

## LOGARITHMIC SOBOLEV INEQUALITIES FOR FINITE MARKOV CHAINS

BY P. DIACONIS AND L. SALOFF-COSTE

*Harvard University and CNRS, Université Paul Sabatier*

This is an expository paper on the use of logarithmic Sobolev inequalities for bounding rates of convergence of Markov chains on finite state spaces to their stationary distributions. Logarithmic Sobolev inequalities complement eigenvalue techniques and work for nonreversible chains in continuous time. Some aspects of the theory simplify considerably with finite state spaces and we are able to give a self-contained development. Examples of applications include the study of a Metropolis chain for the binomial distribution, sharp results for natural chains on the box of side  $n$  in  $d$  dimensions and improved rates for exclusion processes. We also show that for most  $r$ -regular graphs the log-Sobolev constant is of smaller order than the spectral gap. The log-Sobolev constant of the asymmetric two-point space is computed exactly as well as the log-Sobolev constant of the complete graph on  $n$  points.

**1. Introduction.** Logarithmic Sobolev inequalities were introduced in 1975 as a way of isolating smoothing properties of Markov semigroups in infinite-dimensional settings. Pointers to the literature are given at the end of this Introduction. This paper presents a reasonably self-contained treatment of logarithmic Sobolev inequalities in the context of finite Markov chains. It shows how these inequalities can be used to obtain quantitative bounds on the convergence of finite Markov chains to stationary.

We work with a finite state space  $\mathcal{X}$  and an irreducible Markov kernel  $K(x, y) \geq 0$ ,  $\sum_y K(x, y) = 1$ . The continuous time semigroup associated to  $K$  is  $H_t = \exp(-t(I - K))$ . Its kernel is denoted by  $H_t^x(y) = H_t(x, y)$  which is the distribution at time  $t > 0$  of the process started at  $x$ . It has a unique stationary measure  $\pi$  and  $H_t^x(y) \rightarrow \pi(y)$  as  $t$  tends to infinity. The object of this paper is to get quantitative bounds on this convergence, for instance, in total variation distance  $\|H_t^x - \pi\|_{TV}$ . The reader will find the proofs of most of the results stated in this Introduction in Section 2 and in the rest of the paper.

One route that has proved successful in many examples bounds total variation by the  $l^2$  or chi-squared distance with respect to  $\pi$  as

$$(1.1) \quad 2\|H_t^x - \pi\|_{TV} \leq \|(H_t^x/\pi) - 1\|_{2, \pi}.$$

---

Received October 1995; revised March 1996.

AMS 1991 subject classifications. 60J15, 60J27, 60F05.

Key words and phrases. Markov chains, logarithmic Sobolev inequalities.

This  $l^2$ -norm can be represented as an operator norm,

$$(1.2) \quad \max_x \|(H_t^x/\pi) - 1\|_{2, \pi} = \|H_t - E\|_{2 \rightarrow \infty},$$

where  $E: f \rightarrow Ef$  is the operator that associates to  $f$  its mean with respect to  $\pi$ . Now, for any decomposition  $t = t_1 + t_2$ ,  $t_1, t_2 \geq 0$ ,

$$(1.3) \quad \|H_t - E\|_{2 \rightarrow \infty} \leq \|H_{t_1}\|_{2 \rightarrow \infty} \|H_{t_2} - E\|_{2 \rightarrow 2}.$$

Here, we are free to choose an appropriate value for  $t_1$ . The  $2 \rightarrow 2$  norm appearing above can be usefully bounded in terms of the second eigenvalue (i.e., the spectral gap)

$$(1.4) \quad \lambda = \min \left\{ \frac{\mathcal{E}(f, f)}{\text{Var}(f)}; \text{Var}(f) \neq 0 \right\}$$

with the Dirichlet form  $\mathcal{E}$  and the variance defined by

$$\begin{aligned} \mathcal{E}(f, f) &= \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 K(x, y) \pi(x), \\ \text{Var}(f) &= \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 \pi(y) \pi(x). \end{aligned}$$

One widely used bound is

$$(1.5) \quad 4\|H_t^x - \pi\|_{\text{TV}}^2 \leq \frac{1}{\pi_*} e^{-2\lambda t},$$

where  $\pi_* = \min_x \pi(x)$ . This bound follows from (1.1)–(1.3) by choosing  $t_1 = 0$ . We show how to improve upon (1.5) by picking more effective positive  $t_1$ 's. To this end, we need bounds on the decay of

$$\|H\|_{2 \rightarrow \infty} = \max_x \|H_t^x/\pi\|_2$$

as a function of  $t$ . Logarithmic Sobolev inequalities give useful estimates of this decay. A logarithmic Sobolev inequality is an inequality of the form

$$(1.6) \quad \mathcal{L}(f) \leq C\mathcal{E}(f, f),$$

where the entropy-like quantity  $\mathcal{L}(f)$  is defined by

$$\mathcal{L}(f) = \sum |f|^2 \log(|f|^2/\|f\|_2^2) \pi.$$

In analogy with (1.4), define the log-Sobolev constant  $\alpha$  of the chain  $(K, \pi)$  by

$$(1.7) \quad \alpha = \min \left\{ \frac{\mathcal{E}(f, f)}{\mathcal{L}(f)}; \mathcal{L}(f) \neq 0 \right\}.$$

Then  $1/\alpha$  is the smallest constant  $C$  such that (1.6) holds for all  $f$ . This constant  $\alpha$  always satisfies  $\alpha \leq \lambda/2$  and is sometimes equal to  $\lambda/2$ . Section

3 shows that  $\alpha$  can be used to bound convergence to stationarity through the inequality

$$(1.8) \quad 2\|H_t^x - \pi\|_{TV}^2 \leq \left(\log \frac{1}{\pi_*}\right) e^{-2\alpha t}.$$

More precise results are given in Section 3. Roughly speaking, (1.8) is an improvement upon (1.5) when

$$\frac{1}{\lambda} \log \frac{1}{\pi_*} \geq \frac{1}{\alpha} \log \log \frac{1}{\pi_*}.$$

For instance, for the chain  $K$  on the hypercube  $\mathcal{X} = \{-1, 1\}^d$  which, at each step, flips a coordinate chosen at random,  $\lambda = 2/d$ ,  $\alpha = 1/d$ ,  $\pi_* = 2^{-d}$  and (1.8) improves greatly upon (1.5) in this case. Actually a more careful use of  $\alpha$  (see Sections 3 and 4) shows that  $t \sim \frac{1}{4}d \log d$  suffices for approximate stationarity in this example, and this is the right answer.

Another way to express the relation between the log-Sobolev constant  $\alpha$  and convergence to stationarity is to introduce the parameter

$$\tau = \inf\left\{t > 0: \sup_x \|[H_t^x/\pi] - 1\|_2 \leq 1/e\right\}.$$

Classical inequalities assert that, for reversible chains,

$$\frac{1}{\lambda} \leq \tau \leq \frac{2 + \log[1/\pi_*]}{2\lambda},$$

whereas we will prove here that

$$\frac{1}{2\alpha} \leq \tau \leq \frac{4 + \log \log[1/\pi_*]}{4\alpha}.$$

In this precise sense  $\alpha$  is more closely related to convergence to stationarity than  $\lambda$  is.

The variational formula (1.7) defining  $\alpha$  is important because it shows that  $\alpha$  can be bounded by comparison between different chains (see Lemmas 3.3 and 3.4). The papers [12, 14, 17] give examples of complicated chains which are analyzed by comparison with simpler chains. The use of log-Sobolev constants can often produce important improvements on rates of convergence.

One of the keys to understanding the decay of  $H_t^x$  is the equivalence between logarithmic Sobolev inequalities and a certain property of the semi-group called hypercontractivity. Namely, for reversible chains,  $\alpha$  at (1.7) can also be defined as the largest constant  $\beta$  such that

$$\|H_t\|_{2 \rightarrow q(t)} \leq 1 \quad \text{for all } t > 0 \text{ with } q(t) = 1 + e^{4\beta t}.$$

Hypercontractivity will be shown to be equivalent to a log-Sobolev inequality in Section 3. We use this equivalence together with (1.3) to get improved rates for chi-square convergence (and maximal relative error) in Theorem 3.7 and Corollary 3.8. To give a feel for the topic, consider the following illustrative example.

EXAMPLE 1.1. *Metropolis algorithm for the binomial distribution.* The Metropolis algorithm is a widely used tool in simulation. There has been little rigorous analysis of time to stationarity (see [18] for a survey). Consider  $\mathcal{X} = \{0, 1, \dots, n\}$  and  $\pi(x) = 2^{-n} \binom{n}{x}$ . The Metropolis chain begins with a base chain  $K(x, y)$  on  $\mathcal{X}$  which is modified to a new chain  $M(x, y)$  by an auxiliary randomization. The new chain has stationary distribution  $\pi$ . Indeed,  $M(x, y)$  is constructed so that  $\pi(x)M(x, y) = \pi(y)M(y, x)$  for all  $x, y \in \mathcal{X}$ . For the present example, take the base chain to be nearest neighbor random walk

$$K(x, x+1) = K(x, x-1) = 1/2, \quad 1 \leq x \leq n-1$$

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

The standard Metropolis construction (see e.g., [18]) gives

(1.9)

$$M(x, y) = \begin{cases} \frac{1}{2}, & \text{if } \begin{cases} y = x+1 \text{ and } 0 \leq x \leq (n-1)/2, \\ y = x-1 \text{ and } (n+1)/2 \leq x \leq n, \end{cases} \\ \frac{x}{[2(n-x+1)]}, & \text{if } y = x-1 \text{ and } 1 \leq x \leq (n+1)/2, \\ \frac{(n-x)}{[2(x+1)]}, & \text{if } y = x+1 \text{ and } (n-1)/2 \leq x \leq n-1, \\ \frac{(n-2x+1)}{[2(n-x+1)]}, & \text{if } y = x \text{ and } 0 \leq x \leq (n-1)/2, \\ \frac{(2x-n+1)}{[2(x+1)]}, & \text{if } y = x \text{ and } (n+1)/2 \leq x \leq n, \\ \frac{2}{(n+2)}, & \text{if } y = x = n/2 \text{ (} n \text{ even)}. \end{cases}$$

In Section 3 we show  $\lambda = \lambda(M) \geq 1/n$ . Clearly,  $\pi_* = 2^{-n}$ . Using these ingredients, the bound (1.5) shows that  $t$  of order  $n^2$  suffices to have the chain close to stationarity. For this example, we can also show that  $\alpha = \alpha(M) \geq 1/(2n)$  and this gives us the following result.

THEOREM 1.1. *The Metropolis chain (1.9) and the binomial distribution  $\pi(x) = 2^{-n} \binom{n}{x}$  satisfy*

$$\|M_x^l - \pi\|_{\text{TV}} \leq e^{1-c} \quad \text{for } l \geq \frac{n}{2}(\log n + 2c), \quad c > 0.$$

Conversely,

$$\max_x \|M_x^l - \pi\|_{\text{TV}} \geq \frac{1}{4} + o(1) \quad \text{for } l \leq \frac{n}{8} \log n.$$

A short description of the paper is as follows. Section 2 gives a careful development of preliminaries needed from Markov chain theory and elemen-

tary functional analysis. We show how discrete time theorems follow from continuous time results for reversible chains.

Section 3 gives self-contained proofs of the basic results concerning logarithmic Sobolev inequalities. It proves  $\alpha \leq \lambda/2$ , shows that logarithmic Sobolev inequalities are stable under taking products and proves the equivalence between hypercontractivity and log-Sobolev inequalities. Theorem 1.1 is proved as a running example. Finally, the relations between the log-Sobolev constant  $\alpha$  and convergence to stationarity are explained in Theorems 3.6 and 3.7 and Corollary 3.8.

Section 4 describes examples and applications. We present what is known about the log-Sobolev constants for simple random walk on  $\mathbb{Z}_m$ , for random transpositions in the symmetric group  $S_n$  and for random  $r$ -regular graphs and expanders. We also show that natural walks on the box of side length  $n$  in  $d$  dimensions reach stationarity after order  $n^2 d \log d$  steps. This result is sharp and we do not know of any other proof except for some special walks having a product structure. We also present application to the simple exclusion process.

The last section is an Appendix which presents the exact computation of the logarithmic Sobolev constant of the two-point space with stationarity measure  $\theta, 1 - \theta$ . This is probably the simplest example where one can prove that  $\lambda \neq \alpha/2$  (the first example of this sort was given in [30]).

There has not been much previous work on logarithmic Sobolev inequalities on finite state spaces. The work of Stroock and Zegarlinsky [48] can be seen as proving log-Sobolev inequalities for the Metropolis algorithm for simulating from Ising-like models on a finite grid.

A second example is the work of Lu and Yau [32] bounding the log-Sobolev constant of simple exclusion on a grid. Their bound is of the right order of magnitude and improves, in the case of a grid, upon the bound stated at the end of [12] and in Section 4.4, which is more widely applicable.

We have used log-Sobolev techniques to study a challenge problem of Aldous [17]. This problem also appears in a much used algorithm for manipulating elements on large finite groups. Briefly, let  $G$  be a finite group. The walk takes place on a subset of  $G^n$ . One picks a pair of coordinates  $(i, j)$  and multiplies the group element in the  $i$ th coordinate by the group element in  $j$ th coordinate or its inverse. Extensive empirical work suggests that this walk mixes extremely rapidly. We can prove versions of rapid mixing for fixed  $G$  and large  $n$ . The argument compared this chain with a natural walk on the product group. Use of the log-Sobolev constants yields improved rates of convergence in this example.

