Generalization of Discrete-time Geometric Bounds to Convergence Rate of Markov Processes on $\mathbb{R}^n$

by

Wai Kong Yuen[1]

(March 14, 2001.)

Acknowledgment. This work is part of my doctoral research done under the direction of Jeffrey S. Rosenthal. I thank Soren Jarner for sharing some important related results from his working paper and Alfred B. Lehman for solving the graph theory problem in this paper.

Abstract.

Geometric rates of convergence for reversible discrete-time Markov chains are closely related to the spectral gap of the corresponding operator. Quantitative geometric bounds on the spectral gap have been developed using the Cheeger's inequality and some path arguments. We extend the discrete-time results to continuous-time reversible Markov processes. The limit path bounds and the limit Cheeger's bounds are introduced. Two quantitative examples of 1-dimensional diffusions are studied for the limit Cheeger's bounds and a n-dimensional diffusion is studied for the limit path bounds.

1 Introduction

Markov chain Monte Carlo algorithms (Gelfand and Smith [15], Smith and Roberts [36]) have had great impact on simulation theory in recent years. Therefore, quantitative geometric rates of convergence for Markov chains becomes an important topic. For previous results on finite state spaces, see Diaconis [9], Sinclair and Jerrum [35], Jerrum and Sinclair [20], Diaconis and Stroock [12] , Sinclair [34] and Diaconis and Saloff-Coste [10]. For results specifically for Markov chain Monte Carlo, see Hanlon [16], Frieze et al. [13],

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

Frigessi et al. [14], Ingrassia [18] and Belsley [4]. For results on discrete-time general state spaces, see Amit and Grenander [2], Amit [1], Hwang et al. [17], Lawler and Sokal [22], Meyn and Tweedie [23], Rosenthal [29, 30, 31, 32], Baxter and Rosenthal [3] and Roberts and Rosenthal [26, 27, 28].

In particular, there are results related to discrete approximations to a Langevin diffusion (See e.g. Roberts and Rosenthal [28]), which is a continuous time Markov process. Lawler and Sokal [22] took an idea from the literature of differential geometry and 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 for a discrete-time finite space Markov chain. Diaconis and Stroock (1991) and Sinclair (1992) used a similar geometric arguments with paths to bound the second largest eigenvalue of a self-adjoint discrete time Markov chain. Yuen [39] generalized the path argument to bound the spectral gap of a discrete time Markov chain on $\mathbb{R}^n$. In this paper, by considering some form of limits, we generalize some discrete-time results to bound the convergence rate of a reversible homogeneous continuous-time Markov process. We shall derive some quantitative bounds on the convergence rate of two types of diffusions.

There are other techniques to bound the convergence rates of Markov chains and Markov processes such as the coupling method (See e.g. Nummelin [24] for discrete time results) and the application of log-Sobolev inequalities (See Diaconis and Saloff-Coste [11] for an expository article) and direct computation of the spectrum (See e.g. Bhattacharya and Waymire [5]). For some classical results related to diffusions we shall study in the paper, see e.g. Taira [37], Weinberger [38] and Courant and Hilbert [8]. Perhaps those techniques are more powerful mathematically on specific examples, but the techniques in this paper may sometimes be more robust and easier to apply.

The remainder of this section introduces the definitions of $\lambda_0(L)$ and the Cheeger's constants, explains their relation to the convergence rate of a Markov process and some known results from discrete-time.

In Section 2, we will extend some discrete-time results to study the convergence rate of general Markov processes. We first introduce the limit path bounds and the limit Cheeger's bounds. Two examples of 1-dimensional diffusions will be studied in Section 3 using the limit Cheeger's bound. An n-dimensional diffusion will be studied in Section 4 using the limit path bound.

We conclude the paper with Section 5 and discuss some possible development and improvement in the future.

1.1 Basic Notations and the Cheeger's Inequality for Discrete-time Markov Chains

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. By a continuous-time Markov jump process, we mean the process which is piece-wise constant with mean 1 exponential holding times with transition rate kernel $P(x,dy)$. Then P induces a positivity-preserving linear contraction on $L^2(\pi)$ by

$$(Pf)(x) = \int f(y)P(x,dy). \tag{1}$$

P also acts to the left on measures, so that

$$\mu P(A) = \int P(x,A)\mu(dx). \tag{2}$$

Recall that a Markov chain is *reversible* if

$$\pi(dx)P(x,dy) = \pi(dy)P(y,dx). \tag{3}$$

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

$$\lambda_0(P) = \inf \mathrm{spec}(P\mid_{\mathbf{1}^\perp}), \tag{4}$$

$$\lambda_1(P) = \sup \mathrm{spec}(P\mid_{\mathbf{1}^\perp}). \tag{5}$$

We consider the Laplacian $L = I - P$. Since L is also real and self-adjoint, $\lambda_1(P) = 1 - \lambda_0(L)$, where

$$\lambda_0(L) = \inf \mathrm{spec}(L\mid_{\mathbf{1}^\perp}) = \inf\{\frac{(f,Lf)_{L^2(\pi)}}{\|f\|^2_{L^2(\pi)}} : f \in \mathbf{1}^\perp, f \neq 0\}. \tag{6}$$

In Section 1.3, we shall give a class of lower bounds for $\lambda_0(L)$. On the other hand, we define the Cheeger's k-constant as follows:

$$k \equiv \inf_{\substack{A\in\mathcal{F}\\ 0<\pi(A)<1}} k(A) \tag{7}$$

with

$$k(A) \equiv \frac{\int \chi_A(x)P(x,A^c)\pi(dx)}{\pi(A)\pi(A^c)}, \tag{8}$$

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 [22] proved an inequality for reversible Markov chains:

$$\frac{k^2}{8M} \leq 1 - \lambda_1(P) \leq k, \tag{9}$$

where

$$\pi\text{-esssup}\, P(x, \{x\}^c) \leq M. \tag{10}$$

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 inequality says that when k is small, the rate is low, which matches with our intuition.

We also define the Cheeger's h-constant as follows:

$$h \equiv \inf_{\substack{A \in \mathcal{F} \\ 0 < \pi(A) \leq \frac{1}{2}}} h(A) \tag{11}$$

with

$$h(A) \equiv \frac{\int \chi_A(x) P(x, A^c) \pi(dx)}{\pi(A)}, \tag{12}$$

Note that k and h can be bounded by each other by $h \leq k \leq 2h$.

Lawler and Sokal proved another version of the inequality:

$$\frac{h^2}{2M} \leq 1 - \lambda_1(P) \leq 2h. \tag{13}$$

Since the two inequalities (9) and (13) are very similar, we shall refer both to as the *Cheeger's inequality*.

1.2 Basic Notations and L^2 Convergence for Continuous-Time Markov Processes

In Section 2, we shall consider a positive recurrent homogeneous continuous-time Markov process with measurable state space $(S, \mathcal{F})$, transition probability distribution $P^t(x, dy)$ and invariant probability measure π. For each time

$t > 0$, as in (1) and (2) from the discrete-time case, P^t induces a positivity-preserving linear contraction on $L^2(\pi)$ by

$$(P^t f)(x) = \int f(y) P^t(x, dy) \tag{14}$$

and acts to the left on measures, so that

$$\mu P^t(A) = \int P^t(x, A)\mu(dx). \tag{15}$$

A Markov process is reversible if

$$\pi(dx)P^t(x, dy) = \pi(dy)P^t(y, dx) \tag{16}$$

for all t. In other words, for any fixed t, if we treat P^t as a discrete-time transition probability kernel, P^t is a discrete-time reversible Markov chain. In this case, we can define $\lambda_0(P^t)$, $\lambda_1(P^t)$, the Laplacian L^t, $\lambda_0(L^t)$, the Cheeger's k-constant k_t, the Cheeger's h-constant h_t and M_t as in the discrete case for each t.

The following fact will play a significant role in Section 3. A family $\mathcal{F}_0 \subset \mathcal{F}$ is said to be *dense* if for all $A \in \mathcal{F}$ and all $\epsilon > 0$, there exists $B \in \mathcal{F}_0$ such that $\pi(A \triangle B) < \epsilon$. Lawler and Sokal ([22] Lemma 4.1) proved the following.

Proposition 1.1. *Let $\mathcal{F}_0$ be a dense sub-family of $\mathcal{F}$. Then*

$$k = \inf_{\substack{A \in \mathcal{F} \\ 0 < \pi(A) < 1}} k(A) = \inf_{\substack{A \in \mathcal{F}_0 \\ 0 < \pi(A) < 1}} k(A) \tag{17}$$

and likewise for h.

In other words, in order to calculate the infimum, we only need to consider some smaller family of sets A. For example, in a metric space, we can take $\mathcal{F}_0$ as the family of closed sets. This can simplify our computations a lot.

In order to apply the results from the discrete-time, we need to relate the transition probability distribution of a Markov process to the transition probability kernel of a Markov chain. The simplest approach is to fix a small time τ and treat $P^\tau(x, dy)$ as the transitional kernel of a 'discrete-time' Markov chain. Therefore, we can apply the results from discrete-time Markov chains on a general Markov process.

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

$$\|\mu\|_2^2 = \begin{cases} \int_S |\frac{d\mu}{d\pi}|^2 d\pi, & \mu << \pi; \\ \infty, & \text{otherwise.} \end{cases} \tag{18}$$

With this definition, we can 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

$$\|P\ |_{\mathbf{1}^\perp}\ \|_2 = \sup\{\|Pf\|_2; f \in \mathbf{1}^\perp, \|f\|_2 = 1\} \tag{19}$$

where Pf is defined as in (1). We shall use a property of a bounded self-adjoint operator and a result from spectral theory (see e.g. Rudin [33], Kreyszig [21], Conway [7]) to prove the following well-known proposition.

Proposition 1.2. *Let P^t be a positive recurrent time-homogeneous Markov process which is reversible with respect to π, and $\mu \in L^2(\pi)$. Then for any $n \in \mathbb{N}$ and $t > 0$, we have*

$$\|\mu P^t - \pi\|_2 \leq \|\mu - \pi\|_2 \rho^n_{\frac{t}{n}}, \tag{20}$$

where

$$\rho_{\frac{t}{n}} = \max\{|\lambda_0(P^{\frac{t}{n}})|, |\lambda_1(P^{\frac{t}{n}})|\}. \tag{21}$$

Proof. For any $n \in \mathbb{N}$ and $t > 0$, we interpret the Markov process as a Markov chain with n steps. So the 'discrete' transition probability kernel for each step is $P^{\frac{t}{n}}$. It is easy to see that the condition of a reversible Markov process implies that $P^{\frac{t}{n}}$ is reversible with respect to π as a discrete-time Markov chain for any n, t. Define $\rho_{\frac{t}{n}} = \max\{|\lambda_0(P^{\frac{t}{n}})|, |\lambda_1(P^{\frac{t}{n}})|\}$.

