Adding vorticity matrix to a reversible Markov chain

Huang Lu-Jing

Beijing Normal University

(based on a joint work with professor Mao Y-H)

Contents

1. Reversibility and non-reversibility

- We assume that V is a finite state space. Let $K = (K(i,j))_{i,j\in V}$ be a probability transition matrix(PTM), reversible with respect to a probability measure μ:

$$\mu(i)K(i,j) = \mu(j)K(j,i), \quad i,j \in V.$$

- Denote P as a PTM with stationary distribution μ:

$$\sum_i \mu(i)P(i,j) = \mu(j).$$

In general, P is not reversible, but we can get a reversible part from P:

$$K(i,j) = \frac{1}{2}\left[P(i,j) + \mu(j)P(j,i)/\mu(i)\right].$$

At that time,

$$\lim_{n\to\infty} K^n(i,j) = \lim_{n\to\infty} P^n(i,j) = \mu(j).$$

- investigate the properties of chain P and K.

An example: MCMC

- Suppose μ has the form:

$$\mu(i) = Ce^{-\sum_j V_{ij}},$$

where C is unknown and we want to estimate it.

- One method is the Metropolis algorithm. Choose a basic probability matrix J such that

$$J(i,j) > 0 \Leftrightarrow J(j,i) > 0, \quad i,j \in V.$$

Set the acceptance ratio (C disappears)

$$A(i,j) = \frac{\mu(j)J(j,i)}{\mu(i)J(i,j)}, \quad i,j \in V.$$

- A Metropolis chain is formulated as follows:

$$K(i,j) = \begin{cases} J(i,j), & \text{for } i \neq j, A(i,j) \geq 1; \\ J(i,j)A(i,j), & \text{for } i \neq j, A(i,j) \leq 1; \\ 1 - \cdots, & \text{for } i = j. \end{cases}$$

K is reversible w.r.t. μ. Moreover, μ is its unique stationary distribution. Then have $\lim_{n\to\infty} K^n(i,j) = \mu(j)$.
In this case, the better chains is, the faster we will obtain μ.

2. Comparison criteria

- Asymptotic variance related to CLT:
- Spectral gap:
- Mixing times(our comparison criterion).

Asymptotic variance

- Asymptotic variance related to CLT: Let X_k is the Markov chain of P and it has stationary distribution μ. Then for any function $f : V \to \mathbb{R}$,

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (f(X_k) - \mu(f)) \Rightarrow N(0, \nu(f, P, \mu))$$

with

$$\nu(f, P, \mu) = \lim_{n \to \infty} \mathrm{Var}_\mu \left[\sum_{k=0}^{n-1} f(X_k)/\sqrt{n} \right].$$

- Thus the smaller $\nu(f, P, \mu)$, the better the chain is. ([Chen, Hwang 2013], [Bierkens 2015]).

Spectral gap

- In $L^2(\mu)$:
$$\lambda(P) = \sup\left\{\text{real part of } \sigma(P) \setminus \{1\}\right\}.$$
- By total variance:
$$\rho(P) = \inf\left\{\epsilon : |P^n(i,j) - \mu(j)| \le C_{ij}\epsilon^n\right\}.$$
- Also the smaller $\lambda(P)(\rho(P))$, the better the chain is.
([Hwang, Hwang-Ma, Sheu 1993, 2005])
- In the reversible case, $\lambda(P) = \rho(P)$ and we have Poincaré inequality
$$1 - \lambda(P) = \inf\left\{\langle f, (I-P)f \rangle : \mu(f) = 0, \mu(f^2) = 1\right\}.$$

Our criterion: mixing times

- Let
$$\tau_i = \inf\{n \geq 0 : X_n = i\}.$$
- For any pair of points i, j in V, let
$$T_{ij}(P) = \mathbb{E}_i\tau_j + \mathbb{E}_j\tau_i$$
be the commute time between i, j of X.
- For a subset A of V, let
$$T_A(P) = \mathbb{E}_\mu\tau_A$$
be the first hitting time of A from stationary start. When $A = i$, denote $T_A(P)$ as $T_i(P)$.
- The average hitting time
$$T_0(P) := \sum_i \sum_j \mu(i)\mu(j)\mathbb{E}_i\tau_j = \frac{1}{2}\sum_i \sum_j \mu(i)\mu(j)T_{ij}(P)$$
$$= \sum_j T_j(P).$$

