Applications of Cheeger's Constant to The Convergence Rate of Markov Chains on R^n

by

Wai Kong Yuen*
Department of Mathematics
University of Toront

Technical Report No. 9717, September 5, (1997)

TECHNICAL REPORT SERIES
University of Toronto
Department of Statistics

APPLICATIONS OF CHEEGER'S CONSTANT TO THE CONVERGENCE RATE OF MARKOV CHAINS ON $\mathbf{R}^n$

by

Wai Kong Yuen*

(September 5,1997.)

Acknowledgments. This work is part of my doctoral research done under the direction of Jeffrey S. Rosenthal.

Abstract. Quantitative geometric rates of convergence for reversible Markov chains are closely related to the Cheeger's constant, which is hard to calculate for general state spaces. This article describes a geometric argument to bound the Cheeger's constant for chains on bounded subsets of $\mathbf{R}^n$.

1. Introduction. As Markov chain Monte Carlo algorithms (Gelfand and Smith (1990), Smith and Roberts (1993)) are more widely used, quantitative geometric rates of convergence for Markov chains becomes an important topic. Diaconis (1988), Sinclair and Jerrum (1989), Jerrum and Sinclair (1988), Diaconis and Stroock (1991), Sinclair (1992) and Diaconis and Saloff-Coste (1993) proved general results on *finite* state spaces. Hanlon (1992), Frieze et al. (1994), Frigessi et al. (1993), Ingrassia (1994) and Belsley (1993) proved results specifically for Markov chain Montè Carlo. On *general* state spaces, not many results have been found yet. For partial results, see Amit and Grenander (1991), Amit (1991), Hwang et al. (1993), Lawler and Sokal (1988), Meyn and Tweedie (1994), Rosenthal (1995a, 1995b, 1996a, 1996b), Baxter and Rosenthal (1995) and Roberts and Rosenthal (1997a, 1997b).

In particular, Lawler and Sokal (1988) proved the Cheeger's inequality for positive-recurrent discrete-time Markov chains and continuous-time Markov jump process. Jerrum and Sinclair (1988) used a geometric argument with paths to bound the Cheeger's constant

* Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 3G3.Internet: yuen@math.toronto.edu

for a discrete-time finite space Markov chain. In this paper, we shall use an analogous geometric argument to bound the Cheeger's constant for Markov chains on bounded subsets of $\mathbf{R}^n$.

The remainder of this section introduces the definitions of the Cheeger's constant and its relation to the rate of convergence for Markov chains. In section 2, we prove a lower bound for the constant when the Markov chain is defined in a bounded convex subset of $\mathbf{R}^n$ and then extend the result to general bounded subsets of $\mathbf{R}^n$. Section 3 gives examples of both types.

1.1. *Basic notations.*

Consider a discrete-time Markov chain or a continuous time Markov jump process with measurable state space $(S,\mathcal{F})$, transition probability kernel $P(x, dy)$ and invariant probability measure π. Then P induces a positivity-preserving linear contraction on $L^2(\pi)$ by

$$(Pf)(x) = \int f(y)P(x, dy).$$

P also acts to the left on measures, so that

$$\mu P(A) = \int P(x, A)\mu(dx).$$

Recall that a Markov chain is *reversible* if

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx).$$

Hence, a Markov chain is reversible iff the operator P on $L^2(\pi)$ is self-adjoint. In this case, the spectrum is real and we can define

$$\lambda_0(P) = \inf \operatorname{spec}(P \mid_{1^\perp}),$$

$$\lambda_1(P) = \sup \operatorname{spec}(P \mid_{1^\perp}).$$

Define the Cheeger's constant as follows:

$$k \equiv \inf_{\substack{A \in \mathcal{F} \\ 0 < \pi(A) < 1}} k(A)$$

with

$$k(A) \equiv \frac{\int \chi_A(x)P(x, A^c)\pi(dx)}{\pi(A)\pi(A^c)},$$

which is the rate of probability flow, in the stationary Markov chain, from a set A to its complement A^c, normalized by the invariant probabilities of A and A^c. Intuitively, if k is very small, or equivalently, there is a set A s.t. the flow from A to A^c is very small compared to the invariant probabilities of A and A^c, then the Markov chain must converge very slowly to the invariant distribution. In fact, Lawler and Sokal(1988) proved the *Cheeger's inequality* for reversible Markov chains:

$$k^2/8 \leq 1 - \lambda_1(P) \leq k.$$

As we shall see in Section 1.2, $\lambda_1(P)$ is closely related to the convergence rate to the invariant distribution. Roughly speaking, the bigger the $\lambda_1(P)$, the lower the rate. So, the Cheeger's inequality says that when k is small, the rate is low, which matches with our intuition.

1.2. *L^2 convergence to π.*

We extend the definition of L^2 norm for signed measure. Given a signed measure μ on S,define $\|\mu\|_2$ by

$$\|\mu\|_2^2 = \begin{cases} \int_S |\frac{d\mu}{d\pi}|^2 d\pi, & \mu << \pi; \\ \infty, & \text{otherwise.} \end{cases}$$

Hence, we can also represent $L^2(\pi)$ by $\{\mu; \|\mu\|_2 < \infty\}$, so that μ and f represent the same element whenever $f = \frac{d\mu}{d\pi}$ a.e.. Finally, set

$$\|P\,|_{1^\perp}\,\|_2 = \sup\{\|Pf\|_2; f \in \mathbf{1}^\perp, \|f\|_2 = 1\}.$$

We shall use the fact that P is a bounded self-adjoint operator and a result from spectral theory (see e.g. Rudin (1991), Kreyszig (1978), Conway (1985)) to prove the following well-known proposition.