This paper represents a synthesis and translation of a huge body of previous work by others into the language of finite Markov chains. The hypercontractivity literature begins with Nelson's proof of the existence of a time  $t$  at which  $\|H_t\|_{2 \rightarrow 4} < \infty$  for the Ornstein–Uhlenbeck semigroup. Gross developed the general theory, introducing logarithmic Sobolev inequalities as an equivalent method of proving hypercontractivity. From here, there were a huge number of applications, careful work on special cases and technical

improvements. These are reviewed in the following surveys: Gross [25] gives a survey of the entire field with many elegant proofs and a comprehensive bibliography. His paper appears in a volume which also contains a survey of Stroock [47] on his joint work with Zegarlinsky. This work gives remarkable bounds on convergence to stationarity for stochastic Ising models using log-Sobolev constants. Bakry [5] has written a comprehensive survey in course note form, which contains complete proofs and much that is new. It features a development of the Bakry–Emery technique [4] for proving log-Sobolev inequalities using notions of curvature (unfortunately, this technique seems useless in the finite setting). Related material can be found in [7] and [8].

**2. Preliminaries.** This section introduces notation and reviews classical facts that are useful in the sequel. Following basic definitions, some consequences of reversibility are deduced. This is followed by a description of Dirichlet forms for nonreversible chains and the development of various technical lemmas needed in the sequel. Next, a variety of distances to equilibrium are introduced and related. Finally, some elementary results concerning product Markov chains are established. All of this material is elementary, but present versions are difficult to find in available literature.

*2.1. Notation.* A Markov chain on a finite state space  $\mathcal{X}$  with cardinality  $|\mathcal{X}|$  can be described through its kernel  $K$ , which is a function on  $\mathcal{X} \times \mathcal{X}$  satisfying

$$K(x, y) \geq 0, \quad \sum_{y \in \mathcal{X}} K(x, y) = 1.$$

The associated Markov operator, also denoted by  $K$ , acts on any real function  $f$  by

$$Kf(x) = \sum_{y \in \mathcal{X}} K(x, y)f(y).$$

The iterated kernel  $K^n$  is defined by

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

It corresponds to the operator  $K^n$ . We will also use the notation

$$K_x^n(y) = K^n(x, y)$$

when it is convenient. Let  $\pi$  be an invariant probability measure for  $K$ , that is, a probability satisfying

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

Such a measure always exists and, under a mild irreducibility condition,  $\pi$  is unique. We assume throughout that  $\pi$  charges all the points in  $\mathcal{X}$ .

It will be useful to consider the operator  $K$  acting on the space of real-valued functions  $l^p = l^p(\pi)$ ,  $1 \leq p \leq +\infty$ , equipped with the norm

$$\|f\|_{\pi, p} = \|f\|_p = \left( \sum_x |f(x)|^p \pi(x) \right)^{1/p}, \quad \|f\|_\infty = \max_x |f(x)|.$$

Let

$$\|K\|_{p \rightarrow q} = \sup_{\|f\|_p \leq 1} \|Kf\|_q$$

denote the operator norm of  $K$  from  $l^p$  to  $l^q$ . Set

$$Ef = E_\pi f = \sum_x f(x) \pi(x), \quad \text{Var}(f) = \text{Var}_\pi(f) = \|f - Ef\|_2^2.$$

We will often consider  $E$  as an operator acting on functions. Introduce also the entropy

$$\text{Ent}(f) = \text{Ent}_\pi(f) = \sum_x [f(x) \log f(x)] \pi(x)$$

of a nonnegative function  $f$  such that  $E(f) = 1$ . Note that this is equal to the relative entropy of the probability measure  $\mu = f\pi$  with respect to  $\pi$  and we will abusively write

$$\text{Ent}_\pi(f) = \text{Ent}_\pi(\mu) = \sum_x \mu(x) \log \frac{\mu(x)}{\pi(x)}.$$

We now turn to the description of the adjoint of  $K$ . The invariance of  $\pi$  is equivalent to the fact that the adjoint  $K^*$  of  $K$  in  $l^2(\pi)$  is also a Markov operator. Indeed, let

$$k(x, y) = \frac{K(x, y)}{\pi(y)}$$

be the kernel of  $K$  with respect to  $\pi$  and set

$$k^n(x, y) = k_x^n(y) = \frac{K^n(x, y)}{\pi(y)}.$$

The adjoint  $K^*$  is given by

$$K^*f(x) = \sum_y K^*(x, y)f(y) = \sum_y k^*(x, y)f(y)\pi(y),$$

where

$$K^*(x, y) = \frac{K(y, x)\pi(y)}{\pi(x)}, \quad k^*(x, y) = k(y, x).$$

Thus,

$$\begin{aligned} K^*1 &= \sum_y k^*(x, y)\pi(y) = \sum_y k(y, x)\pi(y) \\ &= \frac{1}{\pi(x)} \sum_y K(y, x)\pi(y) = 1. \end{aligned}$$

Because  $K$  and  $K^*$  are Markov operators, they contract  $l^\infty$ ; thus, by duality, they contract  $l^1 = l^1(\pi)$  and, by classical interpolation, any  $l^p(\pi)$ ,  $1 \leq p \leq +\infty$ . This fact will be used throughout without further notice. Alternatively, the last statement can be obtained from Jensen's inequality which implies

$$|Kf|^p(x) \leq K(|f|^p)(x)$$

for all  $p \geq 1$  and  $x \in \mathcal{X}$ . Since  $\pi$  is invariant, it follows that  $\|Kf\|_p \leq \|f\|_p$ .

For important technical reasons, this paper mainly deals with the continuous time semigroup  $H_t$  associated with  $K$  and defined by

$$H_t = e^{-t} \sum_0^\infty \frac{(tK)^n}{n!} = e^{-t(I-K)}.$$

We set

$$H_t(x, y) = H_t^x(y) = e^{-t} \sum_0^\infty \frac{t^n}{n!} K^n(x, y)$$

and

$$h_t(x, y) = h_t^x(y) = \frac{H_t(x, y)}{\pi(y)} = e^{-t} \sum_0^\infty \frac{t^n}{n!} k^n(x, y).$$

REMARKS. (i) Introducing the quantities  $k_x^n(y)$  and  $h_t^x(y)$  is useful and natural as soon as one intends to use functional analytic methods involving the space  $l^p(\pi)$  since these quantities are the *densities* of the probability measures  $K_x^n$  and  $H_t^x$  with respect to  $\pi$ .

(ii) One usually defines the action of  $K$  on probability measures by setting

$$[\mu K](u) = \mu(Ku)$$

for any probability measure  $\mu$  and all functions  $u$ . Now, if  $\mu = f\pi$ , that is, if  $f$  is the density of  $\mu$ , this means that  $\mu K$  has density  $K^*f$ . In particular, if  $\mu = f\pi$ , the measure  $\mu_t = \mu H_t$  has density  $H_t^*f$ .

2.2. *Reversibility.* We say that  $(K, \pi)$  is reversible if  $K(x, y)/\pi(y) = K(y, x)/\pi(x)$ . In other words,  $(K, \pi)$  is reversible if  $k$  is symmetric. This amounts to the fact that  $K$  is a self-adjoint operator on  $l^2(\pi)$ . In this case,  $K$  has real eigenvalues

$$-1 \leq \beta_{\min} = \beta_{|\mathcal{X}|-1} \leq \dots \leq \beta_1 \leq \beta_0 = 1$$

and we fix an orthonormal basis  $(\psi_i)_0^{|\mathcal{X}|-1}$  of real eigenfunctions such that  $K\psi_i = \beta_i\psi_i$  and  $\psi_0 \equiv 1$ . Note that  $\sum_0^{|\mathcal{X}|-1} \psi_i^2(x) = 1/\pi(x)$ . Also, we set  $\lambda_i = 1 - \beta_i$ , so that the eigenvalues of  $H_t$  are the  $\exp -t\lambda_i$ 's with the same corresponding eigenfunctions. Of main interest are the two parameters

$$(2.1) \quad \beta = \max\{|\beta_{\min}|, \beta_1\} \quad \text{and} \quad \lambda = \lambda_1.$$

The latter will be referred to as the spectral gap of the Markov chain  $K$ . With this notation, we have the following lemma.



LEMMA 2.1. *If  $(K, \pi)$  is reversible, it satisfies:*

$$(i) \quad k^n(x, y) = \sum_0^{|\mathcal{X}|-1} \beta_i^n \psi_i(x) \psi_i(y),$$

$$\|k_x^n - 1\|_2^2 = \sum_1^{|\mathcal{X}|-1} \beta_i^{2n} |\psi_i(x)|^2 \leq \frac{1 - \pi(x)}{\pi(x)} \beta^{2n};$$

$$(ii) \quad h_t(x, y) = \sum_0^{|\mathcal{X}|-1} \exp(-t\lambda_i) \psi_i(x) \psi_i(y),$$

$$\|h_t^x - 1\|_2^2 = \sum_1^{|\mathcal{X}|-1} \exp(-2t\lambda_i) |\psi_i(x)|^2 \leq \frac{1 - \pi(x)}{\pi(x)} \exp(-2t\lambda).$$

This is, of course, a classical result. A short proof in the spirit of our presentation is given in [12]. The inequalities on  $\|k_x^n - 1\|_2, \|h_t^x - 1\|_2$  generalize to nonreversible chains; see Lemma 2.3. The following simple result gives a useful way of transferring results between discrete and continuous time.

COROLLARY 2.2. *Assume that  $(K, \pi)$  is reversible and set  $\beta_- = \max\{0, -\beta_{\min}\}$ . Then:*

$$(i) \quad \|h_t^x - 1\|_2^2 \leq \frac{1}{\pi(x)} e^{-t} + \|k_x^{\lfloor t/2 \rfloor} - 1\|_2^2;$$

$$(ii) \quad \|k_x^N - 1\|_2^2 \leq \beta_-^{2n} (1 + \|h_{n'}^x - 1\|_2^2) + \|h_N^2 - 1\|_2^2 \quad \text{for } N = n + n' + 1.$$

PROOF. For part (i), use Lemma 2.1,

$$\beta_i^{2n} = (1 - \lambda_i)^{2n} = \exp(2n \log(1 - \lambda_i))$$

and the inequality  $\log(1 - x) \geq -2x$  for  $0 \leq x \leq 1/2$ . For part (ii), observe that

$$k^{2n+1}(x, x) = \sum_0^{|\mathcal{X}|-1} \beta_i^{2n+1} |\psi_i(x)|^2 \geq 0.$$

This shows that

$$-\sum_{i: \beta_i < 0} \beta_i^{2n+1} |\psi_i(x)|^2 \leq \sum_{i: \beta_i > 0} \beta_i^{2n+1} |\psi_i(x)|^2.$$

Hence

$$\sum_{i: \beta_i < 0} \beta_i^{2n+2} |\psi_i(x)|^2 \leq \sum_{i: \beta_i > 0} \beta_i^{2n} |\psi_i(x)|^2.$$

Now, for those  $\beta_i$  that are positive, we have

$$\beta_i^{2n} = \exp(2n \log(1 - \lambda_i)) \leq \exp(-2n \lambda_i)$$

so that

$$\sum_{i: \beta_i > 0} \beta_i^{2n} |\psi_i(x)|^2 \leq \|h_n^x\|_2^2$$

and

$$\sum_{\substack{i \neq 0 \\ \beta_i > 0}} \beta_i^{2n} |\psi_i(x)|^2 \leq \|h_n^x - 1\|_2^2. \quad \square$$

Putting these pieces together, we get for  $N = n + n' + 1$ ,

$$\begin{aligned} \|k_n^N - 1\|_2^2 &= \sum_1^{|\mathcal{L}|-1} \beta_i^{2N} |\psi_i(x)|^2 \\ &= \sum_{i: \beta_i < 0} \beta_i^{2N} |\psi_i(x)|^2 + \sum_{\substack{i \neq 0 \\ \beta_i > 0}} \beta_i^{2N} |\psi_i(x)|^2 \\ &\leq \beta_-^{2n} \left( \sum_{i: \beta_i < 0} \beta_i^{2n'+2} |\psi_i(x)|^2 \right) + \sum_{\substack{i \neq 0 \\ \beta_i > 0}} \beta_i^{2N} |\psi_i(x)|^2 \\ &\leq \beta_-^{2n} \|h_{n'}^x\|_2^2 + \|h_N^x - 1\|_2^2 \\ &= \beta_-^{2n} (1 + \|h_{n'}^x - 1\|_2^2) + \|h_N^x - 1\|_2^2. \quad \square \end{aligned}$$

This result yields a precise and rather sharp connection between the convergence of  $K^n$  and the convergence of  $H_t$ . Unfortunately, it depends on reversibility. Note, however, that Lemma 2.1 and its corollary apply (with obvious modifications) when  $K$  is normal. As a direct application of the second statement in Corollary 2.2, we have the following corollary which allows us to separate out the effects of the smallest eigenvalue from those of the spectral gap.