One reason: strong ergodicity

- In general, if $T_0(P) < \infty$, then the chain is strong ergodicity: there exist $C < \infty$ and $\rho < 1$ such that

$$\sup_i \sum_j |P^n(i,j) - \mu(j)| \leq C\rho^n.$$

Moreover, we have $\rho \leq 1 - 1/T_0(P)$.

- To see that the smaller average hitting time is, the better the chain is.

another motivation: Aldous-Fill's conjecture

Let P be a ergodic PTM with stationary distribution μ and Z is its fundamental matrix,i.e.

$$Z(i,j) = \sum_{n=0}^{\infty}(P^n(i,j) - \mu(j)), \quad i,j \in V.$$

Let

$$P^*(i,j) = \frac{\mu(j)P(j,i)}{\mu(i)} \quad \text{and} \quad K(i,j) = \frac{1}{2}\left[P(i,j) + P^*(i,j)\right].$$

Aldous-Fill in their book (1995) conjectured that

$$\text{trace}(Z^2(P^* - P)) \geq 0.$$

And they proved that this conjecture can yield that(in fact we can prove they are equivalent)

$$T_0(P_\lambda) \leq T_0(K),$$

where

$$P_\lambda := \lambda P + (1-\lambda)P^* = K + (\lambda - \frac{1}{2})\text{diag}(\mu)^{-1}\Gamma, \ 0 \leq \lambda \leq 1.$$

$$\Gamma(i,j) = \mu(i)P(i,j) - \mu(j)P(j,i).$$

3. Vorticity matrix

- It is easy to see that Γ satisfies:

$$\Gamma(i,j) = -\Gamma(j,i), \quad i,j \in V;$$

$$\sum_j \Gamma(i,j) = 0, \quad i \in V.$$

  In the following, we call the matrix as vorticity matrix if it satisfies above two conditions .

- Conversely, if we have a reversible(w.r.t μ) probability matrix K, choose a vorticity matrix Γ such that:

$$\Gamma(i,j) \leq \mu(i)K(i,j), \quad i,j \in V.$$

  Then $P := K + \text{diag}(\mu)^{-1}\Gamma$ be a PTM and it has same stationary distribution μ.
  For these two chains, we have following results.

4. Main results-(1) The case of reversibility and non-reversibility

Theorem

Let Γ *be a vorticity matrix, such that* $P = K + \mathrm{diag}(\mu)\Gamma$ *be a PTM. Fix any pair of points* $i \neq j$ *in* V *and let* $T_{ij}(K), T_{ij}(P)$ *respectively be the commute time between* i, j *of chains* K *and* P*. Then*

$$T_{ij}(P) \leq T_{ij}(K).$$

Consequently, the average hitting times of the chains satisfy

$$T_0(P) \leq T_0(K).$$

Theorem

Let Γ *be a vorticity matrix, such that* $P = K + \mathrm{diag}(\mu)\Gamma$ *be a PTM. Fix any subset* A *of* V*, let* $T_A(K), T_A(P)$ *respectively be the first hitting time to* A *from stationary start of chains* K *and* P*. Then*

$$T_A(P) \leq T_A(K).$$

Tools

- Capacity representation for the commute time.
- Dirichlet principle of capacity and the first hitting time.

Capacity representation for the commute time

- For two disjoint $i, j \in V$, the capacity of chain P between i and j is

$$C_{ij}(P) = \mu(i)\mathbb{P}_i(\tau_j < \tau_i^+),$$

where $\tau_i^+ = \min\{n \geq 1 : X_n = i\}$ is the first return time to i of chain P.

- The following lemma gives the relation between capacity and the commute time.

