Application of Geometric Bounds to Convergence Rates of Markov Chains and Markov Processes on $\mathbb{R}^n$

by

Wai Kong Yuen

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto

Abstract

Application of Geometric Bounds to Convergence Rate of Markov Chains and Markov Processes on $\mathbb{R}^n$

Ph.D. 2001

Wai Kong Yuen

Department of Mathematics

University of Toronto

Quantitative geometric rates of convergence for reversible Markov chains are closely related to the spectral gap of the corresponding operator, which is hard to calculate for general state spaces. This thesis describes a geometric argument to give different types of bounds for spectral gaps of Markov chains on bounded subsets of $\mathbb{R}^n$ and to compare the rates of convergence of different Markov chains. We also extend the discrete-time results to homogeneous 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.

Acknowledgments

It is a pleasure to thank my supervisor, Professor Jeffrey S. Rosenthal for all his guidance, encouragement and patience throughout my doctoral program.

I thank Professor Peter Rosenthal for helpful discussions about the operator theory issues. I would also like to extend my appreciation to Soren Jarner of Lancaster University for sharing some important related results from his working paper.

I am especially grateful to my wife, Yee Wan Ng. Her love, support and prayers brought me through the hardest time to the completion of my Ph.D.

I am indebted to my parents and my brother for their continuous support and encouragement.

I also wish to thank the staff of the Department of Mathematics at the University of Toronto, in particular Ms Ida Bulat, who have assisted me ever since I came to Toronto.

Many thanks go to all of my friends and my church and for many good times we shared together during the course of my study.

Finally, I would like to thank the Department of Mathematics and the University of Toronto for their financial support.

Contents

Chapter 1

Introduction

1.1 Motivation and Structure of the Thesis

As Markov chain Monte Carlo algorithms (Gelfand and Smith [17], Smith and Roberts [40]) are more widely used, quantitative geometric rates of convergence for Markov chains becomes an important topic. Diaconis [10], Sinclair and Jerrum [39], Jerrum and Sinclair [22], Diaconis and Stroock [13] , Sinclair [38] and Diaconis and Saloff-Coste [11] proved general results on *finite* state spaces. Hanlon [18], Frieze et al. [15], Frigessi et al. [16], Ingrassia [20] and Belsley [5] proved results specifically for Markov chain Monte Carlo. On *general* state spaces, not many results have been found yet. For partial results, see Amit and Grenander [2], Amit [1], Hwang et al. [19], Lawler and Sokal [24], Meyn and Tweedie [25], Rosenthal [33, 34, 35, 36], Baxter and Rosenthal [4] and Roberts and Rosenthal [30, 31].

In particular, Diaconis and Stroock [13] and Sinclair [38] used geometric arguments with paths to bound the second largest eigenvalue of a self-adjoint discrete-time Markov chain. On the other hand, Lawler and Sokal [24] 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. For more on the estimation of eigenvalues that applies to nice continuous situations such as the Laplacian on various domains, see e.g. Bañuelos and Carroll [3]. Jerrum and Sinclair [22] used a geometric argument with paths to bound the Cheeger's constant for a discrete-time finite space Markov chain. In this thesis, we shall use an analogous geometric argument to bound the spectral radius and the Cheeger's constant for Markov chains on bounded subsets of $\mathbb{R}^n$.

In many cases, it is very difficult to give bounds to the convergence rate of a Markov chains. However, it is still possible to compare the rates for two different Markov chains. Quastel [28] introduced a geometric argument to compare the spectral gaps of exclusion processes. Diaconis and Saloff-Coste [11] also used a similar argument to compare the eigenvalues of two discrete-time reversible Markov chains on a discrete state space. Mira [26] introduced the covariance ordering for general Markov chains. In this thesis, we shall use a geometric argument to compare two reversible Markov chains on $\mathbb{R}^n$ quantitatively.

The main results of this thesis require the transition kernel to be of a specified form. Transition kernels arise from the Metropolis-Hasting algorithm (see e.g. Tierney [42]) are of this form. The examples we shall discuss are all applications to random walk chains, a particular case from the Metropolis-Hasting algorithm.

With the knowledge in the discrete-time Markov chains, we shall extend the results to some reversible time-homogeneous Markov processes.

There are other techniques to bound the convergence rates of Markov chains and Markov processes such as the coupling method (See e.g. Nummelin [27] for discrete time results) and the application of log-Sobolev inequalities (See Diaconis and Saloff-Coste [12] for an expository article) and direct computation of the spectrum (See e.g. Bhattacharya and Waymire [6]). For some classical results related to diffusions we shall study in the thesis, see e.g. Taira [41], Weinberger [43] and Courant and Hilbert [9]. Perhaps those techniques are more powerful mathematically on specific examples, but the techniques in this thesis may sometimes be more robust and easier to apply.

The remainder of this chapter introduces the definitions of $\lambda_0(L)$ and explain the meaning of Cheeger's constant. In Chapter 2, we first discuss the relation between the convergence rate of a discrete-time Markov chain and $\lambda_0(L)$. Then we define paths and give lower bounds for both $\lambda_0(L)$ and the Cheeger's constant when the Markov chain is defined in a bounded subset of $\mathbb{R}^n$. We also study some special cases which the paths are easy to find. In Chapter 3, we use similar path arguments from Chapter 2 to compare two reversible Markov chains defined on $\mathbb{R}^n$ quantitatively.

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

We conclude the thesis with Chapter 5 and discuss some possible development and improvement in the future.

Parts of Chapters 1, 2 and 3 were already published in Yuen [44].

1.2 Basic Notations and the Cheeger's Inequality

In this section, we describe the type of Markov chains and Markov processes considered in this thesis.

We first give some informal definitions on Markov chains and Markov processes we considered in this thesis. Let $\mathcal{F}$ be a countably generated σ-algebra on S. A discrete-time Markov chain is a countable collection of random variables $\{X_n : n \in \{0\} \cup \mathbb{N}\}$ defined on a state space $(S, \mathcal{F})$ which has the property that it is forgetful of all but its most immediate past. Probabilistically, this means that there must be a collection of 'transition probabilities' $\{P^n(x, A) : n \in \mathbb{N}\}$ for each $x \in S$ and $A \in \mathcal{F}$ such that

(i) for any fixed $A \in \mathcal{F}$, the function $P^n(\cdot, A)$ is measurable,

(ii) for any fixed $x \in S$, the function $P(x, \cdot)$ is a probability measure on $(S, \mathcal{F})$ and

(iii) for all times m, n,

$$P(X_{m+n} \in A | X_r, r < m; X_m = x) = P^n(x, A);$$

that is, $P^n(x, A)$ denotes the probability that a chain at x will be in the set A after n steps. See Nummelin [27] for formal definitions.

A homogeneous continuous-time Markov processes in this thesis is a collection of random variables $\{X_t : t \geq 0\}$ defined on a metric space S such that there must be a collection of 'transition probabilities' $\{P^t(x, A) : t > 0\}$ for each $x \in S$ and Borel measurable $A \subset S$ such that

(i) for each fixed x and t, $P^t(x, \cdot)$ is a Borel probability measure on S,

(ii) for each fixed x, $P^0(x, \cdot)$ is the unit mass at x,

(iii) for each fixed Borel measurable $A \subset S$, $P^{\cdot}(\cdot, A)$ is a bounded function and

(iv) for all times s, t,

$$P(X_{s+t} \in A | X_r, r \leq t; X_t = x) = P^s(x, A).$$

See Ethier and Kurtz [14] for formal definitions.

In Chapter 2, we 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

$$(Pf)(x) = \int f(y)P(x, dy). \tag{1.1}$$

P also acts to the left on measures, so that

$$\mu P(A) = \int P(x, A)\mu(dx). \tag{1.2}$$

Recall that a Markov chain is *reversible* if

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx). \tag{1.3}$$

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

$$\lambda_0(P) = \inf \operatorname{spec}(P \mid_{\mathbf{1}^\perp}), \tag{1.4}$$

$$\lambda_1(P) = \sup \operatorname{spec}(P \mid_{\mathbf{1}^\perp}). \tag{1.5}$$

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

$$\lambda_0(L) = \inf \operatorname{spec}(L \mid_{\mathbf{1}^\perp}) = \inf\{\frac{(f, Lf)_{L^2(\pi)}}{\|f\|^2_{L^2(\pi)}} : f \in \mathbf{1}^\perp, f \neq 0\}. \tag{1.6}$$

In Section 2.2, 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:

$$k \equiv \inf_{\substack{A \in \mathcal{F} \\ 0<\pi(A)<1}} k(A) \tag{1.7}$$

with

$$k(A) \equiv \frac{\int \chi_A(x) P(x, A^c) \pi(dx)}{\pi(A)\pi(A^c)}, \tag{1.8}$$

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 [24] proved an inequality for reversible Markov chains:

$$\frac{k^2}{8M} \leq 1 - \lambda_1(P) \leq k, \tag{1.9}$$

where

$$\pi\text{-esssup}\, P(x, \{x\}^c) \leq M. \tag{1.10}$$

As we shall see in Section 2.1, $\lambda_1(P)$ 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:

$$h \equiv \inf_{\substack{A \in \mathcal{F} \\ 0 < \pi(A) \leq \frac{1}{2}}} h(A) \tag{1.11}$$

with

$$h(A) \equiv \frac{\int \chi_A(x) P(x, A^c) \pi(dx)}{\pi(A)}, \tag{1.12}$$

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:

$$\frac{h^2}{2M} \leq 1 - \lambda_1(P) \leq 2h. \tag{1.13}$$

Since the two inequalities (1.9) and (1.13) are very similar, we shall refer both to as the *Cheeger's inequality.*

As we shall see in Chapter 2, it is easier to apply the k-constant in the discrete-time Markov chain case.

In Chapter 4, we shall turn our attention to positive recurrent homogeneous continuous-time Markov processes. Since some results are extensions from the discrete-time results, we leave the introduction of notations to Section 4.1.

Chapter 2

Geometric Bounds: Discrete-Time Case

2.1 L^2 Convergence to π in the Discrete-Time Case

Following the discussion in Section 1.2, we assume that $P(x, dy)$ is the transition probability kernel of a discrete time Markov chain with invariant probability measure π. We extend the definition of L^2 norm for signed measure. Given a signed measure μ on S, define $\|\mu\|_2$ by

$$\|\mu\|_2^2 = \begin{cases} \int_S |\frac{d\mu}{d\pi}|^2 d\pi, & \mu << \pi; \\ \infty, & \text{otherwise.} \end{cases} \tag{2.1}$$

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

$$\|P\ |_{\mathbf{1}^\perp}\ \|_2 = \sup\{\|Pf\|_2; f \in \mathbf{1}^\perp, \|f\|_2 = 1\} \tag{2.2}$$

where Pf is defined as in (1.1). We shall use a property of a bounded self-adjoint operator and a result from spectral theory (see e.g. Rudin [37], Kreyszig [23], Conway [8]) to prove the following well-known proposition.