**COROLLARY 2.3.** *Assume that  $(K, \pi)$  is reversible and set  $\lambda_* = \min\{\lambda, 1 + \beta_{\min}\}$ . Then*

$$\|k_x^n - 1\|_2 \leq 3e^{-c} \quad \text{for } n = \frac{1}{2\lambda} \log \frac{1}{\pi(x)} + \frac{c}{\lambda_*} + 1.$$

**EXAMPLE 2.1.** Consider the chain  $K$  on the symmetric group  $\mathcal{S} = S_d$  with  $K(\sigma, \theta) = (1/|S|)1_S(\theta^{-1}\sigma)$ , where  $S = \{\text{id}\} \cup \{(i, j): 1 \leq i < j \leq d\}$ . This corresponds to randomly transposing pairs of cards with the identity thrown in with equal weight. Using results from [19], the lowest eigenvalue  $\beta_{\min}$  is of order  $-1 + 1/d^2$ . The second largest eigenvalue  $\beta_1$  is of order  $1 - 2/d$ . If we apply the bound of Lemma 2.1, we find that order  $d^3 \log d$  steps are sufficient to reach stationarity. If instead we employ Corollary 2.3, we find that order  $d^2 \log d$  steps are sufficient. Indeed, here  $\lambda \sim 2/d$ ,  $\lambda_* \sim 1/d^2$  and  $\pi(x) =$

$1/(d!)$ . The right answer is that order  $d^2$  steps are necessary and sufficient for  $K^n$  to reach stationarity. Observe that the continuous time semigroup  $H_t = e^{-t(I-K)}$  reaches stationarity after a time of order  $d \log d$  (see [16, 41]).

2.3. *Dirichlet forms.* The notion of Dirichlet form is crucial in the sequel. For a given chain  $K$  with invariant measure  $\pi$ , define

$$\mathcal{E}(f, g) = \langle (I - K)f, g \rangle,$$

where  $f$  and  $g$  are two real-valued functions. This satisfies

$$\mathcal{E}(f, f) = \left\langle \left( I - \frac{K + K^*}{2} \right) f, f \right\rangle.$$

Hence,

$$\begin{aligned} \mathcal{E}(f, f) &= \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 \frac{k(x, y) + k(y, x)}{2} \pi(x) \pi(y) \\ &= \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 k(x, y) \pi(x) \pi(y) \\ &= \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 K(x, y) \pi(x). \end{aligned}$$

The first equality is the classical formula for the Dirichlet form of the reversible Markov kernel  $\frac{1}{2}(K + K^*)$ . To summarize, for any real-valued function  $f$ ,

$$\begin{aligned} \mathcal{E}(f, f) &= \langle (I - K)f, f \rangle = \left\langle \left( I - \frac{K + K^*}{2} \right) f, f \right\rangle \\ (2.2) \qquad &= \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 K(x, y) \pi(x). \end{aligned}$$

Strictly speaking,  $\mathcal{E}$  is a Dirichlet form only when  $(K, \pi)$  is reversible. Further, when  $(K, \pi)$  is reversible,  $\mathcal{E}$  also satisfies

$$\begin{aligned} \mathcal{E}(f, g) &= \langle (I - K)f, g \rangle \\ (2.3) \qquad &= \frac{1}{2} \sum_{x, y} (f(x) - f(y))(g(x) - g(y)) K(x, y) \pi(x). \end{aligned}$$

REMARK. When working with complex-valued functions, it is convenient to use the definition

$$\mathcal{E}(f, g) = \operatorname{Re}(\langle (I - K)f, g \rangle).$$

With this definition it is still true that

$$\mathcal{E}(f, f) = \left\langle \left( I - \frac{K + K^*}{2} \right) f, f \right\rangle = \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 K(x, y) \pi(x).$$

To illustrate the use of  $\mathcal{E}$  we give a proof of the following well known result. Define the spectral gap  $\lambda$  of  $(K, \pi)$  by

$$(2.4) \quad \begin{aligned} \lambda &= \min \left\{ \frac{\mathcal{E}(f, f)}{\text{Var}(f)} : \text{Var}(f) \neq 0 \right\} \\ &= \min \{ \mathcal{E}(f, f) : \|f\|_2 = 1, E_\pi(f) = 0 \}. \end{aligned}$$

LEMMA 2.4. *Let  $(K, \pi)$  be a Markov chain on a finite state space  $\mathcal{X}$ . Then*

$$\|(H_t - E)f\|_2^2 \leq e^{-2t\lambda} \text{Var}(f)$$

for any function  $f$ .

PROOF. By elementary calculus,  $\partial_t \|H_t f\|_2^2 = -2\mathcal{E}(H_t f, H_t f)$  and this shows that

$$\partial_t \|(H_t - E)f\|_2^2 = -2\mathcal{E}((H_t - E)f, (H_t - E)f) \leq -2\lambda \|(H_t - E)f\|_2^2.$$

Hence

$$\|(H_t - E)f\|_2^2 \leq e^{-2\lambda t} \text{Var}(f). \quad \square$$

REMARKS. (i) The quantity  $1 - \lambda$  is the second largest eigenvalue of the self-adjoint operator  $\frac{1}{2}(K + K^*)$ .

(ii) Taking the supremum over all functions  $f$  such that  $\|f\|_2 = 1$  in the conclusion of the lemma yields

$$\|H_t - E\|_{2 \rightarrow 2} \leq e^{-\lambda t}.$$

It can be shown that  $\lambda$  is the largest positive number such that such an inequality holds for all  $t > 0$ .

(iii) In the computation above we can either restrict ourselves to real functions or work with complex-valued functions. This does not affect the definition of  $\lambda$ .

To obtain a similar result in discrete time, define  $\beta \geq 0$  by

$$(2.5) \quad \beta = \|K - E\|_{2 \rightarrow 2}.$$

Then we obviously have  $\|K^n - E\|_{2 \rightarrow 2} \leq \beta^n$ . Moreover, one easily shows that  $\beta^2$  is the second largest eigenvalue of the self-adjoint operator  $K^*K$  (or  $KK^*$ ). In other words,  $\beta$  is the second largest singular value of  $K$ . Thus, setting

$$\mathcal{E}_*(f, f) = \langle (I - K^*K)f, f \rangle.$$

we can characterize  $\beta$  by

$$1 - \beta^2 = \min \{ \mathcal{E}_*(f, f) : \|f\|_2 = 1, Ef = 0 \};$$

see [23] and [10]. Section 2 of [10] gives a detailed comparison of the forms  $\mathcal{E}$  and  $\mathcal{E}_*$  and gives examples of the use of  $\mathcal{E}_*$ . Here, we will work only with  $\mathcal{E}$ .

REMARK. When  $(K, \pi)$  is reversible, the definitions of  $\lambda$  and  $\beta$  given in (2.4) and (2.5) are equivalent to (2.1).

In the sequel, we will need some technical results about the Dirichlet form  $\mathcal{E}$ . These are collected here for convenience. They are used throughout (e.g., in showing that a log-Sobolev inequality implies exponential decay of the entropy distance to stationarity; cf. Theorem 3.5).

LEMMA 2.5. For any chain  $(K, \pi)$  and  $p \geq 1$ , the Dirichlet form  $\mathcal{E}$  satisfies

$$\partial_t \|H_t f\|_p^p|_{t=0} = -p\mathcal{E}(f, f^{p-1})$$

for all nonnegative functions  $f$ . Further,

$$\partial_t \text{Ent}_\pi(H_t f)|_{t=0} = -\mathcal{E}(f, \log f).$$

The proofs are obvious.

LEMMA 2.6. Let  $p \geq 2$ . For any chain  $K$  with invariant measure  $\pi$  and any function  $f \geq 0$ ,

$$\mathcal{E}(f, f^{p-1}) \geq \frac{2}{p}\mathcal{E}(f^{p/2}, f^{p/2}).$$

Further, if  $(K, \pi)$  is reversible, then

$$\mathcal{E}(f, f^{p-1}) \geq \frac{4(p-1)}{p^2}\mathcal{E}(f^{p/2}, f^{p/2})$$

for all  $1 < p < \infty$ .

PROOF. When  $p \geq 2$  the function  $t \rightarrow t^{p/2}$  is convex on  $[0, \infty[$ . Now, for any smooth convex function  $\phi$ ,

$$\phi(a) - \phi(b) \geq \phi'(b)(a - b).$$

Hence, we have  $(a^{p/2} - b^{p/2}) \geq (p/2)b^{(p/2)-1}(a - b)$ . Multiplying by  $-b^{p/2}$ , we get

$$(b^{p/2} - a^{p/2})b^{p/2} \leq \frac{p}{2}b^{p-1}(b - a),$$

for all  $a, b \geq 0$ . This gives

$$[(I - K)f^{p/2}]f^{p/2} \leq \frac{p}{2}[(I - K)f]f^{p-1}.$$

Hence

$$\mathcal{E}(f, f^{p-1}) \geq \frac{2}{p}\mathcal{E}(f^{p/2}, f^{p/2}).$$

For the second inequality, write for any  $a > b \geq 0$ ,

$$\begin{aligned} \left( \frac{a^{p/2} - b^{p/2}}{a - b} \right)^2 &= \left( \frac{p}{2(a - b)} \int_b^a t^{p/2-1} dt \right)^2 \\ &\leq \frac{p^2}{4(a - b)} \int_b^a t^{p-2} dt = \frac{p^2}{4(p - 1)} \frac{a^{p-1} - b^{p-1}}{a - b}. \end{aligned}$$

This shows that

$$(a^{p-1} - b^{p-1})(a - b) \geq \frac{4(p - 1)}{p^2} (a^{p/2} - b^{p/2})^2$$

and the second inequality stated in the lemma easily follows from this and (2.3).  $\square$

LEMMA 2.7. *For any chain  $K$  with invariant measure  $\pi$  and any function  $f \geq 0$ ,*

$$\mathcal{E}(\log f, f) \geq 2\mathcal{E}(\sqrt{f}, \sqrt{f}).$$

Further, any reversible chain  $(K, \pi)$  satisfies

$$\mathcal{E}(\log f, f) \geq 4\mathcal{E}(\sqrt{f}, \sqrt{f}).$$

PROOF. Since  $t \rightarrow -\log t^2$  is a convex function, we have

$$-(\log a^2 - \log b^2) \geq -\frac{2}{b}(a - b),$$

for all  $a, b > 0$ . Multiplying by  $-b^2$  yields

$$b^2(\log a^2 - \log b^2) \leq 2b(a - b),$$

for all  $a, b \geq 0$ . This shows that

$$f(x)[(K - I)\log f](x) \leq 2\sqrt{f(x)}[(K - I)\sqrt{f}](x),$$

which yields the first stated result.

To obtain the improved inequality in the reversible case write, for  $a \geq b \geq 0$ ,

$$\begin{aligned} \left( \frac{a^{1/2} - b^{1/2}}{a - b} \right)^2 &= \left( \frac{1}{2(a - b)} \int_b^a \frac{dt}{t^{1/2}} \right)^2 \\ &\leq \frac{1}{4(a - b)} \int_b^a \frac{dt}{t} = \frac{1}{4} \frac{\log a - \log b}{a - b}. \end{aligned}$$

This gives

$$4(a^{1/2} - b^{1/2})^2 \leq (\log a - \log b)(a - b),$$

which, together with (2.3), yields the desired inequality.  $\square$

REMARK. The difficulty in proving the statements of Lemmas 2.5 and 2.7 for nonreversible chains comes from the fact that (2.3) does not hold in general.

2.4. *Distances to equilibrium.* One of the goals of this paper is to study quantitatively the convergence of finite ergodic Markov chains to equilibrium. This can be discussed using various “distances” between probability measures, including total variation, relative entropy or the chi-square distance, the last being the most used in the present paper.

To start with, let us consider a notion of convergence that plays a very distinguished part: convergence in operator norm on  $l^2 = l^2(\pi)$ . Let  $\rho$  denote the spectral radius of the operator  $K - E$  acting on  $l^2$  (i.e., the smallest radius of a circle centered at the origin and containing all the eigenvalues of  $K - E$ ). Because  $K$  preserves the subspace of functions orthogonal to the constants,  $\rho$  is also the spectral radius of  $K$  acting on this subspace. We have

$$\lim_{n \rightarrow +\infty} \|K^n - E\|_{2 \rightarrow 2}^{1/n} = \rho.$$

Also, let  $\tau$  be the maximum of the real part of the spectrum of  $K - E$ . Then

$$\lim_{t \rightarrow +\infty} \|H_t - E\|_{2 \rightarrow 2}^{1/t} = e^{-(1-\tau)}$$

and  $\tau \leq \rho$ . It follows from classical matrix analysis (see [28], page 322) that any norm  $\mathcal{N}$  on matrices also satisfies

$$(2.6) \quad \lim_{n \rightarrow +\infty} \mathcal{N}^{1/n}(K^n - E) = \rho, \quad \lim_{t \rightarrow +\infty} \mathcal{N}^{1/t}(H_t - E) = e^{-(1-\tau)}.$$

As an example, this holds when  $\mathcal{N}(M) = \max_x \sum_y |M_{x,y}|$  which corresponds to total variation. Thus, in this qualitative sense,  $\rho$  and  $\tau$  determine the asymptotic rate of convergence to equilibrium in discrete and continuous time, respectively, whatever norm is chosen. However, this general fact does not give any clue for quantitative problems such as the following:

Given  $\varepsilon > 0$ , find  $N = N(\varepsilon)$  such that  $\mathcal{N}(K^n - E) \leq \varepsilon$  for all  $n \geq N(\varepsilon)$ .

Let us emphasize here that, in general, the answer “it takes roughly  $n = 1/(1 - \rho)$  for  $K^n$  to be close to stationarity” is just wrong.