Proposition 1.1. For a discrete-time Markov chain, if P is reversible w.r.t. π and $\mu \in L^2(\pi)$, then

$$\|\mu P^n - \pi\|_2 \leq \|\mu - \pi\|_2 \rho^n,$$

where

$$\rho = \max\{|\lambda_0(P)|, \lambda_1(P)\}.$$

Proof. For $\mu \in L^2(\pi)$, let $f = \frac{d\mu}{d\pi}$. Since P is reversible w.r.t.π, it is easy to check $\frac{d(\mu P)}{d\pi} = Pf$. So, P acting on $\mu \in L^2(\pi)$ is the same as P acting on $f \in L^2(\pi)$, in the sense that $d\mu = f d\pi$ implies $d(\mu P) = (Pf)d\pi$. Define $\|P\,|_{1^\perp}\,\|_2$ by

$$\|P\,|_{1^\perp}\,\|_2 = \sup\{\|Pf\|_2; f \in \mathbf{1}^\perp \|f\|_2 = 1\}.$$

It is also well known(see e.g. Baxter and Rosenthal (1995)) that $\|P\,|_{1^\perp}\,\|_2 \leq 1$. From a result of spectral theory for bounded self-adjoint operator P,

$$\|P\,|_{1^\perp}\,\|_2 = \rho.$$

Furthermore, we observe that $f - \mathbf{1} \in \mathbf{1}^{\perp}$. Hence,

$$\begin{aligned}\|\mu P^n - \pi\|_2 &= \|\mu P^n - \pi P^n\|_2 \\ &= \|(\mu - \pi)P^n\|_2 \\ &= \|P^n(f - \mathbf{1})\|_2 \\ &\le \|f - \mathbf{1}\|_2 \|P^n\ |_{\mathbf{1}^\perp}\ \|_2 \\ &= \|\mu - \pi\|_2 \|P\ |_{\mathbf{1}^\perp}\ \|_2^n \\ &= \|\mu - \pi\|_2 \rho^n.\end{aligned}$$

Remarks.1. From the Cheeger's inequality, we have an upper bound for $\lambda_1(P)$. However, it is hard to bound $\lambda_0(P)$ in general (for *finite* state spaces, Diaconis and Stroock (1991) gave a geometric lower bound of it). If we assume $P(x,x) \ge a > 0 \forall x \in S$, then it directly implies that $\lambda_0(P) > -1 + 2a$. So, we have $\rho \le \max\{1 - 2a, \lambda_1(P)\}$. In particular, if we consider the chain $\frac{1}{2}(I + P)$, then $a \ge \frac{1}{2}$ and so $\rho = \lambda_1(P)$. In this case, we have an upper bound for the convergence rate in terms of k. In practice, however, it is hard to calculate k numerically for general state spaces. In section 2, we shall use a geometric argument to give a lower bound for it, and hence a upper bound for the convergence rate.

2. For continuous-time Markov jump processes, the result is much simpler. Then the corresponding operator with mean 1 exponential holding times can be written as $\bar{P}^t = \sum_{n=0}^{\infty} e^{-t} \frac{t^n}{n!} P^n$. So, $\sigma(\bar{P}^t) = \{\sum_{n=0}^{\infty} e^{-t} \frac{t^n}{n!} \lambda^n : \lambda \in \sigma(P)\}$ and $\lambda_0(\bar{P}^t) \ge 0$. Hence, we have the following relation:

$$\|\mu \bar{P}^t - \pi\|_2 \le \|\mu - \pi\|_2 e^{-t\lambda_1(P)}.$$

3.For probability measures μ, we have $\|\mu - \pi\|_2^2 = \|\mu\|_2^2 - 1$.

4. We used the L^2 norm to measure the rate of convergence. For the relation between L^2 norm and other norms, see e.g. Roberts and Rosenthal (1997b).

2. Lower Bound for Cheeger's Constant. In this section, we prove a lower bound for k of reversible Markov chains described in 1.1 in the case that S is a bounded subset of $\mathbf{R}^n$ with positive Lebesgue measure.

2.1. *Convex case with uniform paths.*

Consider a reversible Markov chain on S (discrete or continuous time). Let S be a bounded convex subset of $\mathbf{R}^n$ and D be the diameter of S. We assume that the transition kernel is of the form $P(\mathbf{x}, d\mathbf{y}) = \alpha(\mathbf{x})\delta_{\mathbf{x}}(d\mathbf{y}) + p_{\mathbf{x}}(\mathbf{y})d\mathbf{y}$ where $\delta_{\mathbf{x}}$ is the unit point mass on $\mathbf{x}$ for any $\mathbf{x} \in S$. Let π be the invariant distribution on S, which has density $q(\mathbf{y})$ w.r.t. Lebesgue measure.

For any $\mathbf{x},\mathbf{y} \in S$ and $b \in \mathbf{N}$, define a *uniform path* $\eta^b_{\mathbf{xy}} : \{0,\ldots,b\} \to S$ joining $\mathbf{x},\mathbf{y}$ by

$$\eta^b_{\mathbf{xy}}(i) = \frac{(b-i)\mathbf{x} + i\mathbf{y}}{b}.$$

This function is onto the points on the line joining $\mathbf{x},\mathbf{y}$ so that adjacent points have equal spacings. Before stating Theorem 2.1, we define a non-negative constant $g(b)$ by

$$g(b) \equiv \text{essinf}\{p_{\mathbf{x}}(\mathbf{y})q(\mathbf{x}) : \mathbf{x},\mathbf{y} \in S, 0 < \|\mathbf{x}-\mathbf{y}\| \le D/b,\},$$

where essinf is the essential infimum with respect to Lebesgue measure λ on $\mathbf{R}^n \times \mathbf{R}^n$,i.e.,

$$\text{essinf}\{f(x) : x \in X\} = \sup\{a : \lambda\{x : f(x) < a\} = 0\}.$$

Similarly, the essential supremum esssup is defined.