For $\mu \in L^2(\pi)$, let $f = \frac{d\mu}{d\pi}$. Since P^t is reversible w.r.t. π, $P^{\frac{t}{n}}$ is a self-adjoint operator for any $n \in \mathbb{N}$ and $t > 0$. It is also easy to check $\frac{d(\mu P^{\frac{t}{n}})}{d\pi} = P^{\frac{t}{n}} f$. So, $P^{\frac{t}{n}}$ acting on $\mu \in L^2(\pi)$ is the same as $P^{\frac{t}{n}}$ acting on $f \in L^2(\pi)$, in the sense that $d\mu = f d\pi$ implies $d(\mu P^{\frac{t}{n}}) = (P^{\frac{t}{n}} f) d\pi$. It is also well known (see e.g. Baxter and Rosenthal [3]) that $\|P^{\frac{t}{n}}\ |_{\mathbf{1}^\perp}\ \|_2$ defined in (19) satisfies $\|P^{\frac{t}{n}}\ |_{\mathbf{1}^\perp}\ \|_2 \leq 1$. From a result of spectral theory for bounded self-adjoint operator $P^{\frac{t}{n}}$,

$$\|P^{\frac{t}{n}}\ |_{\mathbf{1}^\perp}\ \|_2 = \rho_{\frac{t}{n}}.$$

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

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

Q.E.D.

If we consider $P^{\frac{t}{n}}$ as a discrete-time Markov chain, $\rho_{\frac{t}{n}}$ is known as the *spectral radius* of the operator $P^{\frac{t}{n}}$. In discrete time, it is hard to bound $\lambda_0(P)$ in general. However, for homogeneous continuous-time Markov processes such as diffusions, it is easy to see that since P^t is the square of $P^{t/2}$, $0 \le \lambda_0(P^t)$ (which is $\le \lambda_1(P^t)$), which implies that

$$\rho_t = \max\{|\lambda_0(P^t)|, |\lambda_1(P^t)|\} = |\lambda_1(P^t)|.$$

More generally, we define a useful geometric constant

$$c = \limsup_{t\to 0} \frac{1}{t}\lambda_0(L^t). \tag{22}$$

The following well-known lemma is a continuous version of Proposition 2.1.

Lemma 1.3. *Under the condition in Proposition 1.2,*

$$\|\mu P^t - \pi\|_2 \le \|\mu - \pi\|_2 e^{-ct}. \tag{23}$$

In particular, if $c > 0$, μP^t is converging to π geometrically in the L^2 norm.

Proof. From our discussion, $\rho_t = |\lambda_1(P^t)|$. So,

$$\lim_{n\to\infty} \rho_{\frac{t}{n}}^n = \lim_{n\to\infty} |\lambda_1(P^{\frac{t}{n}})|^n = \lim_{n\to\infty} \lambda_1(P^{\frac{t}{n}})^n.$$

Now,

$$\begin{aligned}
\lim_{n\to\infty} \lambda_1(P^{\frac{t}{n}})^n &= \lim_{n\to\infty} \left(1 - \lambda_0(L^{\frac{t}{n}})\right)^n \\
&= \lim_{n\to\infty} \left(1 - \frac{n\lambda_0(L^{\frac{t}{n}})}{n}\right)^n \\
&\le \lim_{n\to\infty} \left(1 - \frac{\sup_{k\ge n}\{k\lambda_0(L^{\frac{t}{k}})\}}{n}\right)^n.
\end{aligned}$$

Note that in the supremum above, we take k to be any real number greater than n, not just natural number. By assumption,

$$\begin{aligned}
\lim_{n\to\infty} \sup_{k\ge n}\{k\lambda_0(L^{\frac{t}{k}})\} &= t \lim_{n\to\infty}\left(\sup_{k\ge n}\{\frac{k}{t}\lambda_0(L^{\frac{t}{k}})\}\right) \\
&= t \lim_{\tau\to 0} \sup_{s\ge\tau} \frac{1}{s}\lambda_0(L^s) \\
&= ct.
\end{aligned}$$

Hence, by Proposition 1.2 and taking $n \to \infty$, we have

$$\begin{aligned}\|\mu P^t - \pi\|_2 &\le \|\mu - \pi\|_2 \lim_{n\to\infty} \rho^n_{\frac{t}{n}} \\ &\le \|\mu - \pi\|_2 \lim_{n\to\infty} \left(1 - \frac{\sup_{k\ge n}\{k\lambda_0(L^{\frac{t}{k}})\}}{n}\right)^n \\ &= \|\mu - \pi\|_2 e^{-ct}.\end{aligned}$$

Q.E.D.

Remark. By this lemma, we can get a geometric bound for the corresponding Markov process if we know how to calculate c. In the Section 2, we shall derive some lower bounds for c by applying the path bounds we shall explain in Section 1.3 and the Cheeger's inequality.

1.3 Discrete-Time Results: Path Bounds for $\lambda_1(P)$ and Cheeger's Constant

In this section, we describe a class of lower bounds on $\lambda_0(L)$ defined in (6) and an upper bound on the Cheeger's k-constant for a reversible Markov chain (discrete or continuous-time) on $\mathbb{R}^n$ described in 1.2. In both cases, we obtain upper bounds for $\lambda_1(P)$ defined in (5). These bounds are called *path bounds* as each bound is associated with a set of paths. The proofs of the results in this section can be found in Yuen [39].

We assume that the transition kernel is of the form $P(x, dy) = \alpha(x)\delta_x(dy) + p_x(y)dy$ where δ_x is the unit point mass on x for any $x \in \mathbb{R}^n$. Suppose the invariant distribution π has density $q(y)$ w.r.t. Lebesgue measure.

The next requirement is the existence of a set of paths satisfying some regularity conditions. Let $S = \{x \in \mathbb{R}^n : q(x) > 0\}$. $\Gamma = \{\gamma_{xy}\}$ be a *set of paths* for a Markov chain P on $\mathbb{R}^n$ if for any $x, y \in S$, there exists $b_{xy} \in \mathbb{N}$ and a path $\gamma_{xy} : \{0, \dots, b_{xy}\} \to S$ s.t. $\gamma_{xy}(0) = x$ and $\gamma_{xy}(b_{xy}) = y$ and that $p_u(v) > 0$ for any edge (u, v) of any path where (u, v) is an *ith edge* of a path γ_{xy} iff $u = \gamma_{xy}(i-1)$ and $v = \gamma_{xy}(i)$ for some i. In this case, we write $(u, v) \in \gamma_{xy}$. Let E_i be the collection of all ith edges and $E = \cup E_i$. For any $u, v \in \mathbb{R}^n$, define $Q(u, v) = p_u(v)q(u)$ and for $\gamma_{xy} \in \Gamma$ and $\epsilon \in \mathbb{R}$, define $\|\gamma_{xy}\|_\epsilon = \sum_{(u,v)\in\gamma_{xy}} Q(u, v)^{-2\epsilon}$. Note that $Q(u, v) > 0$ for any edge (u, v).

To derive a class of lower bounds for $\lambda_0(L)$, a set of paths Γ has to satisfy the following technical conditions:

Definition 1.1. *Let $\Gamma = \{\gamma_{xy}\}$ be a set of paths for an irreducible Markov chain P on $\mathbb{R}^n$. Let $V = \{(x, y, i) : x, y \in S, 1 \le i \le b_{xy}\}$. Γ satisfies the first regularity condition if the following are true: (i) $T : V \to S^2 \times \mathbb{N}^2$ defined by $T(x, y, i) = (\gamma_{xy}(i-1), \gamma_{xy}(i), b_{xy}, i)$ is a 1-1 map onto $T(V)$.*

(ii) Fix $b, i \in \mathbb{N}$ s.t. $(u, v, b, i) \in T(V)$ for some $(u, v) \in S \times S$ and let $W_{bi} = \{(u, v) : (u, v, b, i) \in T(V)\} \subset E$. The 1-1 map $G_{bi} : W_{bi} \to S \times S$ defined by $G_{bi}(u, v) = (x, y)$ where $T(x, y, i) = (u, v, b, i)$ can be extended to a bijection of open sets and has continuous partial derivatives a.e. with respect to the Lebesgue measure on $\mathbb{R}^n \times \mathbb{R}^n$ for each b, i.

Assuming that the first regularity condition is satisfied, we need the following notation to state Theorem 1.4. Let $J_{bi}(u, v)$ be the Jacobian of the change of variable $(x, y) = G_{bi}(u, v)$, given by

$$J_{bi}(u, v) = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_n} & \frac{\partial x_1}{\partial v_1} & \cdots & \frac{\partial x_1}{\partial v_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_n} & \frac{\partial x_n}{\partial v_1} & \cdots & \frac{\partial x_n}{\partial v_n} \\ \frac{\partial y_1}{\partial u_1} & \cdots & \frac{\partial y_1}{\partial u_n} & \frac{\partial y_1}{\partial v_1} & \cdots & \frac{\partial y_1}{\partial v_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial y_n}{\partial u_1} & \cdots & \frac{\partial y_n}{\partial u_n} & \frac{\partial y_n}{\partial v_1} & \cdots & \frac{\partial y_n}{\partial v_n} \end{vmatrix} \tag{24}$$

where $u = (u_1, \ldots, u_n)$ and so on. We shall represent $J_{bi}(u, v)$ by $J_{xy}(u, v)$ as given $(u, v) \in E$, there is a 1-1 correspondence between (x, y) and (b, i).

The geometric constants that we shall use in the bounds are

$$\begin{aligned} k_\epsilon = k_\epsilon(\Gamma) &= \text{esssup}_{(u,v)\in E}\{Q(u, v)^{-(1-2\epsilon)} \\ &\qquad \times \sum_{\gamma_{xy} \ni (u,v)} \|\gamma_{xy}\|_\epsilon q(x) q(y) |J_{xy}(u, v)|\}, \end{aligned} \tag{25}$$

where esssup is the essential supremum with respect to Lebesgue measure λ on $\mathbb{R}^n \times \mathbb{R}^n$, i.e., $\text{esssup}\{f(x) : x \in X\} = \inf\{a : \lambda\{x : f(x) > a\} = 0\}$, and the sum is taken over all (x, y) s.t. $\gamma_{xy} \ni (u, v)$.

Theorem 1.4. *Let $\Gamma = \{\gamma_{xy}\}$ be a set of paths satisfying the first regularity condition (Definition 1.1) for an irreducible Markov chain P on $\mathbb{R}^n$. Then for any $\epsilon \in \mathbb{R}$,*

$$\lambda_0(L) \geq \frac{1}{k_\epsilon}. \tag{26}$$

Hence, $\lambda_1(P) \leq \inf_\epsilon\{1 - \frac{1}{k_\epsilon}\}$.

To get a bound for the Cheeger's k-constant k, a set of paths Γ has to satisfy a modified version of the first regularity condition:

Definition 1.2. *Let $\Gamma = \{\gamma_{xy}\}$ be a set of paths for an irreducible Markov chain P on $\mathbb{R}^n$. Γ satisfies the second regularity condition if the following* *is true: For any* $A \subset S$ *and* $0 < \pi(A) < 1$, *let* $V_A = A \times A^c$. *Then (i)* $T_A : V_A \to S^2 \times \mathbf{N}^2$ *defined by*

$$T_A(x,y) = (\gamma_{xy}(l-1), \gamma_{xy}(l), b_{xy}, l),$$

where

$$l = l_{xy}^A = \min\{i | \gamma_{xy}(i-1) \in A, \gamma_{xy}(i) \in A^c\}$$

is a 1-1 map onto $T_A(V_A)$. *(ii) Fix* $b, i \in \mathbb{N}$ *s.t.* $(u,v,b,i) \in T_A(V_A)$ *for some* (u,v) *and* A. *Let* $W_{bi} = \{(u,v) : (u,v,b,i) \in T_A(V_A)$ *for some* $A\} \subset E$. *The map* $G_{bi} : W_{bi} \to S \times S$ *defined by* $G_{bi}(u,v) = (x,y)$ *if* $T_A(x,y) = (u,v,b,i)$ *for some* A *is 1-1 and can be extended to a bijection of open sets and has continuous partial derivatives a.e. with respect to the Lebesgue measure on* $\mathbb{R}^n \times \mathbb{R}^n$ *for each* b, i.