Instead of  $\rho$  and  $\tau$ , consider the parameters  $\beta$  and  $\lambda$  introduced in (2.4) and (2.5). Using these parameters, we can replace the asymptotic statements (2.6) by the inequalities

$$(2.7) \quad \|K^n - E\|_{2 \rightarrow 2} \leq \beta^n, \quad \|H_t - E\|_{2 \rightarrow 2} \leq e^{-\lambda t},$$

which are more useful as answers to the question above. For comparison, observe that

$$\rho \leq \beta, \quad \lambda \leq 1 - \tau.$$

Further, it can be shown that  $1 - \lambda \leq \beta$ ; see [29]. When  $(K, \pi)$  is reversible, we have  $\rho = \beta = \max\{|\beta_{\min}|, \beta_1\}$  and  $\lambda = 1 - \tau = 1 - \beta_1$ .

The main flaw of the notion of convergence in  $l^2$  operator norm is that it has no sharp simple interpretation in terms of kernels. We next consider convergence in total variation, which has a clear interpretation. For two probability measures  $\mu, \nu$ , let

$$\|\mu - \nu\|_{\text{TV}} = \sup_{A \subset X} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_x |\mu(x) - \nu(x)|.$$

Thus, in the case at hand,

$$2\|K_x^n - \pi\|_{\text{TV}} = \sum_y |K_x^n(y) - \pi(y)| = \|k_x^n - 1\|_1.$$

Jensen's inequality shows that the total variation distance is dominated by the chi-square distance, namely,

$$2\|K_x^n - \pi\|_{\text{TV}} \leq \|k_x^n - 1\|_2 = \left( \sum_y \left| \frac{K_x^n(y)}{\pi(y)} - 1 \right|^2 \pi(y) \right)^{1/2}.$$

The chi-square distance  $\|k_x^n - 1\|_2$  is, in turn, dominated by the relative error  $\sup_y |k_x^n(y) - 1|$ . Finally, observe that for reversible chains the maximal relative error at time  $2n$ , that is,  $\sup_{x,y} |k_x^{2n}(y) - 1|$ , is *equal to* the square maximal chi-square error  $\sup_x \|k_x^n - 1\|_2^2$  at time  $n$ . These observations hold without changes in continuous time if we replace  $K$  by  $H$ ,  $k$  by  $h$  and  $n$  by  $t$ .

We end this discussion by considering yet another quantity that can be used to analyze convergence to equilibrium. Recall that the (relative) entropy is defined by

$$\text{Ent}_\pi(\mu) = \sum_x \mu(x) \log \frac{\mu(x)}{\pi(x)}.$$

Elementary considerations show that

$$(2.8) \quad 2\|\mu - \pi\|_{\text{TV}}^2 \leq \text{Ent}_\pi(\mu) \leq \|\mu - \pi\|_{\text{TV}} + \frac{1}{2}\|(\mu/\pi) - 1\|_{\pi,2}^2.$$

For the upper bound, consider only the  $x$  where  $\mu/\pi \geq 1$  and use the fact that  $(1+u)\log(1+u) \leq u + \frac{1}{2}u^2$  for  $u \geq 0$ . For the lower bound use the inequality  $\forall u > 0, 3(u-1)^2 \leq (4+2u)(u \log u - u + 1)$ , the Cauchy-Schwarz inequality and the fact that  $u \log u - u + 1 \geq 0$  for  $u > 0$ . In his Ph.D. thesis, Su [49] observed that one also has

$$\text{Ent}_\pi(\mu) \leq \log(1 + \|(\mu/\pi) - 1\|_{\pi,2}^2).$$

Very often, convergence is proved by showing that the chi-square distance tends to zero; this gives bounds on both variation distance and entropy as well as bounds on relative error. Let us now state the simplest and most basic quantitative bounds on the chi-square distance.

**LEMMA 2.8.** *Any finite Markov chain  $K$  with invariant probability  $\pi$  satisfies*

$$\|k_x^n - 1\|_2 \leq \pi(x)^{-1/2} \beta^n, \quad \|h_t^x - 1\|_2 \leq \pi(x)^{-1/2} e^{-t\lambda},$$



with  $\lambda, \beta$  defined in (2.4) and (2.5). In particular,

$$\|k_x^n - 1\|_2 \leq e^{-c} \quad \text{for } n = \frac{1}{1 - \beta} \left( \frac{1}{2} \log \frac{1}{\pi(x)} + c \right), c > 0.$$

Similarly,

$$\|h_t^x - 1\|_2 \leq e^{-c} \quad \text{for } t = \frac{1}{\lambda} \left( \frac{1}{2} \log \frac{1}{\pi(x)} + c \right), c > 0.$$

PROOF. Define  $\delta_x(y)$  to be equal to  $1/\pi(x)$  if  $y = x$  and zero otherwise. Observe that

$$\begin{aligned} (K^{*n} - E)\delta_x(y) &= \sum_z (K^{*n}(y, z) - \pi(z))\delta_x(z) \\ &= \frac{K^n(x, y)}{\pi(y)} - 1 = k^n(x, y) - 1. \end{aligned}$$

Hence,

$$\begin{aligned} \|k_x^n - 1\|_2 &= \|(K^{*n} - E)\delta_x\|_2 \\ &\leq \text{Var}(\delta_x)^{1/2} \|K^n - E\|_{2 \rightarrow 2} \leq \left( \frac{1 - \pi(x)}{\pi(x)} \right)^{1/2} \beta^n. \end{aligned}$$

The proof of the corresponding result for  $h_t^x$  follows the same line of reasoning. Compare with [23, 10]. This proof extends readily to any ergodic Markov chain on a denumerable state space. It only uses the variational definition of  $\beta$  or  $\lambda$  and not the entire spectral decomposition.  $\square$

REMARKS. (i) The formula  $(K^{*n} - E)\delta_x(y) = k^n(x, y) - 1$  and the fact that  $K^*$  contracts  $l^p(\pi)$  for  $1 \leq p \leq \infty$  show that  $n \rightarrow \|k_x^n - 1\|_p$  is a nonincreasing function (similarly for  $t \rightarrow \|h_t^x - 1\|_p$ ). Dividing by 2, we get the same result for total variation.

(ii) The distances considered above belong to a larger family defined by setting

$$d_{p,q}(n) = \left( \sum_x \left( \sum_y |k^n(x, y) - 1|^p \pi(y) \right)^{q/p} \pi(x) \right)^{1/q},$$

where  $1 \leq p, q \leq +\infty$  and obvious modifications when  $p$  or  $q$  are infinite. For instance,  $d_{2,2}$  (which is nothing else than the Hilbert–Schmidt norm of  $K^n - E$ ) can be of special interest because, when  $(K, \pi)$  is reversible,

$$d_{2,2}(n) = \left( \sum_1^{|\mathcal{X}|-1} \beta_i^{2n} \right)^{1/2}$$

and thus depends only on the eigenvalues, whereas  $d_{2,\infty}$  depends on eigenvalues and eigenfunctions. When a group acts transitively on  $X$  and preserves  $K$ ,  $d_{p,q} = d_p$  for all  $p, q$ ; this happens in particular when  $K$  is the simple

random walk on the Cayley graph of a group. Note that  $d_{p,q} \leq d_{p',q'} \leq d_\infty$  for  $p \leq p'$  and  $q \leq q'$ . The distance  $d_{1,1}$  has a natural appearance in long runs of a Markov chain when the chain is run  $n$  steps to stationarity, the output used, the chain run again and so on.

(iii) Of course, the bounds in Lemma 2.8 are bounds on total variation as well. Bounds that are specific to variation distance can also be obtained by probabilistic arguments (e.g., coupling, strong stationary times). In this context, the maximum separation distance

$$d_{\text{sep}}(n) = \sup_{x,y} \{1 - k^n(x,y)\}$$

(no absolute value) is of special interest. See [1] and [9] for details.

(iv) Lemma 2.8 justifies (if necessary) the amount of work devoted to bounding  $\beta$  from above and  $\lambda$  from below. For instance, [21] gives recent results and earlier references. Among others, [12], [23] and [43] describe applications to complex combinatorial examples. See also [44].

(v) Lemma 2.8 tells us that after  $n = (-\log \pi(x)/2(1 - \beta))$  the chain started at  $x$  is close to equilibrium. In many nontrivial examples this is not (even roughly) an optimal result. To illustrate this point, let us consider three basic examples. For simple random walk on the finite circle  $\mathbb{Z}/p\mathbb{Z}$ ,  $p$  odd, we have  $\pi(x) = 1/p$ ,  $1 - \beta \sim 1/p^2$  and the lemma predicts randomness after  $p^2 \log p$  steps. In fact, order  $p^2$  steps are necessary and sufficient. For simple random walk on the hypercube  $\mathbb{Z}_2^d$ , we get  $\pi(x) = 1/2^d$ ,  $1 - \beta = 2/d$  and the prediction is that order  $d^2$  steps are sufficient whereas  $\frac{1}{4}d \log d$  is the right answer. For random transpositions on the symmetric group  $S_d$ ,  $\pi(x) = 1/d!$ ,  $1 - \beta = 2/d$  and the lemma ensures randomness after order  $d^2 \log d$  steps. Diaconis and Shahshahani [19] have shown that the right answer is  $\frac{1}{2}d \log d$ . For a precise analysis of these examples, see [9]. References [10], [11] and [13] describe a host of other cases and develop techniques that yield improved bounds. This is also the goal of the present work.

(vi) The preceding remark should be balanced with the following comment. Fix an integer  $r$  and  $0 < \varepsilon < 1/2$  and consider the nearest-neighbor random walk on  $r$ -regular graphs. It can be shown that “most”  $r$ -regular graphs have  $\lambda \geq \varepsilon$ ; see [2], [33], [42] and the references given there. Also, it obviously takes at least order  $\log_r |\mathcal{X}|$  steps for the nearest-neighbor random walk on an  $r$ -regular graph with vertex set  $\mathcal{X}$  to be close to equilibrium in variation distance. Thus, for “most”  $r$ -regular graphs, Lemma 2.8 gives the correct order of magnitude. See also Section 4.

**2.5. Product chains.** Product chains are of interest both in their own right (the random walk on the hypercube is a product chain) and as a base for comparison in analyzing more complex chains. For  $i = 1, 2, \dots, d$ , let  $K_i$  be a Markov chain on a finite state space  $\mathcal{X}_i$  with invariant probability  $\pi_i$ . We define a chain on the product which corresponds to choosing a coordinate uniformly at random and taking a step in that coordinate. Set  $\mathcal{X} = \prod_1^d \mathcal{X}_i$  and

consider the chain

$$K(x, y) = \frac{1}{d} \sum_{i=1}^d \delta(x_1, y_1) \cdots \delta(x_{i-1}, y_{i-1}) \\ \times K_i(x_i, y_i) \delta(x_{i+1}, y_{i+1}) \cdots \delta(x_d, y_d),$$

where  $x = (x_i)_1^d, y = (y_i)_1^d$  and  $\delta(u, v) = 1$  if  $u = v$  and 0 otherwise. In terms of operators, this means that

$$(2.9) \quad K = \frac{1}{d} \sum_{i=1}^d \underbrace{I \otimes \cdots \otimes I}_{i-1} \otimes K_i \otimes \underbrace{I \otimes \cdots \otimes I}_{d-i}.$$

This  $K$  has invariant distribution

$$(2.10) \quad \pi(x) = \prod_{i=1}^d \pi_i(x_i).$$

Now, let  $H_{i,t}, H_t$  be the semigroups corresponding to  $K_i, K$ . The definitions yield

$$(2.11) \quad H_t(x, y) = \prod_{i=1}^d H_{i,t/d}(x_i, y_i)$$

and the same relation for the relative kernels  $h_t, h_{i,t}$ . From this, we deduce a bound on the rate of convergence of product chains that we found surprisingly difficult.

**THEOREM 2.9.** *Assume that there exists  $b, B_0, B_1 > 0$  such that the factor chains satisfy*

$$\|h_{i,s}^{x_i} - 1\|_2 \leq b e^{-\gamma}$$

for some fixed  $x_i \in \mathcal{X}_i, i = 1, 2, \dots, d$ , and  $s = B_0(B_1 + \gamma), \gamma > 0$ . Set  $x = (x_1, \dots, x_d) \in \mathcal{X} = \prod_1^d \mathcal{X}_i$ . Then the product chain  $(K, \pi)$  defined in (2.9)–(2.11) satisfies

$$\|h_t^x - 1\|_2 \leq b \exp\left(\frac{b^2}{2} - c\right)$$

for  $t = dB_0(B_1 + \frac{1}{2} \log d + c), c > 0$ .

**PROOF.** Let  $x = (x_i)_1^d \in \mathcal{X}$  and write

$$\|h_t^x - 1\|_2^2 = \|h_t^x\|_2^2 - 1 = \prod_{i=1}^d \|h_{i,t/d}^{x_i}\|_2^2 - 1 \\ = \prod_{i=1}^d (\|h_{i,t/d}^{x_i} - 1\|_2^2 + 1) - 1.$$

Now, if  $t = dB_0(B_1 + \frac{1}{2} \log d + c), c > 0$ , we have  $\|h_{i,t/d}^{x_i} - 1\|_2^2 \leq (b^2/d)e^{-2c}$ . Thus, we get

$$\|h_t^x - 1\|_2^2 \leq \left(1 + \frac{b^2}{d} e^{-2c}\right)^d - 1 \leq b^2 \exp(b^2 - 2c). \quad \square$$

Note that, if for some  $s, x_i, \varepsilon$  we have  $\|h_{i,s}^x - 1\|_2 \geq \varepsilon$  for all  $i = 1, \dots, d$ , then the above argument shows that  $\|h_{d,s}^x - 1\|_2 \geq \sqrt{d} \varepsilon$ .