Proposition 2.1. *For a discrete-time Markov chain, if P is reversible w.r.t. π and $\mu \in L^2(\pi)$, then*

$$\|\mu P^n - \pi\|_2 \le \|\mu - \pi\|_2 \rho^n, \tag{2.3}$$

where

$$\rho = \max\{|\lambda_0(P)|, |\lambda_1(P)|\}. \tag{2.4}$$

Proof. For $\mu \in L^2(\pi)$, let $f = \frac{d\mu}{d\pi}$. Since P is reversible w.r.t. π, P is a self-adjoint operator. It is also 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 $d(\mu P) = (Pf)d\pi$. It is also well known (see e.g. Baxter and Rosenthal [4]) that $\|P\ |_{\mathbf{1}^\perp}\ \|_2$ defined in (2.2) satisfies $\|P\ |_{\mathbf{1}^\perp}\ \|_2 \le 1$. From a result of spectral theory for bounded self-adjoint operator P,

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

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

$$\|\mu P^n - \pi\|_2 \quad = \quad \|\mu P^n - \pi P^n\|_2$$

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

Q.E.D.

Remarks. 1. ρ defined in (2.4) is known as the *spectral radius* of the operator P. 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 [13] gave a geometric lower bound of it). If we assume $P(x,x) \geq a > 0 \forall x \in S$, then it directly implies that $\lambda_0(P) > -1+2a$. So, we have $\rho \leq \max\{1-2a, \lambda_1(P)\}$. In particular, if we consider the chain $\frac{1}{2}(I+P)$, then $a \geq \frac{1}{2}$ 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. We can also bound $\lambda_0(L) = 1-\lambda_1(P)$ directly. In section 2.2, we shall use a geometric argument to give a lower bound for k and a class of lower bounds for $\lambda_0(L)$. Hence, two types of upper bounds for the convergence rate will be found.

2. It is easier to get lower bounds on $\lambda_1(P)$. Note that

$$\lambda_1(P) = 1-\lambda_0(L) = 1-\inf\{\frac{(f,Lf)_{L^2(\pi)}}{\|f\|^2_{L^2(\pi)}} : f \in \mathbf{1}^\perp, f \neq 0\}.$$

Therefore, we can get lower bounds on $\lambda_1(P)$ by using test functions f.

3. For continuous-time Markov jump processes, the inequality (2.3) is much simpler. 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 $\lambda_0(\bar{P}^t) \geq 0$. Hence, we have the following relation:

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

In fact, the results can be generalized to more general continuous-time Markov jump processes in which the transition rates are essentially bounded, not necessarily with mean 1 exponential holding times. In this thesis, we only discuss the corresponding continuous-time process of a discrete-time transition kernel.

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

5. 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 [31].

2.2 Path Bounds for $\lambda_1(P)$

In this section, we prove a class of lower bounds for $\lambda_0(L)$ defined in (1.6) and an upper bound of 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 (1.5). These bounds are called *path bounds* as each bound is associated with a set of paths. 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 $S = \{x \in \mathbb{R}^n : q(x) > 0\}$. $\Gamma = \{\gamma_{xy}\}$ is 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 $\gamma_{xy} : \{0, \ldots, b_{xy}\} \to S$ 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 prove a class of lower bounds for $\lambda_0(L)$, a set of paths Γ has to satisfy the following technical conditions:

Definition 2.1. *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 in the proof of Theorem 2.2. Let $J_{bi}(u,v)$ be the Jacobian of the change of variable $(x,y) = G_{bi}(u,v)$, given by

$$J_{bi}(u,v) = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_n} & \frac{\partial x_1}{\partial v_1} & \cdots & \frac{\partial x_1}{\partial v_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_n} & \frac{\partial x_n}{\partial v_1} & \cdots & \frac{\partial x_n}{\partial v_n} \\ \frac{\partial y_1}{\partial u_1} & \cdots & \frac{\partial y_1}{\partial u_n} & \frac{\partial y_1}{\partial v_1} & \cdots & \frac{\partial y_1}{\partial v_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial y_n}{\partial u_1} & \cdots & \frac{\partial y_n}{\partial u_n} & \frac{\partial y_n}{\partial v_1} & \cdots & \frac{\partial y_n}{\partial v_n} \end{vmatrix} \tag{2.5}$$

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).

Intuitively, the condition in Definition 2.1 means that the paths are 'differentiable' in the sense that slight change in the edge can only change a path containing it continuously and the corresponding Jacobian is well defined. With this condition, we can apply a change of variables to prove Theorem 2.2. Note also that b_{xy} need not be constant and the inverse map G_{bi} is only defined for edges which are the ith edges of some paths with $b(\leq b_{xy})$ steps. We shall give some examples in Section A and B following Corollary 2.4 that if we construct some paths 'smooth' enough, the condition is easily verified.

The geometric constants that we shall use in the bounds are

$$\begin{aligned} k_\epsilon = k_\epsilon(\Gamma) &= \operatorname{esssup}_{(u,v)\in E}\{Q(u,v)^{-(1-2\epsilon)} \\ &\qquad \times \sum_{\gamma_{xy}\ni(u,v)} \|\gamma_{xy}\|_\epsilon q(x)q(y)|J_{xy}(u,v)|\}, \end{aligned} \tag{2.6}$$

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

Theorem 2.2. *Let $\Gamma = \{\gamma_{xy}\}$ be a set of paths satisfying the first regularity condition (Definition 2.1) for an irreducible Markov chain P on $\mathbb{R}^n$. Then for any $\epsilon \in \mathbb{R}$,*

$$\lambda_0(L) \geq \frac{1}{k_\epsilon}. \tag{2.7}$$

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

Proof. Since L is real and self-adjoint, we need only to consider real trial functions. For real $f \in \mathbf{1}^\perp$, we have

$$(f, Lf)_{L^2(\pi)} = \frac{1}{2}\iint_{S\times S}(f(v) - f(u))^2 Q(u,v)dudv.$$

Now,

$$
\begin{aligned}
\|f\|^2_{L^2(\pi)} &= \int f^2(x)q(x)dx \\
&= \frac{1}{2}\iint_{S\times S}[f(x)-f(y)]^2 q(x)q(y)dxdy \\
&= \frac{1}{2}\iint_{S\times S}\left[\sum_{(u,v)\in\gamma_{xy}}\left(\frac{Q(u,v)}{Q(u,v)}\right)^{\epsilon}(f(v)-f(u))\right]^2 q(x)q(y)dxdy \\
&\le \frac{1}{2}\iint_{S\times S}\|\gamma_{xy}\|_{\epsilon}q(x)q(y)\sum_{(u,v)\in\gamma_{xy}}Q(u,v)^{2\epsilon}(f(v)-f(u))^2 dxdy \\
&= \frac{1}{2}\iint_{S\times S}\left\{\sum_{i=1}^{b_{xy}}\|\gamma_{xy}\|_{\epsilon}q(x)q(y)Q(\gamma_{xy}(i-1),\gamma_{xy}(i))^{2\epsilon}\right. \\
&\qquad\qquad \left.\times(f(\gamma_{xy}(i))-f(\gamma_{xy}(i-1)))^2\right\}dxdy,
\end{aligned}
$$

where the inequality is Cauchy-Schwartz. Note that

$$
\iint_{S\times S}\sum_{i=1}^{b_{xy}}\cdot dxdy = \iiint_V \cdot\,\delta(di)dxdy,
$$

where δ is the counting measure on $\mathbb{N}$. By the first regularity condition, we can consider the change of variables $(x,y,i)=T^{-1}(u,v,b,i)$ and we have

$$
\begin{aligned}
\iiint_V \cdot\,\delta(di)dxdy &= \iiiint_{T(V)}\cdot\,|J_{bi}(u,v)|\delta(db)\delta(di)dudv \\
&\le \iint_E \sum_{\gamma_{xy}\ni(u,v)}\cdots|J_{xy}(u,v)|dudv
\end{aligned}
$$

as measure since $W_{bi}\subset E$ for each b and each i. Hence, we have

$$
\begin{aligned}
\|f\|^2_{L^2(\pi)} &\leq \frac{1}{2}\iint_E \Bigg\{ \sum_{\gamma_{xy}\ni(u,v)} \|\gamma_{xy}\|_\epsilon q(x)q(y)Q(u,v) \\
&\qquad\qquad \times (f(v)-f(u))^{2\epsilon}|J_{xy}(u,v)| \Bigg\} dudv \\
&\leq \operatorname{esssup}_{(u,v)\in E}\{Q(u,v)^{-(1-2\epsilon)} \sum_{\gamma_{xy}\ni(u,v)} \|\gamma_{xy}\|_\epsilon q(x)q(y)|J_{xy}(u,v)|\} \\
&\qquad\qquad \times \frac{1}{2}\iint_E Q(u,v)(f(v)-f(u))^2 dudv \\
&= k_\epsilon \cdot \frac{1}{2}\iint_E (f(v)-f(u))^2 Q(u,v) dudv \\
&\leq k_\epsilon \cdot \frac{1}{2}\iint_{S\times S} (f(v)-f(u))^2 Q(u,v) dudv \\
&= k_\epsilon (f, Lf)_{L^2(\pi)}.
\end{aligned}
$$

Thus, $\lambda_0(L) = \inf\{\frac{(f,Lf)_{L^2(\pi)}}{\|f\|^2_{L^2(\pi)}} : f \in \mathbf{1}^\perp, f \neq 0\} \geq \frac{1}{k_\epsilon}$ for any $\epsilon \in \mathbb{R}$. Since $\lambda_1(P) = 1 - \lambda_0(L)$, the last statement follows trivially. Q.E.D.

To prove 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 2.1):

Definition 2.2. *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.

Assuming that the second regularity condition is satisfied, we can consider the same change of variable $(x, y) = G_{bi}(u, v)$ and Jacobian $J_{xy}(u, v)$.

The condition is somewhat similar to the first regularity condition except that we require a slight adjustment to use a different change of variables to prove Theorem 2.3.

The geometric constant that we shall use in the bound is

$$K = K(\Gamma) = \text{esssup}_{(u,v)\in E}\{Q(u, v)^{-1} \sum_{\gamma_{xy} \ni (u,v)} q(x)q(y)|J_{xy}(u, v)|\}. \tag{2.8}$$

Theorem 2.3. *Let $\Gamma = \{\gamma_{xy}\}$ be a set of paths satisfying the second regularity condition (Definition 2.2) for an irreducible Markov chain P on $\mathbb{R}^n$. Then the Cheeger's k-constant satisfies*

$$k \geq \frac{1}{K}. \tag{2.9}$$

Hence, if M satisfies (1.10), $\lambda_1(P) \leq 1 - \frac{1}{8K^2M}$.