The geometric constant that we shall use in the bound is

$$K = K(\Gamma) = \operatorname{esssup}_{(u,v)\in E}\{Q(u,v)^{-1} \sum_{\gamma_{xy} \ni (u,v)} q(x)q(y)|J_{xy}(u,v)|\}. \tag{27}$$

Theorem 1.5. *Let* $\Gamma = \{\gamma_{xy}\}$ *be a set of paths satisfying the second regularity condition (Definition 1.2) for an irreducible Markov chain* P *on* $\mathbb{R}^n$. *Then the Cheeger's* k*-constant satisfies*

$$k \geq \frac{1}{K}. \tag{28}$$

Hence, if M *satisfies (10),* $\lambda_1(P) \leq 1 - \frac{1}{8K^2M}$.

In general, there is no definite way to find the paths. However, if S is convex, we shall have at least one way to construct paths satisfying the regularity conditions (Definition 1.1 and 1.2). The geometric constants can be simplified using the following paths:

For any fixed $b \in \mathbb{N}$ and $x, y \in S$, we can choose the *uniform* path $\eta_{xy}^b : \{0, \ldots, b\} \to S$ defined by

$$\eta_{xy}^b(i) = \frac{(b-i)x + iy}{b},$$

the function onto the points on the line joining x, y so that adjacent points have equal spacings. In this case, $T_i(x,y) = (\eta_{xy}^b(i-1), \eta_{xy}^b(i))$ can be easily extended to a bijection on $\mathbb{R}^n \times \mathbb{R}^n$ and that $(T_i)^{-1}$ has continuous partial derivatives for each i. So, both the first and second regularity conditions are satisfied. It can be shown that $|J_i| = b^n$ (since for each j, x_j, y_j depend only on u_j, v_j). We shall apply these paths to the example in Section 4.

The following corollary simplifies the calculation of the geometric constants:

Corollary 1.6. *(a) Under the uniform paths* $\{\eta^b_{xy}\}$ *above, the constants can be simplified as*

$$
\begin{aligned}
k_\epsilon &= b^n \sum_{i=1}^{b} \text{esssup}_{(u,v)\in E_i}\{Q(u,v)^{-(1-2\epsilon)}\|\eta^b_{xy}\|_\epsilon q(x)q(y) : \\
&\qquad\qquad\qquad\qquad (x,y) = (T_i)^{-1}(u,v)\} \qquad (29)
\end{aligned}
$$

and

$$K = b^n \sum_{i=1}^{b} \text{esssup}_{(u,v)\in E_i}\{Q(u,v)^{-1}q(x)q(y) : (x,y) = (T_i)^{-1}(u,v)\}. \qquad (30)$$

(b) Under the linear paths $\{\gamma_{xy}\}$ *above, the constants can be simplified as*

$$k_\epsilon = \text{esssup}_{(u,v)\in E}\{Q(u,v)^{-(1-2\epsilon)} \sum_{\gamma_{xy}\ni(u,v)} \|\gamma_{xy}\|_\epsilon q(x)q(y)b^n_{xy}\}, \qquad (31)$$

and

$$K = \text{esssup}_{(u,v)\in E}\{Q(u,v)^{-1} \sum_{\gamma_{xy}\ni(u,v)} q(x)q(y)b^n_{xy}\}. \qquad (32)$$

2 From Discrete-Time to Continuous-Time

In view of Section 1, we consider positive recurrent time-homogeneous Markov processes such that P^t is reversible with respect to π. We assume that either (i) $\lambda_0(P^t) \geq 0$ for any t in a neighbourhood of 0, or (ii) $\liminf_{t\to 0} \lambda_0(P^t) \geq 0$ and $\liminf_{t\to 0} \lambda_1(P^t) > 0$.

We discuss two approaches in this section to give lower bounds on the constant c defined in (22). Our first approach is to treat P^t as a discrete-time Markov chain, apply the path bounds in Section 1.3 and then take $t \to 0$. The second approach is to use the Cheeger's constant directly without using path bounds and compute the limit as $t \to 0$.

2.1 Limit Path Bounds

In the first approach, we shall apply the path bounds from Section 1.3. We need to modify the path requirements. Bounds derived in this way are called *limit path bounds.* As usual, we assume that the transition probability distribution of the positive recurrent time-homogeneous Markov process is of the form $P^t(x, dy) = \alpha_t(x)\delta_x(dy) + p^t(x, y)dy$ where δ_x is the unit point mass on x for any $x \in \mathbb{R}^n$. Suppose the invariant distribution π has density $q(y)$ w.r.t. Lebesgue measure. As in the discrete case, we assume that

$$\pi\text{-esssup}\, P^t(x, \{x\}^c) \leq M_t. \tag{33}$$

In order to apply the path bounds in Section 1.3, we need to assume the existence of some paths. Since t is varying, instead of having just one set of paths, we need to have a *family* of sets of paths $\{\Gamma_t : 0 < t \leq t_0\}$ for some $t_0 > 0$ so that for each t, the set of paths Γ_t satisfies the first or second regularity condition with respect to the Markov chain P^t. In this case, we say the family $\{\Gamma_t : 0 < t \leq t_0\}$ satisfies the first or second regularity condition. Then we can define the corresponding constants $k_{\epsilon,t}$ (See Theorem 1.4) and K_t (See Theorem 1.5) respectively. In summary, we have the following theorem:

Theorem 2.1. *Let c be the geometric constant defined in (22) and M_t satisfies (33). (a) Let $\{\Gamma_t : 0 < t \leq t_0\}$ be a family of sets of paths so that for each t, the set of paths Γ_t satisfies the first regularity condition (Definition 1.1) for the Markov process described above. Then for any $\epsilon \in \mathbb{R}$,*

$$c \geq \frac{1}{\liminf_{t\to 0} k_{\epsilon,t} t}. \tag{34}$$

(b) If the second regularity condition (Definition 1.2) is satisfied instead, we have

$$c \geq \frac{1}{8 \liminf_{t\to 0} K_t^2 M_t t}. \tag{35}$$

Proof. (a) For any $0 < t \leq t_0$, P^t is a Markov chain and Γ_t is a set of paths satisfying the conditions in Theorem 1.4. So, given any $\epsilon \in \mathbb{R}$,

$$\frac{1}{t}\lambda_0(L^t) \geq \frac{1}{k_{\epsilon,t} t}.$$

Taking $\limsup_{t\to 0}$ on both sides, we get (34).

(b) Similar to (a). Apply Theorem 1.5. Q.E.D.

Remark. We can also generalize Corollary 1.6 in a similar fashion for specific types of paths. We shall not go into the detail here.

2.2 Limit Cheeger's Bounds

In the previous section, the limit path bounds are derived by applying the discrete-time path bounds before taking $t \to 0$. In particular, the path bound in Theorem 2.1(b) is derived from $1/K_t$ which is a path bound of the Cheeger's k-constant. So, one question arises, is it possible to take the limit for the Cheeger's constant directly? In this section, we define the limit Cheeger's bounds and discuss their properties. Then we define two important conditions so that the bounds can be calculated in a different way.

Define the constants k_0 and h_0 as follows.

$$k_0 = \limsup_{t \to 0} \frac{k_t}{\sqrt{M_t t}} \text{ and } h_0 = \limsup_{t \to 0} \frac{h_t}{\sqrt{M_t t}} \tag{36}$$

where k_t and h_t are the Cheeger's k-constant and h-constant for each time t respectively (See the first paragraph of Section 1.2 for formal definition) and M_t satisfies (33). The lower bounds on c obtained in the next theorem are called *limit Cheeger's bounds.*

Theorem 2.2. *Let P^t be a positive recurrent time-homogeneous Markov process which is reversible with respect to π. Then the geometric constant c defined in (22) satisfies*

$$c \geq \frac{k_0^2}{8} \text{ and } c \geq \frac{h_0^2}{2}. \tag{37}$$

Proof. For each fixed t, P^t is a reversible Markov chain with respect to π. So, from the Cheeger's inequality (See Section 1.1),

$$\lambda_0(L^t)(= 1 - \lambda_1(P^t)) \geq \frac{k_t^2}{8M_t} \text{ and } \lambda_0(L^t) \geq \frac{h_t^2}{2M_t}.$$

Dividing by t, we have

$$\frac{\lambda_0(L^t)}{t} \geq \frac{1}{8}(\frac{k_t}{\sqrt{M_t t}})^2 \text{ and } \lambda_0(L^t) \geq \frac{1}{2}(\frac{h_t}{\sqrt{M_t t}})^2.$$

Take $\limsup_{t \to 0}$ and we obtain (37). Q.E.D.

Theorem 2.2 gives two lower bounds for the constant c. However, it is not easy to calculate them in general. Recall the definition of k_t and h_t are

$$k_t = \inf_{\substack{A \in \mathcal{F} \\ 0 < \pi(A) < 1}} k_t(A) \text{ and } h_t = \inf_{\substack{A \in \mathcal{F} \\ 0 < \pi(A) \leq \frac{1}{2}}} h_t(A). \tag{38}$$

To calculate k_0 and h_0, we first take the infimum over A and then take the limit as $t \to 0$. So, one possibility is to interchange the limit and the infimum. However, it is not clear if the value remains unchanged. The next lemma gives a partial answer to the question.

Lemma 2.3. *Let $f(x,y)$ be a function with two variables. Then*

$$\limsup_{x\to x_0}(\inf_y f(x,y)) \le \inf_y(\limsup_{x\to x_0} f(x,y)). \tag{39}$$

Proof. Note that for any y_0,

$$\limsup_{x\to x_0}(\inf_y f(x,y)) \le \limsup_{x\to x_0} f(x,y_0).$$

Taking infimum over y_0 on both sides, we have

$$\limsup_{x\to x_0}(\inf_y f(x,y)) = \inf_{y_0}\limsup_{x\to x_0}(\inf_y f(x,y)) \le \inf_{y_0}(\limsup_{x\to x_0} f(x,y_0)).$$

Q.E.D.

Lemma 2.3 suggests that by interchanging the limit and the infimum, we have

$$k_0 \le \inf_A(\limsup_{t\to 0}\frac{k_t(A)}{\sqrt{M_t t}}) \text{ and } h_0 \le \inf_A(\limsup_{t\to 0}\frac{h_t(A)}{\sqrt{M_t t}}). \tag{40}$$

Unfortunately, since we are looking for lower bounds of c (and hence of k_0 and h_0), it is the reverse inequality we need. It seems to be a difficult technical problem. Nevertheless, we shall discuss some related results in Section 3 for a one-dimensional diffusion problem.