Theorem 2.1. Suppose that $g(b) > 0$ for some $b \in \mathbf{N}$ and $Q \equiv \text{esssup}\{q(\mathbf{x}) : \mathbf{x} \in S\}$ is finite. Then

$$k \ge \frac{g(b)}{Q^2 b^{n+1}}.$$

Proof. By assumption, we can choose $b \in \mathbf{N}$ such that $g(b) > 0$. For any $A \subset S, \mathbf{x} \in A$ and $\mathbf{y} \in A^c$, define $T_A : A \times A^c \to A \times A^c$ by

$$T_A(\mathbf{x},\mathbf{y}) = (\eta^b_{\mathbf{xy}}(l^A_{\mathbf{xy}} - 1), \eta^b_{\mathbf{xy}}(l^A_{\mathbf{xy}})),$$

where

$$l^A_{\mathbf{xy}} = \min\{i | \eta^b_{\mathbf{xy}}(i-1) \in A, \eta^b_{\mathbf{xy}}(i) \in A^c\}.$$

For $i = 1, ...b$, we also define $G^i : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n \times \mathbf{R}^n$ by

$$G^i(\mathbf{x},\mathbf{y}) = (\eta^b_{\mathbf{xy}}(i-1), \eta^b_{\mathbf{xy}}(i)).$$

Obviously, $(G^i)^{-1}$ exists for each i. Now, for any $(\mathbf{x},\mathbf{y}) \in A \times A^c$,, we have $G^{l^A_{\mathbf{xy}}}(\mathbf{x},\mathbf{y}) = T_A(\mathbf{x},\mathbf{y})$ and so

$$(\mathbf{x},\mathbf{y}) \in (G^{l^A_{\mathbf{xy}}})^{-1}(T_A(A \times A^c)),$$

which implies that

$$A \times A^c \subset \bigcup_{i=1}^{b} (G^i)^{-1}(T_A(A \times A^c)).$$

Then

$$\begin{aligned}\pi(A)\pi(A^c) &= \int_A \pi(d\mathbf{x}) \int_{A^c} \pi(d\mathbf{y}) \\ &= \iint_{A\times A^c} q(\mathbf{x})q(\mathbf{y})d\mathbf{x}d\mathbf{y} \\ &\leq \iint_{A\times A^c} Q^2 d\mathbf{x}d\mathbf{y} \\ &\leq Q^2 \iint_{\bigcup_{i=1}^{b}(G^i)^{-1}(T_A(A\times A^c))} d\mathbf{x}d\mathbf{y} \\ &\leq Q^2 \sum_{i=1}^{b} \iint_{(G^i)^{-1}(T_A(A\times A^c))} d\mathbf{x}d\mathbf{y}\end{aligned}$$

Now, for each i, observe that G^i is a 1-1 mapping from $\mathbf{R}^n$ onto $\mathbf{R}^n$ and continuous. So, we can consider the change of variables $(\mathbf{x},\mathbf{y}) = (G^i)^{-1}(\mathbf{u},\mathbf{v})$. Since $(G^i)^{-1}$ has continuous partial derivatives, it is differentiable and we can consider the Jacobian (determinant of the derivative $D(G^i)^{-1}(\mathbf{u},\mathbf{v})$) of the transformation, given by

$$J_i(\mathbf{u},\mathbf{v}) = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_n} & \frac{\partial x_1}{\partial v_1} & \cdots & \frac{\partial x_1}{\partial v_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_n} & \frac{\partial x_n}{\partial v_1} & \cdots & \frac{\partial x_n}{\partial v_n} \\ \frac{\partial y_1}{\partial u_1} & \cdots & \frac{\partial y_1}{\partial u_n} & \frac{\partial y_1}{\partial v_1} & \cdots & \frac{\partial y_1}{\partial v_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial y_n}{\partial u_1} & \cdots & \frac{\partial y_n}{\partial u_n} & \frac{\partial y_n}{\partial v_1} & \cdots & \frac{\partial y_n}{\partial v_n} \end{vmatrix}$$

where $\mathbf{u} = (u_1, \ldots, u_n)$ and so on. It can be shown easily that $|J_i| = b^n$ (since for each j, x_j, y_j depends only on u_j, v_j). So, by the change of variable formula (see e.g. Billingsley (1995), Theorem 17.2), we have

$$\begin{aligned}\pi(A)\pi(A^c) &\leq Q^2 \sum_{i=1}^{b} \iint_{T_A(A\times A^c)} |J_i| d\mathbf{u}d\mathbf{v} \\ &= Q^2 \sum_{i=1}^{b} \iint_{T_A(A\times A^c)} b^n d\mathbf{u}d\mathbf{v} \\ &\leq Q^2 b^{n+1} \iint_{T_A(A\times A^c)} \frac{p_{\mathbf{u}}(\mathbf{v})q(\mathbf{u})}{\operatorname{essinf}\{p_{\mathbf{x}}(\mathbf{y})q(\mathbf{x}) : \mathbf{x},\mathbf{y} \in T_A(A\times A^c)\}} d\mathbf{u}d\mathbf{v} \\ &\leq \frac{Q^2 b^{n+1}}{g(b)} \iint_{T_A(A\times A^c)} p_{\mathbf{u}}(\mathbf{v})q(\mathbf{u})d\mathbf{u}d\mathbf{v} \\ &\leq \frac{Q^2 b^{n+1}}{g(b)} \iint_{A\times A^c} p_{\mathbf{u}}(\mathbf{v})q(\mathbf{u})d\mathbf{u}d\mathbf{v} = \frac{Q^2 b^{n+1}}{g(b)} \int \chi_A(\mathbf{x})P(\mathbf{x}, A^c)\pi(d\mathbf{x}).\end{aligned}$$

Hence,

$$k(A) \geq \frac{g(b)}{Q^2 b^{n+1}}$$

for any $A \in \mathcal{F}$ s.t. $0 < \pi(A) < 1$. Since A is arbitrary, the result follows.