Proof. For any $A \subset S$ and $0 < \pi(A) < 1$,

$$\begin{aligned}
\pi(A)\pi(A^c) &= \int_A q(x)dx \int_{A^c} q(y)dy \\
&= \iint_{A\times A^c} q(x)q(y)dxdy \\
&= \iint_{V_A} q(x)q(y)dxdy.
\end{aligned}$$

By the second regularity condition, we can consider the change of variables $(x,y) = T_A^{-1}(u,v,b,i)$ and we have

$$\begin{aligned}
\iint_{V_A} dxdy &= \iiiint_{T_A(V_A)} |J_{bi}(u,v)|\delta(db)\delta(di)dudv \\
&\leq \iint_{V_A} \sum_{\gamma_{xy}\ni(u,v)} |J_{xy}(u,v)|dudv
\end{aligned}$$

as measure. Hence, we have

$$\begin{aligned}
\pi(A)\pi(A^c) &\leq \iint_{V_A} \sum_{\gamma_{xy}\ni(u,v)} q(x)q(y)|J_{xy}(u,v)|dudv \\
&= \iint_{V_A} Q(u,v)^{-1} \sum_{\gamma_{xy}\ni(u,v)} q(x)q(y)|J_{xy}(u,v)|Q(u,v)dudv \\
&\leq \operatorname{esssup}_{(u,v)\in E}\{Q(u,v) \sum_{\gamma_{xy}\ni(u,v)} q(x)q(y)|J_{xy}(u,v)|\} \\
&\qquad \times \iint_{V_A} Q(u,v)dudv \\
&= K \iint_{A\times A^c} Q(u,v)dudv \\
&= K \int \chi_A(x)P(x,A^c)\pi(dx).
\end{aligned}$$

Thus, the Cheeger's k-constant

$$k = \inf_{\substack{A\in\mathcal{F} \\ 0<\pi(A)<1}} \frac{\int \chi_A(x)P(x,A^c)\pi(dx)}{\pi(A)\pi(A^c)} \geq \frac{1}{K}.$$

By Cheeger's inequality (1.9), $\lambda_1(P) \le 1 - \frac{k^2}{8M} \le 1 - \frac{1}{8K^2M}$. Q.E.D.

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

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

3. There is no guarantee that such paths should exist in Theorem 2.2 and 2.3. All we say is that *if* such paths do exist, *then* the theorem holds. Moreover, we want to choose paths such that some of the geometric constants are finite, so that we have a bound strictly less than 1 for $\lambda_1(P)$. We shall discuss a special type of paths and simplify the constants in Corollary 2.4.

4. In Theorem 2.2, k_ϵ is a bound for *any* ϵ. In particular, when $\epsilon = 0$ or $\frac{1}{2}$, we have the following constants:

$$k_0 = \operatorname{esssup}_{(u,v)\in E}\{Q(u,v)^{-1} \sum_{\gamma_{xy}\ni(u,v)} b_{xy}q(x)q(y)|J_{xy}(u,v)|\}, \tag{2.10}$$

and

$$k_{\frac{1}{2}} = \operatorname{esssup}_{(u,v)\in E}\{ \sum_{\gamma_{xy}\ni(u,v)} \|\gamma_{xy}\|_{\frac{1}{2}} q(x)q(y)|J_{xy}(u,v)|\}. \tag{2.11}$$

This is a continuous version of the results in finite state spaces (see e.g. Diaconis and Stroock [13] and Sinclair [38]).

5. It is not necessary that $(G^{bi})^{-1}$ has continuous partial derivatives for each i on the whole W_{bi} in the regularity conditions. Measure-zero sets can be neglected. The same is true in the following corollaries.

6. For non-reversible Markov chains, Proposition 2.1 does not hold anymore. So, we cannot relate the convergence rate to the spectrum of the corresponding Laplacian. However, we can still say something about the spectrum. In fact, given any non-reversible Markov chains with Laplacian $L = I - P$, by considering the self-adjoint Laplacian $\frac{1}{2}(L + L^*)$, it can be shown that the spectrum of L is contained in the set $\{\lambda : |\lambda| \leq 2, \mathrm{Re}\lambda \geq \frac{k^2}{8M}\}$ (see Lawler and Sokal [24], Theorem 2.3(b)). Note that in the proof of Theorem 2.3, we did not use the fact that L is reversible to prove $k \geq \frac{1}{K}$. So, the spectrum of L is contained in

$$\{\lambda : |\lambda| \leq 2, \mathrm{Re}\lambda \geq \frac{1}{8K^2M}\}.$$

Now we consider a particular case when $b_{xy} = b$, a constant for any $x, y \in S$. Denote the paths by γ^b_{xy}. We can then simplify the regularity conditions as follows. For $i = 1, \ldots, b$, define $T_i : S \times S \to S \times S$ by $T_i(x, y) = (\gamma^b_{xy}(i-1), \gamma^b_{xy}(i))$. Assume T_i can extended to an open set V containing $S \times S$ s.t. T_i is a 1-1 map onto $T_i(V)$ and that $(T_i)^{-1}$ has continuous partial derivatives a.e. for each i. Let $J_i(u, v)$ be the Jacobian of the change of variable $(x, y) = (T_i)^{-1}(u, v)$. From here we can apply Theorem 2.2 and 2.3 directly or we can enjoy a different version by the following corollary.

Corollary 2.4. *Under the above assumptions, the geometric constants in Theorem 2.2 can be simplified as*

$$k_\epsilon(\Gamma) = \sum_{i=1}^{b} \operatorname{esssup}_{(u,v)\in E_i} \Big\{ Q(u,v)^{-(1-2\epsilon)} \|\gamma_{xy}^b\|_\epsilon q(x)q(y)|J_i(u,v)| : (x,y) = (T_i)^{-1}(u,v) \Big\} \tag{2.12}$$

and the geometric constant in Theorem 1.2 can be simplified as

$$K(\Gamma) = \sum_{i=1}^{b} \operatorname{esssup}_{(u,v)\in E_i} \Big\{ Q(u,v)^{-1} q(x)q(y)|J_i(u,v)| : (x,y) = (T_i)^{-1}(u,v) \Big\}. \tag{2.13}$$

Proof. For k_ϵ, we follow the proof of Theorem 1.1 until the change of variables. Since $b_{xy} \equiv b$, we have

$$\iint_{S\times S} \sum_{i=1}^{b_{xy}} dxdy = \sum_{i=1}^{b} \iint_{S\times S} dxdy.$$

For each i, we can consider the change of variables $(x,y) = (T_i)^{-1}(u,v)$ by assumption, which gives

$$\sum_{i=1}^{b} \iint_{S\times S} dxdy = \sum_{i=1}^{b} \iint_{E_i} |J_i(u,v)| dudv.$$

From then on, it is easy to see that k_ϵ can be simplified as in the statement. For K, we follow the proof of Theorem 1.2 until the change of variables. Redefine $T_A : V_A \to S^2$ by

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

where

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

Note that

$$\begin{aligned}\iint_{V_A} dxdy &= \sum_{i=1}^{b} \iint_{(T_i)^{-1}(E_i \cap T_A(V_A))} dxdy \\ &= \sum_{i=1}^{b} \iint_{(E_i \cap T_A(V_A))} |J_i(u,v)| dudv.\end{aligned}$$

From then on, it is easy to see K can be simplified as in the statement.

Q.E.D.

Remark 3 suggests that we have no definite way to find the paths in general. However, in the following special cases of S, we shall have at least one way to construct paths satisfying the regularity conditions (Definition 2.1 and 2.2). The geometric constants can be simplified as follows.

A. *S is convex*

In this case, fix any $b \in \mathbb{N}$ 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).

On the other hand, we can also try the *linear* paths in which $b_{xy} = \lceil \frac{1}{c} \|y - x\| \rceil$ for some constant $c > 0$. Define $\gamma_{xy} : \{0, \ldots, b_{xy}\} \to S$ by

$$\gamma_{xy}(i) = \frac{(b_{xy} - i)x + iy}{b_{xy}}.$$

They are similar to the uniform paths except that the length of any edge $(u, v) \in E$ is less than or equal to c. Intuitively, a majority of edges have length close to c. It is easy to check that both the first and second regularity conditions are satisfied and that $|J_{xy}(u, v)| = b_{xy}^n$. We sum up the results in the following corollary:

Corollary 2.5. *(a) Under the uniform paths $\{\eta_{xy}^b\}$ above, the constants can be simplified as*

$$\begin{aligned} k_\epsilon &= b^n \sum_{i=1}^{b} \operatorname{esssup}_{(u,v)\in E_i} \{Q(u, v)^{-(1-2\epsilon)} \|\eta_{xy}^b\|_\epsilon q(x) q(y) : \\ &\qquad\qquad (x, y) = (T_i)^{-1}(u, v)\} \end{aligned} \tag{2.14}$$

and

$$K = b^n \sum_{i=1}^{b} \operatorname{esssup}_{(u,v)\in E_i} \{Q(u, v)^{-1} q(x) q(y) : (x, y) = (T_i)^{-1}(u, v)\}. \tag{2.15}$$

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

$$k_\epsilon = \operatorname{esssup}_{(u,v)\in E} \{Q(u, v)^{-(1-2\epsilon)} \sum_{\gamma_{xy} \ni (u,v)} \|\gamma_{xy}\|_\epsilon q(x) q(y) b_{xy}^n\}, \tag{2.16}$$

and

$$K = \operatorname{esssup}_{(u,v)\in E}\{Q(u,v)^{-1}\sum_{\gamma_{xy}\ni(u,v)} q(x)q(y)b^n_{xy}\}. \tag{2.17}$$

B. *S is non-convex*

The reason why the uniform paths work so well for convex S is that the line joining any two points lies entirely on S and the Jacobian is relatively simple. For general S, however, we cannot use the uniform paths directly. If S is non-convex but isomorphic to a convex set in a certain sense, we can still use the uniform paths indirectly. Formally, we assume that S satisfies the following conditions. Let U be open in $\mathbb{R}^n$ 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 η^b_{xy} on $\phi(S)$, we can define paths γ_{xy} on S by

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

Define $\Phi : U \times U \to \phi(U) \times \phi(U)$ by $\Phi(x,y) = (\phi(x), \phi(y))$. So, the Jacobian of Φ exists and we can denote it by $J_\Phi(x,y)$.

To compute $J_i(u,v)$ in Corollary 1.2, we notice that $T_i = \Phi^{-1} \circ T^\eta_i \circ \Phi$ where T^η_i represents the T_i for uniform paths. By the chain rule and the inverse function theorem of differentiation,

$$\begin{aligned}
|J_i(u,v)| &= |\det D(T_i)^{-1}(u,v)| \\
&= |\det D(\Phi \circ (T_i^\eta)^{-1} \circ \Phi^{-1})(u,v)| \\
&= |\det D\Phi((T_i^\eta)^{-1} \circ \Phi^{-1}(u,v))| \cdot |\det D(T_i^\eta)^{-1}(\Phi^{-1}(u,v))| \\
&\qquad \cdot |\det D\Phi^{-1}(u,v)| \\
&= |J_\Phi((T_i^\eta)^{-1} \circ \Phi^{-1}(u,v))| \cdot b^n \cdot |\frac{1}{\det D\Phi(\Phi^{-1}(u,v))}| \\
&= \frac{|J_\Phi((T_i^\eta)^{-1} \circ \Phi^{-1}(u,v))| \cdot b^n}{|J_\Phi(\Phi^{-1}(u,v))|}.
\end{aligned}$$

Moreover, if $0 < m \le |J_\Phi| \le M$,

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

If $|J_\Phi|$ is constant, $|J_i(u,v)| = b^n$.