Lemma (Aldous-Fill book, chapter 2, corollary 8)

For $i \neq j$ in V,

$$T_{ij}(P) = \frac{1}{C_{ij}(P)}.$$

Dirichlet principle

- Define $\langle f, g\rangle = \sum_{i\in V} \mu(i)f(i)g(i)$ for $f, g : V \to \mathbb{R}$.

Lemma (Gaudillière-Landim 2014)

Let P be an irreducible PTM with stationary distribution μ. For every pair of points $i \neq j$ in V,

$$C_{ij}(P) = \inf\{\langle f, (I-P)(I-K)^{-1}(I-P)^* f\rangle : f(i) = 1, f(j) = 0\}.$$

- Note that if P is reversible, i.e., $P = K$, then

$$C_{ij}(P) = \inf\{\langle f, (I-P)f\rangle : f(i) = 1, f(j) = 0\}.$$

Lemma

Let P be an irreducible PTM with stationary distribution μ. For any subset A of V,

$$T_A(P) = \inf\{\langle f, (I-P)(I-K)^{-1}(I-P)^* f\rangle : f\,|_A = 1, \langle f, 1\rangle = 0\}.$$

- Note that if P is reversible, i.e., $P = K$, then

$$T_A(P) = \inf\{\langle f, (I-P)f\rangle : f\,|_A = 1, \langle f, 1\rangle = 0\}.$$

Parametrize the vorticity matrix

- Choose a vorticity matrix Γ such that $K + \text{diag}(\mu)^{-1}\Gamma$ be a PTM.
- For $\lambda \in [-1, 1]$, let
$$P_\lambda = K + \lambda \text{diag}(\mu)^{-1}\Gamma.$$
Then P_λ also is a PTM and has the same stationary distribution μ.
- For any pair of point $i \neq j$ and subset A, let $T_{ij}(\lambda), T_A(\lambda)$ be the commute time between i, j and the first hitting time to A from stationary start of chain P_λ respectively. Similarly, we define $T_0(\lambda)$ as the average hitting time of P_λ.

4. Main results-(2) The case of single parameter

Theorem

For any pair of points $i \neq j$ and subset A, let $S(\lambda)$ be any of $T_{ij}(\lambda)$, $T_A(\lambda)$, and $T_0(\lambda)$. Then
(a) (symmetry) For every $\lambda \in [-1,1]$,

$$S(\lambda) = S(-\lambda).$$

(b) (monotone) $S(\lambda)$ increases on $[-1,0]$. In particular,

$$\max_{-1\le\lambda\le 1} S(\lambda) = S(0) \text{ and } \min_{-1\le\lambda\le 1} S(\lambda) = S(1).$$

Moreover if the matrix Γ exists a row that only has two nonzero elements, then $T_0(\lambda)$ is strictly increasing.

- By letting $K = \frac{1}{2}(P + P^*)$ and $\Gamma = (\lambda - 1/2)\text{diag}(\mu)(P - P^*)$, we can prove Aldous-Fill's conjecture([Aldous-Fill book, chapter 9]).

Circle associated to the vorticity matrix

- For the reversible chain K, let $G = (V, E)$ be the directed graph associated to K, where V is the state space and $E = \{(i,j) \in V \times V : K(i,j) > 0\}$ is the set of edges, here we distinguish edges (i,j) and (j,i).
- For different vertices $i_0, i_1, ..., i_{n-1}$ and define $i_n = i_0$. If $(i_k, i_{k+1}) \in E$, $k = 0, 1, ..., n-1$, then we call $c = (i_0, i_1, ..., i_{n-1}, i_n)$ is a cycle of G, and say that $(i_k, i_{k+1}),\ (i_{k+1}, i_k),\ k = 0, 1, ..., n-1$ are the edges of c; $i_k,\ k = 0, 1, ..., n-1$ are the vertices of c.
- Define the unit vorticity matrix $\Gamma^{(c)} = (\Gamma^{(c)}(i,j) : i, j \in V)$ associated with the cycle $c = (i_0, i_1, ..., i_{n-1}, i_n)$ as

$$\Gamma^{(c)}(i,j) = \begin{cases} 1, & i = i_k, j = i_{k+1},\ k = 0, 1, ..., n-1; \\ -1, & i = i_{k+1}, j = i_k,\ k = 0, 1, ..., n-1; \\ 0, & \text{otherwise.} \end{cases}$$