For reversible chains, a discrete time version follows from Theorem 2.9 using Corollary 2.2.

**THEOREM 2.10.** *Assume that there exists  $b, B_0, B_1$  such that  $\|k_{i,x_i}^n - 1\|_2 \leq be^{-\gamma}$  for some fixed  $x_i \in \mathcal{X}_i, i = 1, 2, \dots, d$ , and  $n \geq B_0(B_1 + \gamma), \gamma > 0$ . Assume moreover that the  $(K_i, \pi_i)$ 's are reversible. Set  $x = (x_1, \dots, x_d) \in \mathcal{X} = \prod_1^d \mathcal{X}_i$ . Then the chain  $(K, \pi)$  defined in (2.9) and (2.10) is reversible and satisfies*

$$\|k_x^m - 1\|_2 \leq (1 + 2 \exp(1 + b^2))^{1/2} e^{-c}$$

for  $m \geq 2 dB_0(B_2 + \frac{1}{2} \log d + c) + 1, c > 0$ , where  $B_2 = \max\{B_1, \log(\xi)/2B_0\}, \xi = \max_i(1/\pi_{i,*})$ .

Further details that extend with minor modifications to the above situation can be found in [11], Section 5. We emphasize that we do not know how to extend this result to nonreversible chains. The above results are rather sharp. For instance, Theorems 2.9 and 2.10 show that simple random walk on the torus  $(\mathbb{Z}/p\mathbb{Z})^d, p$  odd, is close to being uniformly distributed after order  $p^2 d \log d$  steps, which is the right answer; see [11].

**3. Logarithmic Sobolev inequalities.** Section 3.1 introduces the log-Sobolev constant  $\alpha$  of a finite Markov chain. This is a constant which always satisfies  $\alpha \leq \lambda/2$ , where  $\lambda$  is the spectral gap of the chain. This constant is shown to behave well under product and comparison (in fact, it behaves exactly as the spectral gap  $\lambda$ ). Section 3.2 contains a self-contained treatment of the relation between the log-Sobolev constant  $\alpha$  and the hypercontractivity of the associated semigroup. Section 3.3 describes how the log-Sobolev constant can be used to bound distance to stationarity. Finally, Section 3.4 gives tools to bound  $\alpha$  from below in terms of upper bounds on the semigroup.

**3.1. The log-Sobolev constant.** Given an irreducible finite Markov chain  $K$  with invariant probability  $\pi$ , consider the Dirichlet form

$$\mathcal{E}(f, g) = \langle (I - K)f, g \rangle$$

introduced in Section 2.3 and set

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

A logarithmic Sobolev (or log-Sobolev) inequality is an inequality of the type

$$\mathcal{L}(f) \leq C \mathcal{E}(f, f)$$

holding for all functions  $f$ . Let  $1/\alpha$  be the smallest constant  $C$  such that this inequality holds. In other words,

$$(3.1) \quad \alpha = \inf \left\{ \frac{\mathcal{E}(f, f)}{\mathcal{L}(f)} : \mathcal{L}(f) \neq 0 \right\}.$$

We say that  $\alpha$  is the log-Sobolev constant of  $K$ . Recall that the spectral gap  $\lambda$  of  $K$  has a similar characterization:

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

where  $\text{Var}(f) = \|f - Ef\|_2^2$  is the  $\pi$ -variance of  $f$ .

The first result compares the log-Sobolev constant to the spectral gap.

**LEMMA 3.1.** *For any chain  $K$  the log-Sobolev constant  $\alpha$  and the spectral gap  $\lambda$  satisfy  $2\alpha \leq \lambda$ .*

**PROOF.** We follow [39]. Set  $f = 1 + \varepsilon g$  and write, for  $\varepsilon$  small enough,

$$\begin{aligned} |f|^2 \log|f|^2 &= 2(1 + 2\varepsilon g + \varepsilon^2|g|^2) \left( \varepsilon g - \frac{\varepsilon^2|g|^2}{2} + O(\varepsilon^3) \right) \\ &= 2\varepsilon g + 3\varepsilon^2|g|^2 + O(\varepsilon^3) \end{aligned}$$

and

$$\begin{aligned} |f|^2 \log\|f\|_2^2 &= (1 + 2\varepsilon g + \varepsilon^2|g|^2)(2\varepsilon Eg + \varepsilon^2\|g\|_2^2 - 2\varepsilon^2(Eg)^2 + O(\varepsilon^3)) \\ &= 2\varepsilon Eg + 4\varepsilon^2gEg + \varepsilon^2\|g\|_2^2 - 2\varepsilon^2(Eg)^2 + O(\varepsilon^3). \end{aligned}$$

Thus,

$$|f|^2 \log \frac{|f|^2}{\|f\|_2^2} = 2\varepsilon(g - Eg) + \varepsilon^2(3|g|^2 - \|g\|_2^2 - 4gEg + 2(Eg)^2) + O(\varepsilon^3)$$

and

$$\begin{aligned} \mathcal{L}(f) &= 2\varepsilon^2(\|g\|_2^2 - (Eg)^2) + O(\varepsilon^3) \\ &= 2\varepsilon^2 \text{Var}(g) + O(\varepsilon^3). \end{aligned}$$

To finish the proof, observe that  $\mathcal{E}(f, f) = \varepsilon^2\mathcal{E}(g, g)$ , multiply by  $\varepsilon^{-2}$ , let  $\varepsilon$  tend to zero and use the variational characterizations of  $\alpha$  and  $\lambda$ .  $\square$

**REMARK.** To what extent can one hope to have  $2\alpha = \lambda$  or at least  $\alpha$  and  $\lambda$  of the same order of magnitude? This is a classic question in the literature on log-Sobolev inequalities. Indeed, in most examples where  $2\alpha$  and  $\lambda$  are explicitly known, they turn out to be equal. In Section 4.5 we show that, in some sense, for simple random walk on a generic regular graph,  $\alpha$  and  $\lambda$  are of different orders of magnitude.

**EXAMPLE 3.1.** Consider the two-point space  $\mathcal{X} = \{-1, 1\}$  with kernel  $K(-1, 1) = K(1, -1) = 1$ . This has stationary measure  $\pi \equiv 1/2$ . The definition (3.1), together with tedious calculus, shows that  $\alpha \geq 1$  in this case. Since it is easy to check that  $\lambda = 2$  we find that, in this case, Lemma 3.1 is sharp and  $\alpha = \lambda/2 = 1$ . For further details, see, for example, Example 2.6 in [24].

The next two lemmas are crucial for the applications we have in mind. Indeed, computing log-Sobolev constants turns out to be extremely difficult. Lemma 3.2 gives a collection of examples (i.e., products) with good log-Sobolev constants. Lemma 3.3 allows comparison of an unknown chain  $K$  with the better known chain  $K'$ .

**LEMMA 3.2.** *Let  $(K_i, \pi_i)$ ,  $i = 1, \dots, d$ , be Markov chains on finite sets  $\mathcal{X}_i$  with spectral gaps  $\lambda_i$  and log-Sobolev constants  $\alpha_i$ . Then the product chain  $(K, \pi)$  on  $\mathcal{X} = \prod_1^d \mathcal{X}_i$  defined at (2.9)–(2.11) satisfies*

$$\lambda = \frac{1}{d} \min_i \lambda_i, \quad \alpha = \frac{1}{d} \min_i \alpha_i.$$

**PROOF.** If  $\mathcal{E}_i$  denotes the Dirichlet form associated with  $K_i$ , then the product chain  $K$  defined in Section 2.5 has Dirichlet form

$$\mathcal{E}(f, f) = \frac{1}{d} \sum_1^d \sum_{x_j: j \neq i} \mathcal{E}_i(f, f) \prod_{l: l \neq i} \pi_l(x_l),$$

where  $\mathcal{E}_i(f, f) = \mathcal{E}_i(f(x_1, \dots, x_d), f(x_1, \dots, x_d))$  has the obvious meaning:  $\mathcal{E}_i$  acts on the  $i$ th coordinate whereas the other coordinates are fixed. It is enough to prove the following statement: Let  $\mathcal{X}_i$ ,  $i = 1, 2$ , be two finite sets. Let  $K_i$  be a Markov chain on  $\mathcal{X}_i$  with invariant measure  $\pi_i$  and Dirichlet form  $\mathcal{E}_i$ . Consider the Dirichlet form on  $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$  defined by

$$\begin{aligned} \mathcal{E}(f, f) &= \theta_1 \sum_{x_2} \mathcal{E}_1(f(\cdot, x_2), f(\cdot, x_2)) \pi_2(x_2) \\ &\quad + \theta_2 \sum_{x_1} \mathcal{E}_2(f(x_1, \cdot), f(x_1, \cdot)) \pi_1(x_1), \end{aligned}$$

where  $\theta_1, \theta_2$  are positive fixed parameters. Let  $\pi(x_1, x_2) = \pi(x_1)\pi(x_2)$ . Then, if  $\lambda_i, \alpha_i$  are the spectral gaps and log-Sobolev constants of  $\mathcal{E}_i$ ,  $i = 1, 2$ , the form  $\mathcal{E}$  has spectral gap

$$\lambda = \min\{\theta_1 \lambda_1, \theta_2 \lambda_2\}$$

and log-Sobolev constant

$$\alpha = \min\{\theta_1 \alpha_1, \theta_2 \alpha_2\}.$$

We will only prove the statement for  $\alpha$ . The proof for  $\lambda$  is similar. Let  $f: \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathbb{R}$  be a nonnegative function and set

$$F(x_2) = \left( \sum_{x_1} f(x_1, x_2)^2 \pi_1(x_1) \right)^{1/2}.$$

Then write

$$\begin{aligned} \mathcal{L}(f) &= \sum_{x_1, x_2} |f(x_1, x_2)|^2 \log \frac{f(x_1, x_2)^2}{\|f\|_{2, \pi}^2} \pi(x_1, x_2) \\ &= \sum_{x_2} |F(x_2)|^2 \log \frac{F(x_2)^2}{\|F\|_{2, \pi_2}^2} \pi_2(x_2) \\ &\quad + \sum_{x_1, x_2} |f(x_1, x_2)|^2 \log \frac{f(x_1, x_2)^2}{F(x_2)^2} \pi(x_1, x_2) \\ &\leq [\theta_2 \alpha_2]^{-1} \theta_2 \mathcal{E}_2(F, F) \\ &\quad + [\theta_1 \alpha_1]^{-1} \sum_{x_2} \theta_1 \mathcal{E}_1(f(\cdot, x_2), f(\cdot, x_2)) \pi_2(x_2). \end{aligned}$$

Now, the triangle inequality

$$\begin{aligned} |F(x_2) - F(y_2)| &= |\|f(\cdot, x_2)\|_{2, \pi_1} - \|f(\cdot, y_2)\|_{2, \pi_1}| \\ &\leq \|f(\cdot, x_2) - f(\cdot, y_2)\|_{2, \pi_1} \end{aligned}$$

implies that

$$\mathcal{E}_2(F, F) \leq \sum_{x_1} \mathcal{E}_2(f(x_1, \cdot), f(x_1, \cdot)) \pi_1(x_1).$$

Hence

$$\begin{aligned} \mathcal{L}(f) &\leq [\theta_2 \alpha_2]^{-1} \sum_{x_1} \theta_2 \mathcal{E}_2(f(x_1, \cdot), f(x_1, \cdot)) \pi_1(x_1) \\ &\quad + [\theta_1 \alpha_1]^{-1} \sum_{x_2} \theta_1 \mathcal{E}_1(f(\cdot, x_2), f(\cdot, x_2)) \pi_2(x_2), \end{aligned}$$

which yields

$$\mathcal{L}(f) \leq \max_i \{1/[\theta_i \alpha_i]\} \mathcal{E}(f, f).$$

This shows that  $\alpha \geq \min_i [\theta_i \alpha_i]$ . Testing on functions that depend on one of the two variables shows that  $\alpha = \min_i [\theta_i \alpha_i]$ .  $\square$

**EXAMPLE 3.2.** Consider the hypercube  $\mathcal{X} = \{-1, 1\}^n$ . Let  $K(x, y) = 1/n$  if  $x, y$  differ at exactly one coordinate and  $K(x, y) = 0$  otherwise. The stationary distribution is the uniform measure  $\pi \equiv 2^{-n}$ . Further,  $K$  is the product chain on  $\mathcal{X}$  coming from the two-point chains on each coordinate. Thus, Lemma 3.2 and Example 3.1 give  $\alpha = \lambda/2 = 1/n$ . This result has several nontrivial corollaries among which we mention the determination of the sharp log-Sobolev constant for the Ornstein–Uhlenbeck semigroup; see [5], [24] and [25] for proofs and historical comments.

The next lemma allows comparison between spectral gaps and log-Sobolev constants of two chains on the same state space in the presence of a comparison between Dirichlet forms and stationary distributions.