We can also use the linear paths instead of the uniform paths. In that case,

$$|J_{xy}(u,v)| = \frac{|J_\Phi((T_{xy}^\eta)^{-1} \circ \Phi^{-1}(u,v))| \cdot b^n}{|J_\Phi(\Phi^{-1}(u,v))|},$$

where $T_{xy}^\eta(u,v)$ represents the corresponding bijection with respect to x, y for linear paths. Moreover, if $0 < m \le |J_\Phi| \le M$,

$$|J_{xy}(u,v)| \le \frac{M \cdot b_{xy}^n}{m}.$$

If $|J_\Phi|$ is constant, $|J_{xy}(u,v)| = b_{xy}^n$.

2.3 Examples

Example 2.3.1 to 2.3.5 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 2.2. They are all random walk chains, a particular case from the Metropolis-Hasting algorithm (see e.g. Tierney [42]).

Note that we choose the stationary distribution of all the following examples to be uniform over the state space in order to simplify the calculations. It is easy to see that we can also get positive lower bounds on the spectral gap for chains with stationary distributions bounded above and below by positive constants (not necessarily uniform). Therefore, our results are quite robust in the sense that we can apply them easily on many different chains instead of just one special chain.

Example 2.3.1. *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. Since S is convex, we can choose the uniform paths or the linear paths and apply Corollary 2.5. For the uniform paths, we need to choose b big enough s.t. the expression inside the esssup is finite for all $(u, v) \in E_i$. So, we require whenever $(x, y) = (T_i)^{-1}(u, v)$, $Q(u, v) > 0$ for K to be finite and in addition $\|\eta^b_{xy}\|_\epsilon$ is finite for k_ϵ to be finite. For the best bounds, we take $b = \lceil 2a \rceil$. In this case, the distance between two adjacent points in a path is less than 1. So, $Q(u,v) = \frac{1}{4a} > 0$ and $\|\eta^b_{xy}\|_\epsilon = \sum_{(u,v)\in\eta^b_{xy}} Q(u,v)^{-2\epsilon} = b(4a)^{2\epsilon}$ which is finite. By Corollary 2.5(a),

$$k_\epsilon = b^1 \sum_{i=1}^{b} (\frac{1}{4a})^{-(1-2\epsilon)} b(4a)^{2\epsilon} (\frac{1}{2a})^2 = \frac{b^3}{a},$$

and

$$K = b^1 \sum_{i=1}^{b} (\frac{1}{4a})^{-1} (\frac{1}{2a})^2 = \frac{b^2}{a}.$$

In particular, if $2a$ is an integer, $b = 2a$ and so $k_\epsilon = 8a^2 = O(a^2)$ and $K = 4a = O(a)$. So, $1 - \frac{1}{k_\epsilon} = 1 - \frac{1}{8a^2}$ and $1 - \frac{1}{8K^2} = 1 - \frac{1}{2^7 a^2}$ are upper bounds of $\lambda_1(P)$ by Theorem 2.2 and 2.3 respectively. In this case, $1 - \frac{1}{k_\epsilon}$ turns out to be a better bound. However, $\frac{1}{k_\epsilon}$ and $\frac{1}{8K^2}$ differ only by a constant.

For the linear paths, we choose $c = 1$ for the same reason as for the uniform paths and apply Corollary 2.5(b). We shall give an upper bound for K. The sum in the essential supremum $\sum_{\gamma_{xy}\ni(u,v)}$ is in fact a double sum: For $(u,v) \in E$ and $I \geq 1$, there are at most I pairs of x, y s.t. $\gamma_{xy} \ni (u,v)$ with $b_{xy} = I$. Moreover, $I \leq (1 - |v-u|)^{-1} \leq \lceil 2a \rceil$ for (u,v) to be an edge. So,

$$\begin{aligned} Q(u,v)^{-1} \sum_{\gamma_{xy}\ni(u,v)} q(x)q(y)b^n_{xy} &\leq (\frac{1}{4a})^{-1} \sum_{I=1}^{\lceil 2a \rceil} \sum_{i=1}^{I} (\frac{1}{2a})^2 I \\ &= \frac{1}{6a} \lceil 2a \rceil (\lceil 2a \rceil + 1)(2\lceil 2a \rceil + 1). \end{aligned}$$

In particular, if $2a$ is an integer, $K \leq \frac{1}{3}(2a+1)(4a+1) = O(a^2)$. Similarly, $k_\epsilon \leq O(a^3)$. Both bounds are much worse (different orders for K and k_ϵ)

than the corresponding ones we get from the uniform paths. This illustrates different paths can give very different bounds.

If we consider $\tilde{P} = \frac{1}{2}(I+P)$ as the transitional kernel for a discrete-time Markov chain, then $\rho \leq 1 - \frac{1}{16a^2}$ (with k_ϵ for the uniform paths). Then by Proposition 1.1, for any $\mu \in L^2(\pi)$,

$$\|\mu\tilde{P}^n - \pi\|_2 \leq \|\mu - \pi\|_2(1 - \frac{1}{16a^2})^n.$$

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

$$\|\mu\bar{P}^t - \pi\|_2 \leq \|\mu - \pi\|_2 e^{-t(1-\frac{1}{16a^2})}.$$

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. We consider the uniform paths η^b_{xy} in Corollary 2.5 for $b \in \mathbb{N}$. For any $(u,v) \in E_i$ and $(x,y) = (T_i)^{-1}(u,v)$,

$$\|\eta^b_{xy}\|_\epsilon = \sum_{i=1}^{b}(\frac{1}{2a}\frac{1}{\sqrt{2\pi}}e^{-\frac{(v-u)^2}{2}})^{-2\epsilon} = b(\frac{1}{2a\sqrt{2\pi}}e^{-\frac{(v-u)^2}{2}})^{-2\epsilon}.$$

Note that $p_x(y)$ is a decreasing function of $|y-x|$. So,

$$k_\epsilon \leq b^1\sum_{i=1}^{b} b(\frac{1}{2a\sqrt{2\pi}}e^{-\frac{(\frac{2a}{b})^2}{2}})^{-1}(\frac{1}{2a})^2 = \frac{\sqrt{2\pi}b^3 e^{\frac{2a^2}{b^2}}}{2a}.$$

To get the best upper bound, we minimize the expression over b and we get $b = \lceil \frac{2a}{\sqrt{3}} \rceil$ or $\lfloor \frac{2a}{\sqrt{3}} \rfloor$. In particular, if $\frac{2a}{\sqrt{3}}$ is an integer,

$$k_\epsilon \leq \frac{4\sqrt{2\pi} a^2 e^{\frac{3}{2}}}{3\sqrt{3}}.$$

We can go through similar calculations for K to get

$$K \leq \frac{\sqrt{2\pi} b^2 e^{\frac{2a^2}{b^2}}}{2a}.$$

Minimizing over b and assuming that $\sqrt{2}a$ is an integer,

$$K \leq e\sqrt{2\pi} a.$$

In this example, the bound we obtained from k_ϵ is better than that from K.

Remarks. From the above calculations, it is easy to show that for random walk chains, i.e., $p_x(y) = p(|y-x|)$, k_ϵ is independent of ϵ with uniform or linear paths. For simplicity, we shall only compute k_0 in the following examples if this is the case.

Example 2.3.2. *Two-dimensional convex case:* $S = B(0,a) \subset \mathbb{R}^2$.

(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 $q(x) = \frac{1}{a^2\pi} \mathbf{1}_S(x)$ As in Example 2.3.1(a), we use the uniform paths η^b_{xy} and take $b = \lceil 2a \rceil$. Since $\frac{2a}{b} \leq 1$, for any $(u,v) \in E_i$, $p_u(v) > 0$. With similar calculations,

$$k_\epsilon = k_0 = b^2 \sum_{i=1}^{b} \left(\frac{1}{\pi}\frac{1}{a^2\pi}\right)^{-1} b \left(\frac{1}{a^2\pi}\right)^2 = \frac{b^4}{a^2}.$$

Similarly, $K = \frac{b^3}{a^2}$. In particular, if $2a$ is an integer, $b = 2a$ and by Theorem 2.2 and 2.3, $\lambda_1(P) \le \min\{1 - \frac{1}{k_0}, 1 - \frac{1}{8K^2}\} = 1 - \frac{1}{16a^2}$. Again, k_ϵ gives a better bound. So, unless otherwise specified, we shall only calculate the most useful bound in the following examples.

(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. Similar to Example 2.3.1(b), we consider the uniform paths and let $b \in \mathbb{N}$. Note that $p_{(u_1,u_2)}(v_1, v_2)$ is constant on any circle on S with center (u_1, u_2) and the value is decreasing as the radius increases. So, the esssup that defines k_ϵ in Corollary 2.5(a) is reached when $\|(v_1, v_2) - (u_1, u_2)\| = \frac{D}{b} = \frac{2a}{b}$ for some $(u, v) \in E_i$. So,

$$k_\epsilon \le \frac{2b^4}{a^2} e^{\frac{2a^2}{b^2}}.$$

Again, minimizing the R.H.S. gives $b = \lceil a \rceil$ or $\lfloor a \rfloor$. In particular, if a is an integer,

$$k_\epsilon \le 2a^2 e^2.$$

Example 2.3.3. *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 2.3.2(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 $S_{r,\theta}$. Take the uniform paths with $b = b_{r,\theta} = \lceil r \rceil$. By Corollary 2.5 (a),

$$k_\epsilon(r, \theta) = k_0(r, \theta) = \frac{4\pi \lceil r \rceil^4}{r^2 \sin 2\theta}.$$

So, $1 - \frac{1}{k_\epsilon(r,\theta)}$, the upper bounds for $\lambda_1(P)$ in Theorem 2.2, go to 1 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 2.3.4. *Two-dimensional non-convex case.*

(a) Suppose $S = \{(x, y) \in (-a, a) \times \mathbb{R} : a - |x| < y < 2(a - |x|)\}$, 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. $q(x) = \frac{1}{a^2}\mathbf{1}_S(x)$. We shall follow the discussion in Section 2B and then apply Corollary 2.4. 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 $S' = \{(x, y) \in [-a, a] \times \mathbb{R} : 0 < y < 2(a - |x|)\}$, 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.. Recall that $\Phi : S \times S \to S' \times S'$ is defined by

$$\begin{aligned}
\Phi((x_1, x_2), (y_1, y_2)) &= (\phi(x_1, x_2), \phi(y_1, y_2)) \\
&= (x_1, 2(x_2 - a + |x_1|), y_1, 2(y_2 - a + |y_1|)).
\end{aligned}$$