Since the equalities are crucial in our discussion in the next section, we say that the *k-condition* is satisfied if

$$k_0 = \inf_A(\limsup_{t\to 0}\frac{k_t(A)}{\sqrt{M_t t}}) \tag{41}$$

and the *h-condition* is satisfied if

$$h_0 = \inf_A(\limsup_{t\to 0}\frac{h_t(A)}{\sqrt{M_t t}}). \tag{42}$$

The above conditions imply that we can interchange the order of taking $t \to 0$ and taking infimum over sets A without changing the values.

3 One-Dimensional Diffusion in a Bounded Interval

In this section, we consider two different one-dimensional diffusions in a bounded interval. In Section 3.1, we shall give a rough lower bound of c defined in (22) for the modulo diffusion by comparing them to a simpler Markov chain which satisfies the h-condition (42). In Section 3.2, we shall work with the original process and explore the possibility of proving the h-condition directly to get an improved bound. In Section 3.3, we use the same technique as in Section 3.1 to study a diffusion with reflecting boundary.

3.1 Modulo Diffusion with No Drift and Constant Diffusion Coefficient

In this section, let P_1^t be the one-dimensional modulo diffusion on $[0,d)$ with drift and diffusion coefficients 0 and $\sigma^2 > 0$ respectively. We can interpret the diffusion as one on a circle. The transition probability distribution is given by

$$P_1^t(x,dy) = p_1^t(x,y)dy = \sum_{m=-\infty}^{\infty} (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(md+y-x)^2}{2\sigma^2 t}\right\} dy. \quad (43)$$

Clearly, $p_1^t(x,y) = p_1^t(y,x)$ for any $t>0$ and $x,y \in [0,d)$. So, P_1^t is reversible with respect to the uniform distribution on $[0,d)$ with $\pi(dx) = q(x)dx = \frac{1}{d}dx$. Moreover, since $P_1^t(x,\{x\}^c) = 1$ for any x, we can take $M_t = 1$. We also define the *modulo distance* between two *points* $x,y \in [0,d)$ by

$$D(x,y) = D(|y-x|) = \min\{|y-x|, d-|y-x|\}. \quad (44)$$

Here, we use the same notation D for convenience. By the periodicity of $p_1^t(x,y)$, p_1^t is a function of $D(x,y)$. We also observe that when $D(x,y)$ increases, $p_1^t(x,y)$ decreases. Therefore, we can write

$$P_1^t(x,dy) = \sum_{m=-\infty}^{\infty} r_m^t(D(x,y))dy \quad (45)$$

where

$$r_m^t(u) = (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(md+u)^2}{2\sigma^2 t}\right\}. \quad (46)$$

Similarly, we define the *modulo distance* between two *subsets* $A,B \subset [0,d)$ by

$$D(A,B) = \inf\{|y-x|, d-|y-x| : x \in A, y \in B\}. \quad (47)$$

In order to calculate $h_t(A)$, we need to estimate $\int_A \int_{A^c} p_1^t(x,y)dydx$, which is the probability 'flow' from set A to set A^c. There are two observations that give us an idea how to do the estimation. First, we observe that p_1^t is dominated by r_0^t. Second, for small enough t, it is almost impossible to 'flow' from A to B if $D(A,B) > 0$. By these two observations, we see that the only significant 'flow' is crossing a boundary of two adjacent sets using the distribution with r_0^t only. We formalize our ideas into the following lemma.

Lemma 3.1. *(a) For any $a < b < c$,*

$$\lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^b \int_b^c (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\} dydx = \frac{\sigma}{\sqrt{2\pi}}. \tag{48}$$

(b) For any $a < b_1 < b_2 < c$,

$$\lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^{b_1} \int_{b_2}^c (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\} dydx = 0. \tag{49}$$

(c) For any $a < b < c$, such that $c - a < d$,

$$\begin{aligned} \lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^b \int_b^c r_0^t(D(x,y))dydx &= \lim_{t\to 0} \frac{1}{\sqrt{t}} \int_b^c \int_a^b r_0^t(D(x,y))dydx \\ &= \frac{\sigma}{\sqrt{2\pi}}. \end{aligned} \tag{50}$$

(d) For any $a < b < c$, such that $c - a = d$,

$$\begin{aligned} \lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^b \int_b^c r_0^t(D(x,y))dydx &= \lim_{t\to 0} \frac{1}{\sqrt{t}} \int_b^c \int_a^b r_0^t(D(x,y))dydx \\ &= \frac{2\sigma}{\sqrt{2\pi}}. \end{aligned} \tag{51}$$

Proof. (a) By the change of variable $u = \frac{y-x}{\sigma\sqrt{t}}$ in the inner integral, we have

$$\int_a^b \int_b^c (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\} dydx = \int_a^b \int_{\frac{b-x}{\sigma\sqrt{t}}}^{\frac{c-x}{\sigma\sqrt{t}}} (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{u^2}{2}\right\} dudx.$$

Let $d_0 = \min\{b-a, c-b\} > 0$ and $a_0 = b - d_0\sigma\sqrt{t}$. For $x \in [a,b]$,

$$\frac{c-x}{\sigma\sqrt{t}} \geq \frac{c-b}{\sigma\sqrt{t}} \geq \frac{d_0}{\sigma\sqrt{t}}.$$

and so

$$\int_{a_0}^b \int_{\frac{b-x}{\sigma\sqrt{t}}}^{\frac{d_0}{\sigma\sqrt{t}}} \leq \int_a^b \int_{\frac{b-x}{\sigma\sqrt{t}}}^{\frac{c-x}{\sigma\sqrt{t}}} \leq \int_{-\infty}^b \int_{\frac{b-x}{\sigma\sqrt{t}}}^{\infty}.$$

Intuitively, the middle integral is over a quadrilateral with two vertical sides. We pick a_0 and d_0 such that the left integral is over the lower triangle of the quadrilateral with a horizontal edge. This explains the first inequality. The second inequality is trivial. Interchanging the integrals, we have

$$\begin{aligned}
\frac{1}{\sqrt{t}}\int_{a_0}^{b}\int_{\frac{b-x}{\sigma\sqrt{t}}}^{\frac{d_0}{\sigma\sqrt{t}}}\frac{1}{\sqrt{2\pi}}e^{\frac{-u^2}{2}}dudx &= \frac{1}{\sqrt{2\pi t}}\int_0^{\frac{d_0}{\sigma\sqrt{t}}}\int_{b-u\sigma\sqrt{t}}^{b}e^{\frac{-u^2}{2}}dxdu \\
&= \frac{1}{\sqrt{2\pi t}}\int_0^{\frac{d_0}{\sigma\sqrt{t}}}e^{\frac{-u^2}{2}}(b-(b-u\sigma\sqrt{t}))du \\
&= \frac{\sigma}{\sqrt{2\pi}}\int_0^{\frac{d_0}{\sigma\sqrt{t}}}ue^{\frac{-u^2}{2}}du \\
&\to \frac{\sigma}{\sqrt{2\pi}}\int_0^{\infty}ue^{\frac{-u^2}{2}}du = \frac{\sigma}{\sqrt{2\pi}}
\end{aligned}$$

as $t \to 0$. Similarly,

$$\frac{1}{\sqrt{t}}\int_{-\infty}^{b}\int_{\frac{b-x}{\sigma\sqrt{t}}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{\frac{-u^2}{2}}dudx = \frac{1}{\sqrt{2\pi t}}\int_0^{\infty}\int_{b-u\sigma\sqrt{t}}^{b}e^{\frac{-u^2}{2}}dudx = \frac{\sigma}{\sqrt{2\pi}}.$$

Hence, by squeeze theorem, we proved (a).

(b) Now,

$$\frac{1}{\sqrt{t}}\int_a^{b_1}\int_{b_2}^{c} = \frac{1}{\sqrt{t}}\int_a^{b_1}\int_{b_1}^{c} - \frac{1}{\sqrt{t}}\int_a^{b_1}\int_{b_1}^{b_2}$$

As $t \to 0$, the right hand side goes to $\frac{\sigma}{\sqrt{2\pi}} - \frac{\sigma}{\sqrt{2\pi}} = 0$

(c) The first equality is trivial by interchanging the order of integrations.

If $c - a \le d/2$,

$$\int_a^b\int_b^c r_0^t(D(x,y))dydx = \int_a^b\int_b^c(2\pi\sigma^2 t)^{-\frac{1}{2}}\exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\}dydx.$$

So, the result follows from (a).

If $c-a > d/2$, without loss of generality, assume that $b-a \le c-a$. Then

$$\begin{aligned}
\int_a^b\int_b^c r_0^t(D(x,y))dydx &= \int_a^b\int_b^{a+d/2}(2\pi\sigma^2 t)^{-\frac{1}{2}}\exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\}dydx \\
&\quad + \int_a^b\int_{a+d/2}^{c} r_0^t(D(x,y))dydx.
\end{aligned}$$

By definition of $D(x,y)$ and a simple change of variable,

$$\int_a^b\int_{a+d/2}^{c} r_0^t(D(x,y))dydx = \int_a^b\int_{a-d/2}^{c-d}(2\pi\sigma^2 t)^{-\frac{1}{2}}\exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\}dydx$$

in which the condition in (b) is satisfied: $a - d/2 < c - d < a < a + d/2$. Hence,

$$\lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^b \int_b^c r_0^t(D(x,y))dydx = \frac{\sigma}{\sqrt{2\pi}} + 0 = \frac{\sigma}{\sqrt{2\pi}}.$$

(d) Similar to (c), except that both terms satisfy the condition in (a) as $a - d/2 < c - d = a < a + d/2$. Thus

$$\lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^b \int_b^c r_0^t(D(x,y))dydx = \frac{\sigma}{\sqrt{2\pi}} + \frac{\sigma}{\sqrt{2\pi}} = \frac{2\sigma}{\sqrt{2\pi}}.$$

Q.E.D.

Recall that we need to prove that either the k-condition (41) or the h-condition (42) is satisfied to get a lower bound on c. One approach to prove the k-condition is to actually find a set $A_{\inf}$ such that it attains the infimum of $k_t(A)$ for any $t > 0$. If this is the case,

$$k_0 = \lim_{t\to 0} \frac{k_t}{\sqrt{t}} = \lim_{t\to 0} \frac{k_t(A_{\inf})}{\sqrt{t}} \geq \inf_A \lim_{t\to 0} \frac{k_t(A)}{\sqrt{t}}$$

which gives the k-condition. The same can be done for the h-condition. In fact, Jarner [19] proved that under certain conditions in a different setting, we need only consider the minimal set $[0, \frac{d}{2})$. We summarize some related results in the following theorem.

Theorem 3.2 (Jarner). *Consider a discrete-time random walk Markov chain on $[0, d)$ with transition probability density $p_x(y) = p(|x-y|)$ with the uniform stationary distribution. Assume that $p(u)$ is decreasing for $u > 0$. Then*

$$h = h([0, \frac{d}{2})). \tag{52}$$

In our problem, the transition probability density $p_1^t(x,y)$ is also of the form $p^t(|x-y|)$ but $p^t(u)$ is *not* decreasing over $[0, d)$. In fact, it is decreasing from 0 to $d/2$ and increasing from $d/2$ to d. Therefore, we cannot apply Theorem 3.2 on the Cheeger's h-constant directly. However, we observe that if we define

$$\bar{p}^t(x,y) \equiv (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\} dy, \tag{53}$$

we have

$$\begin{aligned} p_1^t(x,y) &= \sum_{m=-\infty}^{\infty} (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(md+y-x)^2}{2\sigma^2 t}\right\} dy \\ &> \bar{p}^t(x,y) \\ &> 0. \end{aligned}$$

We also let $\bar{h}_t$ be the Cheeger's h-constant with respect to the the random walk Markov chain Metropolis algorithm with transitional density $\bar{p}^t(x,y)$. We have the following theorem.

