國 立 交 通 大 學

應用數學系

博 士 論 文

有限馬可夫鏈的對數索柏列夫常數

The Logarithmic Sobolev Constant on Finite Markov Chains

研 究 生：陳冠宇

指導教授：許元春

中 華 民 國 九 十 五 年 八 月

有限馬可夫鏈的對數索柏列夫常數
The Logarithmic Sobolev Constant on Finite Markov Chains

研 究 生：陳冠宇　　Student：Guan-Yu Chen

指導教授：許元春　　Advisor：Yuan-Chung Sheu

國 立 交 通 大 學
應 用 數 學 系
博 士 論 文

A Dissertation
Submitted to Department of Applied Mathematics
College of Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in

Applied Mathematics

August 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年八月

有限馬可夫鏈的對數索柏列夫常數

學生：陳冠宇 指導教授：許元春 教授

國立交通大學應用數學系博士班

摘 要

一副撲克牌要洗牌幾次其機率分佈才會接近均勻分佈。數學上，這個問題是屬於有限馬可夫鏈收斂速度的計量分析。在其他的領域裡也有相似的問題，其中包含了統計物理學、計算機科學和生物學。在這篇論文裡，我們討論l^P距離和超皺縮性之間的關係，並介紹兩個與收斂速度相關的常數—譜間隙和對數索柏列夫常數。

我們的目標是要準確地計算出對數索柏列夫常數，其中最主要的結果就是在循環體上簡單隨機運動的對數索柏列夫常數。另外，透過馬可夫鏈的崩塌，我們也得到兩種在直線上隨機運動的對數索柏列夫常數。最後，我們考慮狀態空間為三個點的馬可夫鏈並求得部分的對數索柏列夫常數。

The Logarithmic Sobolev Constant on Finite Markov Chains

Student：Guan-Yu Chen Advisors：Dr. Yuan-Chung Sheu

Department of Applied Mathematics
National Chiao Tung University

ABSTRACT

How many times a deck of cards needed to be shuffled in order to get close to the uniform distribution. Mathematically, this question falls in the realm of the quantitative study of the convergence of finite Markov chains. Similar convergence rate questions for finite Markov chains are important in many fields including statistical physics, computer science, biology and more. In this dissertation, we discuss the relation between the l^P-distance and the hypercontractivity. To bound the convergence rate, we introduced two well-known constants, the spectral gap and the logarithmic Sobolev constant.

Our goal is to compute the logarithmic Sobolev constant for nontrivial models. Diverse tricks in use include the comparison technique and the collapse of Markov chains. One of the main work concerns the simple random walk on the n cycle. For n even, the obtained logarithmic Sobolev constant is equal to half the spectral gap. For n odd, the ratio between the logarithmic Sobolev constant and the spectral gap is not uniform.

Ideally, if the collapse of a chain preserves the spectral gap and the original chain has the logarithmic Sobolev constant equal to a half of the spectral gap, then the logarithmic Sobolev constant of the collapsed chain is known and equal to half the spectral gap. We successfully apply this idea to collapsing even cycles to two different sticks. Throughout this thesis, examples are introduced to illustrate theoretical results. In the last section, we study some three-point Markov chains with introduced techniques.

誌　　謝

首先我要感謝我的指導老師、同時也是我的精神導師許元春教授。在他熱情和耐心的指導下，我在學術上的專業知識和論文寫作都有明顯的進步。其次我要感謝每位博士學位口試委員在口試時給予的寶貴意見，以及交大應用數學系每位老師的指導與教誨。

另外我要感謝長久以來一直支持我的親朋好友。因為他們的鼓勵，讓我有更大的動力去完成博士學位。其中我要特別感謝我的父母親，他們在經濟上的自主讓我能無後顧之憂的專心致力於學業。我要感謝我的兩個弟弟，他們孜孜不倦的向學精神是我在研究遭遇瓶頸時的最大鼓勵。最後，我要感謝敏玉--我的太太。因為她的陪伴與分享，我得以在平凡的日子裡找到自己的興趣並預見未來努力的方向。

目　　　錄

Chapter 1

Introduction

How many times a deck of cards needed to be shuffled in order to get close to the uniform distribution. Mathematically, this question falls in the realm of the quantitative study of the convergence of finite Markov chains. Similar convergence rate questions for finite Markov chains are important in many fields including statistical physics, computer science, biology and more. Many questions posted in these fields are to estimate the average of a function f defined on a finite set Ω with respect to a probability measure π on Ω. From the view point of Markov Chain Monte Carlo method, this is achieved by simulating a Markov chains with limiting distribution π and selecting a state at a random time T as a random sample. Knowing the qualitative behavior of convergence is not enough to determine the sampling time T. A quantitative understanding of the mixing time is essential for theoretical results. In practice, various heuristics are used to choose T.

Diverse techniques have been introduced to estimate the mixing time. Coupling and strong uniform time are discussed by Aldous and Diaconis in [1, 2]. Jerrum and Sinclair use conductance to bound mixing time in [17]. Application of representation theory appears in [8] and Diaconis and Saloff-Coste used comparison techniques in [9, 10]. For lower bound, important techniques are described in [7] and in more recent work of Wilson [27].

In this dissertation, we introduce two well-known constants, the spectral gap and the logarithmic Sobolev constant. Applying fundamental result in calculus and linear algebra, we are able to determine both constants for some specific models.

1.1 Preliminaries

Let $\mathcal{X}$ be a finite set. A discrete time Markov chain is a sequence of $\mathcal{X}$-valued random variables $(X_n)_0^\infty$ satisfying

$$\mathbb{P}\{X_{n+1} = x_{n+1} | X_i = x_i, \forall 0 \le i \le n\} = \mathbb{P}\{X_{n+1} = x_{n+1} | X_n = x_n\}$$

for all $x_i \in \mathcal{X}$ with $0 \le i \le n$ and $n \ge 0$. A Markov chain is *time homogeneous* if the quantity in the right hand side of the above identity is independent of n. In this case, such a Markov chains is specified by the initial distribution (the distribution of X_0) and the one-step transition kernel $K : \mathcal{X} \times \mathcal{X} \to [0, 1]$(also called the Markov kernel) which is defined by

$$\forall x, y \in \mathcal{X}, \quad K(x, y) = \mathbb{P}\{X_{n+1} = y | X_n = x\}.$$

An immediate observation on the Markov kernel K is that $\sum_{y \in \mathcal{X}} K(x, y) = 1$ for all $x \in \mathcal{X}$. Throughout this thesis, all Markov chains are assumed to be time homogeneous. For any Markov chain $(X_n)_0^\infty$ with transition matrix K and initial distribution μ, that is, $\mathbb{P}\{X_0 = x\} = \mu(x)$ for all $x \in \mathcal{X}$, the distribution of X_n is given by

$$\forall x \in \mathcal{X}, \quad \mathbb{P}\{X_n = x\} = (\mu K^n)(x) = \sum_{y \in \mathcal{X}} \mu(y) K^n(y, x),$$

where K^n is a matrix defined iteratively by

$$\forall x, y \in \mathcal{X}, \quad K^n(x, y) = \sum_{z \in \mathcal{X}} K^{n-1}(x, z) K(z, y).$$

In a similar way, one may also consider a continuous-time Markov process. Here we consider only the following specific type. For any Markov kernel K, let $(X_t)_{t \ge 0}$ be a Markov process with infinitesimal generator $K - I$(the Q-matrix defined in [19]). One way to realize this process is to stay in a state for an exponential(1) time and then move to another state according to the Markov kernel

K. In other words, the law of X_t is determined by the initial distribution μ and the continuous-time semigroup $H_t = e^{-t(I-K)}$(a matrix defined formally by $H_t(x,y) = e^{-t}\sum_{n=0}^{\infty}\frac{t^n}{n!}K^n(x,y)$ for $x, y \in \mathcal{X}$ and $t \geq 0$, where $K^0 = I$) through the following formula

$$\forall x \in \mathcal{X},\ t \geq 0, \quad \mathbb{P}\{X_t = x\} = \sum_{y \in \mathcal{X}} \mu(y) H_t(y,x).$$

Note that if $(Y_n)_0^\infty$ is a Markov chain with transition matrix K and N_t is a Poisson process with intensity 1 and independent of $(Y_n)_0^\infty$, then the Markov process $(X_t)_{t\geq 0}$ with infinitesimal generator $K - I$ satisfies $X_t \stackrel{d}{=} Y_{N_t}$(in distribution) for $t \geq 0$. This is because

$$\forall x, y \in \mathcal{X}, \quad H_t(x,y) = \mathrm{E}[K^{N_t}(x,y)] = \mathbb{P}\{Y_{N_t} = y | Y_0 = x\}.$$

For any finite Markov process $(Y_t)_{t\geq 0}$, we may find a constant $c > 0$, a Markov chain $(X_n)_1^\infty$ and a Poisson(1) process independent of $(X_n)_1^\infty$ such that $Y_t = X_{N_{ct}}$ in distribution, or equivalently

$$\mathbb{P}\{Y_t = y | Y_0 = x\} = e^{-ct(I-K)}(x,y), \quad \forall x, y \in \mathcal{X},$$

where K is the Markov kernel of $(X_n)_1^\infty$. To see the details, let Q be the infinitesimal generator of $(Y_t)_{t\geq 0}$, which is a $|\mathcal{X}| \times |\mathcal{X}|$ matrix satisfying

$$Q(x,y) \geq 0, \quad \forall x \neq y,\ x, y \in \mathcal{X},$$

and

$$\sum_{y \in \mathcal{X}} Q(x,y) = 0, \quad \forall x \in \mathcal{X}.$$

Then, for $t \geq 0$, the law of Y_t is given by

$$\mathbb{P}\{Y_t = y | Y_0 = x\} = e^{tQ}(x,y) = \sum_{n=0}^{\infty} \frac{t^n Q^n(x,y)}{n!}.$$

By letting $q = \max\{-Q(x,x) : x \in \mathcal{X}\}$, where we assume its positivity, and $K = q^{-1}Q + I$, one may check $K(x,y) \geq 0$ and $\sum_y K(x,y) = 1$ for all $x, y \in \mathcal{X}$. Then the distribution of Y_t starting from x can be expressed by

$$\mathbb{P}\{Y_t = y | Y_0 = x\} = e^{tQ}(x,y) = e^{-(tq)(I-K)}(x,y), \quad \forall x, y \in \mathcal{X}.$$

Another view point on the continuous-time semigroup H_t is the following. For any Markov kernel K, let $\mathcal{L} = \mathcal{L}_K$ be a linear operator on $\mathbb{R}^{|\mathcal{X}|}$ defined by

$$\forall x \in \mathcal{X}, \quad \mathcal{L}f(x) = (K - I)f(x) = \sum_{y \in \mathcal{X}} K(x,y)f(y) - f(x). \tag{1.1}$$

The operator $\mathcal{L}$ can be viewed intuitively as a Laplacian operator on $\mathcal{X}$. A direct computation shows that, for any real-valued function f on $\mathcal{X}$, the function $u(t,x) = H_t f(x)$ is a solution for the initial value problem of the discrete-version heat equation, i.e.,

$$\begin{cases} (\partial_t + \mathcal{L})u = 0 & u : \mathbb{R}_+ \times \mathcal{X} \to \mathbb{R} \\ u(0,x) = f(x) & \forall x \in \mathcal{X}. \end{cases}$$

For any Markov kernel K, a measure π on $\mathcal{X}$ is called *invariant*(with respect to K) if $\pi K = \pi$ or equivalently

$$\forall x \in \mathcal{X}, \quad \sum_{y \in \mathcal{X}} \pi(y)K(y,x) = \pi(x). \tag{1.2}$$

A measure π on $\mathcal{X}$ is called *reversible* if the following identity holds

$$\forall x, y \in \mathcal{X}, \quad \pi(x)K(x,y) = \pi(y)K(y,x).$$

In this case, K is said to be *reversible* with respect to π. From these definitions, it is obvious that a reversible measure is an invariant measure. Besides, if π is invariant(resp. reversible) with respect to K, then, for all $t \geq 0$, $\pi H_t = \pi$

or equivalently $\sum_{y\in\mathcal{X}} \pi(y)H_t(y,x) = \pi(x)$ for all $x \in \mathcal{X}$(resp. $\pi(x)H_t(x,y) = \pi(y)H_t(y,x)$ for all $x, y \in \mathcal{X}$).

Note that, for any Markov kernel K on $\mathcal{X}$, a constant vector on $\mathcal{X}$ is a right eigenvector of K with eigenvalue 1. This implies the existence of a real-valued function f on $\mathcal{X}$ satisfying $f = fK$, or equivalently $f(x) = \sum_y f(y)K(y,x)$ for all $x \in \mathcal{X}$. By the following computation,

$$\sum_{x\in\mathcal{X}} |f(x)| = \sum_{x\in\mathcal{X}} \left|\sum_{y\in\mathcal{X}} f(y)K(y,x)\right| \leq \sum_{x,y\in\mathcal{X}} |f(y)|K(y,x) = \sum_{y\in\mathcal{X}} |f(y)|,$$

one can find that $|f|$ is also a left eigenvector of K with eigenvalue 1. Hence, for any Markov kernel, there exists a probability measure π, which is invariant with respect to K. In that case, π is called a *stationary* distribution for K.

A Markov kernel K is called *irreducible* if, for any $x, y \in \mathcal{X}$, there exists $n = n(x,y)$ such that $K^n(x,y) > 0$. A state $x \in \mathcal{X}$ is called *aperiodic* if $K^n(x,x) > 0$ for sufficiently large n, and K is called aperiodic if all states are aperiodic. It is known that under the assumption of irreducibility of K, there exists a unique stationary distribution π. In particular, the distribution π is positive everywhere. In addition, if K is irreducible, then K is aperiodic if and only if $\mathcal{X}$ has an aperiodic state.

Proposition 1.1. *Let K be an irreducible Markov kernel on a finite set $\mathcal{X}$ with the stationary distribution π. Then*

$$\forall x, y \in \mathcal{X}, \quad \lim_{t\to\infty} H_t(x,y) = \pi(y).$$

If K is irreducible and aperiodic, then

$$\forall x, y \in \mathcal{X}, \quad \lim_{n\to\infty} K^n(x,y) = \pi(y).$$

Under mild assumptions —irreducibility for continuous-time Markov processes and irreducibility and aperiodicity for discrete-time Markov chains— Proposition 1.1 shows the qualitative result that Markov chains converge to their stationarity as time tends to infinity. If such a convergence happens, the Markov kernel is called *ergodic*.

Note that the irreducibility of a Markov chain is sufficient, by Proposition 1.1, but not necessary for the ergodicity. A counterexample for the necessity is to consider a Markov chain on a two point space $\{0,1\}$ whose kernel is given by

$$K(0,0)=1,\ K(0,1)=0,\ K(1,0)=1-p,\ K(1,1)=p,$$

where $p\in(0,1)$. In this example, K is not irreducible because $K^n(0,1)=0$ for $n\geq 0$. A few computations show that for $n\geq 1$ and $t>0$,

$$K^n=\begin{pmatrix} 1 & 0 \\ 1-p^n & p^n \end{pmatrix},\quad H_t=\begin{pmatrix} 1 & 0 \\ 1-e^{(p-1)t} & e^{(p-1)t} \end{pmatrix}.$$

By the above formulas, the distribution of the Markov chain starting from any fixed state converges to $(1,0)$.

Proposition 1.2. *Let K be a Markov kernel on a finite set $\mathcal{X}$ and π is a positive probability measure on $\mathcal{X}$. If, for all $x,y\in\mathcal{X}$,*

$$\lim_{t\to\infty} H_t(x,y)=\pi(y),$$

then K is irreducible. If the following holds

$$\lim_{n\to\infty} K^n(x,y)=\pi(y),\quad \forall x,y\in\mathcal{X},$$

then K is irreducible and aperiodic.

By Proposition 1.1 and 1.2, if the limiting distribution is assumed positive, then, in continuous-time cases, K is ergodic if and only if K is irreducible, whereas, in discrete-time cases, ergodicity is equivalent to irreducibility and aperiodicity.

In many cases, the state space $\mathcal{X}$ is equipped with a group structure and the Markov kernel K is driven by a probability measure p on $\mathcal{X}$ in the following way.

$$K(x,y) = p(x^{-1}y), \quad \forall x, y \in \mathcal{X}.$$

Let E be the support of p and, for $n \geq 1$, E^n denote the set

$$E^n = \{x_1 x_2 \cdots x_n : x_i \in E, \ \forall 1 \leq i \leq n\}.$$

In the above setting, it is clear that the irreducibility of K is equivalent to the existence of a positive integer n such that

$$\mathcal{X} = \bigcup_{i=1}^{n} E^i.$$

Under the assumption of irreducibility of K, the Markov kernel K is aperiodic if and only if there exists a positive integer n such that $\mathcal{X} = E^n$.

The following proposition characterizes the irreducibility and the aperiodicity of finite Markov chains introduced in the previous paragraph, which has been proved many times by many authors. See [25, 26] for references.

Proposition 1.3 (Proposition 2.3 in [24])**.** *Let $\mathcal{X}$ be a finite group and p be a probability measure on $\mathcal{X}$ with support $E = \{x \in \mathcal{X} : p(x) > 0\}$. Let K be a Markov kernel given by $K(x,y) = p(x^{-1}y)$ for $x, y \in \mathcal{X}$. Then*

(1) K is irreducible if and only if E generates $\mathcal{X}$, that is, any element of $\mathcal{X}$ can be expressed as a product of finitely many elements of E.

(2) Assume that K is irreducible. Then K is aperiodic if and only if E is not contained in a coset of any proper normal subgroup of $\mathcal{X}$.

In particular, if $\mathcal{X}$ is simple and K is irreducible on $\mathcal{X}$, then K is aperiodic.

To determine the reversibility of a Markov chain, by definition, one always needs to compute the stationary distribution first. In the following, we introduce a criterion to inspect the reversibility of a Markov chain without the computation of its stationary distribution.

Proposition 1.4. *Let K be an irreducible Markov kernel on a finite set $\mathcal{X}$ with stationary distribution π. Then (K,π) is reversible if and only if for any sequence $\{x_0,...,x_n\}$ with $x_0 = x_n$,*

$$\begin{aligned} &K(x_0,x_1)K(x_1,x_2)\cdots K(x_{n-1},x_n) \\ &\qquad\qquad = K(x_n,x_{n-1})K(x_{n-1},x_{n-2})\cdots K(x_1,x_0). \end{aligned} \tag{1.3}$$

Proof. Assume first that K is reversible with respect to π, that is, $\pi(x)K(x,y) = \pi(y)K(y,x)$ for all $x,y \in \mathcal{X}$. Let $\{x_0,...,x_n\}$ be a sequence with $x_0 = x_n$. Then

$$\prod_{i=0}^{n-1} K(x_i,x_{i+1}) = \prod_{i=0}^{n-1} \frac{\pi(x_{i+1})K(x_{i+1},x_i)}{\pi(x_i)} = \prod_{i=0}^{n-1} K(x_{i+1},x_i).$$

For the other direction, we assume that (1.3) holds for any sequence $\{x_0,...,x_n\}$ satisfying $x_0 = x_n$. This implies that for $x,y \in \mathcal{X}$ and $n \geq 1$,

$$\begin{aligned} &K^n(x,y)K(y,x) \\ &\quad = \sum_{x_1,\dots,x_{n-1}} K(x,x_1)\prod_{i=1}^{n-2} K(x_i,x_{i+1})K(x_{n-1},y)K(y,x) \\ &\quad = \sum_{x_1,\dots,x_{n-1}} K(x,y)K(y,x_{n-1})\prod_{i=1}^{n-2} K(x_{i+1},x_i)K(x_1,x) \\ &\quad = K(x,y)K^n(y,x). \end{aligned}$$

Applying the above identity to the expansion formula of the continuous-time semigroup, we get

$$H_t(x,y)K(y,x) = H_t(y,x)K(x,y), \quad \forall x,y \in \mathcal{X}.$$

Letting $t \to \infty$, the reversibility of K is then proved by Proposition 1.1. $\square$

The following is an application of the above proposition to random walks on finite trees.

Corollary 1.1. *Let K be an irreducible Markov kernel on a finite set $\mathcal{X}$ and $G = (\mathcal{X}, E)$ be an undirected graph induced from K whose vertex set is $\mathcal{X}$ and the edge set E is given by*

$$E = \{\{x, y\} \in \mathcal{X} \times \mathcal{X} : x \neq y,\ K(x, y) + K(y, x) > 0\}.$$

If G is a tree, then K is reversible.

Remark 1.1. In Corollary 1.1, the induced graph G is connected if and only if K is irreducible. In particular, if G is a tree, we have

$$\forall x, y \in \mathcal{X}, \quad K(x, y) > 0 \iff K(y, x) > 0.$$

Proof of Corollary 1.1. For convenience, any finite sequence of states in $\mathcal{X}$ is called a path. For any path $(x_0, ..., x_n)$, we let $\prod_{i=0}^{n-1} K(x_i, x_{i+1})$ denote its "weight". By the above remark, it suffices to prove the identity (1.3) with paths of positive weight. To show this fact, we define D to be the set of all paths in G and, for all $x, y \in \mathcal{X}$, let $f_{(x,y)}$ be a function on D defined by

$$\forall \gamma = (x_0, ..., x_n) \in D, \quad f_{(x,y)}(\gamma) = \sum_{i=0}^{n-1} \delta_{(x,y)}((x_i, x_{i+1})) - \delta_{(y,x)}((x_i, x_{i+1})).$$

Since G is a tree, it is obvious that, for all $x, y \in \mathcal{X}$, $f_{(x,y)}(\gamma) \in \{1, 0, -1\}$ for any positively weighted path $\gamma \in D$. If $x_0 = x_n$ is assumed further, then $f_{(x,y)}(\gamma) = 0$ for all $x, y \in \mathcal{X}$. This implies that, for such a path γ, the multiplicity of the directed edge (x, y) in γ is the same as that of (y, x). Thus, for all $x, y \in \mathcal{X}$, the multiplicity of (x, y) in γ is the same as that in the inverse path of γ, $(x_n, ..., x_0)$, and hence γ and $(x_n, ..., x_0)$ have the weight. $\square$

1.2 The ℓ^p-distance and the submultiplicativity

As a consequence of Proposition 1.1, irreducible and aperiodic Markov chains converge in distribution to their stationarity. From the view point of the quantitative study, one may arise the following question: *How fast the convergence can be?* To answer this question, we need to specify the function used to measure the distance between the law of a Markov chain and its stationary distribution. In this section, we will introduce some frequently used distances or functions for measuring and give some basic results.