Note that $(T_i)^{-1}$ defined in the discussion before Corollary 2.4 has continuous partial derivatives except on a set M of λ(the Lebesgue measure on $\mathbb{R}^2 \times \mathbb{R}^2$) measure zero. To see this, let $Y = \{(0, y) : 1 \le y \le 2\}$, which is Lebesgue measure zero on $\mathbb{R}^2$). Then M is the union of $Y \times S, S \times Y, T_i(Y \times S), T_i(S \times Y)$, which is clearly of measure zero. It is easy to show that the Jacobian of Φ, $|J_\Phi((x_1, x_2), (y_1, y_2))| \equiv 4$. Since $|J_\Phi|$ is a constant, $|J_i(u, v)| = b^2$ as in Section 2B. To apply Corollary 2.4, we need to find b s.t. $Q(u, v) > 0$ for any edge (u, v). 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 if $b \ge 2\sqrt{2}a$, the horizontal distance between two adjacent points is less than $\frac{1}{\sqrt{2},}$ and so the actual distance is less than 1. Hence, by Corollary 2.4, taking $b = \lceil 2\sqrt{2}a \rceil$,

$$k_\epsilon \le \frac{\pi \lceil 2\sqrt{2}a \rceil^4}{a^2} = O(a^2).$$

(b) Suppose $S = \{(x, y) \in (-a, a) \times \mathbb{R} : \frac{1}{2}\sqrt{a^2 - x^2} < y < \sqrt{a^2 - x^2}\}$, 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. $q(x) = \frac{4}{a^2\pi}\mathbf{1}_S(x)$. Define a bijection $\phi : S \to S'$ by $\phi(x, y) = (x, 2y - \sqrt{a^2 - x^2})$, where $S' = \{(x, y) \in (-a, a) \times \mathbb{R} : 0 < y < \sqrt{a^2 - x^2}\}$, 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 2.4, recall that $\Phi : S \times S \to S' \times S'$ is defined by

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

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. $Q(u,v) > 0$ for any edge (u,v). 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 $(a-\frac{2a}{b}, \sqrt{a^2-(a-\frac{2a}{b})^2})$. So, if $b \geq 4a^2$, or $d(b) = \frac{4a^2}{b} \leq 1$, the distance between any adjacent points is less than 1. Hence, by Corollary 2.4, taking $b = \lceil 4a^2 \rceil$,

$$k_\epsilon \leq \frac{4\lceil 4a^2\rceil^4}{a^2} = O(a^6).$$

This value is much bigger 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. Similarly,

$$K = O(a^4).$$

It is interesting to note that $\frac{1}{k_\epsilon} = O(a^{-6})$ while $\frac{1}{8K^2} = O(a^{-8})$. This is example of using uniform paths and that k_ϵ, K give very different bounds.

(c) Suppose $A > a > 0$ and $S = \{(x,y) \in (-A,A)\times\mathbb{R} : \sqrt{a^2-x^2} < y < \sqrt{A^2-x^2}, y > 0\}$ is the open 'C' shape and the transition kernel is as in (b). Then the invariant p.d.f. $q(x) = \frac{2}{(A^2-a^2)\pi}\mathbf{1}_S(x)$. 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 2.4. 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 $\frac{1}{A^2} < |J_\Phi| < \frac{1}{a^2}$. From the discussion in Section 2B,

$$|J_i(u, v)| \leq \frac{\frac{1}{a^2}b^2}{\frac{1}{A^2}} = \frac{A^2 b^2}{a^2}.$$

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, $Q(u, v) = \frac{2}{(A^2-a^2)\pi}\frac{1}{\pi}$ if $d(b) \leq 1$ for any edge (u, v). For fixed A and a, we can use numerical methods to find the smallest b s.t. $d(b) \leq 1$. For such b, say b_0, by Corollary 2.4,

$$k_\epsilon \leq \frac{2A^2 b_0^4}{a^2(A^2 - a^2)}.$$

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. $d(b) \leq 1$, we have

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

In particular, if $A = 2a$,

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

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 $\frac{\pi}{2.}$

Example 2.3.5. *n-dimensional case: $S = [-a, a]^n$.*

Let $p_{\mathbf{x}}(\mathbf{y}) = 2^{-n}$ (the uniform p.d.f.) on $C(\mathbf{x}, 1) \cap S$ and 0 otherwise, where $C(\mathbf{x}, 1) = \{\mathbf{y} \in \mathbb{R}^n : |y_i - x_i| \leq 1, i = 1, \ldots, n\}$ is the n-cube with width 2 and center $\mathbf{x}$. Then the chain is reversible w.r.t. the invariant distribution with density $q(\mathbf{x}) = (2a)^{-n} \mathbf{1}_S(\mathbf{x})$. By choosing $b = \lceil 2a \rceil$, it is easy to see that for any two adjacent points **u,v** in a uniform path, we have $\mathbf{v} \in C(\mathbf{u}, 1)$. So, by Corollary 2.5(a),

$$k_\epsilon(n) = k_0(n) = \frac{b^{n+2}}{a^n} \approx 2^{n+2} a^2.$$

Chapter 3

Comparison of Spectral Radii

3.1 Comparison of Spectral Radii of Discrete-Time Markov Chains

In this section, we first discuss the comparison of λ_1 for two reversible Markov chains on general state spaces. We use the min-max principle to compare λ_1 of two chains with not necessarily the same invariant probability measure. Then we use a geometric arguments to compare λ_1 for two reversible Markov chains on $\mathbb{R}^n$ quantitatively.

In discrete state spaces, Quastel [28] introduced a geometric argument to compare the spectral gaps of exclusion processes and Diaconis and Saloff-Coste [11] used a similar argument to compare the eigenvalues of two discrete-time reversible Markov chains. In general state spaces, Mira [26] gave an ordering for chains with the same invariant probability measure.

We consider two reversible discrete-time Markov chains (or continuous-time Markov jump processes) with measurable state space $(S,\mathcal{F})$, transition probability kernels $P(x,dy)$, $\tilde{P}(x,dy)$ and invariant probability measures $\pi,\tilde{\pi}$ respectively. Recall that the spectral radius ρ of an operator H is defined in (2.4) as $\rho = \max\{|\lambda_0(H)|,|\lambda_1(H)|\}$. As we mentioned, it is hard to measure $\lambda_0(H)$. We shall compare λ_1 of P and $\tilde{P}$ by the operator form of min-max principle (see e.g. Reed and Simon [29], Theorem XIII.1). Since 0 is the smallest eigenvalue for $L = I - P$ and L is bounded from below ($L \geq cI$ by taking $c = 0$), the min-max principle implies

$$1 - \lambda_1(P) = \lambda_0(L) = \sup_{\phi\in L^2(\pi)} \inf_{\psi\in\phi^{\perp_\pi}} \frac{(\psi, L\psi)_\pi}{||\psi||_\pi^2},$$

where $\phi^{\perp_\pi}$ is a shorthand for $\{\psi|(\psi,\phi)_\pi = 0\}$. The same is true by replacing L,π by $\tilde{L},\tilde{\pi}$ respectively. Then we can extend the result of Diaconis and Saloff-Coste [11] to a more general setting:

Theorem 3.1. *Under the above assumptions, if there exists constants $a, A > 0$ s.t. (i) $(f,\tilde{L}f)_{\tilde{\pi}} \leq A(f,Lf)_\pi$ for all $f \in L^2(\pi)$ and (ii) $\tilde{\pi} \geq a\pi$, then $\lambda_0(L) \geq \frac{a}{A}\lambda_0(\tilde{L})$, and so*

$$\lambda_1(P) \leq 1 - \frac{a}{A}(1 - \lambda_1(\tilde{P})). \tag{3.1}$$

Proof. By condition (ii), we have $a \cdot \int |f|^2 d\pi \leq \int |f|^2 d\tilde{\pi}$, and so $L^2(\tilde{\pi}) \subset L^2(\pi)$. Define $\mathcal{D}(\pi) = \{W \subset L^2(\pi) : \text{codim} W = 1\}$. Then $\mathcal{D}(\tilde{\pi}) \subset \mathcal{D}(\pi)$. It is also easy to see that $\mathcal{D}(\pi) = \{\phi^{\perp_\pi} : \phi \in L^2(\pi)\}$. Hence,

$$\begin{aligned}
1 - \lambda_1(P) = \lambda_0(L) &= \sup_{\phi \in L^2(\pi)} \inf_{\psi \in \phi^{\perp_\pi}} \frac{(\psi, L\psi)_\pi}{\|\psi\|_\pi^2} \\
&= \sup_{W \in \mathcal{D}(\pi)} \inf_{\psi \in W} \frac{(\psi, L\psi)_\pi}{\|\psi\|_\pi^2} \\
&\geq \frac{1}{A} \sup_{W \in \mathcal{D}(\pi)} \inf_{\psi \in W} \frac{(\psi, \tilde{L}\psi)_{\tilde{\pi}}}{\|\psi\|_\pi^2} \\
&\geq \frac{a}{A} \sup_{W \in \mathcal{D}(\pi)} \inf_{\psi \in W} \frac{(\psi, \tilde{L}\psi)_{\tilde{\pi}}}{\|\psi\|_{\tilde{\pi}}^2} \\
&\geq \frac{a}{A} \sup_{W \in \mathcal{D}(\tilde{\pi})} \inf_{\psi \in W} \frac{(\psi, \tilde{L}\psi)_{\tilde{\pi}}}{\|\psi\|_{\tilde{\pi}}^2} \\
&= \sup_{\phi \in L^2(\tilde{\pi})} \inf_{\psi \in \phi^{\perp_{\tilde{\pi}}}} \frac{(\psi, \tilde{L}\psi)_{\tilde{\pi}}}{\|\psi\|_{\tilde{\pi}}^2} \\
&= \frac{a}{A} \lambda_0(\tilde{L}) = \frac{a}{A}(1 - \lambda_1(\tilde{P})),
\end{aligned}$$

which gives the required results. The first inequality follows from condition (i), the second from $a \cdot \int |\psi|^2 d\pi \leq \int |\psi|^2 d\tilde{\pi}$, and the third from $\mathcal{D}(\tilde{\pi}) \subset \mathcal{D}(\pi)$.

Q.E.D.

Remarks. 1. In particular, if $\pi = \tilde{\pi}$, we can take $a = 1$ and the condition is reduced to comparing (f, Lf) and $(f, \tilde{L}f)$. In fact, Mira [26] defines the covariance ordering which preserves the ordering of limiting variance of a function in $L_0^2(\pi)$.

2. If the state space is discrete and finite, we can compare the chains eigenvalue by eigenvalue. See Diaconis and Saloff-Coste [11]. Even in this general setting, we can have more refined comparison. However, for simplic-

ity, we focus on λ_0 which is directly related to the spectral radius.

3. It is usually not hard to satisfy condition (ii). However, this is not the case for condition (i). In what follows, we shall again apply some path arguments to get condition (i) for certain type of chains on $\mathbb{R}^n$.

Now we develop geometric constants that satisfy condition (i) in Theorem 3.1 for reversible Markov chains (discrete or continuous-time) on $\mathbb{R}^n$ described in 1.1. We further assume that the transition kernels are of the form $P(x, dy) = \alpha(x)\delta_x(dy) + p_x(y)dy$ and $\tilde{P}(x, dy) = \tilde{\alpha}(x)\delta_x(dy) + \tilde{p}_x(y)dy$. Suppose the invariant distributions $\pi, \tilde{\pi}$ have density $q(y), \tilde{q}(y)$ w.r.t. Lebesgue measure respectively.