LEMMA 3.3. *Let  $(K, \pi)$  and  $(K', \pi')$  be two Markov chains on the same finite set  $\mathcal{X}$ . Assume that there exist  $A, a > 0$  such that*

$$\mathcal{E}' \leq A\mathcal{E}, \quad a\pi \leq \pi'.$$

Then

$$\lambda' \leq \frac{A}{a}\lambda, \quad \alpha' \leq \frac{A}{a}\alpha.$$

PROOF. The first stated result follows from the variational definition of  $\lambda$  together with the formula

$$\text{Var}(f) = \min_{c \in \mathbb{R}} \sum_x |f(x) - c|^2 \pi(x).$$

The inequality between log-Sobolev constants follows from an observation due to Holley and Stroock [27]:  $\xi \log \xi - \xi \log \zeta - \xi + \zeta \geq 0$  for  $\xi, \zeta > 0$  and

$$\begin{aligned} \mathcal{L}_\pi(f) &= \sum_x (|f(x)|^2 \log |f(x)|^2 - |f(x)|^2 \log \|f\|_2^2 - |f(x)|^2 + \|f\|_2^2) \pi(x) \\ &= \min_{c > 0} \sum_x (|f(x)|^2 \log |f(x)|^2 - |f(x)|^2 \log c - |f(x)|^2 + c) \pi(x). \quad \square \end{aligned}$$

Lemma 3.3 can be extended to allow comparison of chains defined on two different state spaces as in the following lemma.

LEMMA 3.4. *Let  $(K, \pi)$  and  $(K', \pi')$  be two Markov chains defined, respectively, on the finite sets  $\mathcal{X}$  and  $\mathcal{X}'$ . Assume that there exists a linear map*

$$l^2(\mathcal{X}, \pi) \rightarrow l^2(\mathcal{X}', \pi'): f \rightarrow \tilde{f}$$

and constants  $A, B, a > 0$  such that, for all  $f \in l^2(\mathcal{X}, \pi)$ ,

$$\mathcal{E}'(\tilde{f}, \tilde{f}) \leq A\mathcal{E}(f, f) \quad \text{and} \quad a \text{Var}_{\pi'}(\tilde{f}) \leq \text{Var}_{\pi}(f) + B\mathcal{E}(f, f).$$

Then

$$\frac{a\lambda'}{A + B\lambda'} \leq \lambda.$$

Similarly, if

$$\mathcal{E}'(\tilde{f}, \tilde{f}) \leq A\mathcal{E}(f, f) \quad \text{and} \quad a\mathcal{L}_{\pi'}(\tilde{f}) \leq \mathcal{L}_{\pi}(f) + B\mathcal{E}(f, f),$$

then

$$\frac{a\alpha'}{A + B\alpha'} \leq \alpha.$$

REMARK. In [17], we apply this to a simple case where  $\mathcal{X} \subset \mathcal{X}'$ ; functions on  $\mathcal{X}$  are extended to  $\mathcal{X}'$  by an interpolation procedure which allows us to keep control of the different constants. In this particular case, we can take  $B = 0$ .

EXAMPLE 3.3. Fix a positive integer  $n$ . Consider the chain  $K$  on  $\{-1, 1\}^n$  from Example 3.2. This chain induces a birth and death chain on  $\{0, 1, \dots, n\}$



which counts the number of 1's in  $x \in \{-1, 1\}^n$ . This induced chain has kernel

$$P(x, y) = \begin{cases} (n - x)/n, & \text{if } y = x + 1, 0 \leq x \leq n - 1, \\ x/n, & \text{if } y = x - 1, 1 \leq x \leq n, \end{cases}$$

and stationary measure  $\pi = 2^{-n} \binom{n}{x}$ . This is the classical Ehrenfest chain. As a function of the chain on the hypercube, it has  $\alpha(P) \geq 1/n$ . However, the spectral gap is known to be  $\lambda = 2/n$  (the same as for the hypercube). It follows that  $\alpha(P) = 1/n$ . We can now use this result to study the Metropolis chain  $M$  introduced in (1.9), which also has the binomial distribution  $\pi(x) = 2^{-n} \binom{n}{x}$  as stationary measure. Comparing the kernels  $M$  and  $P$ , we get

$$\frac{1}{2}P(x, y) \leq M(x, y) \leq P(x, y)$$

for  $x \neq y$ . Hence,

$$\mathcal{E}_M \leq \mathcal{E}_P \leq 2\mathcal{E}_M.$$

This shows that  $\frac{1}{2}\lambda(P) \leq \lambda(M) \leq \lambda(P)$  and  $\frac{1}{2}\alpha(P) \leq \alpha(M) \leq \alpha(P)$ ; hence,

$$\frac{1}{n} \leq \lambda(M) \leq \frac{2}{n}$$

and

$$\frac{1}{2n} \leq \alpha(M) \leq \frac{1}{n}.$$

**3.2. Hypercontractivity.** We now recall the main result relating log-Sobolev inequalities to the so-called hypercontractivity of the semigroup  $H_t$ . We first state the main result and then attempt some motivation. This is followed by proofs.

**THEOREM 3.5.** *Let  $(K, \pi)$  be a finite Markov chain with log-Sobolev constant  $\alpha$ .*

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

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

(iii) *For nonreversible chains, we still have  $\|H_t\|_{2 \rightarrow q} \leq 1$  for all  $t > 0$  and all  $2 \leq q < +\infty$  satisfying  $e^{2\alpha t} \geq q - 1$ .*

The first two assertions are classic parts of the theory of hypercontractivity and log-Sobolev inequality. For reversible chains, they yield the equivalence between the two notions. The proof of (iii) follows the same classic lines with a little twist. As far as we know, this last result is not in the literature. However, this result is known to Bakry, who intended (but forgot—personal communication) to include it in [5].

Let us start with an informal discussion of hypercontractivity. A basic analytic property of a Markov semigroup  $H_t$  with invariant measure  $\pi$  is the fact that  $H_t$  contracts  $l^2(\pi)$  (and all  $l^p$ ). Now, for any irreducible reversible chain  $K$  with invariant measure  $\pi$  on a finite state space  $\mathcal{X}$ , it is not hard to show that  $\|H_t\|_{2 \rightarrow \infty}^2 = \sup_x h_{2t}(x, x) > 1$  for all finite  $t > 0$ , whereas of course  $\lim_{t \rightarrow \infty} \|H_t\|_{2 \rightarrow \infty} = 1$ . Consider the following strange question: is there a finite  $t > 0$  such that  $\|H_t\|_{2 \rightarrow 4} \leq 1$ ? In view of what has been said for  $\|H_t\|_{2 \rightarrow \infty}$  it is not so clear what the answer should be. It turns out that the answer is “yes.” Namely, for any reversible irreducible  $H_t$  on a finite state space, there exists a finite  $t > 0$  such that  $\|H_t\|_{2 \rightarrow 4} \leq 1$  (and thus equal to 1 because of the constant functions). Further, for any finite  $q \geq 2$ , there exists a finite  $t(q) > 0$  such  $\|H_t\|_{2 \rightarrow q} = 1$  for all  $t \geq t_q$ . This property is called hypercontractivity and any reversible irreducible Markov semigroup on a finite state space is hypercontractive (see Theorem 3.9 for a proof of this well known fact). Let us point out that reversibility is not an issue here and in fact any irreducible Markov semigroup on a finite state space is hypercontractive (see the last remark following Theorem 3.10). Figure 1 illustrates this discussion. For bounding rates of convergence, what we really are interested in is the time  $t_*$  at which  $\|H_{t_*}\|_{2 \rightarrow \infty} \leq 2$ . The constant 2 is arbitrary except that the figure strongly indicates that one should not ask for  $\|H_t\|_{2 \rightarrow \infty} \leq 1$ . At time  $t_*$  the distribution of the continuous time Markov chain is roughly similar to  $\pi$  since  $\|H_{t_*}\|_{2 \rightarrow \infty} = \sup_{x, y} \{H_{2t_*}(x, y)/\pi(y)\}$ . After time  $t_*$ , convergence takes place at the exponential rate  $1/\lambda$ ; see (1.3). What we would like to have is a characterization of  $t_*$  in terms of the Dirichlet form  $\mathcal{E}$ . Unfortunately, such a characterization does not exist at present. The more subtle property,

there exists  $\beta > 0$  such that  $\|H_t\|_{2 \rightarrow q} \leq 1$  for all  $0 < t$ ,  
 $2 \leq q < \infty$ , satisfying  $e^{4\beta t} \geq q - 1$ ,

has the advantage of having an exact translation in terms of Dirichlet forms, namely,  $\beta\mathcal{L} \leq \mathcal{E}$ . This is roughly the content of Theorem 3.5. The largest  $\beta$

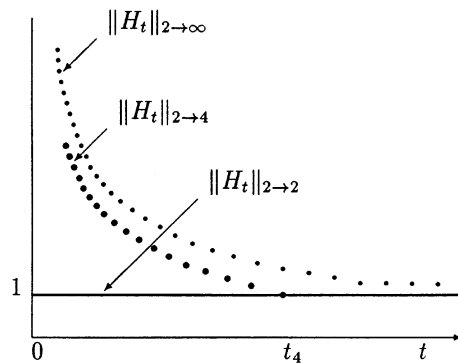


FIG. 1.

for which the above property holds is equal to the log-Sobolev constant  $\alpha$ . This explains the role of hypercontractivity.

To finish this discussion, note that the exact value of  $t_4$  does not give the exact value of  $\alpha$ , but good estimates on  $t_4$  imply good estimates on  $\alpha$  by Theorem 3.9.

PROOF OF THEOREM 3.5. *First statement.* It is convenient to introduce

$$\mathcal{L}_p(f) = \sum_{x \in X} |f(x)|^p \log \left( \frac{|f(x)|^p}{\|f\|_p^p} \right) \pi(x).$$

It is enough to prove the desired log-Sobolev inequality for positive  $f$ . Thus, for  $f > 0$ , set

$$F(t) = \|H_t f\|_{p(t)} \quad \text{where } p(t) = 1 + e^{4\beta t}.$$

We compute the derivative of

$$F(t) = \exp(\log[G(t)]/p(t)) \quad \text{where } G(t) = \|H_t f\|_{p(t)}^{p(t)}.$$

First, we have

$$G'(t) = -p(t)\mathcal{E}(H_t f, (H_t f)^{p(t)-1}) + \frac{p'(t)}{p(t)} \sum_x |H_t f|^{p(t)} \log |H_t f|^{p(t)} \pi(x).$$

Then

$$\begin{aligned} (3.2) \quad F'(t) &= \left[ -\frac{p'(t)\log G(t)}{p(t)^2} + \frac{G'(t)}{p(t)G(t)} \right] F(t) \\ &= F(t)^{-p(t)+1} \left[ \frac{p'(t)}{p(t)^2} \mathcal{L}_p(H_t f) - \mathcal{E}(H_t f, (H_t f)^{p(t)-1}) \right]. \end{aligned}$$

Using the specific formula for  $p(t)$ , we get

$$F'(t) = F(t)^{-p(t)+1} \left[ \frac{4\beta e^{4\beta t}}{(1 + e^{4\beta t})^2} \mathcal{L}_p(H_t f) - \mathcal{E}(H_t f, (H_t f)^{p(t)-1}) \right].$$

Now, since  $\|H_t f\|_{2 \rightarrow p(t)} \leq \|f\|_2$ ,  $H_0 f = f$  and  $p(0) = 2$ , the derivative of  $F(t)$  at  $t = 0$  must be negative. Together with the formula above, this shows that

$$\beta \mathcal{L}(f) \leq \mathcal{E}(f, f),$$

which is the desired inequality.

*Second statement.* Assume that  $(K, \pi)$  is reversible and satisfies the log-Sobolev inequality

$$\alpha \mathcal{L}(f) \leq \mathcal{E}(f, f).$$

For  $f \geq 0$ , Lemma 2.6 gives

$$\alpha \mathcal{L}_p(f) \leq \mathcal{E}(f^{p/2}, f^{p/2}) \leq \frac{p^2}{4(p-1)} \mathcal{E}(f, f^{p-1})$$

for any  $1 < p < \infty$ . If  $p(t) = 1 + e^{4\alpha t}$ , then  $p'(t) = 4\alpha(p(t) - 1)$  and, replacing  $f$  by  $H_t f$ , we obtain

$$\frac{p'(t)}{p(t)^2} \mathcal{L}_{p(t)}(H_t f) - \mathcal{E}(H_t f, (H_t f)^{p(t)-1}) \leq 0.$$

However, using as above the notation  $F(t) = \|H_t f\|_{p(t)}$ , the last inequality and (3.2) yield  $F'(t) \leq 0$  for all  $t \geq 0$ . Since  $F(0) = \|f\|_2$ , this implies

$$\|H_t f\|_{p(t)} \leq \|f\|_2$$

or, taking the supremum over all  $f$  with  $\|f\|_2 = 1$ ,

$$\|H_t\|_{2 \rightarrow p(t)} \leq 1.$$

This is the desired hypercontractivity.

*Third statement.* The proof is almost identical to the one above. The difference comes from the fact that we only have

$$\mathcal{E}(f^{p/2}, f^{p/2}) \leq \frac{p}{2} \mathcal{E}(f, f^{p-1}) \leq \frac{p^2}{2(p-1)} \mathcal{E}(f, f^{p-1})$$

for all  $p \geq 2$ . Thus, we set  $p(t) = 1 + e^{2\alpha t}$ . Proceeding as before, we get again

$$\frac{p'(t)}{p(t)^2} \mathcal{L}_{p(t)}(H_t f) - \mathcal{E}(H_t f, (H_t f)^{p(t)-1}) \leq 0$$

and, together with (3.2), this implies the stated hypercontractivity. This also ends the proof of Theorem 3.5.  $\square$

**3.3. Ergodicity.** One way to use the log-Sobolev constant to discuss ergodicity is through entropy. This is well known and gives a very clean statement. This is not surprising because, if we set  $\mu = f^2 \pi$  for a function  $f \geq 0$ ,  $\|f\|_2 = 1$ , then  $\mathcal{L}(f) = \text{Ent}_\pi(\mu)$ . The following result in the reversible case is contained in [5, 47]. Miclo [35] treats the nonreversible case.