Definition 1.1. Let μ and ν be probability measures on a set $\mathcal{X}$. The *total variation* distance between μ and ν is denoted and defined by

$$d_{\mathrm{TV}}(\mu,\nu)=\|\mu-\nu\|_{\mathrm{TV}}=\max_{A\subset\mathcal{X}}\{\mu(A)-\nu(A)\}.$$

Let π be a positive probability measure on $\mathcal{X}$. For $1\le p\le\infty$ and any (complex-valued) function f on $\mathcal{X}$, the $\ell^p(\pi)$*-norm*(or briefly the ℓ^p*-norm*) of f is defined by

$$\|f\|_p=\|f\|_{\ell^p(\pi)}=\begin{cases}\left(\sum_{x\in\mathcal{X}}|f(x)|^p\pi(x)\right)^{1/p} & \text{if } 1\le p<\infty\\ \max_{x\in\mathcal{X}}|f(x)| & \text{if } p=\infty\end{cases}.$$

Definition 1.2. Let μ, ν and π be finite probability measures on $\mathcal{X}$ and assume that π is positive everywhere. The $\ell^p(\pi)$*-distance*(or briefly the ℓ^p*-distance*) between μ and ν is defined to be

$$d_{\pi,p}(\mu,\nu)=\|f-g\|_{\ell^p(\pi)},$$

where f and g are densities of μ and ν with respect to π, that is, $\mu=f\pi$ and $\nu=g\pi$.

Remark 1.2. From the above two definitions, it is easy to see that, for any probability measures μ and ν,

$$\forall \pi > 0, \quad d_{\pi,1}(\mu, \nu) = 2d_{\mathrm{TV}}(\mu, \nu).$$

Let $(\mathcal{X}, \mu)$ be a measure space. It is well-known that, for $1 \leq p \leq \infty$, if f is ℓ^p-integrable, then

$$\|f\|_p = \sup_{\|g\|_q \leq 1} \int_{\mathcal{X}} f(x)g(x)d\mu(x),$$

where $p^{-1} + q^{-1} = 1$. By this fact, we may characterize the ℓ^p-distance in the following way.

Proposition 1.5. *Let π, μ, ν, f, g be the same as in Definition 1.2. Then, for $1 \leq p \leq \infty$,*

$$d_{\pi,p}(\mu, \nu) = \sup_{\|h\|_q \leq 1} \|(f - g)h\|_1,$$

where $p^{-1} + q^{-1} = 1$.

By Jensen's inequality, if π is a positive probability measure, then

$$\|f\|_p \leq \|f\|_q, \quad \forall 1 \leq p < q \leq \infty.$$

With this fact, we may compare the ℓ^p and ℓ^q distances.

Proposition 1.6. *Let π be a positive probability measure on $\mathcal{X}$. For any two probability measures μ, ν on $\mathcal{X}$, one has*

$$d_{\pi,p}(\mu, \nu) \leq d_{\pi,q}(\mu, \nu), \quad \forall 1 \leq p \leq q \leq \infty.$$

The following fact shows that, for fixed $1 \leq p \leq \infty$, the ℓ^p-distance of Markov chains to their stationarity decays exponentially.

Proposition 1.7. *Let K be an irreducible Markov kernel with stationary distribution π. Then, for $1 \le p \le \infty$, the maps*

$$n \mapsto \max_{x\in\mathcal{X}} d_{\pi,p}(K^n(x,\cdot),\pi) \quad \text{and} \quad t \mapsto \max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi)$$

are non-increasing and submultiplicative. In particular, if there exists $\beta > 0$ such that

$$\max_{x\in\mathcal{X}} d_{\pi,p}(K^m(x,\cdot),\pi) \le \beta \text{ (resp. } \max_{x\in\mathcal{X}} d_{\pi,p}(H_s(x,\cdot),\pi) \le \beta),$$

then for $n \ge m$(resp. $t \ge s$),

$$\max_{x\in\mathcal{X}} d_{\pi,p}(K^n(x,\cdot),\pi) \le \beta^{\lfloor n/m \rfloor} \text{ (resp. } \max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi) \le \beta^{\lfloor t/s \rfloor}).$$

Remark 1.3. By Proposition 1.7, if $\beta \in (0,1)$, then the exponential convergence of ℓ^p-distance has rate at least $m^{-1}\log(1/\beta)$ in discrete-time cases and rate $s^{-1}\log(1/\beta)$ in continuous-time cases.

For any Markov kernel K, we may associate it with a linear operator which is also denoted by K and defined by

$$Kf(x) = \sum_{y\in\mathcal{X}} K(x,y)f(y), \quad \forall x \in \mathcal{X},\ f \in \mathbb{C}^{|\mathcal{X}|}.$$

In a similar way, we can view H_t and π as linear operators on $\mathbb{C}^{|\mathcal{X}|}$ by setting

$$H_t f(x) = \sum_{y\in\mathcal{X}} H_t(x,y)f(y), \quad \pi(f) = \sum_{x\in\mathcal{X}} f(x)\pi(x).$$

To a standard usage, we let L^* denote the adjoint operator of L. The following proposition equates the maximum ℓ^p-distance and the operator norm of the associated linear operator.

Proposition 1.8. *Let K be an irreducible Markov operator with stationary distribution π. For $1 \le p \le \infty$,*

$$\max_{x\in\mathcal{X}} d_{\pi,p}(K^n(x,\cdot),\pi) = \|K^n - \pi\|_{q\to\infty}, \quad \text{for } n \ge 0,$$

and

$$\max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi) = \|H_t - \pi\|_{q\to\infty}, \quad \text{for } t \ge 0,$$

where $p^{-1}+q^{-1}=1$ *and for any linear operator* $L:\ell^r(\pi)\to\ell^s(\pi)$,

$$\|L\|_{r\to s} = \sup_{\|f\|_{\ell^r(\pi)}\le 1} \|Lf\|_{\ell^s(\pi)}. \tag{1.4}$$

Remark 1.4. By Jensen's inequality, for $1\le p\le\infty$, the linear operators K^n and H_t are contractions in ℓ^p, which means that

$$\|K^n\|_{p\to p}\le 1, \quad \|H_t\|_{p\to p}\le 1.$$

This fact implies

$$\|H_{t+s}-\pi\|_{p\to\infty} \le \|H_t\|_{p\to p}\|H_s-\pi\|_{p\to\infty} \le \|H_s-\pi\|_{p\to\infty}$$

and

$$\begin{aligned}\|H_{t+s}-\pi\|_{p\to\infty} &\le \|H_t-\pi\|_{p\to p}\|H_s-\pi\|_{p\to\infty}\\ &\le \|H_t-\pi\|_{p\to\infty}\|H_s-\pi\|_{p\to\infty}.\end{aligned}$$

By Proposition 1.8, these are the monotonicity and the submultiplicativity of the map $t\mapsto \max_x d_{\pi,q}(H_t(x,\cdot),\pi)$, where $p^{-1}+q^{-1}=1$. The same line of reasoning also applies for the discrete-time cases.

Besides the ℓ^p-distance, there are many other functions of interest in measuring how close a Markov chain to its stationarity. We end this section by introducing two other well-known functions which are frequently used in probability theory and statistical physics. Let π be a positive probability measure on a finite set $\mathcal{X}$. For any probability measure μ on $\mathcal{X}$, let h be the density of μ with respect to π. The *separation* of μ with respect to π is defined by

$$d_{\text{sep}}(\mu,\pi) = \max_{x\in\mathcal{X}}\{1-h(x)\},$$

and the (relative) *entropy* of μ with respect to π is defined by

$$d_{\mathrm{ent}}(\mu,\pi) = \mathrm{Ent}_\pi(\mu) = \sum_{x\in\mathcal{X}} [h(x)\log h(x)]\pi(x).$$

(Generally, the entropy of any nonnegative function f on $\mathcal{X}$ with respect to any measure π is defined by $\mathrm{Ent}_\pi(f) = \pi[f\log(f/\pi(f))]$.) The following proposition connects the ℓ^p-distance and the functions introduced above.

Proposition 1.9. *Let π and μ be probability measures on a finite set $\mathcal{X}$ and π is positive everywhere. Then one has*

$$\frac{1}{2}d_{\pi,1}(\mu,\pi) \le d_{\mathrm{sep}}(\mu,\pi) \le d_{\pi,\infty}(\mu,\pi)$$

and

$$\frac{1}{2}d_{\pi,1}(\mu,\pi)^2 \le d_{\mathrm{ent}}(\mu,\pi) \le \frac{1}{2}[d_{\pi,1}(\mu,\pi) + d_{\pi,2}(\mu,\pi)^2].$$

Proof. Let $h = \mu/\pi$. For the first part, it is obvious that $\max_x\{1-h(x)\} \le \|h-1\|_\infty$. For the lower bound, setting $A = \{x \in \mathcal{X} : h(x) < 1\}$ implies that

$$\max_{x\in\mathcal{X}}\{1-h(x)\} = \max_{x\in A}\{1-h(x)\} \ge \sum_{x\in A}\{1-h(x)\}\pi(x) = \|\mu-\pi\|_{\mathrm{TV}}.$$

For the second part, the upper bound is obtained by bounding the positive terms in the summation of the entropy though the following inequality.

$$\forall u > 0, \quad (1+u)\log(1+u) \le u + \frac{u^2}{2}.$$

For the lower bound, applying the fact

$$\forall u > 0, \quad \sqrt{3}|u-1| \le \sqrt{(4u+2)(u\log u - u + 1)}$$

and Cauchy-Schwartz inequality implies that

$$3\|h-1\|_1^2 \le \|4+2h\|_1\|h\log h - h + 1\|_1 = 6\pi(h\log h).$$

□

As in Proposition 1.7, if the distance between a Markov chain and its stationarity is measured by the maximum separation and the maximum entropy, then it is decreasing in time.

Proposition 1.10. *Let* $(\mathcal{X}, K, \pi)$ *be a finite Markov chain and* H_t *be the continuous-time semigroup associated to* K*. Then the following maps*

$$n \mapsto \max_{x\in\mathcal{X}} d_{\text{sep}}(K^n(x,\cdot),\pi), \quad t \mapsto \max_{x\in\mathcal{X}} d_{\text{sep}}(H_t(x,\cdot),\pi), \tag{1.5}$$

and

$$n \mapsto \max_{x\in\mathcal{X}} d_{\text{ent}}(K^n(x,\cdot),\pi), \quad t \mapsto \max_{x\in\mathcal{X}} d_{\text{ent}}(H_t(x,\cdot),\pi). \tag{1.6}$$

are non-increasing. Furthermore, the maps in (1.5) *are submultiplicative.*

Remark 1.5. By definition, if $(\mathcal{X}, K, \pi)$ an irreducible Markov chain, then

$$\max_{x\in\mathcal{X}} d_{\text{sep}}(K(x,\cdot),\pi) = \max_{x\in\mathcal{X}} d_{\text{sep}}(K^*(x,\cdot),\pi).$$

Proof of Proposition 1.10. Let A_1 and A_2 be two stochastic matrices satisfying $\pi = \pi A_1 = \pi A_2$ and set $A = A_1A_2$. For the first part, it suffices to prove that

$$\max_{x\in\mathcal{X}} d_{\text{sep}}(A(x,\cdot),\pi) \le \max_{x\in\mathcal{X}} d_{\text{sep}}(A_1(x,\cdot),\pi)$$

and

$$\max_{x\in\mathcal{X}} d_{\text{ent}}(A(x,\cdot),\pi) \le \max_{x\in\mathcal{X}} d_{\text{ent}}(A_1(x,\cdot),\pi).$$

The first inequality can be easily obtained by the following computation.

$$\begin{aligned}\left(1 - \frac{A(x,y)}{\pi(y)}\right) &= \sum_{z\in\mathcal{X}} \left(1 - \frac{A_1(x,z)}{\pi(z)}\right)\left(\frac{\pi(z)A_2(z,y)}{\pi(y)}\right) \\ &\le \max_{z\in\mathcal{X}} \left\{1 - \frac{A_1(x,z)}{\pi(z)}\right\}, \quad \forall x,y\in\mathcal{X}.\end{aligned}$$

For the second one, note that

$$\forall x,y\in\mathcal{X}, \quad \frac{A(x,y)}{\pi(y)} = \sum_{z\in\mathcal{X}} \left(\frac{A_1(x,z)}{\pi(z)}\right)\left(\frac{\pi(z)A_2(z,y)}{\pi(y)}\right).$$

Since the function $u \mapsto u \log u$ is convex, by Jensen's inequality, one has

$$\frac{A(x,y)}{\pi(y)} \log \left(\frac{A(x,y)}{\pi(y)} \right) \leq \sum_{z \in \mathcal{X}} \frac{A_1(x,z)}{\pi(z)} \log \left(\frac{A_1(x,z)}{\pi(z)} \right) \frac{\pi(z) A_2(z,y)}{\pi(y)}.$$

Multiplying $\pi(y)$ on both sides, summing up all entries y in $\mathcal{X}$ and taking the maximum with respect to x implies the desired inequality.

For the submultiplicativity of the maximum separation, let A_1, A_2, A be the same as in the previous paragraph. We prove this property by following the proof in [3]. Let $c_1 = \max_x d_{\text{sep}}(A_1(x,\cdot),\pi)$ and $c_2 = \max_x d_{\text{sep}}(A_2(x,\cdot),\pi)$. By definition, we may express A_1 and A_2 as follows.

$$A_1(x,y) = (1-c_1)\pi(y) + c_1 B_1(x,y), \quad \forall x, y \in \mathcal{X},$$

and

$$A_2(x,y) = (1-c_2)\pi(y) + c_2 B_2(x,y), \quad \forall x, y \in \mathcal{X},$$

where B_1 and B_2 are stochastic matrices. Furthermore, one may check that $\pi B_1 = \pi B_2 = \pi$. A simple calculation gives

$$\begin{aligned} A(x,y) = \sum_{z \in \mathcal{X}} A_1(x,z) A_2(z,y) &= (1 - c_1 c_2)\pi(y) + c_1 c_2 \sum_{z \in \mathcal{X}} B_1(x,y) B_2(x,y) \\ &\geq (1 - c_1 c_2)\pi(y), \quad \forall x, y \in \mathcal{X}. \end{aligned}$$

This proves the submultiplicativity of the maximum separation. $\square$

1.3 Poincaré inequality and the spectral gap

In this section, we introduce classical tools(the spectral gap of the transition matrix) to bound the ℓ^2-distance of continuous-time Markov chains to their stationary distributions. The following definition fits the classical notion of Dirichlet form if a Markov chain $(\mathcal{X}, K, \pi)$ is reversible.

Definition 1.3. Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain. The quadratic form

$$\mathcal{E}(f, g) = \mathcal{E}_K(f, g) = \mathrm{Re}\langle (I - K)f, g\rangle_\pi, \quad \forall f, g \in \mathbb{C}^{|\mathcal{X}|},$$

is called the *Dirichlet form* associated to the semigroup $H_t = e^{-t(I-K)}$, where $\langle \cdot, \cdot \rangle_\pi$ is the inner product in the complex space $\ell^2(\pi)$.

By definition, if $f = g$, one can rewrite the Dirichlet form as follows.

Lemma 1.1. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and $\mathcal{E}$ be the Dirichlet form associated to the semigroup H_t. Then, for $f \in \mathbb{C}^{|\mathcal{X}|}$,*

$$\begin{aligned}\mathcal{E}(f, f) &= \langle (I - \tfrac{1}{2}(K + K^*))f, f\rangle_\pi = \|f\|_2^2 - \mathrm{Re}\langle Kf, f\rangle_\pi \\ &= \frac{1}{2}\sum_{x,y\in\mathcal{X}} |f(x) - f(y)|^2 K(x, y)\pi(x).\end{aligned}$$

In particular, one has

$$\frac{\partial}{\partial t}\|H_t f\|_2^2 = -2\mathcal{E}(H_t f, H_t f), \quad \forall t > 0. \tag{1.7}$$

From (1.7), one can see that a bound of the ratio $\mathcal{E}(H_t f, H_t f)/\|H_t f\|_2^2$ will give a bound on the ℓ^2-norm of $H_t f$. The following quantity is useful in bounding the rate of the exponential convergence of the ℓ^2-distance.

Definition 1.4. Let $(\mathcal{X}, K, \pi)$ be a Markov chain with Dirichlet form $\mathcal{E}$. The spectral gap denoted by $\lambda = \lambda(K)$ is defined by

$$\lambda = \inf\left\{\frac{\mathcal{E}(f, f)}{\mathrm{Var}_\pi(f)} : \mathrm{Var}_\pi(f) \neq 0\right\},$$

where $\mathrm{Var}_\pi(f)$ is the variance of f, that is, $\mathrm{Var}_\pi(f) = \pi(f - \pi(f))^2$.

By definition, $\lambda(K) = \lambda(K^*)$. Generally, the spectral gap is not an eigenvalue of $I - K$. Note that λ can be characterized by

$$\lambda = \inf\{\mathcal{E}(f, f) : \pi(f) = 0,\ \|f\|_2 = 1\}.$$

If K is irreducible, the first equality in Lemma 1.1 and the minmax theorem in matrix analysis imply that the spectral gap is the smallest non-zero eigenvalue of $I - \frac{1}{2}(K + K^*)$. In particular, if K is reversible, or equivalently, the operator K is self-adjoint in $\ell^2(\pi)$, then λ is the smallest non-zero eigenvalue of $I - K$. Since the operator $K + K^*$ is self-adjoint, the spectral gap can be obtained by taking the infimum of the ratio in Definition 1.4 over all real-valued functions f.

Definition 1.5. Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain. A Poincaré inequality is an inequality of the following type

$$\|f - \pi(f)\|_2^2 \le C\mathcal{E}(f, f), \quad \forall f \in \ell^2(\pi),$$

where C is a positive constant in dependent of f.

From the above definition, if the Poincaré inequality holds for a Markov kernel K with constant C, then $\lambda(K) \ge C^{-1}$. In other words, the spectral gap is the inverse of the smallest C such that the Poincaré inequality holds.

By applying (1.7), we may bound the operator norm $\|H_t - \pi\|_{2\to 2}$ from above by using the spectral gap.

Proposition 1.11. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and λ be the spectral gap of K. Then the continuous-time semigroup H_t satisfies*

$$\forall f \in \ell^2(\pi), \quad \|H_t f - \pi(f)\|_2^2 \le e^{-2\lambda t}\mathrm{Var}_\pi(f).$$

Proof. Let $g = f - \pi(f)$. By Lemma 1.1, one has

$$\frac{\partial}{\partial t}\|H_t g\|_2^2 = -2\mathcal{E}(H_t g, H_t g) \le -2\lambda \mathrm{Var}_\pi(H_t g) = -2\|H_t g\|_2^2.$$

This implies that

$$\|H_t f - \pi(f)\|_2^2 = \|H_t g\|_2^2 \le e^{-2\lambda t}\|H_0 g\|_2^2 = e^{-2\lambda t}\mathrm{Var}_\pi(f).$$

□

Remark 1.6. Since the Dirichlet form and the variance are invariant under the addition of a constant vector, the conclusion in Proposition 1.11 is equivalent to saying that

$$\|H_t - \pi\|_{2\to 2} \le e^{-\lambda t}, \quad \forall t > 0.$$

By considering the spectrum of a Markov kernel, if K is reversible, we have $\|H_t - \pi\|_{2\to 2} = e^{-\lambda t}$ for all $t > 0$. In general, this identity does not hold for all $t > 0$. However, it is proved in [12] that λ is the largest value β such that $\|H_t - \pi\|_{2\to 2} \le e^{-\beta t}$ for all $t > 0$.

By Proposition 1.11, we may derive an upper bound on the ℓ^2-distance for continuous-time Markov chains.

Theorem 1.1. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and λ be the spectral gap of K. One has*

$$\forall x \in \mathcal{X}, \quad d_{\pi,2}(H_t(x,\cdot), \pi) \le \pi(x)^{-1/2} e^{-\lambda t},$$

and

$$\forall x, y \in \mathcal{X}, \quad |H_t(x,y) - \pi(y)| \le \sqrt{\pi(y)/\pi(x)}\, e^{-\lambda t}.$$

Proof. Let H_t^* be the adjoint operator of H_t and set $\delta_x(y) = \pi(x)^{-1}$ if $x = y$ and $\delta_x(y) = 0$ otherwise. Since $\lambda(K) = \lambda(K^*)$, we have, by letting $f = \delta_x$ in Proposition 1.11,

$$d_{\pi,2}(H_t(x,\cdot), \pi)^2 = \|H_t^* \delta_x - \pi(\delta_x)\|_2^2 \le e^{-2\lambda t} \mathrm{Var}_\pi(\delta_x) = \left(\pi(x)^{-1} - 1\right) e^{-2\lambda t}.$$

This proves the first identity.