We also require the existence of a set of paths $\Gamma = \{\gamma_{xy}\}$ satisfying some regularity conditions related to both P, π and $\tilde{P}, \tilde{\pi}$. For each $x \neq y$ with $\tilde{q}(x)\tilde{p}_x(y) > 0$, there exists $b_{xy} \in \mathbb{N}$ and a path $\gamma_{xy} : \{0, \ldots, b_{xy}\} \to \mathbb{R}^n$ 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. It is slightly different from the *set of paths* defined in Section 2. We shall call it a *set of* $(P, \tilde{P})$ *paths.* We also slightly modify the first regularity condition (Definition 2.1) as follows:

Definition 3.1. *Let* $\Gamma = \{\gamma_{xy}\}$ *be a set of* $(P, \tilde{P})$ *paths for irreducible Markov chains* P, π *and* $\tilde{P}, \tilde{\pi}$ *on* $\mathbb{R}^n$. *Let* $V = \{(x, y, i) : \tilde{q}(x)\tilde{p}_x(y) > 0, 1 \leq i \leq b_{xy}\}$. Γ *satisfies the* $(P, \tilde{P})$ *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.*

Again, for paths 'nice' enough, this condition is easily satisfied. For example, the uniform paths defined in Section 2.2 shall be considered.

The geometric constants we shall use are

$$\begin{aligned} A_\epsilon = A_\epsilon(\Gamma) &= \text{esssup}_{(u,v)\in E}\{Q(u, v)^{-(1-2\epsilon)} \\ &\qquad \times \sum_{\gamma_{xy}\ni(u,v)} \|\gamma_{xy}\|_\epsilon \tilde{q}(x)\tilde{p}_x(y)|J_{xy}(u, v)|\}, \end{aligned} \tag{3.2}$$

where the sum is taken over all (x, y) s.t. $\gamma_{xy} \ni (u, v)$.

Theorem 3.2. *Let $\tilde{P}, \tilde{\pi}$ and P, π be reversible Markov chains on $\mathbb{R}^n$ and $\Gamma = \{\gamma_{xy}\}$ be a set of $(P, \tilde{P})$ paths satisfying the $(P, \tilde{P})$ regularity condition (Definition 3.1). Then for any $\epsilon \in \mathbb{R}$ and $f \in L^2(\pi)$,*

$$(f, \tilde{L}f)_{\tilde{\pi}} \leq A_\epsilon(f, Lf)_\pi. \tag{3.3}$$

Proof. Note that

$$(f, \tilde{L}f)_{\tilde{\pi}} = \frac{1}{2}\iint |f(x) - f(y)|^2 \tilde{q}(x)\tilde{p}_x(y)dxdy.$$

With the modifications of the regularity conditions, we can follow the exact same proof of Theorem 2.2 and get the required result. Q.E.D.

Remarks. 1. To apply Theorem 3.1 and 3.2, we require a chain of reference: a chain $\tilde{P}, \tilde{\pi}$ with known $\lambda_1(\tilde{P})$. In the discrete case, we know how to solve the eigenvalues of some chains. In the general case, however, most chains are still unsolved.

2. As in Theorem 2.2, there is no definite procedure to find the *paths.* However, for special cases mentioned in Section 2.2, A_ϵ can be simplified considerably in a similar fashion.

3.2 Examples

The Markov chains we shall compare in Example 3.2.1 and 3.2.2 can be considered as chains described before Theorem 3.2.

Example 3.2.1. *Comparison of different one-dimensional chains with bounded state spaces.*

(a) We consider two Markov chains on $S = [-a, a]$ of the type described in Section 4. $\tilde{p}_x(y) = \frac{1}{2}$ (the uniform p.d.f.) on $[x-1, x+1] \cap S$ and 0 otherwise. $p_x(y) = \frac{b}{2}$ (the uniform p.d.f.) on $[x - \frac{1}{b}, x + \frac{1}{b}] \cap S$ and 0 otherwise for some $b \geq 1$. For simplicity, let b be an integer. As $p_x(y) = p_y(x)$ and $\tilde{p}_x(y) = \tilde{p}_y(x)$, both chains are reversible w.r.t. the invariant distribution $\pi = \tilde{\pi}$ with density $q(x) = \tilde{q}(x) = \frac{1}{2a}\mathbf{1}_S(x)$, uniform over S. We consider the *uniform* paths η^b_{xy} as usual. Then for each $x \neq y$ with $\tilde{q}(x)\tilde{p}_x(y) > 0$ and for any edge (u, v) of η^b_{xy}, $|v - u| \leq \frac{b}{b} = 1$ and so $p_u(v) > 0$. Therefore, it is a set of $(P, \tilde{P})$

paths. Similar to the uniform paths in Section 2.2A, it is easy to show that the $(P, \tilde{P})$ regularity condition is also satisfied with $J_{xy}(u,v) = b$. We are now ready to apply Theorem 3.2. We have

$$\begin{aligned} A_0 &= \text{esssup}_{(u,v)\in E}\{Q(u,v)^{-1} \sum_{\gamma_{xy}\ni(u,v)} b_{xy}\tilde{q}(x)\tilde{p}_x(y)|J_{xy}(u,v)|\} \\ &\leq (\frac{1}{2a}\frac{b}{2})^{-1} b(b\frac{1}{2a}\frac{1}{2}b) = b^2. \end{aligned}$$

So, by Theorem 3.2,

$$(f, \tilde{L}f)_{\tilde{\pi}} \leq b^2(f, Lf)\pi$$

for any $f \in L^2(\pi)$. Since $\tilde{\pi} = \pi$, Theorem 3.1 implies that

$$\lambda_1(P) \leq 1 - \frac{1}{b^2}(1 - \lambda_1(\tilde{P})).$$

Note that from Example 2.3.1(a), $\lambda_1(\tilde{P}) \leq 1 - \frac{1}{8a^2}$. So,

$$\lambda_1(P) \leq 1 - \frac{1}{b^2}(\frac{1}{8a^2}) = 1 - \frac{1}{8(ab)^2}.$$

Observe that the chain P is equivalent to the chain in Example 2.3.1(a), replacing a by ab. So, the bound we get from there is also $1 - \frac{1}{8(ab)^2}$. More importantly, the comparison we obtained is a comparison of the *actual* values of $\lambda_1(P)$ and $\lambda_1(\tilde{P})$.

On the other hand, since if $p_x(y) > 0$, $\tilde{p}_x(y) = \frac{1}{b}p_x(y) > 0$ and $q = \tilde{q}$, we have

$$\begin{aligned} (f, Lf)_\pi &= \frac{1}{2}\int\int |f(x) - f(y)|^2 q(x)p_x(y)dxdy \\ &\leq \frac{1}{2}\int\int |f(x) - f(y)|^2 \tilde{q}(x)b\tilde{p}_x(y)dxdy \\ &= b(f, \tilde{L}f)_{\tilde{\pi}}. \end{aligned}$$

By Theorem 3.1,

$$\lambda_1(\tilde{P}) \leq 1 - \frac{1}{b}(1 - \lambda_1(P)).$$

However, this is not a good comparison as we cannot recover the bound $1 - \frac{1}{8a^2}$ of $\lambda_1(\tilde{P})$ from the fact that $\lambda_1(P) \leq 1 - \frac{1}{8(ab)^2}$. Even so, we still manage to get both lower and upper bound for $\lambda_1(P)$ through $\lambda_1(\tilde{P})$:

$$1 - b(1 - \lambda_1(\tilde{P})) \leq \lambda_1(P) \leq 1 - \frac{1}{b^2}(1 - \lambda_1(\tilde{P})).$$

(b) We consider chains similar to (a). Redefine $\tilde{p}_x(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2}}$ (the normal p.d.f. of $n(x,1)$) and $p_x(y) = \frac{b}{\sqrt{2\pi}} e^{-\frac{b^2(y-x)^2}{2}}$ (the normal p.d.f. of $n(x, \frac{1}{b^2})$) for $y \in S$ and 0 otherwise. Both chains are reversible w.r.t. the invariant distribution $\pi = \tilde{\pi}$ with density $q(x) = \tilde{q}(x) = \frac{1}{2a}\mathbf{1}_S(x)$, uniform over S by Example 2.3.1(b). Again, for simplicity, let b be an integer. As in (a), we consider the set of uniform paths η^b_{xy} which satisfies the $(P, \tilde{P})$ regularity condition (Definition 3.1). For any $(u, v) \in \eta^b_{xy}$,

$$\tilde{q}(x)\tilde{p}_x(y) = \frac{1}{2a}\frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2}} = q(u)\frac{1}{\sqrt{2\pi}} e^{-\frac{b^2(v-u)^2}{2}} = \frac{1}{b} q(u)p_u(v) = \frac{1}{b}Q(u,v).$$

Hence,

$$A_0 \leq \operatorname{esssup}_{(u,v)\in E}\{Q(u,v)^{-1} b(b\frac{1}{b}Q(u,v)b)\} = b^2.$$

By Theorem 3.2 and then 3.1,

$$\lambda_1(P) \leq 1 - \frac{1}{b^2}(1 - \lambda_1(\tilde{P})).$$

In this example, we can try uniform paths with different number of steps. By choosing b steps, the constant A_0 is easy to calculate.

(c) We consider chains similar to (a), (b). Redefine $\tilde{p}_x(y) = \frac{1}{2}$ and $p_x(y) = 1 - |y-x|$ on $[x-1, x+1] \cap S$. The chains are also both reversible w.r.t. $\pi = \tilde{\pi}$ with density $q(x) = \tilde{q}(x) = \frac{1}{2a}\mathbf{1}_S(x)$. Apply the uniform paths η^b_{xy} and we get

$$\begin{aligned}
A_0(b) &\leq \text{esssup}_{(u,v)\in E}\{\frac{1}{1-|v-u|}b(\frac{1}{2}b)\} \\
&= \sup_{0\leq a\leq \frac{1}{b}} \frac{b^2}{2(1-a)} \\
&= \frac{b^3}{2(b-1)}.
\end{aligned}$$

Minimizing over integer b, we have $b = 2$ and $A_0(2) = 4$. Hence,

$$\lambda_1(P) \leq 1 - \frac{1}{4}(1 - \lambda_1(\tilde{P})).$$

Example 3.2.2. *Comparison of different one-dimensional chains with unbounded state space.*

Consider two Markov chains on the real line, both with invariant distribution with standard normal p.d.f.: $q(x) = \tilde{q}(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. The proposal distributions of the chains from a point x are uniform on $[x - \frac{1}{b}, x + \frac{1}{b}]$ and $[x-1, x+1]$ respectively. Formally, by the Metropolis algorithm (see e.g. Tierney [42]), we can define the two Markov chains as in Section 3, where $\tilde{p}_x(y) = \frac{1}{2}\min\{\frac{q(y)}{q(x)}, 1\}$ on $[x-1, x+1]$ and 0 otherwise, and $p_x(y) = \frac{b}{2}\min\{\frac{q(y)}{q(x)}, 1\}$ on $[x - \frac{1}{b}, x + \frac{1}{b}]$ and 0 otherwise. Then both chains are reversible w.r.t. the target density $q = \tilde{q}$. For simplicity, let b be an integer and apply the uniform paths η_{xy}^b. Simplifying the geometric constant, we have

$$A_0 = \operatorname{esssup}_{(u,v)\in E} \frac{b \sum_{\eta_{xy}^b \ni (u,v)} \min\{e^{-\frac{x^2}{2}}, e^{-\frac{y^2}{2}}\}}{\min\{e^{-\frac{u^2}{2}}, e^{-\frac{v^2}{2}}\}}$$

Observe that for any $(u,v) \in \eta_{xy}^b$, $\min\{e^{-\frac{x^2}{2}}, e^{-\frac{y^2}{2}}\} \leq \min\{e^{-\frac{u^2}{2}}, e^{-\frac{v^2}{2}}\}$. So, $A_0 \leq b^2$. By Theorem 3.2 and then 3.1, we have

$$\lambda_1(P) \leq 1 - \frac{1}{b^2}(1 - \lambda_1(\tilde{P})).$$

This is our first example on an unbounded state space. Even though we fail to bound the convergence rate of both chains with the geometric method from Chapter 2, we can still make a comparison with the geometric constants in Chapter 3.