Decomposition of the vorticity matrix

Proposition

Assume that Γ *is a vorticity matrix such that* $P = K + \mathrm{diag}(\mu)^{-1}\Gamma$ *is a transition matrix. Then there exist cycles* $c_1, c_2, ..., c_m$ *on* G *and positive* $\lambda_1, \lambda_2, ..., \lambda_m (m \geq 1)$ *such that*

$$\Gamma = \lambda_1 \Gamma^{(c_1)} + \lambda_2 \Gamma^{(c_2)} + ... + \lambda_m \Gamma^{(c_m)}.$$

Furthermore, $\Gamma^{(c_1)}, \Gamma^{(c_2)}, ..., \Gamma^{(c_m)}$ *can choose be linearly independent in the sense that if there exist* $\alpha_1, \alpha_2, ..., \alpha_m \in \mathbb{R}$ *such that*

$$\alpha_1 \Gamma^{(c_1)} + \alpha_2 \Gamma^{(c_2)} + ... + \alpha_m \Gamma^{(c_m)} = 0,$$

then $\alpha_1 = \alpha_2 = ... = \alpha_m = 0.$

- Unfortunately, this decomposition is not unique, and the circles may intersect.

Multiple parameter

Assume that the graph G associated with chain K has the cycles $c_1, ..., c_r$. For $\lambda = (\lambda_1, \cdots, \lambda_r)$, denote

$$P(\lambda) = K + \sum_k \lambda_k \mathrm{diag}(\mu)^{-1}\Gamma^{(c_k)},$$

where λ such that $P(\lambda)$ be PTM. Then for any λ, let $T_{ij}(\lambda)$ be the commute time between i and j of $P(\lambda)$. Similarly, $T_A(\lambda)$ be the first time to A from stationary start and $T_0(\lambda)$ be the average hitting time respectively.

4. Main results-(3) The case of multiple parameter

Theorem

Assume that G has cycles $c_1, \ldots, c_r$ with not common edges. Fix any pair of points $i \neq j$ in V, let $S(\lambda)$ be any of the mixing times above. Then
(a) (symmetry)

$$S(\lambda_1, \cdots, \lambda_r) = S(|\lambda_1|, \ldots, |\lambda_r|).$$

(b) (monotone) $S(\lambda)$ increases for $\lambda \leq 0$. That is,

$$S(\lambda) \leq S(\hat{\lambda}), \quad \lambda_k \leq \hat{\lambda}_k \leq 0.$$

Reference

- Aldous D.J., Fill J.A. Reversible Markov chains and random walks on graphs. URL www.berkeley.edu/users/aldous/book.html, 1994-2012.
- Bierkens J. Non-reversible Metropolis-Hastings. http://arxiv.org/abs/1401.8087, 2015.
- Chen T.-L., Hwang C.-R. Accelerating reversible Markov chains. Statistics and Probability Letters, 2013, 83(9), 1956-1962.
- Gaudillière A., Landim C. A Dirichlet principle for non reversible Markov chains and some recurrence theorems. Probab. Theory Relat. Fields, 2014, 158, 55-89.
- Hwang C.-R., Hwang-Ma S.-Y., Sheu S.-J. Accelerating diffusions. The Annals of Applied Probability, 2005, 15(2), 1433-1444.
- Hwang C.-R., Hwang-Ma S.-Y., Sheu S.-J. Accelerating Gaussian diffusions. The Annals of Applied Probability, 1993, 3(3), 897-913.
- Huang L.-J., Mao Y.-H. On some mixing times for non-reversible finite Markov chains.

Thank you!