Theorem 3.3. *For the modulo diffusion defined in this section, the geometric constant c defined in (22) satisfies*

$$c \geq \frac{h_0^2}{2} \geq \frac{\sigma^2}{d^2\pi}. \tag{54}$$

Proof. Note that

$$c \geq \frac{h_0^2}{2} = \frac{1}{2}\limsup_{t\to 0}\inf_A\left(\frac{h_t(A)}{\sqrt{M_t t}}\right)^2.$$

By the definition of $p_1^t(x,y)$, we can take $M_t = 1$. Now, for any t and A,

$$\begin{aligned} h_t(A) &= \frac{\int_A\int_{A^c} p_1^t(x,y)dy\frac{1}{d}dx}{\pi(A)} \\ &\geq \frac{\int_A\int_{A^c} \bar{p}^t(x,y)dy\frac{1}{d}dx}{\pi(A)} \\ &= \bar{h}_t(A) \end{aligned}$$

Since $\bar{p}^t(x,y)$ is of the form $\bar{p}^t(|x-y|)$ and $\bar{p}^t(u)$ is decreasing for $u > 0$, by Theorem 3.2, $\bar{h}_t = \bar{h}_t(A_0)$ and so

$$\begin{aligned} c &\geq \frac{1}{2}\limsup_{t\to 0}\inf_A\left(\frac{\bar{h}_t(A)}{\sqrt{t}}\right)^2 \\ &= \frac{1}{2}\limsup_{t\to 0}\left(\frac{\bar{h}_t(A_0)}{\sqrt{t}}\right)^2 \end{aligned}$$

where $A_0 = [0, \frac{d}{2})$. It remains to show that the above limit exists and is equal to $\frac{2\sigma^2}{d^2\pi}$. In fact,

$$\begin{aligned} \lim_{t\to 0}\left(\frac{\bar{h}_t(A_0)}{\sqrt{t}}\right)^2 &= \left(\lim_{t\to 0}\frac{\int_0^{\frac{d}{2}}\int_{\frac{d}{2}}^d \bar{p}^t(x,y)dy\frac{1}{d}dx}{\pi([0,\frac{d}{2}))\sqrt{t}}\right)^2 \\ &= \frac{4}{d^2}\left(\lim_{t\to 0}\frac{1}{\sqrt{t}}\int_0^{\frac{d}{2}}\int_{\frac{d}{2}}^d \bar{p}^t(x,y)dydx\right)^2 \\ &= \frac{4}{d^2}\left(\frac{\sigma}{\sqrt{2\pi}}\right)^2 = \frac{2\sigma^2}{d^2\pi} \end{aligned}$$

by Lemma 3.1(a). Q.E.D.

Remark. The lower bound we derived here is probably not the best lower bound we can get using this method. We expect that $\frac{h_0^2}{2} > \frac{\sigma^2}{d^2\pi}$. Intuitively, the limiting process of the Metropolis algorithm should converge slower than our modulo diffusion process. In the Metropolis algorithm, when the current state is near one boundary, there is almost a half chance that the process will reject the move and stay where it is, whereas in the diffusion process, there is roughly a half chance that the process will jump to the side near the other boundary. Therefore, our process should mix faster and the lower bound should be larger. We shall work with the original transitional density in the next section to get an improved bound.

3.2 On Improving the Lower Bound

As discussed in the remark after Theorem 3.3, we want to prove the lower bounds of c without comparing the process with the corresponding Metropolis algorithm. Theorem 3.4 and Corollary 3.5 suggest that our intuition in the remark after Theorem 3.3 is well founded.

Theorem 3.4. *Consider the modulo diffusion described in Section 3.1. Let* $A = [a_1, b_1) \cup [a_2, b_2) \cup \cdots \cup [a_l, b_l)$, *where* $l \in \mathbb{N}$ *and* $0 = a_1 < b_1 < a_2 < b_2 < \cdots < a_l < b_l < d$. *Then,*

$$\lim_{t \to 0} \frac{k_t(A)}{\sqrt{t}} = \frac{2l\sigma}{d\sqrt{2\pi} \cdot \pi(A)\pi(A^c)} \tag{55}$$

and

$$\lim_{t \to 0} \frac{h_t(A)}{\sqrt{t}} = \frac{2l\sigma}{d\sqrt{2\pi} \cdot \pi(A)}. \tag{56}$$

Proof. We shall only prove (55). (56) follows from (55) as

$$\begin{aligned} \lim_{t \to 0} \frac{h_t(A)}{\sqrt{t}} &= \lim_{t \to 0} \frac{k_t(A)\pi(A^c)}{\sqrt{t}} \\ &= \frac{2l\sigma}{d\sqrt{2\pi} \cdot \pi(A)}. \end{aligned}$$

We now proceed to prove (55). Let $a_{l+1} = d$. Then

$$A^c = [b_1, a_2) \cup [b_2, a_3) \cup \cdots \cup [b_l, a_{l+1}).$$

For any $t > 0$,

$$\begin{aligned}
\int \chi_A(x) P_1^t(x, A^c)\pi(dx) &= \int_A \int_{A^c} p_1^t(x,y)dy\frac{1}{d}dx \\
&= \sum_{i=1}^{l}\sum_{j=1}^{l} \int_{a_i}^{b_i}\int_{b_j}^{a_{j+1}} p_1^t(x,y)dy\frac{1}{d}dx. \\
&= \sum_{i=1}^{l}\sum_{j=1}^{l} \int_{a_i}^{b_i}\int_{b_j}^{a_{j+1}} \sum_{m=-\infty}^{\infty} (2\pi\sigma^2 t)^{-\frac{1}{2}} \\
&\qquad \times \exp\left\{-\frac{(md+y-x)^2}{2\sigma^2 t}\right\} dy\frac{1}{d}dx.
\end{aligned}$$

Let $d_{\min} = \min_{i=1:l}\{b_i - a_i, a_{i+1} - b_i\}$. Consider the 3 cases of (i, j) :
Case 1. $i = j$.

$$\begin{aligned}
\int_{a_i}^{b_i}\int_{b_i}^{a_{i+1}} p_1^t(x,y)dy\frac{1}{d}dx &= \int_{a_i}^{b_i}\int_{b_i}^{a_{i+1}} \sum_{m=-\infty}^{\infty} r_m^t(D(x,y))dy\frac{1}{d}dx \\
&= \int_{a_i}^{b_i}\int_{b_i}^{a_{i+1}} r_0^t(D(x,y))dy\frac{1}{d}dx \\
&\qquad + \int_{a_i}^{b_i}\int_{b_i}^{a_{i+1}} \sum_{m\neq 0}^{\infty} r_m^t(D(x,y))dy\frac{1}{d}dx
\end{aligned}$$

Case 2. $i = j+1$ *or* $\{i = 1, j = l\}$.

First note that by periodicity, the case $\{i = 1, j = l\}$ can be included by letting $b_0 = -(d - b_l)$ and running i from 1 to l. In this case,

$$\begin{aligned}
\int_{a_i}^{b_i}\int_{b_{i-1}}^{a_i} p_1^t(x,y)dy\frac{1}{d}dx &= \int_{a_i}^{b_i}\int_{b_{i-1}}^{a_i} r_0^t(D(x,y))dy\frac{1}{d}dx \\
&\qquad + \int_{a_i}^{b_i}\int_{b_{i-1}}^{a_i} \sum_{m\neq 0}^{\infty} r_m^t(D(x,y))dy\frac{1}{d}dx
\end{aligned}$$

It is also very important to notice that if $l = 1$, $\{i = 1, j = l\}$ is already counted in Case 1. So, we have to single out such a special case.
Case 3. Otherwise. Since $D([a_i, b_i), [b_j, a_{j+1})) > d_{\min}$ and $r_m^t(u)$ decreases as u increases, we have

$$\int_{a_i}^{b_i}\int_{b_j}^{a_{j+1}} p_1^t(x,y)dy\frac{1}{d}dx \le \int_{a_i}^{b_i}\int_{b_j}^{a_{j+1}} \sum_{m=-\infty}^{\infty} r_m^t(d_{\min})dy\frac{1}{d}dx.$$

It is also very important to notice that if $l > 1$, the three cases partition the (i,j)'s into disjoint sets. However, if $l = 1$, $\{i = 1, j = l\}$ of Case 2 is already counted in Case 1. So, we shall revisit such situation after the case $l > 1$.

First we assume that $l > 1$. With the above three cases, we let

$$\Phi(t) = d \cdot \max\left\{ \sum_{m \neq 0}^{\infty} r_m^t(0), \sum_{m=-\infty}^{\infty} r_m^t(d_{\min}) \right\}.$$

Then

$$\begin{aligned} \int \chi_A(x) P_1^t(x, A^c) \pi(dx) &= \frac{1}{d} \sum_{i=1}^{l} \left(\int_{a_i}^{b_i} \int_{b_i}^{a_{i+1}} r_0^t(D(x,y)) dy dx \right. \\ &\qquad \left. + \int_{a_i}^{b_i} \int_{b_{i-1}}^{a_i} r_0^t(D(x,y)) dy dx \right) + R(t), \end{aligned}$$

where

$$\begin{aligned} 0 \leq R(t) &\leq \sum_{i=1}^{l} \sum_{j=1}^{l} \int_{a_i}^{b_i} \int_{b_j}^{a_{j+1}} \Phi(t) \cdot \frac{1}{d^2} dy dx \\ &= \pi(A)\pi(A^c)\Phi(t) \end{aligned}$$

Here,

$$0 \leq \frac{R(t)}{\pi(A)\pi(A^c)\sqrt{t}} \leq \frac{\Phi(t)}{\sqrt{t}}$$

and it is not hard to prove that

$$\lim_{t \to 0} \frac{\Phi(t)}{\sqrt{t}} = 0.$$

Therefore, we can neglect the term $R(t)$ when evaluating $\lim_{t\to 0} \frac{k_t(A)}{\sqrt{t}}$. Since $a_{j+1} - a_i < d$ and $b_i - b_{i-1} < d$, by Lemma 3.1(c),

$$\begin{aligned} \lim_{t \to 0} \frac{k_t(A)}{\sqrt{t}} &= \lim_{t \to 0} \frac{1}{d\pi(A)\pi(A^c)} \sum_{i=1}^{l} \frac{1}{\sqrt{t}} \left(\int_{a_i}^{b_i} \int_{b_i}^{a_{i+1}} r_0^t(D(x,y)) dy dx \right. \\ &\qquad \left. + \int_{a_i}^{b_i} \int_{b_{i-1}}^{a_i} r_0^t(D(x,y)) dy dx \right) \\ &= \lim_{t \to 0} \frac{1}{d\pi(A)\pi(A^c)} \sum_{i=1}^{l} \left(\frac{\sigma}{\sqrt{2\pi}} + \frac{\sigma}{\sqrt{2\pi}} \right) \\ &= \frac{2l\sigma}{d\sqrt{2\pi} \cdot \pi(A)\pi(A^c)} \end{aligned}$$

which is the required result.
Now we assume that $l = 1$. The situation is even simpler as

$$\int \chi_A(x) P_1^t(x, A^c)\pi(dx) = \int_0^{b_1} \int_{b_1}^{d} r_0^t(D(x,y))dydx + R(t)$$

where R(t) is again negligible. In this case, since $c - a = d - 0 = d$, we apply Lemma 3.1(d) and get

$$\begin{aligned}
\lim_{t\to 0} \frac{k_t(A)}{\sqrt{t}} &= \lim_{t\to 0} \frac{1}{d\pi(A)\pi(A^c)} \int_0^{b_1} \int_{b_1}^{d} r_0^t(D(x,y))dydx \\
&= \frac{1}{d\pi(A)\pi(A^c)} \frac{2\sigma}{\sqrt{2\pi}} \\
&= \frac{2l\sigma}{d\sqrt{2\pi} \cdot \pi(A)\pi(A^c)}
\end{aligned}$$

which is the same result. Q.E.D.