Chapter 4

Geometric Bounds: Continuous-Time Case

4.1 L^2 Convergence to π in the Continuous-Time Case

In this chapter, we 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 (1.1) and (1.2) from the discrete-time case, P^t induces a positivity-preserving linear contraction on $L^2(\pi)$ by

$$(P^t f)(x) = \int f(y) P^t(x, dy) \tag{4.1}$$

and acts to the left on measures, so that

$$\mu P^t(A) = \int P^t(x, A)\mu(dx). \tag{4.2}$$

A Markov process is reversible if

$$\pi(dx)P^t(x, dy) = \pi(dy)P^t(y, dx) \tag{4.3}$$

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.

There is a fact that we did not use in the discrete-time case but will play a significant role in Section 4.3. 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 ([24] Lemma 4.1) proved the following.

Proposition 4.1. *Let $\mathcal{F}_0$ be a dense sub-family of $\mathcal{F}$. Then*

$$k = \inf_{\substack{A \in \mathcal{F} \\ 0 < \pi(A) < 1}} k(A) = \inf_{\substack{A \in \mathcal{F}_0 \\ 0 < \pi(A) < 1}} k(A) \tag{4.4}$$

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. Based on Proposition 2.1, we have the following result for a continuous-time Markov process:

Proposition 4.2. *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*

$$\|\mu P^t - \pi\|_2 \leq \|\mu - \pi\|_2 \rho_{\frac{t}{n}}^n, \tag{4.5}$$

where

$$\rho_{\frac{t}{n}} = \max\{|\lambda_0(P^{\frac{t}{n}})|, |\lambda_1(P^{\frac{t}{n}})|\}. \tag{4.6}$$

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 $\rho_{\frac{t}{n}} = \max\{|\lambda_0(P^{\frac{t}{n}})|, |\lambda_1(P^{\frac{t}{n}})|\}$. So, by Proposition 2.1,

$$\|\mu P^t - \pi\|_2 = \|\mu(P^{\frac{t}{n}})^n - \pi\|_2 \leq \|\mu - \pi\|_2 \rho_{\frac{t}{n}}^n.$$

Q.E.D.

In the first remark after Proposition 2.1, we mentioned that 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 $P^{t/2}$, $0 \leq \lambda_0(P^t)$ (which is $\leq \lambda_1(P^t)$), which implies that

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

We define a useful geometric constant

$$c = \limsup_{t \to 0} \frac{1}{t} \lambda_0(L^t). \tag{4.7}$$

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

Lemma 4.3. *Under the condition in Proposition 4.2,*

$$\|\mu P^t - \pi\|_2 \leq \|\mu - \pi\|_2 e^{-ct}. \tag{4.8}$$

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,

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

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

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

Hence, by Proposition 4.2 and taking $n\to\infty$, we have

$$\begin{aligned}\|\mu P^t-\pi\|_2 &\leq \|\mu-\pi\|_2\lim_{n\to\infty}\rho_{\frac{t}{n}}^n\\ &\leq \|\mu-\pi\|_2\lim_{n\to\infty}\left(1-\frac{\sup_{k\geq n}\{k\lambda_0(L^{\frac{t}{k}})\}}{n}\right)^n\\ &= \|\mu-\pi\|_2e^{-ct}.\end{aligned}$$

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 next section, we shall derive some lower bounds for c by applying the path bounds obtained in Chapter 2 and the Cheeger's inequality.

4.2 From Discrete-Time to Continuous-Time

In view of the previous section, 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 (4.7). Our first approach is to treat P^t as a discrete-time Markov chain, apply the path bounds in Section 2.2 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 $t \to 0$.

4.2.1 Limit Path Bounds

In the first approach, we shall apply the path bounds from Section 2.2. 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

$$\pi\text{-esssup}\, P^t(x, \{x\}^c) \leq M_t. \tag{4.9}$$

In order to apply the path bounds in Section 2.2, 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 2.2) and K_t (See Theorem 2.3) respectively. In summary, we have the following theorem:

Theorem 4.4. *Let c be the geometric constant defined in (4.7) and M_t satisfies (4.9). (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 2.1) for the Markov process described above. Then for any $\epsilon \in \mathbb{R}$,*

$$c \geq \frac{1}{\liminf_{t\to 0} k_{\epsilon,t} t}. \tag{4.10}$$

(b) If the second regularity condition (Definition 2.2) is satisfied instead, we have

$$c \geq \frac{1}{8 \liminf_{t\to 0} K_t^2 M_t t}. \tag{4.11}$$

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 2.2. 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 (4.10).

(b) Similar to (a). Apply Theorem 2.3. Q.E.D.

Remark. We can also generalize Corollary 2.4 and Corollary 2.5 in a similar fashion for specific types of paths. We shall not go into the detail here.

4.2.2 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 4.4(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.

$$k_0 = \limsup_{t \to 0} \frac{k_t}{\sqrt{M_t t}} \text{ and } h_0 = \limsup_{t \to 0} \frac{h_t}{\sqrt{M_t t}} \tag{4.12}$$

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 4.1 for formal definition) and M_t satisfies (4.9). The lower bounds on c obtained in the next theorem are called *limit Cheeger's bounds.*

Theorem 4.5. *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 (4.7) satisfies*

$$c \geq \frac{k_0^2}{8} \text{ and } c \geq \frac{h_0^2}{2}. \tag{4.13}$$

Proof. For each fixed t, P^t is a reversible Markov chain with respect to π.

So, from the Cheeger's inequality (See Section 1.2),

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

Dividing by t, we have

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

Take $\limsup_{t\to 0}$ and we obtain (4.13). Q.E.D.

Theorem 4.5 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

$$k_t = \inf_{\substack{A\in\mathcal{F} \\ 0<\pi(A)<1}} k_t(A) \text{ and } h_t = \inf_{\substack{A\in\mathcal{F} \\ 0<\pi(A)\leq\frac{1}{2}}} h_t(A). \tag{4.14}$$

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

$$\limsup_{x\to x_0}(\inf_y f(x,y)) \leq \inf_y(\limsup_{x\to x_0} f(x,y)). \tag{4.15}$$

Proof. Note that for any y_0,

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

Taking infimum over y_0 on both sides, we have

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

Q.E.D.

Lemma 4.6 suggests that by interchanging the limit and the infimum, we have

$$k_0 \leq \inf_A (\limsup_{t \to 0} \frac{k_t(A)}{\sqrt{M_t t}}) \text{ and } h_0 \leq \inf_A (\limsup_{t \to 0} \frac{h_t(A)}{\sqrt{M_t t}}). \tag{4.16}$$

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 4.3 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

$$k_0 = \inf_A (\limsup_{t \to 0} \frac{k_t(A)}{\sqrt{M_t t}}) \tag{4.17}$$

and the *h-condition* is satisfied if

$$h_0 = \inf_A (\limsup_{t \to 0} \frac{h_t(A)}{\sqrt{M_t t}}). \tag{4.18}$$

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.

4.3 One-Dimensional Diffusion in a Bounded Interval

In this section, we consider two different one-dimensional diffusions in a bounded interval. In Section 4.3.1, we shall give a rough lower bound of c defined in (4.7) for the modulo diffusion by comparing them to a simpler Markov chain which satisfies the h-condition (4.18). In Section 4.3.2, we shall work with the original process and explore the possibility of proving the h-condition directly to get an improved bound. In Section 4.3.3, we use the same technique as in Section 4.3.1 to study a diffusion with reflecting boundary.

4.3.1 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

$$P_1^t(x, dy) = p_1^t(x, y)dy = \sum_{m=-\infty}^{\infty} (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(md + y - x)^2}{2\sigma^2 t}\right\} dy. \quad (4.19)$$

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

$$D(x, y) = D(|y - x|) = \min\{|y - x|, d - |y - x|\}. \tag{4.20}$$

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

$$P_1^t(x, dy) = \sum_{m=-\infty}^{\infty} r_m^t(D(x, y))dy \tag{4.21}$$

where

$$r_m^t(u) = (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(md + u)^2}{2\sigma^2 t}\right\}. \tag{4.22}$$

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

$$D(A, B) = \inf\{|y - x|, d - |y - x| : x \in A, y \in B\}. \tag{4.23}$$

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

$$\lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^b \int_b^c (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\} dydx = \frac{\sigma}{\sqrt{2\pi}}. \tag{4.24}$$

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

$$\lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^{b_1} \int_{b_2}^c (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\} dydx = 0. \tag{4.25}$$

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

$$\begin{aligned}\lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^b \int_b^c r_0^t(D(x,y))dydx &= \lim_{t\to 0} \frac{1}{\sqrt{t}} \int_b^c \int_a^b r_0^t(D(x,y))dydx \\ &= \frac{\sigma}{\sqrt{2\pi}}. \end{aligned} \tag{4.26}$$

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

$$\begin{aligned}\lim_{t\to 0} \frac{1}{\sqrt{t}} \int_a^b \int_b^c r_0^t(D(x,y))dydx &= \lim_{t\to 0} \frac{1}{\sqrt{t}} \int_b^c \int_a^b r_0^t(D(x,y))dydx \\ &= \frac{2\sigma}{\sqrt{2\pi}}. \end{aligned} \tag{4.27}$$

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

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

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 (4.17) or the h-condition (4.18) 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,

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

which gives the k-condition. The same can be done for the h-condition. In fact, Jarner [21] 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 4.8 (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*

$$h = h([0, \frac{d}{2})). \tag{4.28}$$

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 4.8 on the Cheeger's h-constant directly. However, we observe that if we define

$$\bar{p}^t(x,y) \equiv (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2 t}\right\} dy, \tag{4.29}$$

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 4.9. *For the modulo diffusion defined in this section, the geometric constant c defined in (4.7) satisfies*

$$c \geq \frac{h_0^2}{2} \geq \frac{\sigma^2}{d^2 \pi}. \tag{4.30}$$

Proof. Note that

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

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 4.8, $\bar{h}_t = \bar{h}_t(A_0)$ and so

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

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 4.7(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.

4.3.2 On Improving the Lower Bound

As discussed in the remark after Theorem 4.9, we want to prove the lower bounds of c without comparing the process with the corresponding Metropolis algorithm. Theorem 4.10 and Corollary 4.11 suggest that our intuition in the remark after Theorem 4.9 is well founded.

Theorem 4.10. *Consider the modulo diffusion described in Section 4.3.1. 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,*

$$\lim_{t \to 0} \frac{k_t(A)}{\sqrt{t}} = \frac{2l\sigma}{d\sqrt{2\pi} \cdot \pi(A)\pi(A^c)} \tag{4.31}$$

and

$$\lim_{t \to 0} \frac{h_t(A)}{\sqrt{t}} = \frac{2l\sigma}{d\sqrt{2\pi} \cdot \pi(A)}. \tag{4.32}$$

Proof. We shall only prove (4.31). (4.32) follows from (4.31) 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 (4.31). 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. $i = j$.