For the second one, note that

$$d_{\pi,2}(H_t^*(x,\cdot), \pi) \le \pi(x)^{-1/2} e^{-\lambda t}.$$

This implies that, for $x, y \in \mathcal{X}$,

$$
\begin{aligned}
|H_t(x,t) - \pi(y)| &= \pi(y) \left| \sum_{z \in \mathcal{X}} \left(\frac{H_{t/2}(x,z)}{\pi(z)} - 1 \right) \left(\frac{H^*_{t/2}(y,z)}{\pi(z)} - 1 \right) \pi(z) \right| \\
&\le \pi(y) d_{\pi,2}(H_{t/2}(x,\cdot), \pi) d_{\pi,2}(H^*_{t/2}(y,\cdot), \pi) \\
&\le \sqrt{\pi(y)/\pi(x)} e^{-\lambda t},
\end{aligned}
$$

where the first inequality applies the Cauchy-Schwartz inequality. □

To relate the spectral gap and the spectrum of K, we define another quantity as follows.

$$
\omega = \omega(K) = \min\{\mathrm{Re}\beta : \beta \neq 0,\ \beta \text{ is an eigenvalue of } I - K\}. \tag{1.8}
$$

Since $H_t = e^{-t(I-K)}$, it follows that the spectral radius of $H_t - \pi$ in $\ell^2(\pi)$ is $e^{-t\omega}$. This implies, for all $1 \le p \le \infty$,

$$
\|H_t - \pi\|_{p \to p} \ge e^{-\omega t}, \quad \forall t > 0. \tag{1.9}
$$

In particular, we have, by applying the operator theory,

$$
\lim_{t \to \infty} \|H_t - \pi\|_{p \to q}^{1/t} = e^{-\omega}, \quad \forall 1 \le p, q \le \infty.
$$

The next theorem summarizes the above fact.

Theorem 1.2. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain, λ be the spectral gap of K and ω be the quantity defined in* (1.8)*. For all $1 \le p \le \infty$,*

$$
\lim_{t \to \infty} \frac{-1}{t} \log \left(\max_{x \in \mathcal{X}} d_{\pi,p}(H_t(x,\cdot), \pi) \right) = \omega.
$$

In particular, $\lambda \le \omega$.

Proof. Immediate from Proposition 1.8, Remark 1.6 and the discussion in the paragraph before this theorem. □

As a consequence of Theorem 1.2, the rate of the exponential convergence of the maximum ℓ^p-distance is asymptotically ω, not the spectral gap λ. However, if an irreducible Markov kernel K is normal, that is, $K^*K = KK^*$, then $\lambda = \omega$. This implies that the asymptotical rate of the exponential convergence is the spectral gap. From this discussion, one can see that the spectral gap is closely related to the long-term behavior of a Markov chain. To reflect the finite-time behavior of the convergence and the notion of "time to equilibrium", we consider the following quantity.

Definition 1.6. Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and H_t be the associated continuous-time semigroup. For $1 \le p \le \infty$, the ℓ^p-mixing time is denoted by $T_p = T_p(K)$ and defined by

$$T_p = \inf\left\{t > 0 : \max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi) \le 1/e\right\}.$$

By (1.9) and Theorem 1.1, we may bound the ℓ^p-mixing time as follows.

Theorem 1.3. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov kernel and set $\pi_* = \min\{\pi(x) : x \in \mathcal{X}\}$. For $1 \le p \le 2$,*

$$\frac{1}{\omega} \le T_p \le \frac{1}{\lambda}\left(1 + \frac{1}{2}\log\frac{1}{\pi_*}\right),$$

and, for $2 < p \le \infty$,

$$\frac{1}{\omega} \le T_p \le \frac{1}{\lambda}\left(1 + \log\frac{1}{\pi_*}\right).$$

Proof. The lower bound is obtained from (1.9) and Proposition 1.8. For the upper bounds, note that, by Proposition 1.6, the identities in Theorem 1.1 imply that

$$\forall 1 \le p \le 2, \quad \max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi) \le \pi_*^{-1/2}e^{-\lambda t}$$

and

$$\forall 1 \le p \le \infty, \quad \max_{x\in\mathcal{X}} d_{\pi,p}(H_t(x,\cdot),\pi) \le \pi_*^{-1}e^{-\lambda t}.$$

This is sufficient to prove the desired upper bounds. □

For an illustration of the above theorem, we consider the following example.

Example 1.1. Fix $n > 1$ and let K_n be a Markov kernel given by

$$K_n(x, x+1) = K_n(x, x-1) = 1/2, \quad \forall x \in \mathbb{Z}_n,$$

where $\mathbb{Z}_n$ is the n-cycle. It is an easy exercise that K_n is irreducible and the stationary distribution π_n is a uniform distribution on $\mathbb{Z}_n$. By a method in Feller [13, p.353], the following functions

$$\phi_{n,i}(j) = \cos(2\pi ij/n), \quad \forall 0 \le j \le n-1,\ 0 \le i \le \lfloor (n-1)/2 \rfloor,$$

and

$$\phi_{n,n-i}(j) = \sin(2\pi ij/n), \quad \forall 0 \le j \le n-1,\ 1 \le i \le \lceil (n-1)/2 \rceil,$$

are eigenfunctions of K_n and the corresponding eigenvalues $\beta_{n,0}, \beta_{n,1}, ..., \beta_{n,n-1}$ are given by

$$\beta_{n,0} = 1,\ \beta_{n,i} = \beta_{n,n-i} = \cos(2\pi i/n), \quad \forall 1 \le i \le \lceil (n-1)/2 \rceil.$$

Since K_n is reversible, the spectral gap is $\lambda(K_n) = \omega(K_n) = 1 - \cos(2\pi/n)$.

Let $H_{n,t}$ be the continuous-time semigroup associated to K_n. Applying Theorem 1.1 and Theorem 1.3 and using (1.9) derives

$$e^{-t(1-\cos(2\pi/n))} \le d_{\pi,2}(H_{n,t}(x,\cdot),\pi) \le \sqrt{n}e^{-t(1-\cos(2\pi/n))}, \quad \forall n \ge 1.$$

and, for $1 \le p \le \infty$,

$$\frac{n^2}{2\pi^2} \sim \frac{1}{1-\cos(2\pi/n)} \le T_p(K_n) \le \frac{1}{1-\cos(2\pi/n)}(1+\log n) \sim \frac{n^2 \log n}{2\pi^2}.$$

This means that, for $1 \le p \le \infty$, the ℓ^p-distance of continuous-time Markov chains asymptotically cannot be too small before the time of order n^2 but is close to 0 after the time $Cn^2 \log n$ for large C. It is worthwhile noting that the correct order for the ℓ^p-mixing time is n^{-2}.

Chapter 2

Hypercontractivity and the logarithmic Sobolev constant

Since Gross introduced the notions of the logarithmic Sobolev constant and of the hypercontractivity, many techniques are developed to compute the logarithmic Sobolev constant. The hypercontractivity is proved useful in bounding the convergence rate of Markov chains to their stationarity. An informative account of the development of logarithmic Sobolev inequalities can be found in the survey paper [14].

In Section 2.1, we define the logarithmic Sobolev constant and use it to bound the entropy of a Markov chain. In Section 2.2, we introduce how the hypercontractivity can be used to bound the ℓ^p-distance and the ℓ^p-mixing time. In Section 2.3, diverse techniques for the estimation of the logarithmic Sobolev constant are introduced. In Section 2.4, we determine the explicit value of the logarithmic Sobolev constant for some examples.

2.1 The logarithmic Sobolev constant

The definition of the logarithmic Sobolev constant is very similar to that of the spectral gap. For a motive of why we concern such a constant, let's start by looking at the relative entropy of the continuous-time Markov chain. Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain, H_t be the associated continuous-time semigroup of K and $\mathcal{E}$ be the Dirichlet form. Recall that the entropy of a probability measure μ

with respect to π is defined by

$$\mathrm{Ent}_\pi(\mu) = \pi(h \log h),$$

where $\mu = h\pi$. Here we abuse the usage of Ent by letting

$$\mathrm{Ent}_\pi(f) = \pi(f \log f),$$

if f is a any nonnegative function but not a probability measure. A simple computation shows that, for any probability measure μ, $\sum_x \mu(x)/\pi(x) \neq 1$ and

$$\mathrm{Ent}_\pi(\mu H_t) = \mathrm{Ent}_\pi\left(\sum_{y\in\mathcal{X}} \frac{\pi(y)H_t(y,\cdot)}{\pi(\cdot)} \frac{\mu(y)}{\pi(y)}\right) = \mathrm{Ent}_\pi(H_t^* h),$$

where $h = \mu/\pi$. In the above setting, we have

$$\forall t > 0, \quad \frac{\partial}{\partial t}\mathrm{Ent}_\pi(H_t^* h) = -\mathcal{E}(H_t^* h, \log(H_t^* h)) \leq -2\mathcal{E}(\sqrt{H_t^* h}, \sqrt{H_t^* h}), \qquad (2.1)$$

where the inequality is proved by Diaconis and Saloff-Coste in [11, Lemma 2.7] and has an improved coefficient 4 instead of 2 if K is assumed reversible. By (2.1), one can see that a bound on the ratio $\mathrm{Ent}_\pi(H_t^* h)/\mathcal{E}(\sqrt{H_t^* h}, \sqrt{H_t^* h})$ suffices to give a bound the rate of the convergence. To define the logarithmic Sobolev constant, we need to replace the variance by the following entropy-like quantity.

$$\mathcal{L}(f) = \mathcal{L}_\pi(f) = \sum_{x\in\mathcal{X}} |f(x)|^2 \log\left(\frac{|f(x)|^2}{\|f\|_2^2}\right)\pi(x). \qquad (2.2)$$

Since $u \mapsto u \log u$ is convex, Jensen's inequality implies that $\mathcal{L}(f)$ is nonnegative. Furthermore, if π is positive everywhere, then $\mathcal{L}(f) = 0$ if and only if f is constant. Note that if $\|f\|_2 = 1$, that is, f^2 is the probability density of $\mu = f^2\pi$ with respective to π, then

$$\mathcal{L}(f) = \mathrm{Ent}_\pi(\mu).$$

Definition 2.1. Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and $\mathcal{L}$ be the functional defined in (2.2). The logarithmic Sobolev constant $\alpha = \alpha(K)$ is defined by

$$\alpha = \inf \left\{ \frac{\mathcal{E}(f,f)}{\mathcal{L}(f)} : \mathcal{L}(f) \neq 0 \right\}.$$

By definition, it is clear that $\alpha(K) = \alpha(K^*)$. Obviously, one has $\mathcal{L}(f) = \mathcal{L}(|f|)$ and $\mathcal{E}(|f|,|f|) \leq \mathcal{E}(f,f)$. By these facts, the logarithmic Sobolev constant can be obtained by taking the infimum of the ratio $\mathcal{E}(f,f)/\mathcal{L}(f)$ over all nonnegative functions f.

Definition 2.2. Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and $\mathcal{E}$ be the Dirichlet form. A logarithmic Sobolev inequality is an inequality of the following type.

$$C\mathcal{L}(f) \leq \mathcal{E}(f,f), \quad \text{for all function } f,$$

where C is a nonnegative constant.

By the above definition, if the logarithmic Sobolev inequality holds for some constant $C \geq 0$, then $\alpha \geq C$. In other words, α is the largest constant C such that the logarithmic Sobolev inequality holds. One may think of the existence of a function f such that the ratio $\mathcal{E}(f,f)/\mathcal{L}(f)$ is equal to 0, which means that the logarithmic Sobolev inequality never holds unless $C = 0$. It has been proved that the irreducibility eliminates such a possibility. Thus, one needs to consider only the case $C > 0$ in Definition 2.2. For a proof of the fact $\alpha > 0$, please see Proposition 2.3. By (2.1), the entropy of a continuous-time Markov chain is bounded from above as follows.

Proposition 2.1. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain, H_t be the associated semigroup and α be the logarithmic Sobolev constant. Let μ be a probability* *measure on $\mathcal{X}$. Then one has*

$$\mathrm{Ent}_\pi(\mu H_t) \le \begin{cases} e^{-2\alpha t}\mathrm{Ent}_\pi(\mu) & \text{in general} \\ e^{-4\alpha t}\mathrm{Ent}_\pi(\mu) & \text{if } K \text{ is reversible.} \end{cases}$$

In particular, for $x \in \mathcal{X}$,

$$\mathrm{Ent}_\pi(H_t(x,\cdot)) \le \begin{cases} e^{-2\alpha t}\log\frac{1}{\pi(x)} & \text{in general} \\ e^{-4\alpha t}\log\frac{1}{\pi(x)} & \text{if } K \text{ is reversible.} \end{cases}$$

Proof. Let $h = \mu/\pi$. By (2.1), one can easily prove that for $t > 0$,

$$\mathrm{Ent}_\pi(\mu H_t) = \mathrm{Ent}_\pi(H_t^* h) \le e^{-2\alpha t}\mathrm{Ent}_\pi(h) = \mathrm{Ent}_\pi(\mu).$$

The same proof as above works for the reversible cases. The second part is followed by letting $\mu = \delta_x$, where $\delta_x(y) = 1$ if $y = x$ and $\delta_x(y) = 0$ otherwise. □

By applying Proposition 1.9 and Proposition 2.1, one may give an upper bound on the total variation distance.

Corollary 2.1. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and α be the logarithmic Sobolev constant of K. Then, for $t > 0$,*

$$d_{\pi,1}(H_t(x,\cdot),\pi) \le \sqrt{2\log(1/\pi(x))}e^{-\alpha t},$$

and, if K is reversible,

$$d_{\pi,1}(H_t(x,\cdot),\pi) \le \sqrt{2\log(1/\pi(x))}e^{-2\alpha t}.$$

In particular, one has

$$T_1 \le \frac{1}{2\alpha}\left(3 + \log_+ \log\frac{1}{\pi_*}\right),$$

and, for reversible chains,

$$T_1 \le \frac{1}{4\alpha}\left(3 + \log_+ \log\frac{1}{\pi_*}\right),$$

where $\pi_* = \min_x \pi(x)$ and $\log_+ t = \max\{0, \log t\}$.

2.2 Hypercontractivity

In the previous section, the entropy and the ℓ^1-distance of a continuous-time Markov chain are proved to converge exponentially with rate at least the logarithmic Sobolev constant. It is natural to consider using the logarithmic Sobolev constant to bound the ℓ^p-distance. The following theorem is the well-known hypercontractivity introduced in [14], which is sufficient to derive a bound on the ℓ^p-distance.

Theorem 2.1. (Theorem 3.5 in [11]) *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and α be the logarithmic Sobolev constant of K.*

(1) Assume that there exists $\beta > 0$ such that $\|H_t\|_{2\to q} \leq 1$ for all $t > 0$ and $2 \leq q < \infty$ satisfying $e^{4\beta t} \geq q - 1$. Then $\beta\mathcal{L}(f) \leq \mathcal{E}(f,f)$ for all f, and thus $\alpha \geq \beta$.

(2) Assume that (K, π) is reversible. Then $\|H_t\|_{2\to q} \leq 1$ for all $t > 0$ and $2 \leq q < \infty$ satisfying $e^{4\alpha t} \geq q - 1$.

(3) For non-reversible chains, we have $\|H_t\|_{2\to q} \leq 1$ for all $t > 0$ and $2 \leq q < \infty$ satisfying $e^{2\alpha t} \geq q - 1$.

Proof. See the proof given in [11]. □

Remark 2.1. Note that if (K, π) is reversible, then the first two assertions in Theorem 2.1 characterize the logarithmic Sobolev constant as follows.

$$\alpha = \max\{\beta : \|H_t\|_{2\to q} \leq 1, \forall t \geq \tfrac{1}{4\beta}\log(q-1), 2 \leq q < \infty\}.$$

To point out a surprising observation from the hypercontractivity, we recall the following fact in [23].

Lemma 2.1. *Assume that K is a normal operator on $\ell^2(\pi)$ and $\beta_0 = 1, \beta_1, ..., \beta_{|\mathcal{X}|-1}$ are the eigenvalues of K with corresponding eigenvectors $\phi_0 \equiv 1, \phi_1, ..., \phi_{|\mathcal{X}|-1}$. Then, for all $x \in \mathcal{X}$, one has*

$$\|H_t(x,\cdot)/\pi\|_2^2 = \sum_{i=0}^{|\mathcal{X}|-1} e^{-2t(1-\mathrm{Re}\beta_i)}|\phi_i(x)|^2.$$

It follows from the above lemma that $\|H_t\|_{2\to\infty} > 1$ if K is normal. Since H_t is a contraction in ℓ^2 and has eigenvalue 1 with corresponding eigenvector $\mathbf{1}$, we have $\|H_t\|_{2\to 2} = 1$. A nontrivial observation, even in the discrete setting of a state space, from the hypercontractivity is the existence of $0 < t_q < \infty$, for any $2 < q < \infty$, such that $\|H_t\|_{2\to q} = 1$ when $t \geq t_q$.

By Theorem 2.1, we may bound the ℓ^p-distance from above by using the logarithmic Sobolev constant.

Theorem 2.2. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and λ and α be the spectral gap and the logarithmic Sobolev constant of K. Then, for $\epsilon, \theta, \sigma \geq 0$ and $t = \epsilon + \theta + \sigma$,*

$$d_{\pi,2}(H_t(x,\cdot),\pi) \leq \begin{cases} \|H_\epsilon(x,\cdot)/\pi\|_2^{2/(1+e^{4\alpha\theta})} e^{-\lambda\sigma} & \text{if } K \text{ is reversible} \\ \|H_\epsilon(x,\cdot)/\pi\|_2^{2/(1+e^{2\alpha\theta})} e^{-\lambda\sigma} & \text{in general} \end{cases}.$$

In particular, for $c \geq 0$, one has

$$d_{\pi,2}(H_t(x,\cdot),\pi) \leq e^{1-c},$$

as

$$t = \begin{cases} (4\alpha)^{-1}\log_+ \log(1/\pi(x)) + c\lambda^{-1} & \text{if } K \text{ is reversible} \\ (2\alpha)^{-1}\log_+ \log(1/\pi(x)) + c\lambda^{-1} & \text{in general} \end{cases},$$

where $\log_+ t = \max\{0, \log t\}$.

Proof. We consider only the reversible case by using Theorem 2.1(2), while the general case can be proved by applying Theorem 2.1(3). Let $\theta > 0$ and $q(\theta) = 1 + e^{4\alpha\theta}$. By Theorem 2.1(2), it is clear that $\|H_\theta\|_{2\to q(\theta)} \le 1$, and by the duality given in Lemma A.1, it follows that $\|H_\theta^*\|_{q'(\theta)\to 2} \le 1$, where $q(\theta)^{-1} + (q'(\theta))^{-1} = 1$. For convenience, let h_t^x denote the density of $H_t(x,\cdot)$ with respect to π. Note that $h_{t+s}^x = H_t^* h_s^x$. This implies

$$\begin{aligned}\|h_t^x - 1\|_2 &= \|(H_\sigma^* - \pi)(H_\theta^* h_\epsilon^x)\|_2 \\ &\le \|H_\sigma^* - \pi\|_{2\to 2}\|H_\theta^*\|_{q'(\theta)\to 2}\|h_\epsilon^x\|_{q'(\theta)} \le e^{-\lambda\sigma}\|h_\epsilon^x\|_2^{2/q(\theta)},\end{aligned}$$

where the last inequality uses Remark 1.6 and the following Hölder inequality.

$$\|f\|_{q'} \le \|f\|_1^{1-2/q}\|f\|_2^{2/q},$$

for all $1 \le q' \le 2$ and $q^{-1} + (q')^{-1} = 1$. This proves the first inequality.

For the second part, note that $\|h_0^x\|_2 = \pi(x)^{-1/2}$ for $x \in \mathcal{X}$. By letting $\epsilon = 0$, we obtain

$$\|h_t^x - 1\|_2 \le \left(\frac{1}{\pi(x)}\right)^{1/(1+e^{4\alpha\theta})} e^{-\lambda\sigma}.$$

Let $\sigma = c\lambda^{-1}$. To get the desired upper bound for the ℓ^2-distance, we let $\sigma = c\lambda^{-1}$, choose $\theta = 0$ if $\pi(x) > e^{-1}$, and put

$$\theta = \frac{1}{4\alpha}\log\log\frac{1}{\pi(x)},$$

if $\pi(x) < e^{-1}$. □

Using the Cauchy-Schwartz inequality, the ℓ^∞-distance can be bounded from above by the ℓ^2-distance. In fact, for $t > 0$, one has

$$\begin{aligned}|h_t(x,y) - 1| &= \left|\sum_{z\in\mathcal{X}}(h_{t/2}(x,z) - 1)(h_{t/2}^*(y,z) - 1)\pi(z)\right| \\ &\le \|h_{t/2}(x.\cdot) - 1\|_2\|h_{t/2}^*(y,\cdot) - 1\|_2.\end{aligned}$$

This implies the following corollary.

Corollary 2.2. *Let* $(\mathcal{X}, K, \pi)$ *be an irreducible Markov chain and* λ *and* α *be the spectral gap and logarithmic Sobolev constant of* K*. Then, for* $c > 0$*, one has*

$$|H_t(x,y)/\pi(y) - 1| \le e^{2-c},$$

if

$$t = \begin{cases} \frac{1}{4\alpha}\left(\log_+ \log \frac{1}{\pi(x)} + \log_+ \log \frac{1}{\pi(y)}\right) + c\lambda^{-1} & \text{if } K \text{ is reversible} \\ \frac{1}{2\alpha}\left(\log_+ \log \frac{1}{\pi(x)} + \log_+ \log \frac{1}{\pi(y)}\right) + c\lambda^{-1} & \text{in general} \end{cases},$$

where $\log_+ t = \max\{0, \log t\}$.

Summing up Theorem 2.2 and Corollary 2.2, we may bound the ℓ^p-mixing time by using the logarithmic Sobolev constant.