Remark. 1. The idea of the above proof is that we reduce the double integral over $A \times A^c$ into a sum of integrals over squares (product of intervals). It turns out that the integrals that do not vanish as $t \to 0$ are those over product of adjacent intervals in the periodic sense. Mathematically, $|md + D(x,y)| \geq \alpha > 0$ for some α for x, y from non-adjacent intervals. This concept will be used again in Section 3.3.

2. We singled out the case $l = 1$ in the proof due to the periodicity of the diffusion. Geometrically, even though there is only one pair of adjacent intervals, the 'flow' is counted twice as there are two crossings over the two contact boundaries of the intervals. This shall not be a factor in the reflecting boundaries example as we shall see in the proof of Theorem 3.12.

Corollary 3.5. *Suppose that either the k-condition (41) or the h-condition (42) is satisfied. Then the geometric constant c defined in (22) satisfies*

$$c \geq \frac{4\sigma^2}{d^2\pi}. \tag{57}$$

Proof. Since $S = [0, d)$, by Proposition 1.1, we need only consider sets of the form

$$A = [a_1, b_1) \cup [a_2, b_2) \cup \cdots \cup [a_l, b_l),$$

where $l \equiv l_A \in \mathbb{N}$ and $0 = a_1 < b_1 < a_2 < b_2 < \cdots < a_l < b_l < d$. If the

k-condition is satisfied, then by Theorem 2.2,

$$\begin{aligned} c \geq \frac{k_0^2}{8} &= \frac{1}{8} \inf_A (\limsup_{t \to 0} \frac{k_t(A)}{\sqrt{M_t t}})^2 \\ &= \frac{1}{8} \inf_A (\frac{2 l_A \sigma}{d\sqrt{2\pi} \cdot \pi(A)(1-\pi(A))})^2 \\ &= \frac{1}{8} (\frac{2 l_{A_0} \sigma}{d\sqrt{2\pi} \cdot \pi(A_0)(1-\pi(A_0))})^2 \end{aligned}$$

where $A_0 = [0, \frac{d}{2})$, since for any A such that $0 < \pi(A) < 1$, $l_A \geq 1 = l_{A_0}$ and

$$\pi(A)(1-\pi(A)) \leq \frac{1}{2}(1 - \frac{1}{2}) = \pi(A_0)(1-\pi(A_0)).$$

Hence,

$$c \geq \frac{1}{8}(\frac{2(1)\sigma}{d\sqrt{2\pi} \cdot \frac{1}{4}})^2 = \frac{4\sigma^2}{d^2\pi}.$$

Similarly, if the h-condition is satisfied, we can take the same A_0 and by Theorem 2.2 again,

$$c \geq \frac{h_0^2}{2} = \frac{1}{2}(\frac{2(1)\sigma}{d\sqrt{2\pi} \cdot \frac{1}{2}})^2 = \frac{4\sigma^2}{d^2\pi}.$$

Q.E.D.

Remarks. Corollary 3.5 is based on the condition that either the k-condition or the h-condition is satisfied. If this is indeed the case, we get a greater lower bound $\frac{4\sigma^2}{d^2\pi}$ than $\frac{\sigma^2}{d^2\pi}$ derived in Theorem 3.3. IN the remainder of this section, we shall investigate the validity of the conditions.

Recall that the $k(h)$-condition is satisfied if we can interchange the order of limit and the infimum. By Theorem 3.4, the limits already exist. One may suspect that the convergence is uniform over all A. In fact, if we could prove such a claim, the order can be interchanged and the condition is satisfied. However, it turns out that this is not the case. Consider the following counterexample.

Example 3.6.

Let $A_0 = [0, \frac{d}{2})$ and $A_n = [0, \frac{d}{4}) \cup [\frac{d}{4} + \frac{1}{n}, \frac{d}{2} + \frac{1}{n})$. So, $\pi(A_0) = \pi(A_n) = \frac{1}{2}$. If the convergence is uniform, given $\epsilon = \frac{\sigma}{d\sqrt{2\pi}}$, there exist $t > 0$, independent of n, such that

$$\left| \frac{k_t(A_0)}{\sqrt{t}} - \frac{8\sigma}{d\sqrt{2\pi}} \right| < \epsilon$$

and

$$\left|\frac{k_t(A_n)}{\sqrt{t}} - \frac{16\sigma}{d\sqrt{2\pi}}\right| < \epsilon.$$

So, for any n,

$$\frac{k_t(A_0)}{\sqrt{t}} < \frac{9\sigma}{d\sqrt{2\pi}} \text{ and } \frac{k_t(A_n)}{\sqrt{t}} > \frac{15\sigma}{d\sqrt{2\pi}}.$$

On the other hand, for this fixed t, it is clear that we can take n big enough so that $\frac{k_t(A_n)}{\sqrt{t}}$ is arbitrarily close to $\frac{k_t(A_0)}{\sqrt{t}}$, which is a contradiction.

Another approach to prove the k(or h, same below)-condition is to actually find a set A_{inf} such that it attains the infimum of $k_t(A)$ for any $t > 0$. If this is the case,

$$k_0 = \lim_{t\to 0}\frac{k_t}{\sqrt{t}} = \lim_{t\to 0}\frac{k_t(A_{\mathrm{inf}})}{\sqrt{t}} \geq \inf_A \lim_{t\to 0}\frac{k_t(A)}{\sqrt{t}}$$

which gives the k-condition. This approach is similar to the one by Jarner [19]. Recall that the conditions of Theorem 3.2 require $p_1^t(x,y)$ to be of the form $p_1^t(|x-y|)$ and that $p_1^t(u)$ is decreasing for $u > 0$. Note that we can interpret our $p_1^t(x,y)$ as $p_1^t(D(|x-y|))$. Since $D(|x-y|) \leq \frac{d}{2}$ and $p_1^t(u)$ is also decreasing over $[0, \frac{d}{2}]$, we may want to imitate Jarner's proof for the h condition. We start with the following lemma. It concerns how big a set A we should choose to minimize $h_t(A)$ if we restrict ourselves to intervals with 0 as one end point.

Lemma 3.7. *Consider a discrete-time random walk Markov chain on $[0,d)$ with transition probability density $p_x(y) = p(D(|x-y|))$ with the uniform stationary distribution. Then for any $a \leq b \leq \frac{d}{2}$,*

$$h([0,a)) \geq h([0,b)). \tag{58}$$

Proof. We shall prove this lemma with a more elementary and intuitive approach. Note that it suffices to prove the inequality for any a, b of the form $a = \frac{i}{2n+1}d$ and $b = \frac{j}{2n+1}d$ where $0 < i < j \leq n$ are integers. Indeed, for any $a \leq \frac{d}{2}$, we can choose sufficiently large i, n to approximate $h([0,a))$ by $h([0, \frac{i}{2n+1}))$ and so the inequality still holds for any $a \leq b \leq \frac{d}{2}$. In fact, it is clear that we only need to show that for any integers $0 < k \leq n-1$,

$$h([0, \frac{k}{2n+1}d)) \geq h([0, \frac{k+1}{2n+1}d)).$$

For $i = 1, \ldots, n$, define

$$p_i = \int_0^{\frac{1}{2n+1}d}\int_{\frac{i}{2n+1}d}^{\frac{i+1}{2n+1}d} p_x(y)\,dy\,dx.$$

Since $p_x(y) = p(D(|x-y|))$, it is easy to see that if $A_1 = [\frac{i_1}{2n+1}d, \frac{i_1+1}{2n+1}d)$, $A_2 = [\frac{i_2}{2n+1}d, \frac{i_2+1}{2n+1}d)$ for $0 \le i_1 \ne i_2 \le 2n$ and $D(A_1, A_2) = k-1$,

$$p_k = \int_{A_1}\int_{A_2} p_x(y)dydx = \int_{A_2}\int_{A_1} p_x(y)dydx.$$

So,

$$\begin{aligned}
\int_0^{\frac{k}{2n+1}d}\int_{\frac{k}{2n+1}d}^{d} p_x(y)dydx &= \sum_{i=1}^{k}\int_{\frac{i-1}{2n+1}d}^{\frac{i}{2n+1}d}\int_{[0,d)-[\frac{i-1}{2n+1}d,\frac{i}{2n+1}d)} p_x(y)dydx \\
&\quad -\sum_{i=1}^{k}\int_{\frac{i-1}{2n+1}d}^{\frac{i}{2n+1}d}\int_{[0,\frac{k}{2n+1}d)-[\frac{i-1}{2n+1}d,\frac{i}{2n+1}d)} p_x(y)dydx \\
&= \sum_{i=1}^{k} 2\sum_{j=1}^{n} p_j - 2\sum_{i=1}^{k}\int_{\frac{i-1}{2n+1}d}^{\frac{i}{2n+1}d}\int_{\frac{i}{2n+1}d}^{\frac{k}{2n+1}d} p_x(y)dydx \\
&= 2k\sum_{j=1}^{n} p_j - 2\sum_{i=1}^{k}\sum_{j=1}^{k-i} p_j \\
&= 2k\sum_{j=1}^{n} p_j - 2\sum_{j=1}^{k}\sum_{i=1}^{k-j} p_j \\
&= 2k\sum_{j=1}^{n} p_j - 2\sum_{j=1}^{k}(k-j)p_j.
\end{aligned}$$

Therefore,

$$\begin{aligned}
h([0, \frac{k}{2n+1}d)) &= \frac{\int_0^{\frac{k}{2n+1}d}\int_{\frac{k}{2n+1}d}^{d} p_x(y)dydx}{\pi([0,\frac{k}{2n+1}d))} \\
&= \frac{2n+1}{k}\left(2k\sum_{j=1}^{n} p_j - 2\sum_{j=1}^{k}(k-j)p_j\right) \\
&= 2(2n+1)\left(\sum_{j=1}^{n} p_j - \sum_{j=1}^{k} p_i + \frac{1}{k}\sum_{j=1}^{k} jp_j\right)
\end{aligned}$$

Hence,

$$
\begin{aligned}
& h([0, \frac{k}{2n+1}d)) - h([0, \frac{k+1}{2n+1}d)) \\
& = 2(2n+1)\left(p_{k+1} + \frac{1}{k}\sum_{j=1}^{k} jp_j - \frac{1}{k+1}\sum_{j=1}^{k+1} jp_j\right) \\
& = 2(2n+1)\left(\frac{1}{k}\sum_{j=1}^{k} jp_j - \frac{1}{k+1}\sum_{j=1}^{k} jp_j\right) \\
& > 0
\end{aligned}
$$

which is the required result. Q.E.D.