**THEOREM 3.6.** *Let  $K$  be a finite Markov chain with invariant measure  $\pi$  and log-Sobolev constant  $\alpha$ . Then for any probability measure  $\mu$  on  $X$ , we have*

$$\text{Ent}_\pi(\mu H_t) \leq \text{Ent}_\pi(\mu) e^{-2\alpha t}, \quad t > 0.$$

Here  $\mu H_t$  is defined by  $\mu H_t(y) = \sum_x H_t(x, y) \mu(x)$ .

Further, if we assume that  $(K, \pi)$  is reversible, then

$$\text{Ent}_\pi(\mu H_t) \leq \text{Ent}_\pi(\mu) e^{-4\alpha t}, \quad t > 0.$$

**PROOF.** The proof follows readily from Lemmas 2.5 and 2.7. Indeed, write  $\mu = f\pi$ . Then the density of  $\mu H_t$  with respect to  $\pi$  is  $H_t^* f$ . Lemma 2.5 gives

$$\partial_t \text{Ent}(H_t^* f) = -\langle (I - K^*) H_t^* f, \log H_t^* f \rangle$$

and Lemma 2.7 yields

$$\begin{aligned} \langle (I - K^*)H_t^*f, \log H_t^*f \rangle &\geq 2\langle (I - K^*)(H_t^*f)^{1/2}, (H_r^*f)^{1/2} \rangle \\ &= 2\mathcal{E}((H_t^*f)^{1/2}, (H_r^*f)^{1/2}). \end{aligned}$$

Thus,

$$\partial_t \text{Ent}(H_t^*f) \leq -2\mathcal{E}((H_t^*f)^{1/2}, (H_t^*f)^{1/2}) \leq -2\alpha \text{Ent}(H_t^*f).$$

This immediately yields

$$\text{Ent}(H_t^*f) \leq e^{-2\alpha t} \text{Ent}(f).$$

The improvement in the reversible case follows from the corresponding improvement appearing in Lemma 2.7.  $\square$

REMARK. As a corollary of the last result we obtain the inequality

$$(3.3) \quad 2\|H_t^x - \pi\|_{\text{TV}}^2 \leq \left( \log \frac{1}{\pi(x)} \right) e^{-2\alpha t},$$

which may be compared with the bound

$$(3.4) \quad 4\|H_t^x - \pi\|_{\text{TV}}^2 \leq \frac{1}{\pi(x)} e^{-2\lambda t},$$

which follows from Lemma 2.8. The reversible chain (3.3) holds with  $2\alpha$  replaced by  $4\alpha$ .

The next result shows that the chi-square distance can also be bounded in terms of  $\alpha$ .

THEOREM 3.7. *Let  $(K, \pi)$  be a finite Markov chain. Assume that  $\pi(x) \leq 1/e$ . Then*

$$\|h_t^x - 1\|_2 \leq e^{1-c} \quad \text{for } t = \frac{1}{2\alpha} \log \log \frac{1}{\pi(x)} + \frac{c}{\lambda}, \quad c > 0.$$

*For reversible chains, the inequality holds for  $t = (4\alpha)^{-1} \log \log(1/\pi(x)) + c/\lambda$ ,  $c > 0$ .*

PROOF. For  $s > 0$ , set  $q(s) = 1 + e^{2\alpha s}$ . The third statement of Theorem 3.5 gives  $\|H_s\|_{2 \rightarrow q(s)} \leq 1$ . By duality, it follows that  $\|H_s^*\|_{q'(s) \rightarrow 2} \leq 1$ , where  $q'(s)$  is the Hölder conjugate of  $q(s)$  defined by  $1/q'(s) + 1/q(s) = 1$ . Consider the function  $\delta_x$  defined by  $\delta_x(x) = 1/\pi(x)$  and  $\delta_x(y) = 0$  for  $x \neq y$  and recall that

$$(H_s^* - E)\delta_x(y) = h_s(x, y) - 1.$$

Then write

$$\begin{aligned} \|h_{t+s}^x - 1\|_2 &= \|(H_{t+s}^* - E)\delta_x\|_2 \leq \|H_s^*\delta_x\|_2 \|H_t^* - E\|_{2 \rightarrow 2} \\ &\leq \|\delta_x\|_{q'(s)} \|H_s^*\|_{q'(s) \rightarrow 2} \|H_t^* - E\|_{2 \rightarrow 2} \\ &= \pi(x)^{-1/q(s)} \|H_s\|_{2 \rightarrow q(s)} \|H_t - E\|_{2 \rightarrow 2} \\ &\leq \pi(x)^{-1/q(s)} e^{-\lambda t}. \end{aligned}$$

Choosing  $s = (1/(2\alpha))\log \log(1/\pi(x))$ , we have  $q(s) = 1 + \log(1/\pi(x))$  and thus

$$\|h_{t+s}^x - 1\|_2 \leq e^{1-\lambda t}.$$

This gives the stated inequality. For reversible chains, we use the second statement of Theorem 3.5 instead.  $\square$

REMARK. Both (3.3) and Theorem 3.7 can be used to bound total variation in terms of  $\alpha$ . However, Theorem 3.7 is more precise and yields a result which is better by a factor of 2. Further, Theorem 3.7 yields a bound in maximal relative error:

$$\sup_{x,y} |h_t(x,y) - 1| \leq e^{2-c} \quad \text{for } t = \frac{1}{2\alpha} \log \log \frac{1}{\pi_*} + \frac{c}{\lambda},$$

where  $\pi_* = \min_x \pi(x)$ . For reversible chains, this follows readily from Theorem 3.7 and  $\sup_{x,y} |h_t(x,y) - 1| = \sup_x \|h_{t/2}^x - 1\|_2^2$ . For nonreversible chains, replace  $2\alpha$  by  $\alpha$  and use

$$\sup_{x,y} |h_t(x,y) - 1| \leq \left[ \sup_x \|h_{t/2}^x - 1\|_2 \right] \left[ \sup_x \|h_{t/2}^{*,x} - 1\|_2 \right]$$

instead and observe that  $K$  and  $K^*$  have the same log-Sobolev constant.

For reversible chains, Theorem 3.7 and Corollary 2.2 yield the following Corollary.

COROLLARY 3.8. *Assume that  $(K, \pi)$  is reversible and  $\pi(x) \leq 1/e$ . Set  $\lambda_* = \min\{\lambda, 1 + \beta_{\min}\}$ . Then*

$$\|k_x^n - 1\|_2 \leq (1 + 2e^2)^{1/2} e^{-c} \quad \text{for } n \geq \frac{1}{4\alpha} \log \log \frac{1}{\pi(x)} + \frac{c}{\lambda_*} + 1, c > 0.$$

Further, setting  $\pi_* = \min_x \pi(x)$ , we have

$$\sup_{x,y} |k^{2n}(x,y) - 1| \leq (1 + 2e^2) e^{-2c}$$

$$\text{for } n \geq \frac{1}{4\alpha} \log \log \frac{1}{\pi_*} + \frac{c}{\lambda_*} + 1, c > 0.$$

EXAMPLE 3.4. We can now apply the results obtained in Example 3.3 to prove the first statement in Theorem 1.1. Namely, Example 3.3 shows that

the Metropolis chain (1.9) on  $\{0, \dots, n\}$  which has stationary measure  $\pi(x) = 2^{-n} \binom{n}{x}$  satisfies

$$\lambda(M) \geq \frac{1}{n}, \quad \alpha(M) \geq \frac{1}{2n}.$$

Further, (1.9) shows that  $M(x, x) \geq 2/(n + 3)$  for all  $x \in \mathcal{X}$  and it easily follows that  $\beta_{\min} \geq -1 + 4/(n + 3)$ . Hence,

$$\left\| \frac{M_x^l}{\pi} - 1 \right\|_2 \leq (1 + 2e^2)^{1/2} e^{-c} \quad \text{for } l \geq \frac{n}{2}(\log n + 2c) + 1, c > 0.$$

For the lower bound stated in Theorem 1.1, see [18].

Further examples are discussed in Section 4.

REMARK. After this paper was submitted for publication Miclo discovered a discrete time version of Theorems 3.7 and 3.8. He kindly authorized us to present some of his results. The discrete time version of the decay of entropy reads as follows: for any finite irreducible chain  $K$  with invariant measure  $\pi$ , let  $\alpha_*$  be the log-Sobolev constant of the chain  $KK^*$  and  $\alpha^*$  that of  $K^*K$ , that is,

$$\alpha_* = \min \left\{ \frac{\langle (I - KK^*)f, f \rangle}{\mathcal{L}(f)}; \mathcal{L}(f) \neq 0 \right\},$$

$$\alpha^* = \min \left\{ \frac{\langle (I - K^*K)f, f \rangle}{\mathcal{L}(f)}; \mathcal{L}(f) \neq 0 \right\}.$$

(These constants can well be zero because  $KK^*$  and  $K^*K$  need not be irreducible.) Miclo [36] proves that, for any probability measure  $\mu$ ,

$$\text{Ent}_\pi(\mu K^n) \leq (1 - \alpha_*)^n \text{Ent}_\pi(\mu).$$

To obtain a statement analogous to Theorem 3.7, we will use Miclo’s discrete version of hypercontractivity which reads

$$\|K^n\|_{2 \rightarrow q} \leq 1 \quad \text{for all } n, q \geq 2 \text{ such that } q = 2[1 + \alpha^*]^n.$$

The proofs of these two results are elementary but subtle. For instance, Miclo’s entropy bound is based on the inequality

$$(t + s)\log(t + s) \geq t \log(t) + (1 + \log(t))s + (\sqrt{t + s} - \sqrt{t})^2,$$

which holds for all  $t, s + t \geq 0$ . See [36].

Miclo’s hypercontractivity result and the line of reasoning used to prove Theorem 3.7 yield the

**THEOREM 3.7’.** *Let  $K$  be a finite irreducible chain with invariant measure  $\pi$  satisfying  $\alpha^* > 0$  and  $\lambda_* = \min\{\lambda, 1 + \beta_{\min}\} > 0$ . Then*

$$\|k_x^n - 1\|_2 \leq e^{1-c}$$

for

$$n \geq \frac{1}{\log(1 + \alpha^*)} \log \log \frac{1}{\sqrt{\pi(x)}} + \frac{c}{-\log(1 - \lambda_*)} + 2.$$

As mentioned above,  $\alpha_*$ ,  $\alpha^*$  can well be zero even when  $K$  is irreducible (in which case  $\alpha > 0$ ). To cope with this difficulty, one can use the log-Sobolev constants of  $K^l K^{l,*}$  and  $K^{l,*} K^l$  for some large enough  $l$ . See [10, 16, 36].

3.4. *Bounding  $\alpha$  from below.* We now give a result which is useful in bounding  $\alpha$  from below. The idea is as follows. For reversible chains, Theorem 3.5 gives a characterization of  $\alpha$  in terms of the function

$$t(q) = \inf\{s > 0: \|H_s\|_{2 \rightarrow q} \leq 1\}, \quad q \in [2, +\infty[.$$

Namely,

$$(3.5) \quad \alpha = \inf_{q > 2} \frac{\log(q - 1)}{4t(q)}.$$

Theorem 3.9 bounds  $\alpha$  in terms of just  $t(q)$  for fixed  $q > 2$  or, more generally, in terms of  $\lambda$  and  $q$ ,  $M_q, t_q$  where  $M_q, t_q$  satisfy  $\|H_{t_q}\|_{2 \rightarrow q} \leq M_q$ . The result can be applied either with a finite  $q$  (e.g.,  $q = 4$ ) or with  $q = \infty$ . In the first case, the bound on  $\alpha$  is potentially of the right order of magnitude. However, precise bounds on  $\|H_t\|_{2 \rightarrow q}$  for a finite  $q$  seem difficult to obtain. The case  $q = \infty$  is appealing because good bounds on  $\|H_t\|_{2 \rightarrow \infty}$  are often available. However, even the best bound on  $\|H_t\|_{2 \rightarrow \infty}$  can produce bounds on  $\alpha$  that are off; see Example 3.5.