Corollary 2.3. *Let* K *be a reversible and irreducible Markov chain with stationary distribution* π *and* α *be the logarithmic Sobolev constant. For* $1 \le p \le \infty$*, let* T_p *be the* ℓ^p*-mixing time of* K*. Then, for* $1 < p \le 2$*,*

$$\frac{1}{2m_p\alpha} \le T_p \le \frac{1}{4\alpha}\left(4 + \log_+ \log \frac{1}{\pi_*}\right)$$

and for $2 < p \le \infty$*,*

$$\frac{1}{2\alpha} \le T_p \le \frac{1}{2\alpha}\left(3 + \log_+ \log \frac{1}{\pi_*}\right)$$

where $\log_+ t = \max\{0, \log t\}$*,* $\pi_* = \min_x \pi(x)$ *and* $m_p = 1 + \lceil (2-p)/(2p-2) \rceil$.

Proof. The upper bounds are obtained immediately from Theorem 2.2 and Corollary 2.2. For the lower bound, Theorem 3.9 in [11] proves the case $2 < p \le \infty$. For $1 < p \le 2$, we use the fact

$$T_2 \le m_p T_p, \quad \forall 1 < p \le 2.$$

□

Remark 2.2. For general cases, Theorem 2.2 and Corollary 2.2 derive an upper bound of the ℓ^p-mixing time which is twice of that in Corollary 2.3.

Remark 2.3. Comparing Corollary 2.3 with Theorem 1.3, one may find that to bound the ℓ^p-mixing time of a reversible continuous-time Markov chain, the logarithmic Sobolev constant is more closely related to T_p than the spectral gap.

2.3 Tools to compute the logarithmic Sobolev constant

It follows from Theorem 1.3 and Corollary 2.3 that the logarithmic Sobolev constant provides a tighter bound(in the sense of order) for the time to equilibrium T_p than the spectral gap. Based on Corollary 2.3, to bound the ℓ^p-mixing time by using the logarithmic Sobolev constant, we need to determine its value. For this view point, it is natural to ask: can one compute explicitly or estimate the constant α? In this section, we introduce several established tools to help determine the logarithmic Sobolev constant.

1. Bounding α from above by using the spectral gap λ. The following proposition establishes a relation between the spectral gap and the logarithmic Sobolev constant.

Proposition 2.2. (Lemma 2.2.2 in [23]) *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain. Then the spectral gap λ and the logarithmic Sobolev constant α of K satisfy $\alpha \leq \lambda/2$. Furthermore, let ϕ be an eigenvector of the matrix $\frac{1}{2}(K + K^*)$ whose corresponding eigenvalue is $(1 - \lambda)$. If $\pi(\phi^3) \neq 0$, then $\alpha < \lambda/2$.*

Proof. We show by following the proof in [23] whose original idea comes from [22]. Let g be a real function on $\mathcal{X}$ and set $f = \mathbf{1} + \epsilon g$. Then for small enough ϵ, we have

$$
\begin{aligned}
f^2 \log f^2 &= (\mathbf{1} + 2\epsilon g + \epsilon^2 g^2)\left(2\epsilon g - \epsilon^2 g^2 + \tfrac{2}{3}\epsilon^3 g^3 + O(\epsilon^4)\right) \\
&= 2\epsilon g + 3\epsilon^2 g^2 + \tfrac{2}{3}\epsilon^3 g^3 + O(\epsilon^4),
\end{aligned}
$$

and

$$
\begin{aligned}
f^2 \log \|f\|_2^2 &= (\mathbf{1} + 2\epsilon g + \epsilon^2 g^2)\big[2\epsilon\pi(g) + \epsilon^2(\pi(g^2) - 2\pi(g)^2) \\
&\qquad + \epsilon^3\left(\tfrac{8}{3}\pi(g)^3 - 2\pi(g)\pi(g^2)\right) + O(\epsilon^4)\big] \\
&= 2\epsilon\pi(g) + \epsilon^2(4g\pi(g) + \pi(g^2) - 2\pi(g)^2) \\
&\qquad + \epsilon^3\big[\tfrac{8}{3}\pi(g)^3 - 2\pi(g)\pi(g^2) + 2g\pi(g^2) \\
&\qquad - 4g\pi(g)^2 + 2g^2\pi(g)\big] + O(\epsilon^4)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
f^2 \log \frac{f^2}{\|f\|_2^2} &= 2\epsilon\left[g - \pi(g)\right] + \epsilon^2\left[3g^2 - 4g\pi(g) - \pi(g^2) + 2\pi(g)^2\right] \\
&\qquad + \epsilon^3\big[\tfrac{2}{3}g^3 - \tfrac{8}{3}\pi(g)^3 + 2\pi(g)\pi(g^2) - 2g\pi(g^2) \\
&\qquad + 4g\pi(g)^2 - 2g^2\pi(g)\big] + O(\epsilon^4)
\end{aligned}
$$

and

$$
\mathcal{L}(f) = 2\epsilon^2 \mathrm{Var}_\pi(g) + \epsilon^3\left[\tfrac{2}{3}\pi(g^3) + \tfrac{4}{3}\pi(g)^3 - 2\pi(g)\pi(g^2)\right] + O(\epsilon^4),
$$

where $O(\cdot)$ depends only on $\|g\|_\infty$.

To finish the proof, note that $\mathcal{E}(f,f) = \epsilon^2\mathcal{E}(g,g)$. Let ϕ be an eigenfunction of $\frac{1}{2}(K + K^*)$ whose eigenvalue is $1 - \lambda$. By definition, it is clear that $\mathcal{E}(\phi,\phi) = \lambda\mathrm{Var}_\pi(\phi)$ and $\pi(\phi) = 0$. Letting $g = \phi$ implies

$$
\alpha \le \frac{\mathcal{E}(f,f)}{\mathcal{L}(f)} = \frac{\lambda \mathrm{Var}_\pi(\phi)}{2\mathrm{Var}_\pi(\phi) + \frac{2}{3}\epsilon\pi(\phi^3) + O(\epsilon^2)}
$$

The first inequality is obtained by letting $\epsilon \to 0$. For the second part, since $\pi(\phi^3) \neq 0$, we may choose $|\epsilon| > 0$ such that $\epsilon\pi(\phi^3) > 0$ and $\frac{2}{3}\epsilon\pi(\phi^3) > O(\epsilon^2)$. This proves the second inequality. □

2. One sufficient condition for $\alpha = \lambda/2$. As a consequence of Proposition 2.2, the logarithmic Sobolev constant α is bounded from above by $\lambda/2$. Furthermore, a sufficient condition for the case $2\alpha < \lambda$ is also given in that proposition. In the following, we give a necessary condition for the situation $2\alpha < \lambda$ to happen.

Proposition 2.3. (Theorem 2.2.3 in [23]) *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and λ and α be the spectral gap and the logarithmic Sobolev constant of K. Then either $\alpha = \lambda/2$ or there exists a positive non-constant function u which is a solution of*

$$2u \log u - 2u \log \|u\|_2 = \frac{1}{\alpha}(I - K)u = 0, \tag{2.3}$$

where $\alpha = \mathcal{E}(u, u)/\mathcal{L}(u)$. In particular, $\alpha > 0$.

Proof. We prove by considering the minimizer of the infimum in Definition 2.1. Note that we may restrict ourselves to non-negative vectors with mean 1(under π). By definition, either α is attained by a nonnegative non-constant vector, say u, or the infimum is attained at the constant vector $\mathbf{1}$. In the latter case, one may choose a minimizing sequence $(\mathbf{1} + \epsilon_n g_n)_1^\infty$ satisfying

$$\epsilon_n \to 0 \quad \text{and} \quad \pi(g_n) = 0, \ \|g_n\|_2 = 1, \quad \forall n \geq 1.$$

This implies that the sequence $\{\|g_n\|_\infty\}$ is bounded from above and below by positive numbers. Then, by the proof of Proposition 2.2, we get

$$\begin{aligned}
\alpha &= \lim_{n\to\infty} \frac{\mathcal{E}(\mathbf{1} + \epsilon_n g_n, \mathbf{1} + \epsilon_n g_n)}{\mathcal{L}(\mathbf{1} + \epsilon_n g_n)} \\
&= \lim_{n\to\infty} \frac{\mathcal{E}(g_n, g_n)}{2\mathrm{Var}_\pi(g_n) + O(\epsilon_n)} \geq \liminf_{n\to\infty} \frac{\lambda}{2 + O(\epsilon_n)} = \frac{\lambda}{2}
\end{aligned}$$

This proves $\alpha = \lambda/2$.

If α is attained by a nonnegative non-constant vector f, then by viewing $\mathcal{E}(f,f)/\mathcal{L}(f)$ as a function defined on $\mathbb{R}^{|\mathcal{X}|}$, we have the following Euler-Lagrange equation

$$\nabla\left(\frac{\mathcal{E}(f,f)}{\mathcal{L}(f)}\right)\Bigg|_{f=u} = \mathbf{0},$$

which is identical to (2.3). To show the positiveness of u, observe that if $u(x) = 0$ for some $x \in \mathcal{X}$, then (2.3) implies that $Ku(x) = 0$, or equivalently, $u(y) = 0$ if $K(x,y) > 0$. Thus, by the irreducibility of K, one has $u \equiv 0$, which contradicts the assumption that u is not constant. $\square$

Remark 2.4. Note that a constant function is always a solution of (2.3).

Corollary 2.4. *Let $(\mathcal{X}, K, \pi)$ be an irreducible Markov chain and λ and α be the spectral gap and logarithmic Sobolev constant of K. If a non-constant function u on $\mathcal{X}$ and a positive number β satisfy the following system of equations*

$$(I-K)u = 2\beta(u\log u - u\log\|u\|_2), \tag{2.4}$$

then $\beta = \mathcal{E}(u,u)/\mathcal{L}(u)$. In particular, (2.4) has no non-constant solution for $\beta \in (0,\alpha)$. Moreover, if (2.4) has no non-constant solution for $\beta \in (0,\lambda/2)$, then $\alpha = \lambda/2$.

3. Comparison technique. In many cases, the model of interest is complicated but can be replaced by a simpler one. In that case, the tradeoff of the replacement can be the loss of the accuracy of α(up to a constant) but the advantage is the simplicity of the new chain and, mostly, α is of the same order as the logarithmic Sobolev constant of the new Markov chain.

Proposition 2.4. (Lemma 2.2.12 in [23]) *Let* $(\mathcal{X}_1, K_1, \pi_1)$ *and* $(\mathcal{X}_2, K_2, \pi_2)$ *be irreducible Markov chains and* $\mathcal{E}_1$ *and* $\mathcal{E}_2$ *be respective Dirichlet forms. Assume that there exists a linear map*

$$T : \ell^2(\pi_1) \to \ell^2(\pi_2)$$

and constant $A > 0, B \geq 0, a > 0$ *such that, for all* $f \in \ell^2(\pi_1)$*,*

$$\mathcal{E}_2(Tf, Tf) \leq A\mathcal{E}_1(f, f), \quad a\mathrm{Var}_{\pi_1}(f) \leq \mathrm{Var}_{\pi_2}(Tf) + B\mathcal{E}_1(f, f).$$

Then the spectral gaps $\lambda_1 = \lambda(K_1)$ *and* $\lambda_2 = \lambda(K_2)$ *satisfy*

$$\frac{a\lambda_2}{A + B\lambda_2} \leq \lambda_1.$$

Similarly, if

$$\mathcal{E}_2(Tf, Tf) \leq A\mathcal{E}_1(f, f), \quad a\mathcal{L}_{\pi_1}(f) \leq \mathcal{L}_{\pi_2}(Tf) + B\mathcal{E}_1(f, f),$$

then the logarithmic Sobolev constants $\alpha_1 = \alpha(K_1)$ *and* $\alpha_2 = \alpha(K_2)$ *satisfy*

$$\frac{a\alpha_2}{A + B\alpha_2} \leq \alpha_1.$$

In particular, if $\mathcal{X}_1 = \mathcal{X}_2$*,* $\mathcal{E}_2 \leq A\mathcal{E}_1$ *and* $a\pi_1 \leq \pi_2$*, then*

$$\frac{a\lambda_2}{A} \leq \lambda_1, \quad \frac{a\alpha_2}{A} \leq \alpha_1.$$

Proof. The proof follows from the variational definitions of the spectral gap and the logarithmic Sobolev constant. For the spectral gap, we have

$$\begin{aligned} a\mathrm{Var}_{\pi_1}(f) &\leq \mathrm{Var}_{\pi_2}(Tf) + B\mathcal{E}_1(f, f) \leq \frac{\mathcal{E}_2(Tf, Tf)}{\lambda_2} + B\mathcal{E}_1(f, f) \\ &\leq \left(\frac{A}{\lambda_2} + B\right)\mathcal{E}_1(f, f). \end{aligned}$$

The proof for the logarithmic Soboloev constant goes in the same way.

To show the last part, consider the following characterizations of λ and α.

$$\mathrm{Var}_\pi(f) = \min_{c\in\mathbb{R}} \|f-\mathbf{c}\|_2^2 = \min_{c\in\mathbb{R}} \sum_{x\in\mathcal{X}} [f(x)-c]^2\pi(x), \tag{2.5}$$

and

$$\begin{aligned}\mathcal{L}_\pi(f) &= \sum_{x\in\mathcal{X}} \left[f^2(x)\log f^2(x) - f^2(x)\log\|f\|_2^2 - f^2(x) + \|f\|_2^2\right]\pi(x) \\ &= \min_{c>0} \sum_{x\in\mathcal{X}} \left[f^2(x)\log f^2(x) - f^2(x)\log c - f^2(x) + c\right]\pi(x).\end{aligned} \tag{2.6}$$

Letting $T = I$ implies that

$$a\mathrm{Var}_{\pi_1}(f) \le \mathrm{Var}_{\pi_2}(f), \quad a\mathcal{L}_{\pi_1}(f) \le \mathcal{L}_{\pi_2}(f),$$

where the second one use the fact, $t\log t - t\log s - t + s \ge 0$ for $t, s \ge 0$. □

The following is a simple but useful tool which involves collapsing a chain to that with a smaller state space.

Corollary 2.5. *Let $(\mathcal{X}_1, K_1, \pi_1)$ and $(\mathcal{X}_2, K_2, \pi_2)$ be irreducible Markov chains and $\mathcal{E}_1$ and $\mathcal{E}_2$ be respective Dirichlet forms. Assume that there exists a surjective map $p : \mathcal{X}_2 \to \mathcal{X}_1$ such that*

$$\mathcal{E}_2(f\circ p, f\circ p) \le A\mathcal{E}_1(f,f), \quad \forall f \in \mathbb{R}^{|\mathcal{X}_1|},$$

and

$$a\pi_1(f) \le \pi_2(f\circ p), \quad \forall f \ge 0.$$

Then the spectral gaps $\lambda_1 = \lambda(K_1)$, $\lambda_2 = \lambda(K_2)$ and the logarithmic Sobolev constants $\alpha_1 = \alpha(K_1)$, $\alpha_2 = \alpha(K_2)$ satisfy

$$\frac{a\lambda_2}{A} \le \lambda_1, \quad \frac{a\alpha_2}{A} \le \alpha_1.$$

In particular, if $a = A = 1$, $\alpha_2 = \lambda_2/2$ and $\lambda_1 = \lambda_2$, then $\alpha_1 = \lambda_1/2$.

Proof. Let $T : \ell^2(\pi_1) \to \ell^2(\pi_2)$ be a linear map defined by $Tf = f \circ p$. In this setting, the assumption becomes

$$\mathcal{E}_2(Tf, Tf) \leq A\mathcal{E}_1(f, f), \quad \forall f \in \mathbb{R}^{|\mathcal{X}_1|},$$

and

$$a\pi_1(f) \leq \pi_2(Tf), \quad \forall f \geq 0.$$

By (2.5) and (2.6), the second inequality implies

$$a\mathrm{Var}_{\pi_1}(f) \leq \mathrm{Var}_{\pi_2}(Tf), \quad a\mathcal{L}_{\pi_1}(f) \leq \mathcal{L}_{\pi_2}(Tf), \quad \forall f \in \mathbb{R}^{|\mathcal{X}_1|}.$$

The desired identity is then proved by Proposition 2.4. □

Remark 2.5. Note that, in Corollary 2.5, if π_1 is a pushforward of π_2, that is,

$$\sum_{y:p(y)=x} \pi_2(y) = \pi_1(x), \quad \forall x \in \mathcal{X}_1,$$

then $\pi_1(f) = \pi_2(f \circ p)$ for all $f \in \mathbb{R}^{|\mathcal{X}_1|}$.

The following is a further corollary of Corollary 2.5 and the above remark which gives a sufficient condition on $a = A = 1$ in Corollary 2.5.

Corollary 2.6. *Let $(\mathcal{X}_1, K_1, \pi_1)$ and $(\mathcal{X}_2, K_2, \pi_2)$ be irreducible Markov chains and $p : \mathcal{X}_2 \to \mathcal{X}_1$ be a surjective map. Assume that, for all $x, y \in \mathcal{X}_1$,*

$$\sum_{\substack{z:p(z)=x \\ w:p(w)=y}} \pi_2(z)K_2(z, w) = \pi_1(x)K_1(x, y). \tag{2.7}$$

Let λ_1, λ_2 and α_1, α_2 be respectively the spectral gaps and logarithmic Sobolev constants of K_1 and K_2. Then

$$\lambda_2 \leq \lambda_1, \quad \alpha_2 \leq \alpha_1.$$

In particular, if $\lambda_1 = \lambda_2$ and $\alpha_2 = \lambda_2/2$, then $\alpha_1 = \lambda_1/2$.

Proof. It suffices to show that both constants a, A in Corollary 2.5 are equal to 1. Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be the Dirichlet forms of $(\mathcal{X}_1, K_1, \pi_1)$ and $(\mathcal{X}_2, K_2, \pi_2)$. By Lemma 1.1, a simple computation shows

$$\begin{aligned}\mathcal{E}_1(f,f) &= \frac{1}{2}\sum_{x,y\in\mathcal{X}_1}|f(x)-f(y)|^2\pi_1(x)K_1(x,y)\\ &= \frac{1}{2}\sum_{x,y\in\mathcal{X}_1}\sum_{\substack{z:p(z)=x\\ w:p(w)=y}}|f\circ p(z)-f\circ p(w)|^2\pi_2(z)K_2(z,w)\\ &= \frac{1}{2}\sum_{z,w\in\mathcal{X}_2}|f\circ p(z)-f\circ p(w)|^2\pi_2(z)K_2(z,w) = \mathcal{E}_2(f\circ p, f\circ p).\end{aligned}$$

By the definition of a stationary distribution in (1.2), summing up each side of (2.7) over all $x \in \mathcal{X}_1$ implies

$$\sum_{w:p(w)=y}\pi_2(w) = \pi_1(y), \quad \forall y \in \mathcal{X}_1.$$

By Remark 2.5, $\pi_1(f) = \pi_2(f \circ p)$ for all real f. Thus, by Corollary 2.5, $\lambda_2 \le \lambda_1$ and $\alpha_2 \le \alpha_1$. □

In some models, we may "collapse" the state space into a smaller one by partitioning the state space into several subsets and viewing each of them as a new state. In the induced state space, the stationary distribution of the new Markov chain is a lumped probability of the original one in the sense of Remark 2.5. The following proposition provides a sufficient condition for collapsing Markov chains.

Proposition 2.5. *Let $(\mathcal{X}_2, K_2, \pi_2)$ be an irreducible Markov chain and $p : \mathcal{X}_2 \to \mathcal{X}_1$ be a surjective map. Assume that*

$$K_2(f\circ p)(x) = K_2(f\circ p)(y), \quad \forall p(x)=p(y),\ f\in\mathbb{R}^{|\mathcal{X}_1|}. \tag{2.8}$$

Set, for $z, w \in \mathcal{X}_1$,

$$K_1(z,w) := \sum_{t:p(t)=w}K_2(s,t),$$

where $p(s) = z$. Then K_1 is irreducible and the stationary distribution π_1 is given by

$$\pi_1(x) = \sum_{y:p(y)=x} \pi_2(y).$$

Furthermore, if λ_1, λ_2 and α_1, α_2 are respectively the spectral gaps and logarithmic Sobolev constants of K_1, K_2. Then $\lambda_2 \leq \lambda_1$ and $\alpha_2 \leq \alpha_1$.

Remark 2.6. Note that (2.8) is equivalent to

$$\sum_{z:p(z)=w} K_2(x,z) = \sum_{z:p(z)=w} K_2(y,z)$$

for all $x, y \in \mathcal{X}_2$ satisfying $p(x) = p(y)$ and $w \in \mathcal{X}_1$.