2.2. *Extension to general paths.*

Now suppose that S is a bounded subset of $\mathbf{R}^n$, not necessarily convex. So, Theorem 2.1 cannot be directly applied since the paths we used no longer exist in general. In this case, the idea is to choose new paths depending on the shape of S so that any two adjacent points are 'close' enough. Formally, for any $\mathbf{x}, \mathbf{y} \in S$ and $b \in \mathbf{N}$, let $\gamma^b_{\mathbf{xy}} : \{0, \ldots, b\} \to S$ be some path (not necessarily the uniform path $\eta^b_{\mathbf{xy}}$ in the previous section) s.t. $\gamma^b_{\mathbf{xy}}(0) = \mathbf{x}$ and $\gamma^b_{\mathbf{xy}}(b) = \mathbf{y}$. Extend the definition of the non-negative constant $g(b)$ by

$$g(b) \equiv \text{essinf}\{p_{\mathbf{u}}(\mathbf{v})q(\mathbf{u}) : \exists \mathbf{x}, \mathbf{y}, i \text{ s.t. } \mathbf{u} = \gamma^b_{\mathbf{xy}}(i), \mathbf{v} = \gamma^b_{\mathbf{xy}}(i+1)\}.$$

As in the proof of Theorem 2.2, we define T_A and G^i in terms of $\gamma^b_{\mathbf{xy}}$. To generalize the theorem, we further assume that G^i is a 1-1 mapping of an open set V containing S onto an open set $G^i(V)$and that $(G^i)^{-1}$ has continuous partial derivatives for each i. Let $J_i(\mathbf{u}, \mathbf{v})$ be the Jacobian of the change of variable $(\mathbf{x}, \mathbf{y}) = (G^i)^{-1}(\mathbf{u}, \mathbf{v})$. Then we have the following theorem:

Theorem 2.2. Suppose that $g(b) > 0$ for some $b \in \mathbf{N}$, $Q \equiv \text{esssup}\{q(\mathbf{x}) : \mathbf{x} \in S\}$ is finite and $0 < |J_i(\mathbf{u}, \mathbf{v})| \leq J_i$ for each i. Then

$$k \geq \frac{g(b)}{Q^2 \sum_{i=1}^b J_i}.$$

Proof. We can follow the proof of Theorem 2.1 until the change of variables. We have

$$\begin{aligned}
\pi(A)\pi(A^c) &\leq Q^2 \sum_{i=1}^b \iint_{T_A(A \times A^c)} |J_i(\mathbf{u}, \mathbf{v})| d\mathbf{u} d\mathbf{v} \\
&\leq Q^2 \sum_{i=1}^b \iint_{T_A(A \times A^c)} J_i d\mathbf{u} d\mathbf{v} \\
&\leq \frac{Q^2 \sum_{i=1}^b J_i}{g(b)} \iint_{A \times A^c} p_{\mathbf{u}}(\mathbf{v}) q(\mathbf{u}) d\mathbf{u} d\mathbf{v} \\
&= \frac{Q^2 \sum_{i=1}^b J_i}{g(b)} \int \chi_A(\mathbf{x}) P(\mathbf{x}, A^c) \pi(d\mathbf{x}).
\end{aligned}$$

This leads to the result.

Remarks.1.In the proofs of the theorems, $\alpha(\mathbf{x})$ doesn't play a role except making $P(\mathbf{x}, d\mathbf{y})$ a probability measure. In the following examples, unless otherwise specified, we only define $p_{\mathbf{x}}(\mathbf{y})$, and then $\alpha(\mathbf{x}) = 1 - \int_S p_{\mathbf{x}}(\mathbf{y})\mathbf{dy}$ is assumed.

2. For transition kernel of the form we considered in the theorems, the reversibility condition is equivalent to $q(\mathbf{x})p_{\mathbf{x}}(\mathbf{y}) = q(\mathbf{y})p_{\mathbf{y}}(\mathbf{x})$ for almost all $\mathbf{x}, \mathbf{y} \in S$.

3. There is no guarantee that such paths should exist in Theorem 2.2. All we say is that *if* such paths do exist, *then* the theorem holds.

4. It is not necessary that $(G^i)^{-1}$ has continuous partial derivatives for each i on the whole S. Measure-zero sets can be neglected. See Corollary 2.3 below.

Remark 3 suggests that we have no definite rules to find the paths in general. In particular, if S is non-convex, such paths may not even exist. However, if S satisfies the following conditions, we shall have at least one way to construct paths. Let U be open in $\mathbf{R}^n$ and $S \subset U$. Suppose there exists $\phi : U \to \phi(U)$ s.t. $\phi(U)$ is open and $\phi(S)$ is convex. Assume that ϕ^{-1} exists and both ϕ, ϕ^{-1} are continuously differentiable a.e.. Since we already have uniform paths $\eta^b_{\mathbf{xy}}$ on $\phi(S)$, we can define paths $\gamma^b_{\mathbf{xy}}$ on S by

$$\gamma^b_{\mathbf{xy}}(i) = \phi^{-1}(\frac{(b-i)\phi(\mathbf{x}) + i\phi(\mathbf{y})}{b}).$$

Following the discussions before Theorem 2.2, $g(b)$, T_A and G^i are defined in terms of $\gamma^b_{\mathbf{xy}}$. Define $\Phi : U \times U \to \phi(U) \times \phi(U)$ by $\Phi(\mathbf{x}, \mathbf{y}) = (\phi(\mathbf{x}), \phi(\mathbf{y}))$. So, the Jacobian of Φ exists and we can denote it by $J_\Phi(\mathbf{x}, \mathbf{y})$. Then we have the following corollary.

