity
   - Convergence to equilibrium
   - Separation and L^2-cutoff
   - New non-reversible examples with known spectral expansion, eigenfunction and stationary distribution
2. Metropolis-Hastings reversiblizations
   - Weyl's inequality
   - Comparison theorem
   - MH-spectral gap
   - Variance bounds

References

This talk is based on

- M.C.H. Choi and P. Patie. Skip-free Markov chains. Submitted.
- M.C.H. Choi and P. Patie. Analysis of non-reversible Markov chains via similarity orbit. Submitted.
- M.C.H. Choi. Metropolis-Hastings reversibilizations of non-reversible chains. Submitted.

Thank you! Question(s)?