Proof of Proposition 2.5. By choosing $f = \delta_w$(the function taking value 1 at w and 0 otherwise) in (2.8), the quantity $K_1(z,w)$ is well-defined for all $z, w \in \mathcal{X}_1$. It is clear that the irreducibility of K_1 is obtained immediately from that of K_2. By a simple computation, we have

$$\sum_{\substack{t:p(t)=w \\ s:p(s)=z}} \pi_2(s)K_2(s,t) = \sum_{s:p(s)=z} \pi_2(s)K_1(z,w) = \pi_1(z)K_1(z,w).$$

Summing up both sides over all $z \in \mathcal{X}_1$ implies that π_1 is the stationary distribution of K_1 and the remaining part is implied by Corollary 2.6. □

4. The product chains. In the following, we consider the logarithmic Sobolev constant of a product chain. For $1 \leq i \leq n$, let $(\mathcal{X}_i, K_i, \pi_i)$ be an irreducible Markov chain. Let μ be a probability measure on $\{0, 1, 2, ..., n\}$ and K be a Markov kernel on the product space $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$ defined by

$$K_\mu(x,y) = K(x,y) = \mu(0)\delta(x,y) + \sum_{i=1}^n \mu(i)\delta_i(x,y)K_i(x_i,y_i) \tag{2.9}$$

where $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$ and

$$\delta_i(x, y) = \prod_{\substack{j=1 \\ j \neq i}}^{n} \delta(x_i, y_i), \quad \delta(s, t) = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases}.$$

In the above setting, it is obvious that K is irreducible and the stationary distribution is $\pi = \bigotimes_1^n \pi_i$, where

$$\pi(x) = \prod_{i=1}^{n} \pi_i(x_i), \quad \forall x = (x_1, ..., x_n) \in \mathcal{X}. \tag{2.10}$$

Proposition 2.6. (Lemma 2.2.11 in [23]) *Let $\{(\mathcal{X}_i, K_i, \pi_i)\}_1^n$ be a sequence of irreducible Markov chains and $(\lambda_i)_1^n$ and $(\alpha_i)_1^n$ be their spectral gaps and logarithmic Sobolev constants. Let μ be a probability measure on the set $\{0, 1, ..., n\}$ and $(\mathcal{X}, K, \pi)$ be a product chain, where $\mathcal{X} = \prod_1^n \mathcal{X}_i$ and K and π are defined in (2.9) and (2.10). Then $\lambda = \lambda(K)$ and $\alpha = \alpha(K)$ are given by*

$$\lambda = \min\{\mu(i)\lambda_i\}, \quad \alpha = \min\{\mu(i)\alpha_i\}.$$

Proof. See P.339 in [23]. □

2.4 Some examples

Since the logarithmic Sobolev constant was introduced in 1975, many people dedicate to estimating its value. Their experiences show that it is not an easy job even though the computation of the logarithmic Sobolev constant is to find its correct order. As one can see in [11, Theorem A.2], the computation of the logarithmic Sobolev constant for asymmetric Markov kernels on a two point space is tough and complicated. Up to now, the explicit computation of the logarithmic Sobolev constant is still restricted to very simple examples and few of them are determined. By Proposition 2.6, the computation of the logarithmic Sobolev constants for Markov chains with small state spaces is not futile. In this section, we introduce some examples whose exact logarithmic Sobolev constants are known.

1. Random walk on a two point space. We first consider the simplest case where the state space has only two points, say 0 and 1. Let $\mathcal{X} = \{0,1\}$ and K be a Markov kernel on $\mathcal{X}$ defined by $K(0,0) = p_1$, $K(0,1) = q_1$, $K(1,0) = p_2$ and $K(1,1)q_2$, where $p_1 + q_1 = p_2 + q_2 = 1$. Equivalently, the matrix form of K is given by

$$K = \begin{pmatrix} p_1 & q_1 \\ p_2 & q_2 \end{pmatrix}. \tag{2.11}$$

The following theorem treats the case $p_1 = p_2$.

Theorem 2.3 ([11, Theorem A.2])**.** *Fix $p, q \in (0,1)$, $p + q = 1$. For the two-point space $\mathcal{X} = \{0,1\}$ equipped with the chain*

$$K(0,0) = K(1,0) = q,\ K(0,1) = K(1,1) = p,\ \pi(0) = q,\ \pi(1) = p. \tag{2.12}$$

we have $\lambda = 1$ and $\alpha = 1/2$ if $p = q = 1/2$ and

$$\alpha = \frac{p-q}{\log(p/q)} \quad \text{if } p \neq q.$$

Proof. The fact $\lambda = 1$ is an easy exercise. We prove the statement concerning α using Corollary 2.4. Setting $\phi(0) = b, \phi(1) = a$ and normalizing $qb^2 + pa^2 = 1$, we look for triplets (β, a, b) of positive numbers that are solutions of (2.4), that is,

$$\begin{cases} p(b-a) & = 2\beta b \log b \\ q(a-b) & = 2\beta a \log a \\ pa^2 + qb^2 & = 1. \end{cases}$$

Luckily, β can be eliminated by using the first two equations. This yields the system

$$\begin{cases} pa\log a + qb\log b &= 0 \\ p(a^2-1)+q(b^2-1) &= 0. \end{cases}$$

Setting aside the solution $a = b = 1$, we can assume $a, b \in (0,1)\cup(1,+\infty)$ and write this system as

$$\begin{cases} pa\log a + qb\log b &= 0 \\ \frac{a-a^{-1}}{\log a} = \frac{b-b^{-1}}{\log b}. \end{cases}$$

Calculus shows that the function $x \mapsto (x - x^{-1})/\log x$ is decreasing on $(0,1)$ and increasing on $(1,\infty)$. As it obviously satisfies $f(x) = f(1/x)$, it follows that the second equation can only be satisfied if $b = 1/a$. Reporting in the first equation yields $pa - q/a = 0$, that is, $a = \sqrt{q/p}$. It follows that the solutions of our original system are the triplets $(\beta, 1, 1)$ (β arbitrary) and, when $p \neq q$,

$$\left(\frac{p-q}{\log(p/q)}, \sqrt{q/p}, \sqrt{p/q}\right).$$

As $\frac{p-q}{\log(p/q)} < 1/2$ when $p \neq q$, we conclude from Corollary 2.4 that the logarithmic Sobolev constant of the asymmetric two-point space at (2.12) is

$$\alpha = \frac{p-q}{\log(p/q)}, \; p \neq q$$

and that, in the symmetric case $p = q = 1/2$, we have $2\alpha = \lambda = 1$. □

Remark 2.7. The proof of Theorem 2.3 given above is outlined without details in [4]. It is much simpler than the two different proofs given in [11, 23]. Here, we have been careful to treat both the symmetric and the asymmetric cases at once. In fact, the proof in [23] is incorrect (it can however be corrected with additional pain but without changing the main ideas). On the one hand, in the case $p = q = 1/2$, the proof above consists in showing that no nonconstant minimizers exist, leading to the conclusion that $\alpha = \lambda/2$. This is the main line of reasoning that will be used in this work to treat other examples. On the other hand, in the case $p \neq q$, we were able to find a unique normalized nonconstant solution of (2.3) with $\alpha < \lambda/2$ leading to the explicit computation of α. To the best of our knowledge, this is the only case with $\alpha < \lambda/2$ where α has been computed by solving (2.3). Our study of other small examples indicate that such computation is typically extremely difficult if not impossible.

Remark 2.8. As a consequence of Theorem 2.3 and Definition 2.1, we have

$$f(p) = \inf \left\{ \frac{pq(x-y)^2}{px^2 \log x^2 + qy^2 \log y^2} : x \neq y,\ px^2 + qy^2 = 1 \right\}$$

where

$$f(p) = \begin{cases} \frac{2p-1}{\log p - \log(1-p)} & \text{if } p \neq 1/2 \\ 1/2 & \text{if } p = 1/2 \end{cases}.$$

Let K be the Markov kernel in (2.11). A computation shows that the stationary distribution is equal to $\pi = (\frac{p_2}{p_2+q_1}, \frac{q_1}{p_2+q_1})$ and, for any function $f = (x, y)$ satisfying $\|f\|_2 = 1$,

$$\mathcal{E}(f,f) = \frac{p_2 q_1 (x-y)^2}{p_2 + q_1}, \quad \mathcal{L}_\pi(f) = \frac{p_2 x^2 \log x^2 + q_1 y^2 \log y^2}{p_2 + q_1}.$$

By the above identities and Remark 2.8, the logarithmic Sobolev constant of a general two point Markov chain is then a corollary of Theorem 2.3.

Corollary 2.7. *Let $(\{0,1\}, K, \pi)$ be an irreducible Markov chain where K is given by $K(0,0) = p_1$, $K(0,1) = q_1$, $K(1,0) = p_2$ and $K(1,1) = q_2$ with $p_2 q_1 \neq 0$. Then the spectral gap λ and the logarithmic Sobolev constant α are given by*

$$\lambda = p_2 + q_1, \quad \alpha = \begin{cases} \frac{p_2 - q_1}{\log p_2 - \log q_1} & \textit{if } p_2 \neq q_1 \\ p_2 & \textit{if } p_2 = q_1 \end{cases}$$

Remark 2.9. Let K be the Markov kernel given by (2.11). By Corollary 2.7, $\alpha = \lambda/2$ if and only if $K(0,1) = K(1,0)$, that is, K is symmetric.

2. Finite Markov chain with kernel $K(x,\cdot) \equiv \pi$. Let $\mathcal{X}$ be a finite set and π be a positive probability measure on $\mathcal{X}$. Consider a Markov kernel K, where $K(x,y) = \pi(y)$ for $x, y = \mathcal{X}$. In this setting, such a chain perfectly reaches its stationarity once the transition starts. Clearly, the spectral gap λ is 1 and the stationary distribution of K is π. To determine the logarithmic Sobolev constant α, we need the following computation.

Assume that $|\mathcal{X}| > 2$. Let $\pi_* = \min_x \pi(x) < 1/2$ and $x_0 \in \mathcal{X}$ be such that $\pi(x_0) = \pi_*$. Consider the projection $p : \mathcal{X} \mapsto \{0,1\}$ where $p(x_0) = 0$ and $p(x) = 1$ for $x \in \mathcal{X} \setminus \{x_0\}$. Let $\widetilde{K}$ be a Markov kernel on $\{0,1\}$ obtained by collapsing K through the map p, that is,

$$\widetilde{K} = \begin{pmatrix} \pi_* & 1-\pi_* \\ \pi_* & 1-\pi_* \end{pmatrix},$$

and $\widetilde{\alpha}$ be the logarithmic Sobolev constant of $\widetilde{K}$. Then, by Proposition 2.5 and Theorem 2.3, one has

$$\alpha \le \widetilde{\alpha} = \frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*])} < \lambda/2. \tag{2.13}$$

Theorem 2.4. ([11, Theorem A.1]) *Let $\mathcal{X}$ be a finite set with cardinality at least 3 and π be a positive probability measure on $\mathcal{X}$. Let K be a Markov kernel given by $K(x,y) = \pi(y)$ for $x, y \in \mathcal{X}$. Then the spectral gap is 1 and the logarithmic Sobolev constant α is equal to*

$$\alpha = \frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*]},$$

where $\pi_* = \min_x \pi(x)$.

Proof. The upper bound of α is given by (2.13) and it remains to prove its lower bound. By (2.13) and Proposition 2.3, there exists a non-constant positive function u on $\mathcal{X}$ satisfying (2.3). Without loss of generality, we assume that $\|u\|_2 = 1$. Then the Euler-Lagrange equation becomes

$$\forall x \in \mathcal{X},\ 2\alpha u(x) \log u(x) = u(x) - \pi(u).$$

Since the map $t \mapsto 2\alpha t \log t - t$ is convex, the function u has exactly two values, say s and t. Let $A = \{x \in \mathcal{X} : u(x) = s\}$. By Corollary 2.4, we get

$$\alpha = \frac{\mathcal{E}(u,u)}{\mathcal{L}(u)} = \frac{\pi(A)(1-\pi(A))(s-t)^2}{\pi(A)s^2 \log s^2 + (1-\pi(A))t^2 \log t^2}$$

where $\pi(A)s^2 + (1-\pi(A))t^2 = 1$. Thus, by Remark 2.8,

$$\alpha \geq \frac{1-2\pi(A)}{\log[(1-\pi(A))/\pi(A)]} \geq \frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*]},$$

where the last inequality comes from the monotonicity of the function $t \mapsto \frac{1-2t}{\log(1/t-1)}$ on the interval $[0, 1/2]$ and the fact $\pi(A) \geq \pi_*$. $\square$

3. Simple random walk on the 3-cycle. Let $(\mathbb{Z}_3, K, \pi)$ be a simple random walk on the 3-cycle, where the Markov kernel K is defined by $K(i, i+1) = K(i, i-1) = 1/2$ for $i \in \mathbb{Z}_3$ and the addition and substraction are understood modulo 3. A calculation shows that the spectral gap is 3/2. We compute the logarithmic Sobolev constant of K by considering a general case.

Corollary 2.8. ([11, Corollary A.5]) *Let $(\mathcal{X}, K, \pi)$ be a finite Markov chain, where $|\mathcal{X}| \geq 3$ and $K(x,y) = \frac{1}{|\mathcal{X}|-1}$ if $x \neq y$ and $K(x,x) = 0$ for $x \in \mathcal{X}$. Let α be the logarithmic Sobolev constant of K. Then*

$$\alpha = \frac{|\mathcal{X}|-2}{(|\mathcal{X}|-1)\log(|\mathcal{X}|-1)}.$$

In particular, the logarithmic Sobolev constant for the simple random walk on the 3-cycle is $\frac{1}{2\log 2}$.

Proof. Let $\widetilde{K}$ be a Markov kernel on $\mathcal{X}$ where $\widetilde{K}(x,y) \equiv |\mathcal{X}|^{-1}$. Let $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ be the Dirichlet forms of K and $\widetilde{K}$. Then, for any function f on $\mathcal{X}$, one has $\mathcal{E}(f,f) = \frac{|\mathcal{X}|}{|\mathcal{X}|-1}\widetilde{\mathcal{E}}(f,f)$. Since K and $\widetilde{K}$ have the same stationary distribution, the logarithmic Sobolev constants $\alpha, \widetilde{\alpha}$ of K and $\widetilde{K}$ are related by

$$\alpha = \frac{|\mathcal{X}|\widetilde{\alpha}}{|\mathcal{X}|-1} = \frac{|\mathcal{X}|-2}{(|\mathcal{X}|-1)\log(|\mathcal{X}|-1)},$$

where the last equality applies Theorem 2.4. □

4. Simple random walk on the 4-cycle.

Theorem 2.5. *Let $(\mathbb{Z}_4, K, \pi)$ be a simple random walk on a 4-cycle, where $K(i,i+1) = K(i,i-1) = 1/2$ for $i \in \mathbb{Z}_4$. Then the spectral gap is 1 and the logarithmic Sobolev constant is 1/2.*

Proof. Let K_1, K_2 be two independent Markov chains on $\{0,1\}$, where

$$K_i(0,1) = K_i(1,0) = 1,\ K_i(0,0) = K_i(1,1) = 0, \quad \forall i = 1,2.$$

Consider a uniform probability measure μ on $\{1,2\}$ and the product chain with kernel K_μ given by (2.9). Let $p : \{0,1\}^2 \to \mathbb{Z}_4$ be a bijective map defined by

$$p(0,0) = 0,\ p(0,1) = 1,\ p(1,1) = 2,\ p(1,0) = 3.$$

Then for $x, y \in \{0,1\}^2$, we have $K_\mu(x,y) = K(p(x),p(y))$. This implies that K and K_μ share the same spectral gap and logarithmic Sobolev constant. By Corollary 2.7, the spectral gap and logarithmic Sobolev constant of K_1 are 2 and 1. By Proposition 2.6, the spectral gap and logarithmic Sobolev constant are 1 and 1/2. □

Chapter 3

Logarithmic Sobolev constants for some finite Markov chains

From the computation of the logarithmic Sobolev constants in Section 2.4, one can see that different models need different tricks. In this chapter, we concentrate on the calculation of the logarithmic Sobolev constant for the simple random walks on the n-cycle. In Section 3.1, we focus on the even cycles and explicitly determine their logarithmic Sobolev constants. Thereafter, an application for collapsing a cycle is introduced. In Section 3.2, we implement another trick to determine the logarithmic Sobolev constant of the 5-cycle.

3.1 The simple random walk on an even cycle

For $n \geq 2$, consider a simple random walk on the n-cycle $\mathbb{Z}_n = \{1, 2, ..., n\}$. Clearly, the corresponding Markov kernel K_n is given by $K_n(x, x \pm 1) = \frac{1}{2}$ and the uniform distribution on $\mathbb{Z}_n$ is its unique stationary distribution.(For $n = 2$, we consider the case $K(1, 2) = K(2, 1) = 1$ and $K(1, 1) = K(2, 2) = 0$. By Corollary 2.7, $\alpha = \frac{\lambda}{2} = 1$.) Throughout this section, we assume that $n \geq 3$.

Let λ_n and α_n be the spectral gap and logarithmic Sobolev constant of K_n. It has been shown in Example 1.1 that $\lambda_n = 1 - \cos(2\pi/n)$ and in Corollary 2.8 and Theorem 2.5 that

$$\alpha_3 = \frac{1}{2 \log 2} < \frac{\lambda_3}{2} = \frac{3}{4}, \quad \alpha_4 = \frac{\lambda_4}{2} = \frac{1}{2}.$$

3.1.1 The main result

The following is the main result of this section. This is a joint work with Yuan-Chung Sheu and has been polished in [6].

Theorem 3.1. *For $n > 2$, let K_n be the Markov kernel of the simple random walk on the n-cycle. Assume that n is even. Then the spectral gap $\lambda_n = \lambda(K_n)$ and the logarithmic Sobolev constant $\alpha_n = \alpha(K_n)$ satisfy $\alpha_n = \lambda_n/2 = \frac{1}{2}(1 - \cos\frac{2\pi}{n})$.*

The following is a simple application of the above theorem.

Corollary 3.1. *For $n \geq 3$, let K_n be a Markov kernel on $\mathbb{Z}_n$ defined by $K_n(i, i-1) = p$, $K_n(i,i) = r$ and $K_n(i, i+1) = q$ for $i \in \mathbb{Z}_n$, where $p + q + r = 1$. Then the spectral gap λ_n and the logarithmic Sobolev constant α_n satisfy $\alpha_n = \lambda_n/2 = \frac{1-r}{2}(1 - \cos\frac{2\pi}{n})$*

Proof. Let $\widetilde{K}_n$ be the Markov of the simple random walk on $\mathbb{Z}_n$ and $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ be the Dirichlet forms of K_n and $\widetilde{K}_n$. Obviously, both K_n and $\widetilde{K}_n$ have the same stationary distribution, the uniform distribution on $\mathbb{Z}_n$. By Lemma 1.1, one has $\mathcal{E}(f,f) = (1-r)\widetilde{\mathcal{E}}(f,f)$ for any function f on $\mathbb{Z}_n$ and then, by definition, $\lambda_n = (1-r)\widetilde{\lambda}_n$ and $\alpha_n = (1-r)\widetilde{\alpha}_n$. □

We will prove this theorem in the next subsection. Here, we consider first the ratio $\mathcal{E}(f,f)/\mathcal{L}(f)$ and, by studying the Dirichlet form, restrict the minimizer, if any, for the logarithmic Sobolev constant to a specific class of functions. For any function $f = (f(1), ..., f(n)) = (x_1, ..., x_n)$, we have

$$\mathcal{L}(f) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \log\frac{x_i^2}{\|f\|_2^2} \tag{3.1}$$

and

$$\mathcal{E}(f,f) = \frac{1}{2n}(|x_1 - x_2|^2 + |x_2 - x_3|^2 + \cdots + |x_{n-1} - x_n|^2 + |x_n - x_1|^2). \tag{3.2}$$

It is obvious that the uniformity of the stationary distribution π_n of K_n implies the invariance of $\mathcal{L}(f)$ under the permutation of the components of f. We now investigate the extreme value of $\mathcal{E}$ over all permutations on the components of f.

Consider the function

$$F(x) = |x_1 - x_2|^2 + |x_2 - x_3|^2 + \cdots + |x_{n-1} - x_n|^2 + |x_n - x_1|^2 \tag{3.3}$$

where $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$. To every $x = (x_1, x_2, ..., x_n)$ with $0 \le x_1 \le x_2 \le \cdots \le x_n$, there corresponds an element $\widetilde{x} \in \mathbb{R}^n$ given by the formula

$$\widetilde{x} = \begin{cases} (x_1, x_3, x_5, ..., x_{2k+1}, x_{2k}, ..., x_4, x_2) & \text{if } n = 2k+1 \\ (x_1, x_3, x_5, ..., x_{2k-1}, x_{2k}, ..., x_4, x_2) & \text{if } n = 2k. \end{cases} \tag{3.4}$$

Denote by S_n the set of all permutations on $\{1, 2, ..., n\}$ and write $\theta x = (x_{\theta(1)}, x_{\theta(2)}, ..., x_{\theta(n)})$ for $\theta \in S_n$ and $x \in \mathbb{R}^n$.

Proposition 3.1. *For every $x = (x_1, x_2, ..., x_n)$ with $0 \le x_1 \le x_2 \le \cdots \le x_n$, we have $F(\theta x) \ge F(\widetilde{x})$ for all $\theta \in S_n$.*

Proof. We prove this by induction on n. There is nothing to prove in the case $n = 2$. Assume that it is also true for $n = k$. We consider the case $n = k+1$ and fix $x = (x_1, x_2, ..., x_{k+1})$ where $0 \le x_1 \le x_2 \le \cdots \le x_{k+1}$.