Corollary 2.3. Suppose $g(b) > 0$ for some $b \in \mathrm{N}$, $Q \equiv \mathrm{esssup}\{q(\mathbf{x}) : \mathbf{x} \in S\}$ is finite and $0 < m \leq |J_\Phi(\mathbf{x}, \mathbf{y})| \leq M$ a.e. for each i. Then

$$k \geq \frac{g(b) \cdot m}{Q^2 b^{n+1} M}.$$

In particular, if $|J_\Phi(\mathbf{x}, \mathbf{y})|$ is constant a.e.,

$$k \geq \frac{g(b)}{Q^2 b^{n+1}}.$$

Proof. Observe that Φ^{-1} exists and

$$G^i(\mathbf{x}, \mathbf{y}) = (\gamma^b_{\mathbf{xy}}(i-1), \gamma^b_{\mathbf{xy}}(i)) = \Phi^{-1} \circ G^i_\eta \circ \Phi(\mathbf{x}, \mathbf{y}),$$

where G^i_η denote the G^i in the proof of Theorem 2.1 corresponding to the uniform paths $\eta^b_{\mathbf{xy}}$. Let D denote the derivative operator. Then $|\det D(G^i_\eta)^{-1}| \equiv b^n$. Obviously, G^i satisfies the required conditions and $(G^i)^{-1} = \Phi \circ (G^i_\eta)^{-1} \circ \Phi^{-1}$ has continuous partial derivatives a.e. for each i. To apply Theorem 2.2, we need to show that the Jacobian

$J_i(\mathbf{u},\mathbf{v})$ of the change of variable $(\mathbf{x},\mathbf{y}) = (G^i)^{-1}(\mathbf{u},\mathbf{v})$ satisfies the following inequalities for almost all $(\mathbf{u},\mathbf{v})$:

$$0 < |J_i(\mathbf{u},\mathbf{v})| \le \frac{b^n M}{m}.$$

Indeed, by the chain rule and the inverse function theorem of differentiation,

$$\begin{aligned}
|J_i(\mathbf{u},\mathbf{v})| &= |\det D(G^i)^{-1}(\mathbf{u},\mathbf{v})| \\
&= |\det D(\Phi \circ (G^i_\eta)^{-1} \circ \Phi^{-1})(\mathbf{u},\mathbf{v})| \\
&= |\det D\Phi((G^i_\eta)^{-1} \circ \Phi^{-1}(\mathbf{u},\mathbf{v}))| \cdot |\det D(G^i_\eta)^{-1}(\Phi^{-1}(\mathbf{u},\mathbf{v}))| \cdot |\det D\Phi^{-1}(\mathbf{u},\mathbf{v})| \\
&= |J_\Phi((G^i_\eta)^{-1} \circ \Phi^{-1}(\mathbf{u},\mathbf{v}))| \cdot b^n \cdot |\frac{1}{\det D\Phi(\Phi^{-1}(\mathbf{u},\mathbf{v}))}| \\
&\le \frac{M \cdot b^n}{|J_\Phi(\Phi^{-1}(\mathbf{u},\mathbf{v}))|} \\
&\le \frac{b^n M}{m}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
|J_i(\mathbf{u},\mathbf{v})| &= |J_\Phi((G^i_\eta)^{-1} \circ \Phi^{-1}(\mathbf{u},\mathbf{v}))| \cdot b^n \cdot |\frac{1}{\det D\Phi(\Phi^{-1}(\mathbf{u},\mathbf{v}))}| \\
&\ge \frac{m \cdot b^n}{|J_\Phi(\Phi^{-1}(\mathbf{u},\mathbf{v}))|} \\
&\ge \frac{b^n m}{M} > 0.
\end{aligned}$$

Hence, by Theorem 2.2,

$$k \ge \frac{g(b)}{Q^2 \sum_{i=1}^{b} \frac{b^n M}{m}} = \frac{g(b) \cdot m}{Q^2 b^{n+1} M}.$$

In particular, if $|J_\Phi(\mathbf{x},\mathbf{y})|$ is constant a.e., $M = m$ and the result follows.

3.Examples. All the examples below can be considered as reversible discrete-time Markov chains or continuous-time Markov jump processes on S and transition probability kernel defined as in Section 2. They are all random walk chains, a particular case from the Metropolis-Hasting algorithm (see e.g. Tierney (1994)).

Example 3.1.One dimensional case: $S = [-a, a]$.

(a) Let $p_x(y) = \frac{1}{2}$ (the uniform p.d.f.) on $[x-1, x+1] \cap S$ and 0 otherwise. As $p_x(y) = p_y(x)$, the chain is reversible w.r.t. the invariant distribution π with density $q(x) = \frac{1}{2a}\mathbf{1}_S(x)$, uniform over S. So, $Q = \frac{1}{2a}$. To apply Theorem 2.1, we need to find b s.t. $g(b) > 0$. Since the essinf of $g(b)$ is taken over $x, y \in S, 0 < |x-y| \le \frac{D}{b}$ and $p_x(y) = 0$

for $|y-x|>1$, we need to choose b big enough s.t. $\frac{D}{b}\leq 1$. For the best lower bound, we take $b=\lceil 2a\rceil$. Then $\frac{D}{b}=\frac{2a}{b}\leq 1$ and $g(b)=\frac{1}{2}\frac{1}{2a}$. By Theorem 2.1,

$$k\geq \frac{\frac{1}{4a}}{(\frac{1}{2a})^2 b^{1+1}}=\frac{a}{b^2}.$$

In particular, if $2a$ is an integer, $b=2a$ and so $k\geq \frac{1}{4a}$. If we consider $\tilde{P}=\frac{1}{2}(I+P)$ as the transitional kernel for a discrete time Markov chain, then $\rho\leq 1-\frac{(\frac{1}{8a})^2}{8}=1-\frac{1}{8^3a^2}$. Then by Proposition 1.1, for any $\mu\in L^2(\pi)$,

$$\|\mu\tilde{P}^n-\pi\|_2\leq \|\mu-\pi\|_2(1-\frac{1}{8^3a^2})^n.$$

We can also consider the continuous-time Markov jump process with mean 1 holding times with operator $\bar{P}^t=\sum_{n=0}^{\infty}e^{-t}\frac{t^n}{n!}P^n$. By Remark 2 after Proposition 1.1, we have

$$\|\mu\bar{P}^t-\pi\|_2\leq \|\mu-\pi\|_2 e^{-t(1-\frac{1}{8^3a^2})}.$$

The following examples can all be treated in the same way.