Remarks. 1. Note that we did not use the fact that $p(u)$ is decreasing over $[0, \frac{d}{2}]$ in our problem. It will be used in the Lemma 3.9.

2. From the definition of the Cheeger's h-constant, we only consider sets with measure less than or equal to $\frac{1}{2}$. So, if only single intervals are considered, the infimum is reached when $A = [0, \frac{d}{2})$. So, it remains to show that single intervals are the only important sets to consider in the infimum. Intuitively, if A is a set which has the same measure as an interval, it is always easier to escape from A than from the interval. It is also closely related to saying that the 'total distance' from A to A^c is smaller. The following theorem from graph theory is a simplified discrete version of this distance concept.

Theorem 3.8 (Lehman). *Consider an n-cycle with vertices $\{1, 2, \ldots, n\}$ and distance function $D(x, y)$ as usual. Let A be a subset of the vertices with $|A| = k$. Then for any natural number j,*

$$
\begin{aligned}
& |\{(x,y) \in A \times A^c : D(x,y) \leq j\}| \\
& \geq |\{(x,y) \in \{1, \ldots k\} \times \{k+1, \ldots n\} : D(x,y) \leq j\}|.
\end{aligned} \tag{59}
$$

Proof. See Appendix.

Remark. If $\{1, 2, \ldots, n\}$ is on a straight line and the distance function is $|x - y|$, the above inequality can be easily proved by induction, which is the key to the proof of Theorem 3.2.

Lemma 3.9. *Consider a discrete-time random walk Markov chain on $[0, d)$ with transition probability density $p_x(y) = p(D(x, y))$ with the uniform stationary distribution. Assume that $p(u)$ is decreasing over $[0, \frac{d}{2}]$. Let $A \subset [0, d)$ such that $\pi(A) = \frac{a}{d} \leq \frac{1}{2}$. Then $h(A) \geq h([0, a))$.*

This result is very similar to Theorem 6.16 of Jarner [19]. Therefore, we shall only explain the idea of the proof. Interested reader should refer to that paper.

In order to understand the proof, we need some extra notations and a well known result. Let ν_1 and ν_2 be two distributions on $(\mathbb{R}, \mathcal{B})$. We say that ν_1 *is stochastically smaller than* ν_2, written $\nu_1 \leq \nu_2$, if for all x,

$$\int_{-\infty}^{x} d\nu_1 \geq \int_{-\infty}^{x} d\nu_2. \tag{60}$$

We can further show that

$$\nu_1 \leq \nu_2 \text{ iff } \int f d\nu_1 \geq \int f d\nu_2 \tag{61}$$

for all decreasing measurable functions $f \geq 0$ (see Lemma 6.8 of Jarner [19]).

With this result and Theorem 3.8, we can give a sketch of the proof.

Proof of Lemma 3.9. Let A be a proper subset of an n-cycle with vertices $\{1, 2, \ldots, n\}$ and $|A| = k$. Assume that $X \perp Y$, $U \perp V$ where $X \sim U_A$, $Y \sim U_{A^c}$, $U \sim U_{\{1,\ldots k\}}$ and $V \sim U_{\{k+1,\ldots n\}}$. Let the distributions of $D(X, Y)$ and $D(U, V)$ be ν_1 and ν_2 respectively. Then Theorem 3.8 implies that for any j,

$$P(D(X, Y) \leq j) \geq P(D(U, V) \leq j)$$

and so

$$\nu_1 \leq \nu_2.$$

Now, we replace the n-cycle by $[0, d)$, A by any subset of measure $\frac{a}{d}$, $\{1, \ldots k\}$ by $[0, a)$ and $\{k+1, \ldots n\}$ by $[a, d)$. Then by a simple discrete approximation, we can prove that the inequality still holds. So, $\pi(A) = \pi([0, a)) = \frac{a}{d}$ and

$$\begin{aligned} h(A) &= \frac{\int_A \int_{A^c} p_x(y) dy \frac{1}{d} dx}{\pi(A)} \\ &= \frac{1}{a} \int_A \int_{A^c} p(D(x, y)) dy dx \\ &= \frac{1}{a} \int p(u) \nu_1(du) \end{aligned}$$

Since $\nu_1 \leq \nu_2$ and p is decreasing, by the equivalent condition (61),

$$\begin{aligned} h(A) &\geq \frac{1}{a} \int p(u) \nu_2(du) \\ &= \frac{\int_0^a \int_a^d p_x(y) dy \frac{1}{d} dx}{\pi([0, a))} \\ &= h([0, a)) \end{aligned}$$

Q.E.D.

Theorem 3.10. *Consider a discrete-time random walk Markov chain that satisfies the conditions in Lemma 3.9. Then*

$$h = h([0, \frac{d}{2})). \tag{62}$$

Therefore, for the diffusion defined in this section, the h-condition (42) is satisfied and so the geometric constant c defined in (22) satisfies

$$c \geq \frac{4\sigma^2}{d^2\pi}. \tag{63}$$

Proof. By Lemma 3.9,

$$h = \inf_A h(A) \geq \inf_{a \leq \frac{1}{2}} h([0, a)).$$

So, by Lemma 3.7,

$$\inf_{a \leq \frac{1}{2}} h([0, a)) \geq h([0, \frac{d}{2}))$$

and so we have the required result.

Now, consider the diffusion described in this section. For each fixed $t > 0$, $p^t(x, y)$ is of the form $p^t(D(|x - y|))$ and $p^t(u)$ is decreasing over $[0, d/2]$. So,

$$\begin{aligned} h_0 &= \limsup_{t \to 0} \frac{h_t}{\sqrt{M_t t}} \\ &= \limsup_{t \to 0} \frac{h_t([0, \frac{d}{2}))}{\sqrt{M_t t}} \\ &\geq \inf_A \limsup_{t \to 0} \frac{h_t(A)}{\sqrt{M_t t}} \end{aligned}$$

which is the h-condition. So, by Corollary 3.5,

$$c \geq \frac{4\sigma^2}{d^2\pi}.$$

Remarks. 1. We proved (63) by proving the h-condition. However, since $h_t = h_t([0, \frac{d}{2}))$, it is not necessary to compute $h_t(A)$ for all A. On the other hand, Theorem 3.4 shows that the infimum of $\lim(h_t(A)/\sqrt{t})$ is attained when $A = [0, \frac{d}{2})$. This strongly motivates us to consider the same A when we try to find the infimum of $h_t(A)$.

2. By directly computing the spectral gap associated to the reversible Markov semigroup P^t, it is known that the best possible c that satisfy equation (23) is $2\pi^2\sigma^2/d^2$ (See e.g. Bhattacharya and Waymire [5]).

3.3 Diffusion with Reflecting Boundaries, No Drift and Constant Diffusion Coefficient

In this section, let P_2^t be the one-dimensional diffusion on $[0, d]$ with reflecting boundaries $0, d$ and the drift and diffusion coefficients 0 and $\sigma^2 > 0$ respectively. So, the transition probability distribution is given by

$$\begin{aligned} P_2^t(x, dy) &= p_2^t(x, y)dy \\ &= \sum_{m=-\infty}^{\infty} (2\pi\sigma^2 t)^{-\frac{1}{2}} \Bigg(\exp\left\{ - \frac{(2md + y - x)^2}{2\sigma^2 t} \right\} \\ &\qquad\qquad + \exp\left\{ - \frac{(2md + y + x)^2}{2\sigma^2 t} \right\} \Bigg) dy \\ &= \sum_{m=-\infty}^{\infty} \left(r_{2m}^t(D(x, y)) + s_{2m}^t(x, y) \right) dy \end{aligned} \tag{64}$$

where

$$s_m^t(u) = (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{ - \frac{(md + x + y)^2}{2\sigma^2 t} \right\}. \tag{65}$$

Reversibility with respect to the uniform distribution is easy to see. As in the modulo diffusion,

$$p_2^t(x, y) > \bar{p}^t(x, y).$$

So, we have a similar theorem.

Theorem 3.11. *For the diffusion with reflecting boundaries defined in this section, the geometric constant c defined in (22) satisfies*

$$c \geq \frac{h_0^2}{2} \geq \frac{\sigma^2}{d^2\pi}. \tag{66}$$

Proof. Replace $p_1^t(x, y)$ by $p_2^t(x, y)$ in the proof of Theorem 3.3. Q.E.D.

One may think that as in Section 3.3, we could find an improved bound for the diffusion. However, the following results show that the bound in Theorem 3.11 is as good as the one derived directly from the process.

Theorem 3.12. *Consider the diffusion with reflecting boundaries described in this section. Let $A = [a_1, b_1) \cup [a_2, b_2) \cup \cdots \cup [a_l, b_l)$, where $l \in \mathbb{N}$ and $0 = a_1 < b_1 < a_2 < b_2 < \cdots < a_l < b_l < d$. Then,*

$$\lim_{t \to 0} \frac{k_t(A)}{\sqrt{t}} = \frac{(2l - 1)\sigma}{d\sqrt{2\pi} \cdot \pi(A)\pi(A^c)} \tag{67}$$

and

$$\lim_{t\to 0}\frac{h_t(A)}{\sqrt{t}} = \frac{(2l-1)\sigma}{d\sqrt{2\pi}\cdot\pi(A)}. \tag{68}$$

Proof. We shall not give a detail proof as it is very similar to Theorem 3.4. There are two points that we have to clarify in order to get $(2l-1)$ instead of $2l$ in the corresponding equations from Theorem 3.4. Recall that

$$p_2^t(x,y) = \sum_{m=-\infty}^{\infty}\left(r_{2m}^t(D(x,y)) + s_{2m}^t(x,y)\right).$$

From the remark after Theorem 3.4, we need to find the number of integrals in the sum that do not vanish as $t\to 0$.

First, we notice that the first term in the sum is very similar to that $p_1^t(x,y)$ except that we only run through even integers. The only difference is in Case 2, in which the case $\{i=1, j=l\}$ has to be excluded. Recall that in the proof, the associated integral does not vanish due to periodicity. However, in the reflecting boundary case, the intervals $[a_1,b_1)$ and $[b_l,d)$ are no longer adjacent to each other. Hence, we have one less integral which does not vanish.

Second, it is easy to see that in the second term, for any $x\in A$ and $y\in A^c$, $|2md+x+y|\geq \min\{a_1, d-b_l\}>0$. By the same remark again, the sum vanishes. Hence, we have the required result. Q.E.D.