Step1. Set $y = (x_1, x_2, ..., x_k)$ and consider the corresponding vector $\widetilde{y}$ given by (3.4). For every $i = 1, 2, ..., k-2$, set

$$\widetilde{y}_{i,i+2} = \begin{cases} (x_1, x_3, ..., x_i, x_{k+1}, x_{i+2}, ..., x_4, x_2) & \text{if } i \text{ is odd} \\ (x_1, x_3, ..., x_{i+2}, x_{k+1}, x_i, ..., x_4, x_2) & \text{if } i \text{ is even.} \end{cases} \tag{3.5}$$

Thus $\widetilde{y}_{i,i+2}$ is obtained by inserting x_{k+1} in $\widetilde{y}$ between x_i and x_{i+2}. We also set

$$\widetilde{y}_{1,2} = (x_1, x_3, ..., x_4, x_2, x_{k+1})$$

and

$$\widetilde{y}_{k-1,k} = \begin{cases} (x_1, x_3, ..., x_k, x_{k+1}, x_{k-1}, ..., x_4, x_2) & \text{if } k \text{ is odd} \\ (x_1, x_3, ..., x_{k-1}, x_{k+1}, x_k, ..., x_4, x_2) & \text{if } k \text{ is even.} \end{cases} \tag{3.6}$$

We claim that

$$F(\widetilde{y}_{1,2}) \geq F(\widetilde{y}_{k-1,k}) \tag{3.7}$$

and

$$F(\widetilde{y}_{i,i+2}) \geq F(\widetilde{y}_{k-1,k}) \text{ for all } i = 1, 2, ..., k-2. \tag{3.8}$$

Note that for $1 \leq i \leq k-2$, a simple computation shows

$$F(\widetilde{y}_{i,i+2}) = F(\widetilde{y}) + (x_i - x_{k+1})^2 + (x_{k+1} - x_{i+2})^2 - (x_i - x_{i+2})^2. \tag{3.9}$$

Therefore for $1 \leq i \leq k-4$, we get

$$\begin{aligned} F(\widetilde{y}_{i,i+2}) - F(\widetilde{y}_{i+2,i+4}) &= [(x_i - x_{k+1})^2 + (x_{k+1} - x_{i+2})^2 - (x_i - x_{i+2})^2] \\ &\quad - [(x_{i+2} - x_{k+1})^2 + (x_{k+1} - x_{i+4})^2 - (x_{i+2} - x_{i+4})^2] \\ &= 2(x_{k+1} - x_{i+2})(x_{i+4} - x_i) \geq 0. \end{aligned} \tag{3.10}$$

Besides, we also have

$$\begin{aligned} F(\widetilde{y}_{k-2,k}) - F(\widetilde{y}_{k-1,k}) &= [(x_{k+1} - x_{k-2})^2 + (x_{k+1} - x_k)^2 - (x_{k-2} - x_k)^2] \\ &\quad - [(x_{k+1} - x_{k-1})^2 + (x_{k+1} - x_k)^2 - (x_k - x_{k-1})^2] \\ &= 2(x_{k+1} - x_k)(x_{k-1} - x_{k-2}) \geq 0 \end{aligned} \tag{3.11}$$

and

$$F(\widetilde{y}_{k-3,k-1}) - F(\widetilde{y}_{k-1,k}) = 2(x_{k+1} - x_{k-1})(x_k - x_{k-3}) \geq 0. \tag{3.12}$$

Combining (3.10), (3.11) and (3.12) gives (3.8). To prove (3.7), it suffices to show that $F(\widetilde{y}_{1,2}) \geq F(\widetilde{y}_{1,3})$, whereas this follows easily from the following computation.

$$\begin{aligned} F(\widetilde{y}_{1,2}) - F(\widetilde{y}_{1,3}) &= [(x_1 - x_{k+1})^2 + (x_{k+1} - x_2)^2 - (x_1 - x_2)^2] \\ &\quad - [(x_1 - x_{k+1})^2 + (x_{k+1} - x_3)^2 - (x_1 - x_3)^2] \\ &= 2(x_{k+1} - x_1)(x_3 - x_2) \geq 0. \end{aligned}$$

Step2. We prove that for every $\theta \in S_{n+1}$,

$$F(\theta x) \geq F(\widetilde{y}_{k-1,k}) = F(\widetilde{x}). \tag{3.13}$$

Fix $\theta \in S_{n+1}$ and set $c = \theta x$. It loses no generality to write $c = (..., x_i, x_{k+1}, x_j, ...)$ for some $i < j$ and let $z = (..., x_i, x_j, ...) \in \mathbb{R}^n$ be obtained by removing the component x_{k+1} from the vector c. Then, for $1 \leq j \leq k-2$, we have

$$\begin{aligned} &F(c) - F(\widetilde{y}_{j,j+2}) \\ &\quad = [F(z) + (x_i - x_{k+1})^2 + (x_j - x_{k+1})^2 - (x_i - x_j)^2] \\ &\qquad - [F(\widetilde{y}) + (x_j - x_{k+1})^2 + (x_{k+1} - x_{j+2})^2 - (x_j - x_{j+2})^2] \\ &\quad = F(z) - F(\widetilde{y}) + 2(x_{k+1} - x_j)(x_{j+2} - x_i) \geq 0. \end{aligned} \tag{3.14}$$

(The last inequality applies the inductive assumption that $F(z) \geq F(\widetilde{y})$.) For $j = k-1$, we have

$$\begin{aligned} &F(c) - F(\widetilde{y}_{k-1,k}) \\ &\quad = [F(z) + (x_i - x_{k+1})^2 + (x_{k-1} - x_{k+1})^2 - (x_i - x_{k-1})^2] \\ &\quad - [F(\widetilde{y}) + (x_k - x_{k+1})^2 + (x_{k+1} - x_{k-1})^2 - (x_k - x_{k-1})^2] \\ &\quad = F(z) - F(\widetilde{y}) + 2(x_k - x_i)(x_{k+1} - x_{k-1}) \geq 0. \end{aligned}$$

For $j = k$, we have

$$\begin{aligned} F(c) - F(\widetilde{y}_{k-1,k}) \\ &= [F(z) + (x_k - x_{k+1})^2 + (x_i - x_{k+1})^2 - (x_i - x_k)^2] \\ &\quad - [F(\widetilde{y}) + (x_k - x_{k+1})^2 + (x_{k+1} - x_{k-1})^2 - (x_k - x_{k-1})^2] \\ &= F(z) - F(\widetilde{y}) + 2(x_{k-1} - x_i)(x_{k+1} - x_k) \geq 0. \end{aligned} \tag{3.15}$$

Therefore (3.13) follows from (3.14)-(3.15) and (3.8). □

The following is a consequence of Proposition 3.1 and is critical in computing the logarithmic Sobolev constant of the simple random walk on the n-cycle.

Corollary 3.2. *For $n \geq 3$, let α_n be the logarithmic Sobolev constant of the simple random walk on the n cycle. Assume that there exists a positive non-constant function f such that $\alpha_n = \mathcal{E}(f,f)/\mathcal{L}(f)$. Let $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n$ be the components of f and $\widetilde{f} = (x_1, x_3, ..., x_4, x_2)$. Then the Euler-Lagrange equation* (2.3) *is satisfied with $\alpha = \alpha_n$ and $u = \widetilde{f}$. Furthermore, $\alpha_n = \mathcal{E}(\widetilde{f},\widetilde{f})/\mathcal{L}(\widetilde{f})$.*

3.1.2 Proof of Theorem 3.1

In this subsection, we dedicate in proving Theorem 3.1. Throughout this section, n is even and $n \geq 4$. The way we prove Theorem 3.1 is first to verify by contradiction that there is no positive non-constant function u and $\alpha < \frac{\lambda_n}{2}$ satisfying the Euler-Lagrange equation (2.3). Our main result then follows from Corollary 2.4. Before starting to prove the main result, we derive a series of lemmas using combinatorial arguments.

Define the shifting operator σ by

$$\sigma(x_1, x_2, ..., x_n) = (x_n, x_1, x_2, ..., x_{n-1}), \tag{3.16}$$

where $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$. Set $\sigma^j(x) = \sigma(\sigma^{j-1}(x))$ for $j \geq 2$ and write σ^{-j} for the inverse of σ^j.

Lemma 3.1. *Consider a vector of the form*

$$u = (x_1, x_3, ..., x_{2k-1}, x_{2k}, ..., x_4, x_2)$$

where $x_1 \leq x_2 \leq ... \leq x_{2k}$ *and write* $\sigma^j(u) = ((\sigma^j(u))_1, (\sigma^j(u))_2, ..., (\sigma^j(u))_{2k})$. *Then for every* $1 \leq j \leq k-1$, *we have*

$$(\sigma^j(u))_i \leq (\sigma^j(u))_{2k-i+1}, \quad \text{for } i = 1, ..., k \tag{3.17}$$

and

$$(\sigma^{-j}(u))_i \geq (\sigma^{-j}(u))_{2k-i+1}, \quad \text{for } i = 1, ..., k. \tag{3.18}$$

Proof. Assume $1 \leq j \leq k-1$. Then we have

$$(\sigma^j(u))_i = \begin{cases} x_{2(j-i+1)}, & \text{if } 1 \leq i \leq j; \\ x_{2(i-j)-1}, & \text{if } j+1 \leq i \leq j+k; \\ x_{2k-2[i-(j+k+1)]}, & \text{if } j+k+1 \leq i \leq 2k. \end{cases}$$

Case 1: $1 \leq i \leq j \wedge (k-j)$**.** Since $i \leq (k-j)$, we get $2k-i+1 \geq k+j+1$ and $(\sigma^j(u))_{2k-i+1} = x_{2(i+j)}$, which implies

$$(\sigma^j(u))_i = x_{2(j-i+1)} \leq x_{2(i+j)} = (\sigma^j(u))_{2k-i+1}.$$

Case 2: $j \vee (k-j) < i \leq k$**.** Note that $(k-j) < i \leq k$ implies $k+1 \leq (2k-i+1) \leq (k+j)$. Hence, we have

$$(\sigma^j(u))_i = x_{2(i-j)-1}, \quad (\sigma^j(u))_{2k-i+1} = x_{2(2k-i-j)+1}.$$

Since $2(2k-i-j)+1 \geq 2(i-j)-1$, we get $(\sigma^j(u))_i \leq (\sigma^j(u))_{2k-i+1}$.

Case 3: $j \wedge (k-j) < i \le j \vee (k-j)$**.** It is obvious that only the situation $j \neq k-j$ is needed to be considered. On one hand, if $j < k-j$, then $j < i \le (k-j)$ and $2k - i + 1 \ge j - k + 2k + 1 = k + j + 1$. This implies

$$(\sigma^j(u))_i = x_{2(i-j)-1} \le x_{2(i+j)} = (\sigma^j(u))_{2k-i+1}.$$

On the other hand, if $k - j < j$, then $k - j < i \le j$. By this fact, we have

$$(\sigma^j(u))_i = x_{2(j-i+1)} \le x_{2(2k-i-j)+1} = (\sigma^j(u))_{2k-i+1}.$$

Combining all above proves (3.17). The proof of (3.18) can be done by similar arguments. □

Lemma 3.2. *Let $u = (u_1, u_2, ..., u_{2k-1}, u_{2k})$ be a vector with $u_i > 0$ for all $1 \le i \le 2k$. Assume that there exist two positive constants, c and d, such that*

$$2u_i - (u_{i-1} + u_{i+1}) = cu_i \log du_i^2 \tag{3.19}$$

for all $i = 1, ..., 2k$.(Here we write $u_0 = u_{2k}$ and $u_{2k+1} = u_1$.) Then:

(a) If $u_i \le u_{2k-i+1}$ for all $1 \le i \le k$, then we have

$$u_1^2 - u_{2k}^2 + u_k^2 - u_{k+1}^2 \ge c[(u_1^2 + \cdots + u_k^2) - (u_{k+1}^2 + \cdots + u_{2k}^2)].$$

(b) If $u_i \ge u_{2k-i+1}$ for all $1 \le i \le k$, then we have

$$u_{2k}^2 - u_1^2 + u_{k+1}^2 - u_k^2 \ge c[(u_{k+1}^2 + \cdots + u_{2k}^2) - (u_1^2 + \cdots + u_k^2)].$$

Proof. For (a), assume that $u_i \le u_{2k-i+1}$ for all $1 \le i \le k$. For every $1 \le i \le k$, we rewrite (3.19) as

$$2 - \frac{u_{i-1} + u_{i+1}}{u_i} = c \log du_i^2.$$

Then a simple computation implies

$$\begin{aligned}
&\frac{u_{2k-i}+u_{2k-i+2}}{u_{2k-i+1}} - \frac{u_{i-1}+u_{i+1}}{u_i} \\
&\qquad = \frac{u_i(u_{2k-i}+u_{2k-i+2}) - u_{2k-i+1}(u_{i-1}+u_{i+1})}{u_i u_{2k-i+1}} \\
&\qquad = c(2\log\frac{u_i}{u_{2k-i+1}}) \geq c(\frac{u_i}{u_{2k-i+1}} - \frac{u_{2k-i+1}}{u_i}),
\end{aligned} \tag{3.20}$$

where the last inequality uses the fact that $2\log t \geq t - \frac{1}{t}$ for every $0 < t \leq 1$. Hence, by (3.20), we have

$$(u_i u_{2k-i+2} - u_{i-1}u_{2k-i+1}) + (u_i u_{2k-i} - u_{i+1}u_{2k-i+1}) \geq c(u_i^2 - u_{2k-i+1}^2)$$

for all $i = 1, ..., k$. The desired identity is obtained by summing up both sides of the above inequality over all $1 \leq i \leq k$.

For (b), assume that $u_i \geq u_{2k-i+1}$ for all $1 \leq i \leq k$. For every i, set $v_i = u_{2k-i+1}$. Then our result follows by applying (a) to the vector $v = (v_1, v_2, ..., v_{2k})$. □

Lemma 3.3. *Consider the following $k \times k$ matrices:*

$$A = \begin{pmatrix}
2 & 1 & 0 & \cdots & \cdots & 0 \\
1 & 2 & 1 & \ddots & & \vdots \\
0 & 1 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 2 & 1 & 0 \\
\vdots & & \ddots & 1 & 2 & 1 \\
0 & \cdots & \cdots & 0 & 2 & 2
\end{pmatrix}$$

and

$$B = \begin{pmatrix} 2 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 2 & 1 & \ddots & & \vdots \\ 0 & 1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 2 & 1 & 0 \\ \vdots & & \ddots & 1 & 2 & 1 \\ 0 & \cdots & \cdots & 0 & 1 & 1 \end{pmatrix}.$$

Then:

(a) *If* $t < 2(1 - \cos\frac{\pi}{2k})$, *then* $P_A(t) = \det(A - tI) > 0$.

(b) *If* $t < 2(1 - \cos\frac{\pi}{2k+1})$, *then* $P_B(t) = \det(B - tI) > 0$.

Proof. For (a), let $\theta_l = \frac{(2l-1)\pi}{2k}$ for $1 \le l \le k$ and

$$v_l = \begin{pmatrix} \sin\theta_l \\ \sin 2\theta_l \\ \vdots \\ \sin k\theta_l \end{pmatrix}.$$

Routine calculation shows that $Av_l = 2(1 + \cos\theta_l)v_l$ for $1 \le l \le k$. Therefore $\{2(1+\cos\theta_l) | 1 \le l \le k\}$ is the set of all real roots of the characteristic polynomial $P_A(t)$. Note that $(-t)^k$ is the highest order term of $P_A(t)$. This implies that $\lim_{t\to-\infty} P_A(t) = \infty$. Since $2(1 - \cos\frac{\pi}{2k})$ is the smallest real root of $P_A(t)$, we observe that $P_A(t) > 0$ for all $t < 2(1 - \cos\frac{\pi}{2k})$.

The proof of (b) is the same as that of (a) except the replacement of θ_l with $\frac{2l\pi}{2k+1}$. □

Lemma 3.4. *(a) Consider the following system of inequalities:*

$$\begin{cases} A_j - A_{j+1} \geq 4t(A_1 + \cdots + A_j), j = 1, ..., k-1 \\ A_k \geq 2t(A_1 + \cdots + A_k). \end{cases} \tag{3.21}$$

If $t < \frac{1}{2}(1 - \cos\frac{\pi}{2k})$*, then the system* (3.21) *has no solution* $(A_1, A_2, ..., A_k)$ *with* $A_1 < 0$.

(b) Consider the following system of inequalities:

$$\begin{cases} A_j - A_{j+1} \geq 4t(A_1 + \cdots + A_j), j = 1, ..., k-1 \\ A_k \geq 4t(A_1 + \cdots + A_k). \end{cases} \tag{3.22}$$

If $t < \frac{1}{2}(1 - \cos\frac{\pi}{2k+1})$*, then the system* (3.22) *has no solution* $(A_1, A_2, ..., A_k)$ *with* $A_1 < 0$.

Proof. For (a), let $f_1(t) = 2 - 4t$ and $g_1(t) = 4t$. For every $1 \leq l \leq k-1$, put

$$f_{l+1}(t) = (1 - 4t)f_l(t) - g_l(t) \tag{3.23}$$

and

$$g_{l+1}(t) = 4tf_l(t) + g_l(t). \tag{3.24}$$

Clearly, (3.23) and (3.24) imply

$$g_{l+1}(t) - g_l(t) = 4tf_l(t) = f_l(t) - g_l(t) - f_{l+1}(t).$$

Hence we have $f_l(t) = g_{l+1}(t) + f_{l+1}(t)$ for $1 \leq l \leq k-1$. By this fact, we obtain, for $2 \leq l \leq k-1$,

$$\begin{aligned} f_{l+1}(t) &= (2 - 4t)f_l(t) - (f_l(t) + g_l(t)) \\ &= (2 - 4t)f_l(t) - f_{l-1}(t). \end{aligned}$$

Note that

$$f_1(t) = 2 - 4t, \quad f_2(t) = (1-4t)f_1(t) - g_1(t) = (2-4t)^2 - 2,$$

and, therefore,

$$f_l(t) = \det(M_l - 4tI_l), \quad 1 \le l \le k \tag{3.25}$$

where I_l is the $l \times l$ identity matrix and M_l is the $l \times l$ matrix of the same form as that in Lemma 3.3(a).

Assume that $t < \frac{1}{2}(1 - \cos\frac{\pi}{2k})$ and $(A_1, A_2, ..., A_k)$ satisfies the system of inequalities (3.21). Since $t < \frac{1}{2}(1 - \cos\frac{\pi}{2l})$ for $1 \le l \le k$, Lemma 3.3(a) and (3.25) imply that $f_l(t) > 0$ for all $l = 1, 2, ..., k$.

For $1 \le i \le k-1$, we have, by (3.21),

$$A_{k-i} - A_{k-i+1} \ge 4t(A_1 + \cdots + A_{k-i}).$$

We claim that

$$f_j(t)A_{k-j+1} \ge g_j(t)(A_1 + \cdots + A_{k-j}), \quad \forall 1 \le j \le k. \tag{3.26}$$

Clearly (3.26) holds for $j = 1$. Suppose that it also holds for some i with $1 \le i \le k-1$. Since $f_i(t) > 0$, we get

$$\begin{aligned} f_i(t)A_{k-i} &= f_i(t)(A_{k-i} - A_{k-i+1}) + f_i(t)A_{k-i+1} \ge (4tf_i(t) + g_i(t))(A_1 + \cdots + A_{k-i}) \\ &= g_{i+1}(t)(A_1 + \cdots + A_{k-i-1}) + (4tf_i(t) + g_i(t))A_{k-i}. \end{aligned}$$

The above inequality implies that (3.26) also holds for $j = i+1$ and hence is true for $1 \le j \le k$. Plugging $j = k$ into (3.26) gives $f_k(t)A_1 \ge 0$. Since $f_k(t) > 0$, we have $A_1 \ge 0$. This proves part (a).

The same line of reasoning as above applies for part (b) and the proof goes word for word except the replacement of $f_1(t)$ with $1 - 4t$. □

Proof of Theorem 3.1. By Corollary 2.4, it suffices to show that there is no positive non-constant function u and $0 < \beta < \lambda_n/2$ satisfying (2.4). We prove this fact by contradiction. Suppose the inverse, that is, (2.4) is satisfied for some $\beta < \frac{\lambda}{2} = \frac{1}{2}(1 - \cos\frac{2\pi}{n})$ and a positive non-constant function u. Without loss of generality, we assume $\sum_i (u(i))^2 = 1$. By Corollary 3.2, we may assume further that $u = (x_1, x_3, ..., x_{n-1}, x_n, ..., x_4, x_2)$, where $0 < x_1 \le x_2 \le \cdots \le x_n$ and $x_1 < x_n$. In this case, the Euler-Lagrange equation in (2.4) is given by

$$2x_i - (x_i^{(1)} + x_i^{(2)}) = 2\beta x_i \log n x_i^2, \quad 1 \le i \le n.$$

where $x_i^{(1)}$ and $x_i^{(2)}$ are the two nearest neighbors of x_i.

Recall the shifting operator σ defined in (3.16) and $\sigma^j = \sigma(\sigma^{j-1})$ for $j \ge 2$. Note that we may write $n = 4k$ or $n = 4k + 2$. For $j = 1, ..., k$, we have

$$\sigma^j(f) = (x_{2j}, ..., x_2, x_1, ..., x_{n-2j-1}, x_{n-2j+1}, ..., x_{n-1}, x_n, ..., x_{2j+2})$$

and

$$\sigma^{-j}(f) = (x_{2j+1}, ..., x_{n-1}, x_n, ..., x_{n-2j+2}, x_{n-2j}, ..., x_2, x_1, ..., x_{2j-1}).$$

By Lemma 3.1 and Lemma 3.2(a), we get

$$\begin{aligned}(x_{2j}^2 &- x_{2j+2}^2 + x_{n-2j-1}^2 - x_{n-2j+1}^2)\\ &\ge 2\beta[(x_2^2 + x_4^2 + \cdots + x_{2j}^2 + x_1^2 + x_3^2 + \cdots + x_{n-2j-1}^2)\\ &\quad - (x_{n-2j+1}^2 + x_{n-2j+3}^2 + \cdots + x_{n-1}^2 + x_{2j+2}^2 + x_{2j+4}^2 + \cdots + x_n^2)].\end{aligned}$$

Similarly Lemma 3.1 and Lemma 3.2(b) imply that

$$\begin{aligned}(x_{2j-1}^2 &- x_{2j+1}^2 + x_{n-2j}^2 - x_{n-2j+2}^2)\\ &\ge 2\beta[(x_1^2 + x_3^2 + \cdots + x_{2j-1}^2 + x_2^2 + x_4^2 + \cdots + x_{n-2j}^2)\\ &\quad - (x_{2j+1}^2 + x_{2j+3}^2 + \cdots + x_{n-1}^2 + x_{n-2j+2}^2 + x_{n-2j+4}^2 + \cdots + x_n^2)].\end{aligned}$$

Note that $n-2j-1 \geq 2j+1$ and $n-2j \geq 2j+2$ for $1 \leq j \leq k$. Summing up the above two inequalities gives

$$\begin{aligned}
&(x_{2j-1}^2 + x_{2j}^2 - x_{2j+1}^2 - x_{2j+2}^2) + (x_{n-2j-1}^2 + x_{n-2j}^2 - x_{n-2j+1}^2 - x_{n-2j+2}^2) \\
&\geq 4\beta[(x_1^2 + x_2^2 + \cdots + x_{2j}^2) - (x_{n-2j+1}^2 + x_{n-2j+2}^2 + \cdots + x_n^2)].
\end{aligned}$$

Letting $A_i = x_{2i-1}^2 + x_{2i}^2 - x_{n-2i+1}^2 - x_{n-2i+2}^2$ for $1 \leq i \leq k$ implies, for $n = 4k$,

$$\begin{cases} A_j - A_{j+1} \geq 4\beta(A_1 + A_2 + \cdots + A_j), \quad j = 1, ..., k-1 \\ A_k \geq 2\beta(A_1 + A_2 + \cdots + A_k) \end{cases}$$

and for $n = 4k+2$,

$$\begin{cases} A_j - A_{j+1} \geq 4\beta(A_1 + A_2 + \cdots + A_j), \quad j = 1, ..., k-1 \\ A_k \geq 4\beta(A_1 + A_2 + \cdots + A_k) \end{cases}$$

Note that $\beta < \frac{1}{2}(1 - \cos\frac{2\pi}{n})$ and $A_1 = x_1^2 + x_2^2 - x_{n-1}^2 - x_n^2 \leq x_1^2 - x_n^2 < 0$. This contradicts Lemma 3.4. □

3.1.3 An application: collapse of cycles and product of sticks

In this section, we discuss some applications of Theorem 3.1. This is a joint work with Laurent Saloff-Coste and Wai-Wai Liu in [5]. We first consider the following two ways of collapsing even cycles.