(b) Let $p_x(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-x)^2}{2}}$ (the normal p.d.f. of $n(x,1)$) for $y\in S$ and 0 otherwise. As $p_x(y)=p_y(x)$, the chain is reversible w.r.t. the invariant distribution with density $q(x)=\frac{1}{2a}\mathbf{1}_S(x)$, uniform over S. So $Q=\frac{1}{2a}$. Note that $p_x(y)$ is a decreasing function of $|y-x|$. For any $b\in\mathbf{N}$, the essinf that defines $g(b)$ is reached when $|y-x|=\frac{D}{b}=\frac{2a}{b}$. This implies $g(b)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(\frac{2a}{b})^2}{2}}\frac{1}{2a}$. By Theorem 2.1,

$$k\geq \frac{2a}{b^2\sqrt{2\pi}}e^{-\frac{(\frac{2a}{b})^2}{2}}.$$

To get the best lower bound, we maximize the R.H.S. over b and we get $b=\lceil\sqrt{2}a\rceil$ or $\lfloor\sqrt{2}a\rfloor$. In particular, if $\sqrt{2}a$ is an integer,

$$k\geq \frac{1}{e\sqrt{2\pi}a}.$$

Example 3.2. Two dimensional convex case: $S=B(0,a)\subset\mathbf{R}^2$.

(a) Let $p_{(u_1,u_2)}(v_1,v_2)=\frac{1}{\pi}$ (the uniform p.d.f.) on $B(x,1)\cap S$ and 0 otherwise. Then the chain is reversible w.r.t. the invariant distribution with density $q(x)=\frac{1}{a^2\pi}\mathbf{1}_S(x)$ and so $Q=\frac{1}{a^2\pi}$. As in Example 3.1(a), take $b=\lceil 2a\rceil$. Since $\frac{D}{b}=\frac{2a}{b}\leq 1$, $g(b)=\frac{1}{\pi}\frac{1}{a^2\pi}$. By Theorem 2.1,

$$k\geq \frac{\frac{1}{a^2\pi^2}}{(\frac{1}{a^2\pi})^2 b^{2+1}}=\frac{a^2}{b^3}.$$

In particular, if $2a$ is an integer, $b = 2a$ and so $k \geq \frac{1}{8a}$.

(b) Let $p_{(u_1,u_2)}(v_1, v_2) = \frac{1}{2\pi} e^{-\frac{(v_1-u_1)^2+(v_2-u_2)^2}{2}}$ (the bivariate normal p.d.f. for two independent normal distributions $n(u_1, 1)$,$n(u_2, 1)$) for $(v_1, v_2) \in S$ and 0 otherwise. It is also clear that $p_{(u_1,u_2)}(v_1, v_2) = p_{(v_1,v_2)}(u_1, u_2)$. Then the invariant p.d.f. is $q(x) = \frac{1}{a^2\pi} \mathbf{1}_S(x)$, uniform over S, and so $Q = \frac{1}{a^2\pi}$. Similar to Example 3.1(b), $p_{(u_1,u_2)}(v_1, v_2)$ is constant on any circle on S with center (u_1, u_2) and the value is decreasing as the radius increases. So, for any $b \in \mathbf{N}$, the essinf that defines $g(b)$ is reached when $\|(v_1, v_2) - (u_1, u_2)\| = \frac{D}{b} = \frac{2a}{b}$. This implies $g(b) = \frac{1}{2\pi} e^{-\frac{(\frac{2a}{b})^2}{2}} \frac{1}{a^2\pi}$. By Theorem 2.1,

$$k \geq \frac{a^2}{2b^3} e^{-\frac{(\frac{2a}{b})^2}{2}}.$$

Again, maximizing the R.H.S. gives $b = \lceil \frac{2a}{\sqrt{3}} \rceil$ or $\lfloor \frac{2a}{\sqrt{3}} \rfloor$. In particular, if $\frac{2a}{\sqrt{3}}$ is an integer,

$$k \geq \frac{3\sqrt{3}}{16ae^{\frac{3}{2}}}.$$

Example 3.3. Family of right-angled triangle

Let $S_{r,\theta}$ be the right-angled triangle with hypotenuse r and an angle θ with transition kernel as in 3.2(a). Then the invariant p.d.f. is $q_{r,\theta}(x) = \frac{4}{r^2 \sin 2\theta} \mathbf{1}_{S_{r,\theta}}(x)$, uniform over $S_{r,\theta}$ and so $Q_{r,\theta} = \frac{4}{r^2 \sin 2\theta}$ The diameter of $S_{r,\theta}$ is r and take $b = b_{r,\theta} = \lceil r \rceil$. By Theorem 2.1,

$$k_{r,\theta} \geq \frac{r^2 \sin 2\theta}{4\pi \lceil r \rceil^3},$$

which goes to 0 as $\theta \to 0$. Intuitively, as the angle becomes sharper, the convergence rate (for the continuous time Markov chain or the discrete time Markov chain with transition kernel $\frac{1}{2}(I + P)$) becomes slower.

Example 3.4. Two dimensional non-convex case.

(a) Suppose $S = \{(x, y) \in (-a, a) \times \mathbf{R} : a - |x| < y < 2(a - |x|)\}$, an open inverted 'V' shape with vertices $(a, 0), (0, 2a), (-a, 0), (0, a)$, which is a non-convex set. Let $p_{(u_1,u_2)}(v_1, v_2) = \frac{1}{\pi}$ (the uniform p.d.f.) on $B(x, 1) \cap S$ and 0 otherwise. Then the invariant p.d.f. $q(x) = \frac{1}{a^2} \mathbf{1}_S(x)$ and so $Q = \frac{1}{a^2}$. We shall apply Corollary 2.3. So, we need to construct a bijection ϕ from S(which is open already) to a convex set, in which we already have the uniform paths. In this example, we define $\phi : S \to S'$ s.t. $\phi(x, y) = (x, 2(y - a + |x|))$ where $S' = \{(x, y) \in [-a, a] \times \mathbf{R} : 0 < y < 2(a - |x|)\}$, an open triangle with vertices $(a, 0), (0, 2a), (-a, 0)$. So, $\phi^{-1}(x, y) = (x, a - |x| + \frac{y}{2})$. It is obvious that both ϕ, ϕ^{-1} are continuously differentiable a.e.. To apply Corollary 2.3, recall that $\Phi : S \times S \to S' \times S'$ is defined by

$$\Phi((x_1, x_2), (y_1, y_2)) = (\phi(x_1, x_2), \phi(y_1, y_2)) = (x_1, 2(x_2 - a + |x_1|), y_1, 2(y_2 - a + |y_1|)).$$