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

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,

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

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

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

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

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

where

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

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 4.7(c),

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

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 4.7(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 4.3.3.

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 4.18.

Corollary 4.11. *Suppose that either the k-condition (4.17) or the h-condition (4.18) is satisfied. Then the geometric constant c defined in (4.7) satisfies*

$$c \geq \frac{4\sigma^2}{d^2\pi}. \tag{4.33}$$

Proof. Since $S = [0, d)$, by Proposition 4.1, 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 4.5,

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

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 4.5 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 4.11 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 4.9. 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 4.10, 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 4.12.

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,

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

which gives the k-condition. This approach is similar to the one by Jarner [21]. Recall that the conditions of Theorem 4.8 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 4.13. *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* $a \leq b \leq \frac{d}{2}$,

$$h([0, a)) \geq h([0, b)). \tag{4.34}$$

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

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

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 $0 \leq i_1 \neq i_2 \leq 2n$ 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,

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

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 4.15.

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 conjecture from graph theory is a simplified discrete version of this distance concept.

Conjecture 4.14. *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,

$$\begin{aligned} &|\{(x,y) \in A \times A^c : D(x,y) \le j\}| \\ &\ge |\{(x,y) \in \{1,...k\} \times \{k+1,...n\} : D(x,y) \le j\}|. \end{aligned} \tag{4.35}$$

Remarks. 1. This conjecture has JUST been proved by Alfred B. Lehman, Department of Computer Science, University of Toronto. We shall assume this conjecture is true for the rest of the section.

2. 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 4.8.

Lemma 4.15. *Assume Conjecture 4.14 is true. 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* $\pi(A) = \frac{a}{d} \le \frac{1}{2}$. *Then* $h(A) \ge h([0, a))$.

This result is very similar to Theorem 6.16 of Jarner [21]. 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,

$$\int_{-\infty}^{x} d\nu_1 \geq \int_{-\infty}^{x} d\nu_2. \tag{4.36}$$

We can further show that

$$\nu_1 \leq \nu_2 \text{ iff } \int f d\nu_1 \geq \int f d\nu_2 \tag{4.37}$$

for all decreasing measurable functions $f \geq 0$ (see Lemma 6.8 of Jarner [21]).

With this result and Conjecture 4.14, we can give a sketch of the proof.

Proof of Lemma 4.15. 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 Conjecture 4.14 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 (4.37),

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

Q.E.D.

Theorem 4.16. *Assume Conjecture 4.14 is true. Consider a discrete-time random walk Markov chain that satisfies the conditions in Lemma 4.15. Then*

$$h = h([0, \frac{d}{2})). \tag{4.38}$$

Therefore, for the diffusion defined in this section, the h-condition (4.18) is satisfied and so the geometric constant c defined in (4.7) satisfies

$$c \geq \frac{4\sigma^2}{d^2\pi}. \tag{4.39}$$

Proof. By Lemma 4.15,

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

So, by Lemma 4.13,

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

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 4.11,

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

Remarks. 1. We proved (4.39) 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 4.10 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 (4.8) is $2\pi^2\sigma^2/d^2$ (See e.g. Bhattacharya and Waymire [6]).

4.3.3 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

$$\begin{aligned}
P_2^t(x,dy) &= p_2^t(x,y)dy \\
&= \sum_{m=-\infty}^{\infty} (2\pi\sigma^2 t)^{-\frac{1}{2}} \Bigg(\exp\left\{ -\frac{(2md+y-x)^2}{2\sigma^2 t} \right\} \\
&\qquad\qquad + \exp\left\{ -\frac{(2md+y+x)^2}{2\sigma^2 t} \right\} \Bigg) dy \\
&= \sum_{m=-\infty}^{\infty} \left(r_{2m}^t(D(x,y)) + s_{2m}^t(x,y) \right) dy \qquad (4.40)
\end{aligned}$$

where

$$s_m^t(u) = (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left\{ -\frac{(md+x+y)^2}{2\sigma^2 t} \right\}. \qquad (4.41)$$

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 4.17. *For the diffusion with reflecting boundaries defined in this section, the geometric constant c defined in (4.7) satisfies*

$$c \geq \frac{h_0^2}{2} \geq \frac{\sigma^2}{d^2\pi}. \qquad (4.42)$$

Proof. Replace $p_1^t(x,y)$ by $p_2^t(x,y)$ in the proof of Theorem 4.9. Q.E.D.

One may think that as in Section 4.3.3, we could find an improved bound for the diffusion. However, the following results show that the bound in Theorem 4.17 is as good as the one derived directly from the process.

Theorem 4.18. *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,*

$$\lim_{t\to 0} \frac{k_t(A)}{\sqrt{t}} = \frac{(2l-1)\sigma}{d\sqrt{2\pi}\cdot \pi(A)\pi(A^c)} \tag{4.43}$$

and

$$\lim_{t\to 0} \frac{h_t(A)}{\sqrt{t}} = \frac{(2l-1)\sigma}{d\sqrt{2\pi}\cdot \pi(A)}. \tag{4.44}$$

Proof. We shall not give a detail proof as it is very similar to Theorem 4.10. There are two points that we have to clarify in order to get $(2l-1)$ instead of $2l$ in the corresponding equations from Theorem 4.10. 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 4.10, 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 4.19. *The second inequality in Theorem 4.17 is actually an equality, i.e.,*

$$c \geq \frac{h_0^2}{2} = \frac{\sigma^2}{d^2\pi}. \tag{4.45}$$

Proof. It suffices to show the other inequality

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

By Lemma 4.6 and Theorem 4.18 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 4.18. 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.

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

In Section 4.2, 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 4.4(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 4.8 and Lemma 4.15. It is not clear how to generalize the results to higher dimension.

In this section, we shall apply Theorem 4.4 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 4.3.3 by setting $n = 1$.

Let $P^t_{(n)}$ be the n-dimensional diffusion on $S = [0, d]^n \subset \mathbb{R}^n$ with reflecting boundaries and 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:

$$P^t_{(n)}(x,dy) \geq (2\pi\sigma^2 t)^{-\frac{n}{2}} \exp\left\{-\frac{||y-x||^2}{2\sigma^2 t}\right\} dy \tag{4.46}$$

where

$$||y-x||^2 = \sum_{i=1}^{n}(y_i - x_i)^2. \tag{4.47}$$

For each $t > 0$, we have to find a set of paths Γ_t that satisfies the path conditions. As in a similar one-dimensional case in Example 2.3.1(a), we apply the uniform paths $\eta^{b_t}_{xy}$ where the number of steps b_t in each path is to be determined later. By going through similar calculations, we have

$$k_{\epsilon,t} \leq (2\pi\sigma^2 t)^{\frac{n}{2}} \exp\left\{\frac{nd^2}{2\sigma^2 t b_t^2}\right\} b_t^{n+2} d^{-n} \tag{4.48}$$

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 4.4(a), we have the following result:

Theorem 4.20. *For the n-dimensional diffusion described in this section, the geometric constant c defined in (4.7) satisfies*

$$c \geq \frac{\sigma^2}{d^2}(2\pi)^{-\frac{n}{2}} e^{-\frac{n+2}{2}} \left(\frac{n}{(n+2)}\right)^{-\frac{n+2}{2}}. \tag{4.49}$$

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

Corollary 4.21. *For the one-dimensional diffusion with reflecting boundaries described in Section 4.3.3, the geometric constant c defined in (4.7) satisfies*

$$c \geq \frac{3^{\frac{3}{2}}\sigma^2}{d^2\sqrt{2\pi}e^{\frac{3}{2}}}. \tag{4.50}$$

Remarks. 1. Comparing this bound with the limit Cheeger's bound in (4.45), 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 Chapter 2, it seems that the path bound k_ϵ defined in (2.6) is always better than the Cheeger's path bound K defined in (2.8). 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.

2. As in Remark 2 after Theorem 4.16, it is known that the best possible c that satisfy equation (4.8) is $\pi^2\sigma^2/(2d^2)$ (See e.g. Bhattacharya and Waymire [6]).

Chapter 5

Conclusion and Future Research

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

There are three main results in this thesis. The first main result is the introduction of path bounds in Chapter 2. The second main result is the comparison of the spectral radii of different Markov chains. The third main result is the introduction of the limit path bounds and the limit Cheeger's bounds for some specific Markov processes in Chapter 4.

The first main result is focused mainly on discrete-time Markov chains. Although the result is also applicable to continuous-time Markov jump processes by simple modification, we did not extend the result to jump processes with killing. It should be worth considering a more general setting as these continuous processes are more useful in approximating more general continuous Markov processes such as diffusions. In fact, Roberts and Rosenthal [32] used a jump process to approximate the Langevin diffusion.

The second main result is only for comparison of discrete-time Markov chains. With the techniques developed in Chapter 4, 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 4.3 when we compared the diffusion with a Metropolis algorithm to calculate the limit Cheeger's bound.

The last main result is on the convergence rates of some continuous-time Markov processes. We introduced two techniques to compute geometric bounds. The first technique is to extend the definition of paths. In Section 4.4, 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.

We also notice that in the examples of Chapter 2, we always found that the first path bound k_ϵ is better that the second path bound K (for the Cheeger's constant). Is this always true in general? More surprisingly, in Section 4.4, we showed that the first limit path bound is also better than the limit Cheeger's bound (without the path approximation). So, is the first limit path bound always better than the limit Cheeger's bound? These are important questions which remain unanswered.

The modulo diffusion studied in Section 4.3.1 and Section 4.3.2 induces more possible research topics. First, Conjecture 4.14 is still unsolved. This problem is similar to a graph bipartitioning problem for a circulant graph. Some special cases have been proved ($k \leq 2$) but it is far from being complete. Second, 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. Third, 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 [7]) to get around the technicality.

So far, we are unable to find a positive lower bound on the spectral radius $\lambda_0(L)$ of a Markov chain over an unbounded domain using the theorems in Chapter 2. In Example 3.2.2, we were only able to compare the spectral radii of two Markov chains over an unbounded domain, without knowing the corresponding bounds. One possible research direction is to find suitable paths (other than the uniform or linear paths) for an unbounded example so that the lower bound on the spectral radius is positive. Of course, the Jacobian in the geometric constant will be more complicated and make the computation harder.

Bibliography

[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] R. Bañuelos and T. Carroll, Brownian motion and the fundamental frequency of a drum, Duke Math. J. 75, (1994) 575-602.

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

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

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

[7] 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

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

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

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

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

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

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

[14] S.N. Ethier and T.G. Kurtz, Markov processes: characterization and convergence (John Wiley & Sons, New York, 1986).

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

[16] 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.

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

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

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

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

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

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

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

[24] 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.

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

[26] A. Mira, Ordering, slicing and splitting Monte Carlo Markov chains, Ph.D. dissertation, School of Statistics, University of Minnesota (1998).

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

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

[29] M. Reed and B. Simon, Methods of mathematical physics IV (Academic Press 1978).

[30] 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.

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

[32] 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.

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

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

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

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

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

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

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

[40] 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.

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

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

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

[44] 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.