1. Collapsing the $2n$-cycle to the n-stick with loops at the ends. Fix $n \geq 2$ and let K_1 and K_2 be Markov kernels on $\mathbb{Z}_n$ and $\mathbb{Z}_{2n}$ defined by

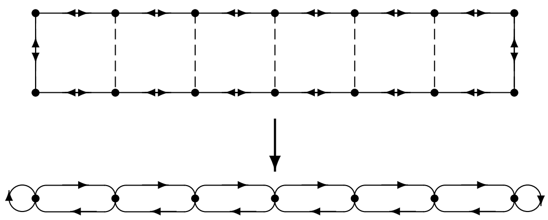
$$K_1(0,0) = K_1(n-1, n-1) = K_1(i, i+1) = K_1(i+1, i) = 1/2,$$

for all $i = 0, ..., n-2$, and

$$K_2(i, i+1) = K_2(i, i-1) = 1/2, \quad \forall i = 0, ..., 2n-1.$$

Let $p : \mathbb{Z}_{2n} \to \mathbb{Z}_n$ be a surjective map defined by $p(i) = p(2n-1-i) = i$ for $i = 0, ..., n-1$. A simple computation(checking the requirement in Proposition 2.5) shows that the Markov kernel K_2 collapses to K_1 via the projection p. See Figure 3.1.

Figure 3.1: The 14-cycle collapses to the 7-stick with loops at the ends. All edges have weight 1/2.



Let λ_1 and λ_2 be the spectral gaps of K_1 and K_2. By Proposition 2.5, $\lambda_2 \le \lambda_1$. It has been shown in Example 1.1 that $\lambda_2 = 1 - \cos\frac{\pi}{n}$. To see λ_1, note that K_2 has eigenvalue $1 - \lambda_2$ with multiplicity 2 and the two dimensional eigenspace contains the function $f(x) = \cos(\frac{\pi}{n}(x + \frac{1}{2}))$, which has the property $f(x) = f(2n-1-x)$. By letting $g(x) = f(x)$ for $0 \le x \le n-1$, one has

$$\mathcal{E}_2(f, f) = \mathcal{E}_2(g \circ p, g \circ p) = \mathcal{E}_1(g, g),$$

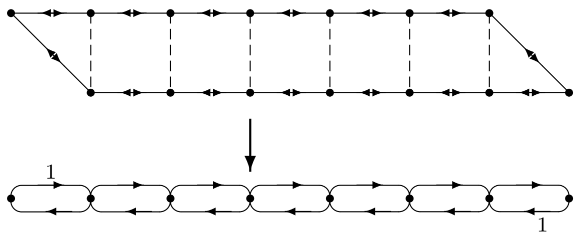
and $\mathrm{Var}_{\pi_2}(f) = \mathrm{Var}_{\pi_1}(g)$. Then, by the minmax theorem, $\lambda_1 \le \lambda_2$.

Proposition 3.2. *Fix $n \ge 2$. Let K be a Markov kernel on $\mathbb{Z}_n$ given by $K(0,0) = K(n-1, n-1) = K(i, i+1) = K(i+1, i) = 1/2$ for all $0 \le i \le n-2$. Then the spectral gap λ and the logarithmic Sobolev constant α satisfy $2\alpha = \lambda = 1 - \cos\frac{\pi}{n}$.*

2. Collapsing the $2n$-cycle to the $n+1$-stick with reflecting barriers. Fix $n \ge 2$ and let K_2 be the simple random walk on $\mathbb{Z}_{2n}$. Consider a Markov kernel K_1

on $\mathbb{Z}_{n+1}$ given by $K_1(0,1) = K_1(n, n-1) = 1$ and $K_1(i, i+1) = K_1(i, i-1) = 1/2$ for all $1 \leq i \leq n-1$. Let $p : \mathbb{Z}_{2n} \to \mathbb{Z}_{n+1}$ be a map defined by $p(i) = p(2n-i) = i$ for $0 \leq i \leq n$. Then K_1 is obtained by collapsing K_2 through the projection p. See Figure 3.2.

Figure 3.2: The 14-cycle collapses to a 8-stick with reflecting barriers. All edges have weight 1/2 except those marked which have weight 1.



It is easy to check that $f(x) = \cos\frac{\pi x}{n}$ is an eigenvector of K_2 with corresponding eigenvalue $\cos\frac{\pi}{n}$. The same line of reasoning as in case 1 implies that both K_1 and K_2 have the same spectral gap. Then, by Proposition 2.5 and Theorem 3.1, we have the following proposition.

Proposition 3.3. *Let $n \geq 2$ and K be a Markov kernel on $\mathbb{Z}_n$ defined by $K(0,1) = K(n-1, n-2) = 1$ and $K(i, i+1) = K(i, i-1) = 1/2$ for all $1 \leq i \leq n-2$. Then the spectral gap λ and the logarithmic Sobolev constant α are given by $2\alpha = \lambda = 1 - \cos\frac{\pi}{n-1}$.*

Proof. Note that the case $n = 2$ is part of the result in Corollary 2.7 and the case $n > 2$ is given by the discussion in front of this proposition. □

3. Product of sticks. In this case, we consider an application of Proposition 3.2. Fix $d \geq 1$ and let $b = (b_1, ..., b_d)$ be an integer vector, where $b_i \geq 2$ for all

$1 \le i \le d$. In $\mathbb{Z}^d$ with basis $\{e_1, ..., e_d\}$, consider a rectangular box

$$R_b = \{x = (x_1, ..., x_d) \in \mathbb{Z}^d : x_i \in \{1, ..., b_i\},\ 1 \le i \le d\}. \tag{3.27}$$

The first application deals with a Markov kernel K on R_b, where

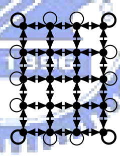
$$\forall x, y \in R_b,\ x \neq y, \quad K(x, y) = \frac{1}{2d}, \tag{3.28}$$

and

$$\forall x \in R_d, \quad K(x, x) = 1 - \frac{1}{2d} \sum_{i=1}^{d} \left(K(x, x + e_i) + K(x, x - e_i) \right), \tag{3.29}$$

where, in the summation, $K(x, y) = 0$ if $y \notin R_d$. See Figure 3.3 for an example with $d = 2$ and $(b_1, b_2) = (4, 5)$.

Figure 3.3: The box R_b with its Dirichlet form structure, $b = (b_1, b_2) = (4, 5)$. All edges have weight 1/4 except the corner loops which have weight 1/2. The stationary measure is uniform.



It is obvious that K is a direct product of Markov chains $\{(\mathbb{Z}_{b_i}, K_i, \pi_i)\}_1^d$ through the formula (2.9) with $\mu(i) = \frac{1}{d}$ for $1 \le i \le d$ and

$$K_i(0, 0) = K_i(b_i - 1, b_i - 1) = K_i(j, j + 1) = K_i(j + 1, j) = \frac{1}{2},$$

for all $0 \le j \le b_i - 2$. By Proposition 2.6, one may generalize Proposition 3.2 as follows.

Theorem 3.2. *Let $d \ge 1$ be an integer and $b = (b_1, ..., b_d)$ be an integer vector with $2 \le b_1 \le \cdots \le b_d$. Let R_b be a subset of $\mathbb{Z}^d$ defined in (3.27) and K be a*

Markov kernel on R_b given by (3.28) *and* (3.29)*. Then the spectral gap λ and the logarithmic Sobolev constant α of K satisfy*

$$2\alpha = \lambda = \frac{1-\cos(\pi/b_d)}{d}.$$

3.2 The simple random walk on the 5-cycle

Referring to Theorem 3.1 and Corollary 2.7, the logarithmic Sobolev constant for the simple random walk on an even cycle is a half of the spectral gap but this is not true for the simple random walk on the 3-cycle. It is not sure how the spectral gap and the logarithmic Sobolev constant are related if the simple random walk is considered on an odd cycle. A numerical result for the cases $n = 5, 7$ and 9, where n denotes the n-cycle, shows that the logarithmic Sobolev constant should be a half of the spectral gap. However, a mathematical proof is not available yet. The goal of this section is to clarify the fact $\alpha = \lambda/2$ for the case $n = 5$, whereas a similar proof is proposed by Wai-Wai Liu, Laurent Saloff-Coste and the author of this dissertation.

Theorem 3.3. *Let K be the Markov kernel of the simple random walk on the 5-cycle and λ and α be the spectral gap and the logarithmic Sobolev constant of K. Then $2\alpha = \lambda = 1 - \cos\frac{2\pi}{5}$.*

Remark 3.1. In the section, what will be proved is a stronger result than the above theorem which says $\mathcal{E}(f,f) \geq \frac{\lambda}{2}\mathcal{L}(f)$ for all functions f and the equality holds if and only if f is constant.

Before proving this theorem, we consider the following application.

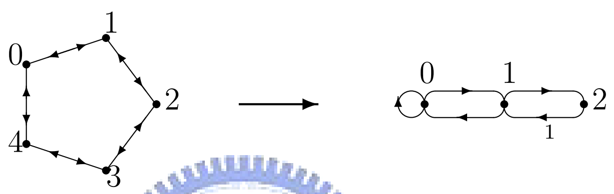
Corollary 3.3. *Let $\widetilde{K}$ be a Markov kernel on $\mathbb{Z}_3$ given by*

$$\widetilde{K}(0,0) = \widetilde{K}(0,1) = \widetilde{K}(1,2) = \widetilde{K}(1,0) = 1/2, \quad \widetilde{K}(2,1) = 1,$$

and $\widetilde{\lambda}$ *and* $\widetilde{\alpha}$ *be the spectral gap and logarithmic Sobolev constant of* $\widetilde{K}$*. Then* $2\widetilde{\alpha} = \widetilde{\lambda} = 1 - \cos\frac{2\pi}{5}$.

Proof. Let K be the Markov kernel of the simple random walk on the 5-cycle with spectral gap λ and logarithmic Sobolev constant α. Consider the map $p : \mathbb{Z}_5 \to \mathbb{Z}_3$ defined by $p(i) = p(4-i) = i$ for $i = 0, 1, 2$. It is clear that K collapses to $\widetilde{K}$ through p. See Figure 3.4.

Figure 3.4: The 5 cycle collapses to the 3-point stick with a loop at one end. All edges have weight 1/2 except marked otherwise.



Note that $f(x) = \cos(\frac{\pi}{n}(x + \frac{1}{2}))$ is an eigenfunction of K corresponding to the eigenvalue $1-\lambda$. It is easy to see that the function $f|_{\{0,1,2\}}$ is also an eigenfunction of $\widetilde{K}$. Thus $\lambda = \widetilde{\lambda} = 1 - \cos\frac{2\pi}{5}$ and the identity $\widetilde{\alpha} = \widetilde{\lambda}/2$ is then proved by Theorem 3.3 and Proposition 2.5. □

To prove Theorem 3.3, we need the following two lemmas.

Lemma 3.5. *Consider the function* $g_\beta(t) = 2t - 4\beta t \log t$ *for* $t > 0$ *and* $g_\beta(0) = 0$*. Assume that* $\beta > 0$*. Then for* $0 \le s < t < \infty$*, one has*

$$g_\beta(t) - g_\beta(s) > (t-s)\left[2 - 4\beta - 4\beta \log\left(\frac{t+s}{2}\right)\right].$$

Proof. Fix $t > 0$, $\beta \geq 0$ and let h be a function on $[0,t]$ defined by

$$h(s) = \frac{1}{4\beta}\left\{g_\beta(t) - g_\beta(s) - (t-s)\left[2 - 4\beta - 4\beta\log\left(\frac{t+s}{2}\right)\right]\right\}$$
$$= s\log s - t\log t - (s-t)\left[1 + \log\left(\frac{s+t}{2}\right)\right]$$

Then the first derivative of h is given by

$$h'(s) = 1 - \frac{2s}{s+t} + \log\left(\frac{2s}{s+t}\right) < 0, \quad \forall s \in (0,t),$$

where the inequality uses the fact $\log u < u - 1$ for $u > 0$ and $u \neq 1$. This implies that h is strictly decreasing in $[0,t]$ and hence proves this lemma since $h(t) = 0$. □

Lemma 3.6. *For $\beta > 0$, let g_β be the function defined in Lemma 3.5 and D_β be the following region*

$$D_\beta = \{(s,t) : 0 \leq s \leq t,\ s+t \leq 2,\ 0 \leq g_\beta(s) - t\}. \tag{3.30}$$

Consider the following function

$$F_\beta(s,t) = g_\beta(g_\beta(t) - s) - g_\beta(g_\beta(s) - t) - (t-s), \quad \forall (s,t) \in D_\beta.$$

Assume that $0 < \beta \leq \frac{1}{2}(1 - \cos\frac{2\pi}{5})$, then $F_\beta(s,t) \geq 0$ on D_β and the equality holds if and only if $s = t$.

Remark 3.2. Note that $F(\beta, t)$ is well-defined on D_β since one has, by Lemma 3.5,

$$(g_\beta(t) - s) - (g_\beta(s) - t) \geq (t-s)\left[3 - 4\beta - 4\beta\log\left(\frac{t+s}{2}\right)\right] \geq 0,$$

for all $t + s \leq 2$ and $0 \leq s \leq t$.

Proof. Obviously, $F_\beta(t,t) = 0$ for all $t \geq 0$ and $\beta > 0$. We now assume that $0 < \beta \leq \frac{1}{2}(1 - \cos\frac{2\pi}{5})$ and the pair $(s,t) \in D_\beta$ satisfies $s < t$. It remains to show $F_\beta(s,t) > 0$. By Lemma 3.5, one has

$$(g_\beta(t) - s) - (g_\beta(s) - t) > (t-s)[3 - 4\beta - 4\beta f_1(s,t)] > 0,$$

where $f_1(s,t) = \log\left(\frac{t+s}{2}\right)$. This implies, by using Lemma 3.5 twice,

$$\begin{aligned} F_\beta(s,t) &> [g_\beta(t) - g_\beta(s) + t - s][2 - 4\beta - 4\beta f_2(s,t)] - (t-s) \\ &> (t-s)\left\{[3 - 4\beta - 4\beta f_1(s,t)][2 - 4\beta - 4\beta f_2(s,t)] - 1\right\}, \end{aligned} \tag{3.31}$$

where

$$\begin{aligned} f_2(s,t) &= \log\left(\frac{g_\beta(t) + g_\beta(s) - (t+s)}{2}\right) \\ &= \log\frac{(t+s) - 4\beta(t\log t + s\log s)}{2}. \end{aligned}$$

Note that the second inequality in (3.31) uses the convexity of the map $u \mapsto u\log u$ for $u > 0$ to get

$$f_2(s,t) < \log\left(\frac{t+s}{2}\right) + \log\left(1 - 4\beta\log\left(\frac{t+s}{2}\right)\right), \tag{3.32}$$

and then applies the fact $r - 4\beta r\log r < e^{\frac{1-2\beta}{2\beta}}$ for all $r > 0$ and $0 < \beta \le \frac{1}{2}(1 - \cos\frac{2\pi}{5})$. A simple computation shows that

$$(2-4\beta)(3-4\beta) - 1 = 16\beta^2 - 20\beta + 5 \ge 0,$$

for $0 \le \beta \le \frac{1}{2}(1 - \cos\frac{2\pi}{5})$. To finish this proof, it suffices to show that

$$(2-4\beta)f_1(s,t) + [3 - 4\beta - 4\beta f_1(s,t)]\, f_2(s,t) \le 0.$$

Since $(t+s)/2 < 1$, it remains to prove, by using (3.32), that

$$h(x) = (2-4\beta)x + (3 - 4\beta - 4\beta x)[x + \log(1 - 4\beta x)] < 0, \quad \forall x < 0.$$

Taking the first derivative of h, we get

$$\begin{aligned} h'(x) &= 5 - 12\beta + \frac{4\beta(4\beta - 2)}{1 - 4\beta x} - 4\beta[2x + \log(1 - 4\beta x)] \\ &> 5 - 12\beta + 4\beta(4\beta - 2) = 16\beta^2 - 20\beta + 5 \ge 0, \end{aligned}$$

where the first inequality is implies by the facts that the mapping $x \mapsto \frac{4\beta(4\beta-2)}{1-4\beta x}$ for $x < 0$ is decreasing and

$$\forall x < 0, \quad 2x + \log(1 - 4\beta x) \leq 2x(1 - 2\beta) < 0.$$

Therefore, h is strictly increasing. In addition to the fact $h(0) = 0$, we get $h(x) < 0$ for $x < 0$. □

Proof of Theorem 3.3. By Proposition 2.2, one always has $0 < \alpha \leq \lambda = \frac{1}{2}(1 - \cos\frac{2\pi}{5})$. We prove this theorem by showing that there is no nonconstant solution u for the Euler-Lagrange equation

$$2\alpha u \log(u/\|u\|_2) = (I - K)u.$$

Assume the inverse, that is, the above equation is satisfied with nonconstant u whose entries are $0 < x_0 \leq x_1 \leq x_2 \leq x_3 \leq x_4$. There is not loss of generality to assume that $\|u\|_2 = 1$, or equivalently, $x_0^2 + x_1^2 + \cdots + x_4^2 = 5$. By Corollary 3.2, we may assume further that $u = (x_0, x_2, x_4, x_3, x_1)$. In the above setting, the minimizing equation is equal to

$$\begin{aligned} &x_1 + x_2 = g_\alpha(x_0), \quad x_0 + x_3 = g_\alpha(x_1), \quad x_0 + x_4 = g_\alpha(x_2), \\ &x_1 + x_4 = g_\alpha(x_3), \quad x_2 + x_3 = g_\alpha(x_4), \end{aligned} \tag{3.33}$$

where $g_\alpha(x) = 2x - 4\alpha x \log x$.

Note that the assumption of nonconstant u derives $x_0 < x_4$, and the normalization of u implies $x_0 < 1$. Since g_α is a concave function with derivative $g'_\alpha(1) = 2 - 4\alpha > 0$, we have $g_\alpha(x) \in (0, 2)$ for $x \in (0, 1)$. On one hand, by this observation, the equality $x_1 + x_2 = g_\alpha(x_0)$ implies $x_1 < 1$ and then the identity $x_0 + x_3 = g_\alpha(x_1)$ implies $x_0 + x_3 \leq 2$. On the other hand, by (3.33), one can obtain the following equation

$$F_\alpha(x_0, x_2) = g_\alpha(g_\alpha(x_2) - x_0) - g_\alpha(g_\alpha(x_0) - x_2) - (x_2 - x_0) = 0.$$

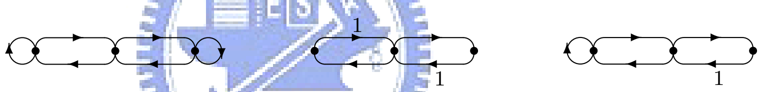
Since u is a solution of (3.33), it is clear that $(x_0, x_2) \in D_\alpha$, the region defined in (3.30). Thus, by Lemma 3.6, we have $x_0 = x_1 = x_2$ and, by the first equality of (3.33), we get $x_1 = 1$. This contradicts $x_0 < 1$.

Since there is no nonconstant solution for the equation (2.3) with $0 < \alpha \leq \lambda/2$, Proposition 2.3 implies that $2\alpha = \lambda$. □

3.3 Some other 3-points chains

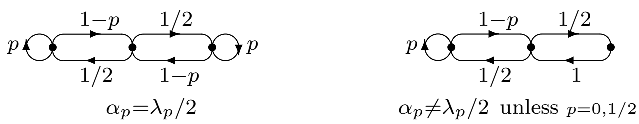
By collapsing 4, 5 and 6 cycles, we have obtained in Sections 3.1.3 and 3.2 the equality $\alpha = \lambda/2$ for the three chains on the 3-point stick described in figure 3.5. The first two theorems in this section concern the variants (depending on a

Figure 3.5: Three chains on the 3-point stick. All edges have weight 1/2 except when marked otherwise. In all cases $\alpha = \lambda/2$.



parameter $p \in [0, 1)$) described in Figure 3.6.

Figure 3.6: The families of Theorems 3.4 and 3.5, $p \in [0, 1)$.



Theorem 3.4. *For $0 \leq p < 1$, let K_p be the Markov kernel on the 3-point space*

$\{1,2,3\}$ *defined by*

$$K_p = \begin{pmatrix} p & 1-p & 0 \\ .5 & 0 & .5 \\ 0 & 1-p & p \end{pmatrix}$$

with stationary distribution $\mu_p = (\frac{1}{4-2p}, \frac{2-2p}{4-2p}, \frac{1}{4-2p})$. *Then* $\alpha_p = \lambda_p/2 = (1-p)/2$.