Note that $(G^i)^{-1}$ defined in the proof of Corollary 2.3 has continuous partial derivatives except on a set M of λ(the Lebesgue measure on $\mathbf{R}^2 \times \mathbf{R}^2$) measure zero. To see this, let $Y = \{(0, y) : 1 \le y \le 2\}$, which is Lebesgue measure zero on $\mathbf{R}^2$). Then M is the union of $Y \times S, S \times Y, G^i(Y \times S), G^i(S \times Y)$, which is clearly of measure zero. It is easy to show that the Jacobian of Φ, $|J_\Phi((x_1, x_2), (y_1, y_2))| \equiv 4$, a constant. To apply Corollary 2.3, we need to find b s.t. $g(b) > 0$. Now, the essinf that defines $g(b)$ is over the set of pair of points which are adjacent to each other. Since $p_{(u_1,u_2)}(v_1, v_2) = 0$ when $\|(v_1, v_2) - (u_1, u_2)\| > 1$, we have to choose b big enough that the distance between any two such points is less than or equal to 1. According to the definition of ϕ, it is easy to see that

$$g(b) = \begin{cases} \frac{1}{a^2}\frac{1}{\pi}, & \text{if } b \ge 2\sqrt{2}a; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, by Corollary 2.3, taking $b = \lceil 2\sqrt{2}a \rceil$,

$$k \ge \frac{a^2}{\pi \lceil 2\sqrt{2}a \rceil^3}.$$

(b) Suppose $S = \{(x, y) \in (-a, a) \times \mathbf{R} : \frac{1}{2}\sqrt{a^2 - x^2} < y < \sqrt{a^2 - x^2}\}$, an open crescent shape, which is non-convex. Let $p_{(u_1,u_2)}(v_1, v_2) = \frac{1}{\pi}$ (the uniform p.d.f.) on $B(x, 1) \cap S$ and 0 otherwise. Then the invariant p.d.f. $q(x) = \frac{4}{a^2\pi}\mathbf{1}_S(x)$ and so $Q = \frac{4}{a^2\pi}$. Define bijection $\phi : S \to S'$ by $\phi(x, y) = (x, 2y - \sqrt{a^2 - x^2})$, where $S' = \{(x, y) \in (-a, a) \times \mathbf{R} : 0 < y < \sqrt{a^2 - x^2}\}$, an open semi-disc. So, $\phi(S) = S'$ is convex. It is also obvious that both ϕ, ϕ^{-1} are continuously differentiable. To apply Corollary 2.3, recall that $\Phi : S \times S \to S' \times S'$ is defined by

$$\Phi((x_1, x_2), (y_1, y_2)) = (\phi(x_1, x_2), \phi(y_1, y_2)) = (x_1, 2x_2 - \sqrt{a^2 - x_1^2}, y_1, 2y_2 - \sqrt{a^2 - y_1^2}).$$

It is easy to show that the Jacobian of Φ, $|J_\Phi((x_1, x_2), (y_1, y_2))| \equiv 4$, a constant. Similar to (a), we have to choose b s.t. $g(b) > 0$. According to the definition of ϕ, it is not hard to see that for any fixed b, the distance between any two adjacent points must be less than the distance $d(b)$ between $(a, 0)$ and $(a - \frac{2a}{b}, \sqrt{a^2 - (a - \frac{2a}{b})^2})$. Now, for $b \ge 4a^2$, or $d(b) = \frac{4a^2}{b} \le 1$,

$$g(b) = \frac{4}{a^2\pi}\frac{1}{\pi},$$

Hence, by Corollary 2.3, taking $b = \lceil 4a^2 \rceil$,

$$k \ge \frac{a^2}{4\lceil 4a^2 \rceil^3}.$$

This value is much smaller than that in part (a). Intuitively, since the angles at the vertices are 0, it is relatively hard to 'escape' from the vertices. This accounts for the slower convergence rate.

(c) Suppose $A > a > 0$ and $S = \{(x,y) \in (-A,A) \times \mathbf{R} : \sqrt{a^2 - x^2} < y < \sqrt{A^2 - x^2}, y > 0\}$ is the open 'C' shape and the transition kernel is as in (b). Then the invariant p.d.f. $q(x) = \frac{2}{(A^2-a^2)\pi}\mathbf{1}_S(x)$ and so $Q = \frac{2}{(A^2-a^2)\pi}$. Instead of using the type of bijections in (a) and (b), we consider the natural polar bijection ϕ. Define bijection $\phi : S \to (a,A) \times (0,\pi)$ by $\phi(x,y) = (\sqrt{x^2+y^2}, \theta)$, where $\theta \in (0,\pi)$ is given $\tan\theta = \frac{y}{x}$. Hence, $\phi^{-1}(r,\theta) = (r\cos\theta, r\sin\theta)$. Similar to (a) and (b), we can apply Corollary 2.3. By direct calculations, $|\det D\phi^{-1}(r,\theta)| = r$, and so $|\det D\phi(x,y)| = \frac{1}{\sqrt{x^2+y^2}}$. Hence,

$$|J_\Phi((x_1,x_2),(y_1,y_2))| = \frac{1}{\sqrt{x_1^2+x_2^2}\sqrt{y_1^2+y_2^2}},$$

and so $\frac{1}{A^2} < |J_\Phi| < \frac{1}{a^2}$. Now, we have to choose b s.t. $g(b) > 0$. For any fixed b, it is easy to see that the distance $d(b)$ between $(A,0)$ and $\frac{(b-1)A+a}{b}(\cos\frac{\pi}{b}, \sin\frac{\pi}{b})$ is the longest among the distances between adjacent points. Therefore,

$$g(b) = \begin{cases} \frac{2}{(A^2-a^2)\pi}\frac{1}{\pi}, & \text{if } d(b) \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