Corollary 3.13. *The second inequality in Theorem 3.11 is actually an equality, i.e.,*

$$c \geq \frac{h_0^2}{2} = \frac{\sigma^2}{d^2\pi}. \tag{69}$$

Proof. It suffices to show the other inequality

$$h_0 \leq \frac{\sqrt{2}\sigma}{d\sqrt{\pi}}.$$

By Lemma 2.3 and Theorem 3.12 we have

$$\begin{aligned} h_0 &\leq \inf_A(\limsup_{t\to 0}\frac{h_t(A)}{\sqrt{M_t t}}) \\ &= \inf_A\frac{(2l-1)\sigma}{d\sqrt{2\pi}\cdot\pi(A)} \end{aligned}$$

where the infimum is taken over all A with $\pi(A) \leq \frac{1}{2}$ of the form described in Theorem 3.12. Take $A_0 = [0, \frac{d}{2})$. Then for any A,

$$l_{A_0} = 1 \leq l_A \text{ and } \pi(A_0) = \frac{1}{2} \geq \pi(A).$$

Hence,

$$h_0 \leq \frac{(2(1)-1)\sigma}{d\sqrt{2\pi}\cdot\frac{1}{2}} = \frac{\sqrt{2}\sigma}{d\sqrt{\pi}}.$$

Q.E.D.

4 n-Dimensional Diffusion in a Bounded Domain with Reflecting Boundaries

In Section 2, we discussed two types of bounds for the constant c. We first looked at the limit path bounds and then the limit Cheeger's bound. The advantage of the former approach is that it is able to handle higher dimensional problem as we shall see. The disadvantage is that by introducing the paths, we end up finding a suboptimal constant for the limit Cheeger's bound in Theorem 2.1(b). On the other hand, the advantage of the latter approach is that we are calculating the tight limit of the Cheeger's constant. The disadvantage is that the calculation can be very tedious and relies on results like Theorem 3.2 and Lemma 3.9. It is not clear how to generalize the results to higher dimension.

In this section, we shall apply Theorem 2.1 to derive the limit path bound for a n-dimensional diffusion in a bounded domain with reflecting boundaries. We shall then compare it with the limit Cheeger's bound obtained in Section 3.3 by setting $n = 1$.

Let $P^t_{(n)}$ be the n-dimensional diffusion on $S = [0, d]^n \subset \mathbb{R}^n$ with drift vector $\mathbf{0}$ and diffusion matrix $\sigma^2 I$ where I is the identity matrix. The diffusion is again reversible with respect to the uniform distribution on $[0, d]^n$ with $\pi(dx) = q(x)dx = \frac{1}{d^n}dx$. We shall not write down the explicit transition probability distribution. A more important fact is the following inequality:

$$P^t_{(n)}(x, dy) \geq (2\pi\sigma^2 t)^{-\frac{n}{2}} \exp\left\{-\frac{||y-x||^2}{2\sigma^2 t}\right\} dy \tag{70}$$

where

$$||y-x||^2 = \sum_{i=1}^{n} (y_i - x_i)^2. \tag{71}$$

For each $t > 0$, we have to find a set of paths Γ_t that satisfies the path conditions. We apply the uniform paths $\eta_{xy}^{b_t}$ where the number of steps b_t in each path is to be determined later. By going through a routine calculation using Corollary 1.6, we have

$$k_{\epsilon,t} \leq (2\pi\sigma^2 t)^{\frac{n}{2}} \exp\left\{\frac{nd^2}{2\sigma^2 t b_t^2}\right\} b_t^{n+2} d^{-n} \tag{72}$$

as $||y-x|| \leq \frac{\sqrt{n}d}{b_t}$ for any two adjacent points in a path. For each t, we minimize the expression over b_t and get

$$b_t = \frac{d}{\sigma}\sqrt{\frac{n}{(n+2)t}}.$$

We shall ignore the fact that b_t must be a natural number because it does not affect the result as $t \to 0$. Therefore,

$$\begin{aligned} k_{\epsilon,t} &\leq (2\pi\sigma^2 t)^{\frac{n}{2}} \exp\left\{\frac{n+2}{2}\right\} \left(\frac{d}{\sigma}\sqrt{\frac{n}{(n+2)t}}\right)^{n+2} d^{-n} \\ &= \frac{d^2}{t\sigma^2}(2\pi)^{\frac{n}{2}} e^{\frac{n+2}{2}} \left(\frac{n}{(n+2)}\right)^{\frac{n+2}{2}} \end{aligned}$$

Hence, applying Theorem 2.1(a), we have the following result:

Theorem 4.1. *For the n-dimensional diffusion described in this section, the geometric constant c defined in (22) satisfies*

$$c \geq \frac{\sigma^2}{d^2}(2\pi)^{-\frac{n}{2}} e^{-\frac{n+2}{2}} \left(\frac{n}{(n+2)}\right)^{-\frac{n+2}{2}}. \tag{73}$$

In particular, when $n = 1$, we have the following corollary:

Corollary 4.2. *For the one-dimensional diffusion with reflecting boundaries described in Section 3.3, the geometric constant c defined in (22) satisfies*

$$c \geq \frac{3^{\frac{3}{2}}\sigma^2}{d^2\sqrt{2\pi}e^{\frac{3}{2}}}. \tag{74}$$

Comparing this bound with the limit Cheeger's bound in (69), we find that $\frac{3^{\frac{3}{2}}\sigma^2}{d^2\sqrt{2\pi}e^{\frac{3}{2}}} > \frac{\sigma^2}{d^2\pi}$. In other words, the limit path bound is better than the limit Cheeger's bound. From the examples in Yuen [39], it seems that the path bound k_ϵ defined in (25) is always better than the Cheeger's path bound K defined in (27). We did not, however, compare k_ϵ with the *exact* Cheeger's constant h or k. It shows that in this example, after taking $t \to 0$, the exact bound obtained from the Cheeger's argument is worse than that obtained from a simpler path argument.

5 Conclusion and Future Research

We conclude our paper with some overall remarks on the main result and some possible future research topics.

Our main result is the introduction of the limit path bounds and the limit Cheeger's bounds. Both are techniques generalized from the discrete time to compute geometric bounds on the convergence rates of some homogeneous continuous-time Markov processes. The first technique is to extend the definition of paths. In Section 4, we considered the uniform paths and found that the optimal b_t is a continuous function of time t if we ignore the fact that b_t must be natural numbers. In fact, in the example, it is not even important that b_t must be natural numbers when we take $t \to 0$. Therefore, it should be interesting to generalize the family of sets of paths to a more continuous setting.

The second technique is to compute the Cheeger's bound explicitly and then take $t \to 0$. We managed only to calculate some one-dimensional examples. The problem lies on the fact that it is very difficult to prove the $k(h)$-condition. Our approach is to find a particular set $A_{\inf}$ so that the infimum is attained. In the one-dimensional bounded state space case, it involves some graph theory problems. One may want to generalize the problem to higher dimensions.

There are other applications of path bounds for the comparison of discrete-time Markov chains (See e.g. Quastel [25], Diaconis and Saloff-Coste [10] and Yuen [39]). With the techniques developed in this paper, it should be interesting to generalize the results to the comparison between Markov processes or even between a Markov chain and a Markov process. Another possible research direction is to generalize the kind of comparison we used in Section 3 when we compared the diffusion with a Metropolis algorithm to calculate the limit Cheeger's bound.

The modulo diffusion studied in Section 3.1 and Section 3.2 induces more possible research topics. First, we can think of the diffusion as a diffusion on a circle. In fact, we may even want to consider a diffusion on a compact Riemannian manifold. Second, since the interchanging of the limit and the infimum in the $k(h)$-condition is so technical, we can perhaps imitate the original Cheeger's argument (see [6]) to get around the technicality.

Appendix

The following theorem is proved by Alfred B. Lehman, Department of Computer Science, University of Toronto.

Theorem 3.8 (Lehman). *Consider an n-cycle with vertices $\{1, 2, \ldots, n\}$ and distance function $D(x, y)$ as usual. Let A be a subset of the vertices* *with $|A| = k$. Then for any natural number j,*

$$|\{(x,y) \in A \times A^c : D(x,y) \leq j\}|$$
$$\geq |\{(x,y) \in \{1,...k\} \times \{k+1,...n\} : D(x,y) \leq j\}|.$$

Proof. Fix a natural number j. Let $B = A^c$. In what follows, we say that two vertices $x \in A$ and $y \in B$ form an A, B *link* if the distance $D(x,y)$ between them is less than or equal to j. We also label A, B such that $k \leq n/2$. With this notation, it suffices to show that the sequential arrangement yields a minimum number of A, B links. We prove the assertion in three cases:
Case 1. $j \geq (n-1)/2$.

All A, B vertex pairs are within distance j and hence there are $k(n-k)$ links regardless of the A, B arrangement. For the rest of the proof assume that $j \leq (n-1)/2$.

In order to proceed, we make the following observation. If we remove an element of A or B, the number of links between A and B cannot increase under the j assumption. To see this, suppose an element of A is removed. A set of new (vertex) disjoint links can be created - all of distance j. The B elements involved were all linked previously to the removed element of A. Thus the number of created links cannot exceed the number removed. Symmetrically this also holds for the removal of an element of B.
Case 2. $j \leq k($ *and so* $n-k)$.

For the sequential arrangement there are $j(j+1)$ links. Remove elements of A until A is of size j and elements of B until B is of size $j+1$. From the $j = (n-1)/2$ instance of Case 1 there are now $j(j+1)$ links regardless of arrangement. Thus the sequential arrangement gives a minimum.
Case 3. $k \leq j$.

For the sequential arrangement there are $k(2j-k+1)$ links. Remove elements of B until B is of size $2j-k+1$ and the minimality again follows from the $j = (n-1)/2$ instance of Case 1. Again the sequential arrangement gives a minimum. Q.E.D.

References

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

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

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

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

[5] R.N. Bhattacharya and E.C. Waymire, Stochastic processes with applications (John Wiley & Sons, New York, 1990).

[6] J. Cheeger, A lower bound for the lowest eigenvalue of the Laplacian, Problems in Analysis: A Symposium in Honor of S. Bochner (R. C. Gunning, ed.), Princeton Univ. Press, Princeton, N.J., (1970) 195-199

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

[8] R. Courant and D. Hilbert, Methods of mathematical physics, Vol. I, II (New York Interscience Publishers 1953-62).

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

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

[11] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains, Ann. Appl. Probab. 3 (1996) 695-750.

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

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

[14] 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.

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

[16] P. Hanlon, A Markov chain on the symmetric group and Jack symmetric functions. Discrete Math. 99 (1992) 123-140.

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

[18] S. Ingrassia, On the rate of convergence of the metropolis algorithm and Gibbs sampler by geometric bounds, Ann. Appl. Probab. 2 (1994) 347-389.

[19] S.F. Jarner, Conductance of Metropolis Algorithm, working paper.

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

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

[22] 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.

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

[24] E. Nummelin, General irreducible Markov chains and nonnegative operators (Cambridge University Press, Cambridge 1984).

[25] J. Quastel, Diffusion of colour in the simple exclusion process, Comm. Pure Appl. Math. 45 (1992) 623-679.

[26] 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.

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

[28] G.O. Roberts and J.S. Rosenthal, Optimal scaling of discrete approximations to Langevin diffusions. J. R. Stat. Soc. Ser. B Stat. Methodol. 60 , no. 1 (1998), 255–268.

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

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

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

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

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

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

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

[36] 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.

[37] K. Taira, Diffusion processes and partial differential equations (Academic Press 1988).

[38] H.F. Weinberger, Variational methods for eigenvalue approximation (SIAM Philadelphia 1974).

[39] W.K. Yuen, Applications of geometric bounds to the convergence rate of Markov chains on $\mathbb{R}^n$, Stochastic processes and their Applications 87 (2000) 1-23.