Theorem 3.5. *For* $0 \le p < 1$, *Let* K'_p *be the Markov kernel on the* 3*-point space* $\{1,2,3\}$ *defined by*

$$K'_p = \begin{pmatrix} 0 & 1 & 0 \\ .5 & 0 & .5 \\ 0 & 1-p & p \end{pmatrix}$$

with stationary measure $\mu'_p = (\frac{1-p}{4-3p}, \frac{2-2p}{4-3p}, \frac{1}{4-3p})$. *Then the log Sobolev constant* α_p *satisfies* $\alpha_p = \lambda_p/2 = \frac{1}{4}(3-p-\sqrt{p^2+1})$ *only when* $p=0$ *or* $p=1/2$.

Remark 3.3. Both K_p in Theorem 3.4 and K'_p in Theorem 3.5 are reversible with respect to their stationary distributions.

To prove the above two theorems, we need the following elementary lemma.

Lemma 3.7. *Consider the continuous function* $u : [0,\infty) \to \mathbb{R}$ *defined by*

$$u(s) = \begin{cases} 0 & \text{if } s = 0 \\ s\log s & \text{if } s \in (0,\infty). \end{cases} \tag{3.34}$$

The function u *has the following properties:*

$$\forall\, t \in [0,\infty), \quad u(t) \ge t-1. \tag{3.35}$$

$$\forall\, s,t \in [0,\infty) \text{ with } s \le t \text{ and } s+t \le 2, \quad u(t)-u(s) \le t-s. \tag{3.36}$$

$$\forall\, s,t \in [1,\infty) \text{ with } s \le t, \quad u(t)-u(s) \ge t-s. \tag{3.37}$$

Proof. The function $s \mapsto s \log s - s + 1$ has derivative $s \mapsto \log s$ on $(0, \infty)$. Hence it attains its minimum at $s = 1$. As the value at $s = 1$ is 0, (3.35) follows.

To prove (3.36), fix $s \geq 0$ and set, for $t \geq s$,

$$\begin{aligned} g(t) &= u(t) - u(s) - (t-s)u'((t+s)/2) \\ &= t \log t - s \log s - (t-s)(1 + \log((t+s)/2)). \end{aligned}$$

Compute the derivatives

$$g'(t) = \log\left(\frac{2t}{t+s}\right) - \frac{t-s}{t+s}, \quad g''(t) = \frac{s(s-t)}{t(t+s)^2}.$$

It follows that g is non-increasing on $[s, \infty)$. Hence $g(t) \leq g(s) = 0$ on $[s, \infty)$, that is,

$$u(t) - u(s) \leq (t-s)(1 + \log((t+s)/2)).$$

The inequality (3.36) obviously follows when $s + t \leq 2$.

Finally, (3.37) follows from the Mean Value Theorem applied to the function u since $u' \geq 1$ on $[1, \infty)$. □

Proof of Theorem 3.4. First observe that an easy computation gives $\lambda_p = 1 - p$. By Corollary 2.4, it suffices to show that for $\beta < \lambda_p/2$, the system (2.4) has no non-constant positive solution $u = (a, b, c)$. Suppose the contrary. By symmetry, we can assume that $a \geq c$. There is no loss of generality to assume further the normalization

$$a^2 + (2 - 2p)b^2 + c^2 = 4 - 2p. \tag{3.38}$$

Then (2.4) is equivalent to (using the function u defined at (3.34))

$$\begin{aligned} \frac{2\beta}{1-p}u(a) &= a - b & (3.39) \\ 4\beta u(b) &= 2b - a - c & (3.40) \\ \frac{2\beta}{1-p}u(c) &= c - b. & (3.41) \end{aligned}$$

We prove by considering two subcases, $a > c$ and $a = c$.

Case 1: $a > c$. Subtract (3.41) from (3.39) to obtain

$$u(a) - u(c) = \frac{1-p}{2\beta}(a-c) > a - c.$$

By (3.36), it follows that $a + c > 2$. This implies

$$a^2 + c^2 > 2 \tag{3.42}$$

and thus, by (3.38),

$$b < 1. \tag{3.43}$$

Now, add (3.39) divided by a to (3.41) divided by c and subtract (3.40) divided by b to obtain

$$\frac{2\beta}{1-p}\log ac - 4\beta \log b = \frac{a}{b} - \frac{b}{a} + \frac{c}{b} - \frac{b}{c}.$$

Rearranging the terms yields

$$\frac{4p\beta}{1-p}\log b = \left(\frac{a}{b} - \frac{b}{a} - \frac{2\beta}{1-p}\log\frac{a}{b}\right) - \left(\frac{b}{c} - \frac{c}{b} - \frac{2\beta}{1-p}\log\frac{b}{c}\right). \tag{3.44}$$

Consider the function $h(t) = t - t^{-1} - k\log t$ on $(0,\infty)$ and note that $h'(t) = t^{-2}(t-1)^2 + t^{-1}(2-k)$ is positive on $(0,\infty)$ if $k < 2$. In the present case, we take $k = 2\beta/(1-p)$ which, by hypothesis, is less than 1. Hence h is increasing. The left-hand side of (3.44) is negative since $b < 1$ by (3.43). Hence $h(a/b) - h(b/c) < 0$ and thus $a/b < b/c$ or, equivalently,

$$ac < b^2 < 1.$$

By (3.38) and (3.42), we have

$$\begin{aligned} 4 - 2p &= a^2 + 2(1-p)b^2 + c^2 > a^2 + 2(1-p)ac + c^2 \\ &= (a+c)^2 - 2pac > 4 - 2pac > 4 - 2p, \end{aligned}$$

a contradiction. Hence, we must have $\alpha_p = \lambda_p/2 = (1-p)/2$.

Case 2: $a = c$**.** In the case, we may rewrite (3.44) as follows.

$$g(b) = \frac{4p\beta}{1-p}\log b + \left(\frac{b}{a} - \frac{a}{b} - \frac{2\beta}{1-p}\log\frac{b}{a}\right) - \left(\frac{a}{b} - \frac{b}{a} - \frac{2\beta}{1-p}\log\frac{a}{b}\right) = 0.$$

Note that g is strictly increasing on $(0,\infty)$. If $a < 1$, then $b < 1$ since $g(1) > 0$ and if $a > 1$, then $b > 1$ since $g(1) < 0$. Both cases contradict (3.38). If $a = 1$, then $b = 1$, which contradicts the assumption that u is nonconstant. □

Proof of Theorem 3.5. Referring to the family of chains in Theorem 3.5, the facts that $\alpha_p = \lambda_p/2$ when $p = 0$ and $p = 1/2$ are contained respectively in Theorem 3.4 and in Corollary 3.3. To prove $\alpha_p < \lambda_p/2$ when $p \neq 0, 1/2$, we use the criteria contained in Proposition 2.2. A simple computation yields

$$\lambda_p = \frac{3 - p - \sqrt{1+p^2}}{2}$$

with eigenfunction

$$\phi = \left(1, \tfrac{p-1+\sqrt{1+p^2}}{2}, (p-1)(p+\sqrt{1+p^2})\right).$$

Thus, we compute

$$\mu_p'(\phi^3) = \frac{p(1-p)(p-1/2)[3 - 3p + 6p^2 - 4p^3 + \sqrt{1+p^2}(-1+6p-4p^2)]}{4-3p}.$$

On one hand, the map $p \mapsto 3 - 3p + 6p^2 - 4p^3$ is (strictly) decreasing with value 2 at $p = 1$. This implies $3 - 3p + 6p^2 - 4p^3 > 2$ for $p \in (0,1)$. On the other hand, the map $p \mapsto \sqrt{1+p^2}(-1+6p-4p^2)$ is (strictly) increasing on $(0,1/2)$ with value -1 at $p = 0$ and the map $p \mapsto -1 + 6p - 4p^2$ is concave with value 1 at $p = 1/2$ and $p = 1$. This implies $\sqrt{1+p^2}(-1+6p-4p^2) > -1$ for $p \in (0,1)$. Combining both observations, it is easy to see that

$$3 - 3p + 6p^2 - 4p^3 + \sqrt{1+p^2}(-1+6p-4p^2) > 0, \quad \forall p \in (0,1),$$

which implies $\mu_p'(\phi^3) \neq 0$ unless $p = 0$ or $p = 1/2$. By Proposition 2.2, we must have $\alpha_p < \lambda_p/2$ for $p \neq 0, 1/2$. $\square$

We end this section with the study of one of the most natural chain on a 3-point stick where transitions are to the left with probability $q = 1 - p$ and to the right with probability p.

Theorem 3.6. *For $0 < p < 1$ and set $q = 1 - p$. Let $K_p : \{1,2,3\} \times \{1,2,3\}$ be the Markov kernel defined by*

$$K_p = \begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix}$$

with stationary distribution

$$\mu_p = \left(c_p, c_p(p/q), c_p(p/q)^2\right), \quad c_p = \left(1 + (p/q) + (p/q)^2\right)^{-1}.$$

Then the spectral gap λ_p and the logarithmic Sobolev constant α_p are given by

$$\lambda_p = 1 - \sqrt{pq}, \quad \alpha_p = \frac{p - q}{2(\log p - \log q)}.$$

In particular, a minimizer of α_p is $\psi = (p/q, 1, q/p)$.

Remark 3.4. Let $p \in (0,1)$ and α_p and λ_p be as in Theorem 3.6. Recall in Theorem 2.3 that $\alpha_p \leq 1/4$ for $p \in (0,1)$ and the equality holds only if $p = 1/2$. A simple computation shows that $\lambda_p \geq 1/2$ and the equality holds only if $p = 1/2$. Combining both bounds, we have $\alpha_p \leq \lambda_p/2$ and the equality holds only if $p = 1/2$.

Proof of Theorem 3.6. Since K_p is reversible, the spectral gap is obtained by a direct computation of the eigenvalues of K_p. For the logarithmic Sobolev constant, we compare this chain with another 3-point chain

$$\widetilde{K}_p = \begin{pmatrix} q & p & 0 \\ q/2 & 1/2 & p/2 \\ 0 & q & p \end{pmatrix}$$

with stationary distribution

$$\widetilde{\mu}_p = \left(\widetilde{c}_p, 2\widetilde{c}_p(p/q), \widetilde{c}_p(p/q)^2\right), \quad \widetilde{c}_p = \left(1 + 2(p/q) + (p/q)^2\right)^{-1}.$$

The Dirichlet forms associated with (K_p, μ_p) and $(\widetilde{K}_p, \widetilde{\mu}_p)$ are respectively

$$\mathcal{E}_p(u,u) = c_p p \left((u_1 - u_2)^2 + (p/q)(u_2 - u_3)^2\right)$$

and

$$\widetilde{\mathcal{E}}_p(u,u) = \widetilde{c}_p p \left((u_1 - u_2)^2 + (p/q)(u_2 - u_3)^2\right).$$

Hence we have

$$\widetilde{\mathcal{E}}_p = (\widetilde{c}_p/c_p)\mathcal{E}_p \text{ and } (\widetilde{c}_p/c_p)\mu_p \le \widetilde{\mu}_p. \tag{3.45}$$

By Proposition 2.4, it follows that

$$\alpha_p \ge \widetilde{\alpha}_p. \tag{3.46}$$

Next, on $\{0,1\}^2$, we consider the product chain (with weights $(1/2, 1/2)$) of two copies of 2-point asymmetric chain in Theorem 2.3. This product chain has transitions given by

$$K((0,0),(0,0)) = q, \; K((1,1),(1,1)) = p,$$

$$K((0,0),(0,1)) = K((0,0),(1,0)) = p/2,$$

$$K((1,0),(1,1)) = K((0,1),(1,1)) = p/2,$$

and

$$K((1,1),(0,1)) = K((1,1),(1,0)) = q/2,$$

$$K((0,1),(0,0)) = K((1,0),(0,0)) = q/2,$$

$$K((0,1),(0,1)) = K((1,0),(1,0)) = 1/2.$$

By Proposition 2.6 and Theorem 2.3, its logarithmic Sobolev constant is $\frac{p-q}{2\log(p/q)}$. This chain projects to the 3-point space $\{1,2,3\}$ using the map

$$p : \{0,1\}^2 \to \{1,2,3\}, \quad (x,y) \mapsto 1 + |x| + |y|$$

and the projected chain is $\widetilde{K}_p$. Hence, by Proposition 2.5 and (3.46), we get

$$\alpha_p \geq \widetilde{\alpha}_p \geq \frac{p-q}{2(\log p - \log q)}. \tag{3.47}$$

To show that this is in fact an equality, it suffices to find a good test function. Letting $\psi = (p/q, 1, q/p)$ derives

$$\alpha_p \leq \frac{\mathcal{E}_p(\psi,\psi)}{\mathcal{L}_{\mu_p}(\psi)} = \frac{p-q}{2(\log p - \log q)}.$$

Thus $\alpha_p = \frac{p-q}{2(\log p - \log q)}$. □

Remark 3.5. Fix $p \in (0,1)$ and let K and K_p be the Markov kernels in the proof of Theorem 3.6. As the proof shows, K collapses to K_p and the logarithmic Sobolev constant of K_p is the same as that of K. However, the spectral gap of K_p, which is equal to $1-\sqrt{pq}$, is not the same as the spectral gap of K, which is equal to $1/2$. The main reason is that the eigenfunction of K corresponding to eigenvalue $1/2$ has different values at $(0,1)$ and $(1,0)$ if $p \neq 1/2$. This makes the projection p fail to collapse the eigenfunction onto the three point space $\{1,2,3\}$.

The following corollary is an observation based on the inequality (3.47) obtained in the proof of Theorem 3.6.

Corollary 3.4. *Let $p \in (0,1)$ and set $q = 1-p$. Consider the following Markov kernels.*

$$K_p = \begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix}, \quad \widetilde{K}_p = \begin{pmatrix} q & p & 0 \\ q/2 & 1/2 & p/2 \\ 0 & q & p \end{pmatrix}$$

Let α_p and $\widetilde{\alpha}_p$ be their logarithmic Sobolev constants. Then

$$\alpha_p = \widetilde{\alpha}_p = \frac{p-q}{2\log(p/q)}.$$

In particular, $\psi = (p/q, 1, q/p)$ is a minimizer for both constants.

Proof. By (3.47) and Theorem 3.6, it remains to show that ψ is a minimizer of $\widetilde{\alpha}_p$. By (2.6), the fact (3.45) derived in the proof of Theorem 3.6 implies

$$\widetilde{\mathcal{E}}_p(\psi,\psi) \le (\widetilde{c}_p/c_p)\mathcal{E}_p(\psi,\psi), \quad (\widetilde{c}_p/c_p)\mathcal{L}_{\mu_p}(\psi) \le \mathcal{L}_{\widetilde{\mu}_p}(\psi).$$

Since ψ is not constant, taking the ratio the Dirichlet form to the entropy implies

$$\alpha_p = \widetilde{\alpha}_p \le \frac{\widetilde{\mathcal{E}}_p(\psi,\psi)}{\mathcal{L}_{\widetilde{\mu}_p}(\psi)} \le \frac{\mathcal{E}_p(\psi,\psi)}{\mathcal{L}_{\mu_p}(\psi)} = \alpha_p.$$

□

Remark 3.6. Both K_p and $\widetilde{K}_p$ in Corollary 3.4 are reversible and the spectral gap $\widetilde{\lambda}_p$ of $\widetilde{K}_p$ is equal to $1/2$. Let $\widetilde{\alpha}_p$ be the logarithmic Sobolev constant of $\widetilde{K}_p$. By Corollary 3.4 and Theorem 2.3, $\widetilde{\alpha}_p \le \widetilde{\lambda}_p/2$ and the equality holds only if $p = 1/2$.

Appendix A

Techniques and proofs

A.1 Fundamental results of analysis

Lemma A.1. *Let* $(\mathcal{X}, \mu)$ *and* $(\mathcal{Y}, \nu)$ *be measure spaces and* $T : L^p(\mu) \to L^r(\nu)$ *be a bounded linear operator with* $1 \le p, r \le \infty$. *Let* $T^* : (L^r(\nu))^* \to (L^p(\mu))^*$ *be the adjoint operator of* T. *Then the operator norms of* T *and* T^*, *denoted by* $\|T\|_{p\to r}$ *and* $\|T^*\|_{s\to q}$ *with* $p^{-1} + q^{-1} = 1$ *and* $r^{-1} + s^{-1} = 1$, *satisfy*

$$\|T^*\|_{s\to q} = \|T\|_{p\to r}.$$

Proof. Note that for $f \in (L^r(\nu))^*$ and $u \in L^p(\mu)$,

$$|(T^* f)(u)| = |f(Tu)| \le \|T\|_{p\to r} \|f\|_{(L^r(\nu))^*} \|u\|_p,$$

which implies $\|T^*\|_{s\to q} \le \|T\|_{p\to r}$.

Conversely, for $v \in L^s(\nu)$, define $T_v(w) = \int_{\mathcal{Y}} v(y)w(y)d\nu(y)$ for all $w \in L^r(\nu)$. It is obvious that $T_v \in (L^r(\nu))^*$, $\|T_v\|_{(L^r(\nu))^*} = \|v\|_s$ and for $u \in L^p(\mu)$,

$$\int_{\mathcal{Y}} v(y)(Tu)(y)d\nu(y) = T_v(Tu) = (T^*T_v)(u) \le \|T^*\|_{s\to q} \|v\|_s \|u\|_p,$$

which implies $\|T\|_{p\to r} \le \|T^*\|_{s\to q}$. □

BIBLIOGRAPHY

[1] Aldous, D. *Random walks on finite groups and rapidly mixing Markov chains.* Seminar on probability, XVII, 243–297, Lecture Notes in Math., 986, Springer, Berlin, 1983.

[2] Aldous, D. and Diaconis, P. *Shuffling cards and stopping times.* Amer. Math. Monthly **93**, 333–348, 1986.

[3] Aldous, D. and Diaconis, P. *Strong uniform times and finite random walks.* Adv. in Appl. Math. 8, no. 1, 69–97, 1987.

[4] Bobkov, S. and Tetali, P. *Modified log-Sobolev inequalities, mixing and hypercontractivity.* Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, 287–296, ACM, New York, 2003.

[5] Chen, G.-Y., Liu, W.-W. and Saloff-Coste, L. *The logarithmic Sobolev constant of some finite Markov chains* Preprint.

[6] Chen, G.-Y. and Sheu, Y.-C. *On the log-Sobolev constant for the simple random walk on the n-cycle: the even cases.* J. Funct. Anal. **202** 473–485, 2003.

[7] Diaconis, P. *representations in probability and statistics.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

[8] Diaconis, P. and Shahshahani, M. *Generating a random permutation with random transpositions.* Z. Wahrsch. Verw. Gebiete **57**, no. 2, 159–179, 1981.

[9] Diaconis, P. and Saloff-Coste, L. *Comparison theorem for reversible Markov chains* Ann. Appl. Probab. **3** 696–730, 1993.

[10] Diaconis, P. and Saloff-Coste, L. *Comparison techniques for random walks on finite groups* Ann. Probab. **21** 2131–2156, 1993.

[11] Diaconis, P. and Saloff-Coste, L. *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** 695–750, 1996.

[12] Diaconis, P. and Saloff-Coste, L. *Nash inequalities for finite Markov chains.* J. Theoret. Probab. **9**, no. 2, 459–510, 1996.

[13] Feller, W. *An introduction to probability theory and its applications*, New York : Wiley, 3rd ed. 1968.

[14] L.Gross, *Logarithmic Sobolev inequalities and contractivity properties of semigroups*, Lecture Notes in Math. **1563**, 54–88, Springer, Berlin, 1993.

[15] Horn, P. and Johnson, C. *Matrix Analysis.* Cambridge Univ. Press, 1985.

[16] Horn, P. and Johnson, C. *Topics in Matrix Analysis.* Cambridge Univ. Press, 1990.

[17] Jerrum, M. and Sinclair, A. *Aproximating the perminant* SIAM Jorurnal on computing **18**, 1149–1178, 1989.

[18] Karlin, S. and Taylor, H. *A first course in stochastic processes* New York : Academic Press, 2d ed, 1975.

[19] Norris, J. R. *Markov chains.* New York : Cambridge University Press 1997.

[20] Rothaus, O. *Logarithmic Sobolev inequalities and the spectrum of Sturm-Liouville operators*, J. Funct. Anal. **39**, 42–56, 1980.

[21] Rothaus, O. *Logarithmic Sobolev inequalities and the spectrum of Schrödinger operators*, J. Funct. Anal. **42**, 110-120, 1981.

[22] Rothaus, O. *Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities.* J. Funct. Anan., **42**, 102–109, 1981.

[23] Saloff-Coste, L. *Lectures on finite Markov chains.* Lectures on probability theory and statistics (Saint-Flour Summer School, 1996), 301–413, Lecture Notes in Math., 1665, Springer, Berlin, 1997.

[24] Saloff-Coste, L. *Random walks on finite groups.* in Probability on discrete structures, 263–346, Encyclopaedia Math. Sci., 110, Springer, Berlin, 2004. (H. Kesten, ed.)

[25] Woess, W. *Aperiodische Wahrscheinlichkeitsmasse auf topologischen* Gruppen. Mh. Math. **90**, 339–345, 1980.

[26] Woess, W. *Périodicité de mesures de probabilité sur les topologiques groups.* In Marches Aléatories et Processus Stochastiques sur le Groupe de Lie. Inst. Élie Cartan, **7**, 170–180. Univ. Nancy. 1983.

[27] Wilson, D. *Mixing times of lozenge tiling and card shuffling Markov chains.* Ann. Appl. Probab. **14**, 2004, 274–325.