For fixed A and a, we can use numerical methods to find the smallest b s.t. $d(b) \le 1$. For such b, say b_0, by Corollary 2.3,

$$k \ge \frac{a^2(A^2-a^2)}{2A^2 b_0^3}.$$

For very large A and a, b_0 will also be large. So, the triangle with vertices $(A,0)$, $\frac{(b_0-1)A+a}{b_0}(\cos\frac{\pi}{b_0}, \sin\frac{\pi}{b_0})$ and $A(\cos\frac{\pi}{b_0}, \sin\frac{\pi}{b_0})$ is approximately right-angled at the last vertex and so

$$d(b_0) \approx \sqrt{(\frac{A-a}{b_0})^2 + 2A^2(1-\cos\frac{\pi}{b_0})} \approx \sqrt{(\frac{A-a}{b_0})^2 + (A\frac{\pi}{b_0})^2}.$$

Since b_0 is the smallest b s.t. $d(b) \le 1$, we have

$$b_0 \approx \sqrt{(A-a)^2 + (A\pi)^2}.$$

In particular, if $A = 2a$,

$$k \ge \frac{a^2(A^2-a^2)}{2A^2 b_0^3} \approx \frac{3}{8a(1+4\pi^2)^{\frac{3}{2}}}.$$

This value differs that in (a) by only a constant. Intuitively, S has no 'sharp' ends. In fact, all the angles at the vertices are $\frac{\pi}{2}$.

Remarks. Markov chains on state spaces of higher dimensions can be treated in a similar fashion.

REFERENCES

Y. Amit, On the rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions, J. Multivar. Anal. 38 (1991) 89-99.

Y. Amit and U. Grenander, Comparing sweep strategies for stochastic relaxation, J. Multivar. Anal. 37(2), (1991) 197-222.

J.R. Baxter and J.S. Rosenthal, Rates of convergence for everywhere-positive Markov chains, Statist. Probab. Lett. 22 (1995) 333-338.

E.D. Belsley, Rates of convergence of Markov chains related to association schemes, Ph.D. dissertation, Department of Mathematics, Harvard University (1993).

P. Billingsley, Probability and measure, 3^{nd} ed. (John Wiley & Sons, New York, 1995).

J.B. Conway, A course in functional analysis (Springer-Verlag, New York, 1985).

P. Diaconis, Group representations in probability and statistics, IMS Lecture Series 11, Institute of Mathematical Statistics, Hayward (California, 1988).

P. Diaconis and L. Saloff-Coste, Comparison theorems for reversible Markov chains, Ann. Appl. Probab. 3 (1993) 696-730.

P. Diaconis and D. Stroock, Geometric bounds for eigenvalues of Markov chains, Ann. Appl. Probab. 1 (1991) 36-61.

A. Frieze, R. Kannan and N.G. Polson, Sampling from log-concave distributions, Ann. Appl. Probab. 4 (1994) 812-837.

A. Frigessi, C.-R. Hwang, S.J. Sheu and P. Di Stefano, Convergence rates of the Gibbs sampler, the metropolis algorithm, and other single-site updating dynamics, J. Royal Statist. Soc. Ser. B 55 (1993) 205-220.

A.E. Gelgand and A.F.M. Smith, Sampling based approaches to calculating marginal densities, J. Amer. Statist. Assoc. 85 (1990) 398-409.

C.-R. Hwang, S.-Y. Hwang-Ma and S.-J. Sheu, Accelerating Gaussian diffusions, Ann. Appl. Probab. 3 (1993) 897-913.

S. Ingrassia, On the rate of convergence of the metropolis algorithm and Gibbs sampler by geometric bounds, Ann. Appl. Probab. 4 (1994) 347-389.

M. Jerrum and A. Sinclair, Approximating the permanent, SIAM J. Comput. 18 (1988) 1149-1178.

E. Kreyszig, Introductory functional analysis with applications (John Wiley & Sons, Toronto, 1978).

G.F. Lawler and A.D. Sokal, Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality, Trans. Amer. Math. Soc. 309 (1988) 557-580.

S.P. Meyn and R.L. Tweedie, Computable bounds for convergence rates of Markov chains, Ann. Appl. Prob. 4 (1994) 981-1011.

G.O. Roberts and J.S. Rosenthal, Shift-coupling and convergence rates of ergodic averages, Communications in Statistics – Stochastic Models 13, No. 1 (1997a) 147-165.

G.O. Roberts and J.S. Rosenthal, Geometric ergodicity and hybrid Markov chains, Electronic Communications in Probability 2 (1997b) 13-25.

J.S. Rosenthal, Rates of convergence for Gibbs sampler for variance components models, Ann. Statist. 23 (1995a) 740-761.

J.S. Rosenthal, Minorization conditions and convergence rates for Markov chain Monte Carlo, J. Amer. Statist. Assoc. 90 (1995b) 558-566.

J.S. Rosenthal, Analysis of Gibbs samples for a model related to James-Stein estimators, Statist. Comput. 6 (1996a), 269-275.

J.S. Rosenthal, Markov chain convergence: From finite to infinite, Stochastic processes and their Applications 62 (1996b) 55-72.

W. Rudin, Functional Analysis, 2^{nd} ed. (McGraw-Hill, New York, 1991).

A. Sinclair, Improved bounds for mixing rates of Markov chains and multicomodity flow, Combinatories, Prabab. Comput. 1 (1992) 351-370.

A. Sinclair and M. Jerrum, Approximate counting, uniform generation and rapidly mixing Markov chains, Inform. Comput. 82 (1989) 93-133.

A.F.M. Smith and G.O. Roberts, Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion), J. Roy. Statist. Soc. Ser. B 55 (1993) 3-24.

L. Tierney, Markov chains for exploring posterior distributions, Ann. Statist. 22 (1994) 1701-1762.