**THEOREM 3.9.** *Assume that  $(K, \pi)$  is reversible. Fix  $2 < q \leq +\infty$  and assume that  $t_q, M_q$  satisfy  $\|H_{t_q}\|_{2 \rightarrow q} \leq M_q$ . Then*

$$\alpha \geq \frac{(1 - 2/q)\lambda}{2(\lambda t_q + \log M_q + (q - 2)/q)}.$$

**PROOF.** The proof is based on Stein's interpolation theorem for analytic families of operators: see [46], [7] or [45], page 385. Note that in order to apply Stein's interpolation theorem, we need to work with complex-valued functions. Consider the complex time semigroup

$$H_z = e^{-z(I-K)} = e^{-z} \sum_0^{\infty} \frac{z^n K^n}{n!}.$$

In the present elementary setting, this is clearly a well defined analytic family of operators. Set  $T_z = H_{zt_q}$ . Because  $(K, \pi)$  is reversible, we can use spectral theory to show that, for all real  $a$ ,

$$\|T_{ia}\|_{2 \rightarrow 2} \leq 1 \quad \text{and} \quad \|T_{1+ia}\|_{2 \rightarrow q} \leq M_q.$$



Here we have used the hypothesis  $\|H_{t_q}\|_{2 \rightarrow q} \leq M_q$  to obtain the second inequality. Now, Stein's interpolation yields

$$(3.6) \quad \|T_s\|_{2 \rightarrow p_s} \leq M_q^s \quad \text{for} \quad \frac{1}{p_s} = \frac{s}{q} + \frac{1-s}{2}, \quad 0 \leq s \leq 1.$$

From here, we can again restrict ourselves to real-valued functions. If we express (3.6) in terms of  $H_t = T_s$ , where  $t = st_q$ , and set

$$p(t) = p_s = \frac{2qt_q}{(2-q)t + qt_q},$$

we obtain

$$\|H_t\|_{2 \rightarrow p(t)} \leq \exp\left(\frac{t}{t_q} \log M_q\right).$$

Let  $f$  be a function in  $l^2(\pi)$ . From the last inequality, we deduce that

$$\exp\left(-\frac{t}{t_q} \log M_q\right) \|H_t f\|_{p(t)} \leq \|f\|_2.$$

Since at  $t = 0$  the left-hand side is equal to  $\|f\|_2$ , this shows that the derivative of

$$t \rightarrow U(t) = \exp\left(-\frac{t}{t_q} \log M_q\right) \|H_t f\|_{p(t)}$$

at  $t = 0$  is less than or equal to zero. Using (3.2), we get

$$U'(0) = -\frac{\log M_q}{t_q} \|f\|_2 + \|f\|_2^{-1} \left( \frac{p'(0)}{p(0)^2} \mathcal{L}(f) - \mathcal{E}(f, f) \right) \leq 0.$$

Hence

$$\frac{q-2}{2qt_q} \mathcal{L}(f) \leq \mathcal{E}(f, f) + \frac{1}{t_q} (\log M_q) \|f\|_2^2$$

or

$$\mathcal{L}(f) \leq \frac{2q}{q-2} \left( t_q \mathcal{E}(f, f) + (\log M_q) \|f\|_2^2 \right).$$

This inequality can be found in [24] as well as in [5] and [8]. See also the references given in these papers. The lower bound on  $\alpha$  given in Theorem 3.9 now follows from this inequality and

$$(3.7) \quad \mathcal{L}(f) \leq \mathcal{L}(f - Ef) + 2\|f - Ef\|_2^2.$$

The proof of (3.7) is surprisingly tricky; see [8], page 246, or [5], page 47.  $\square$

Theorem 3.9 has a useful variant.

**THEOREM 3.10.** *Assume that  $(K, \pi)$  is reversible. Fix  $2 < q \leq +\infty$  and assume that  $t_q, M_q$  satisfy  $\|H_{t_q} - E\|_{2 \rightarrow q} \leq M_q$ . Then*

$$\alpha \geq \frac{(1 - 2/q)\lambda}{2(\lambda t_q + \log M_q + (q - 2)/2)}.$$

**PROOF.** The same proof as above yields

$$\mathcal{L}(f - Ef) \leq \frac{2q}{q-2} \left( t_q \mathcal{E}(f, f) + (\log M_q) \|f - Ef\|_2^2 \right).$$

By (3.7), the result follows.  $\square$

**COROLLARY 3.11.** *Assume that  $(K, \pi)$  is reversible. Define*

$$\tau = \inf \left\{ t > 0 : \sup_x \|h_t^x - 1\|_2 \leq 1/e \right\}.$$

Then

$$\frac{1}{2\alpha} \leq \tau \leq \frac{1}{4\alpha} \left( 4 + \log \log \frac{1}{\pi_*} \right).$$

**PROOF.** For the lower bound, use  $q = \infty$ ,  $t_\infty = \tau$ ,  $M_\infty = 1/e$  in Theorem 3.10. For the upper bound, use Theorem 3.7.  $\square$

**EXAMPLE 3.5.** For the chain  $K$  on the hypercube  $\mathcal{X} = \{-1, 1\}^n$  described in Example 3.2, the eigenvalues of  $I - K$  are  $2j/n$  with multiplicity  $\binom{n}{j}$ ,  $0 \leq j \leq n$ . Further,

$$\max_x \|h_t^x - 1\|_2^2 = \|h_t^{\text{id}} - 1\|_2^2 = \sum_1^n \exp(-4jt/n) \binom{n}{j}.$$

Thus,

$$\max_x \|h_t^x - 1\|_2^2 \leq 2 \sum_1^{n/2} \exp(-4jt/n) \frac{n^j}{j!} \leq 2(\exp(n \exp(-4t/n)) - 1).$$

It follows that

$$\|H_t - E\|_{2 \rightarrow \infty} = \max_x \|h_t^x - 1\|_2 \leq 2 \quad \text{for } t = \frac{n}{4} \log n.$$

Using this information in Theorem 3.10 gives

$$\alpha \geq \frac{2}{n(4 + \log n)},$$

which has to be compared to the known value  $\alpha = 1/n$ .

**EXAMPLE 3.6.** Consider the nearest-neighbor chain  $K$  on  $\{0, \dots, n\}$  with lops at the ends. The eigenvalues and eigenfunctions of  $I - K$  are

$$\begin{aligned} \lambda_0 &= 0, & \psi_0(x) &\equiv 1, \\ \lambda_j &= 1 - \cos \frac{\pi_j}{n+1}, & \psi_j(x) &= \sqrt{2} \cos \left( \frac{\pi j(x - 1/2)}{n+1} \right) \quad \text{for } j = 1, \dots, n. \end{aligned}$$

See [22], page 436. Using this information, we show in [10] that

$$\|H_t - E\|_{2 \rightarrow \infty}^2 = \max_x \|h_t^x - 1\|_2^2 \leq 2 \exp(-4t/(n + 1)^2) \left(1 + \sqrt{(n + 1)^2/4t}\right).$$

Thus, for  $t = \frac{1}{2}(n + 1)^2$ ,

$$\|H_t - E\|_{2 \rightarrow \infty} \leq 1.$$

Using this and  $\lambda \geq 2/(n + 1)^2$  in Theorem 3.10 gives

$$(3.8) \quad \frac{1}{2(n + 1)^2} \leq \alpha \leq \frac{1}{2} \left(1 - \cos \frac{\pi}{n + 1}\right) = \frac{\pi^2}{4(n + 1)^2} + O\left(\frac{1}{n^4}\right).$$

To the best of our knowledge, the exact value of  $\alpha$  is not known.

REMARKS. (i) The lower bounds on  $\alpha$  given by Theorems 3.9 and 3.10 are nondecreasing functions of  $\lambda$ . Thus, any lower bound on  $\lambda$  can be used in these estimates.

(ii) Since Lemma 2.8 always gives  $\|H_t - E\|_{2 \rightarrow \infty} \leq (1/\sqrt{\pi_*})e^{-\lambda t}$ , where  $\pi_* = \min_x \pi(x)$ , we can apply Theorem 3.10 with  $q = \infty$  for each  $t > 0$ . This yields

$$(3.9) \quad \alpha \geq \frac{\lambda}{2 + \log(1/\pi_*)}.$$

This bound will be improved to

$$(3.10) \quad \alpha \geq \frac{(1 - 2\pi_*)\lambda}{\log(1/\pi_* - 1)}$$

in the Appendix (see Corollary 5.4). Clearly, these bounds have little value in practice, but they give universal quantitative lower bounds. We will also show that (3.10) is sharp in the case of the complete graph.

(iii) Let  $(K, \pi)$  be a nonreversible chain with stationary measure  $\pi$ . Consider the reversible chain with kernel  $Q = \frac{1}{2}(K + K^*)$  and stationary measure  $\pi$  (this is often called the additive reversibilization of  $K$ ; see [23, 10]). Note that  $K$  is irreducible if and only if  $Q$  is. Observe also that (2.2) says that  $(K, \pi)$  and  $(Q, \pi)$  share the same Dirichlet form (when restricted to  $f = g$ ):

$$\mathcal{E}(f, f) = \langle (I - K)f, f \rangle = \langle (I - Q)f, f \rangle.$$

Consider the semigroups  $H_t = e^{-t(I-K)}$  and  $S_t = e^{-t(I-Q)}$ . The relation between the hypercontractivity of  $H_t$  and  $S_t$  is shown in the following comments.

- (a) By definition,  $K$  and  $Q$  have the same log-Sobolev constant  $\alpha$ .
- (b) If one knows that  $\|S_t\|_{2 \rightarrow q} \leq 1$  for all  $t \geq 0$  and all  $q \geq 2$  such that  $q \leq 1 + e^{4\beta t}$  (for some  $\beta \geq 0$ ), then  $\|H_t\|_{2 \rightarrow q} \leq 1$  for all  $t \geq 0$  and all  $q \geq 2$  such that  $q \leq 1 + e^{2\beta t}$ . For the proof, use all the assertions of Theorem 3.5.
- (c) If instead we know  $\|H_t\|_{2 \rightarrow q} \leq 1$  for all  $t \geq 0$  and all  $q \geq 2$  such that  $q \leq 1 + e^{4\beta t}$  (for some  $\beta \geq 0$ ), then the same statement holds for  $S_t$ .
- (d) If we know that  $\|S_{t_0}\|_{2 \rightarrow 4} \leq 1$  for some fixed  $t_0 > 0$ , Theorem 3.9 says that  $\alpha \geq 1/(4t_0 + 2/\lambda)$ . This shows that, for  $t \geq (\log 3)(2t_0 + 1/\lambda)$ ,  $\|H_t\|_{2 \rightarrow 4} \leq 1$ .

(e) There is no useful converse to the last statement. It might well happen that  $\|H_{t_0}\|_{2 \rightarrow 4} \leq 1$  for some  $t_0$ , but that  $\|S_t\|_{2 \rightarrow 4} \leq 1$  only for  $t$ 's much larger than  $t_0$ .

(f) Using the above comments and (3.9), one can show that any irreducible  $K$  with invariant measure  $\pi$  on a finite set is hypercontractive.

**4. Examples and applications.** This section begins with a study of what is known about the log-Sobolev constant in three well studied examples: the hypercube  $\mathbb{Z}_2^n$ , the circle  $\mathbb{Z}_m$  and random transpositions. Following this we study several classes of examples where good approximations to log-Sobolev constants are available. For graphs with “moderate growth” we show that the log-Sobolev constant is of the same order as the spectral gap. For expander graphs, they are of different orders. Finally, we treat two examples previously announced: the simple exclusion process and the Metropolis algorithm in a box. Here, present techniques give considerable improvement over previously available rates of convergence.

**4.1. The hypercube.** Let  $\mathcal{X} = \mathbb{Z}_2^n$  and, for  $i = 1, \dots, n$ , let  $e_i$  be the element of  $\mathbb{Z}_2^n$  with all coordinates 0 except for the  $i$ th, which is 1. Define a probability  $Q$  by setting  $Q(0) = Q(e_i) = 1/(n+1)$  for  $i = 1, \dots, n$  and  $Q(x) = 0$  otherwise. The associated random walk on  $\mathbb{Z}_2^n$  has kernel  $K(x, y) = Q(x-y)$ ,  $K_l(x, y) = Q^{(l)}(x-y)$ , where  $Q^{(l)}$  denotes the  $l$ th convolution power of  $Q$ . The invariant measure is  $\pi \equiv 1/2^n$ . It is well known that the spectral gap is  $\lambda = 2/(n+1)$ . See [9] for instance. We will show that here the log-Sobolev constant equals half the spectral gap.

To find the log-Sobolev constant  $\alpha$  we proceed as follows. Consider the product chain  $\tilde{K}$  on  $\mathbb{Z}_2^n$  with kernel (the notation is as in Section 2.5):

$$\begin{aligned} \tilde{K}(x, y) &= \frac{1}{n} \sum_{i=1}^n \delta(x_1, y_1) \cdots \delta(x_{i-1}, y_{i-1}) \\ &\quad \times K_i(x_i, y_i) \delta(x_{i+1}, y_{i+1}) \cdots \delta(x_n, y_n), \end{aligned}$$

where  $K_i(0, 1) = K_i(1, 0) = 1$  for each  $i$ . As in Example 3.2, the spectral gap  $\tilde{\lambda}$  and the log-Sobolev constant  $\tilde{\alpha}$  of this chain satisfy  $\tilde{\alpha} = \tilde{\lambda}/2 = 1/n$ . Since our original chain  $K$  on  $\mathbb{Z}_2^n$  can be written  $K = (n+1)^{-1}n\tilde{K} + (n+1)^{-1}I$ , the corresponding Dirichlet forms satisfy  $\mathcal{E} = (n+1)^{-1}n\tilde{\mathcal{E}}$ . This shows that

$$\alpha = \frac{n\tilde{\alpha}}{n+1} = \frac{1}{n+1} = \frac{\lambda}{2}.$$

Using this, the fact that  $\beta_{\min} = -1 + 2/(n+1)$  and Corollary 3.8, we get the bound

$$\|k_x^l - 1\|_2 \leq (1 + 2e^2)^{1/2} e^{-c} \quad \text{for } l = \frac{n+1}{2} \left( \frac{1}{2} \log \log 2^n + c \right) + 1$$

for all  $n \geq 2$ . This is essentially the same as the sharp result given in [9], which reads

$$(4.1) \quad \|k_x^l - 1\|_2 \leq \sqrt{2} e^{-c} \quad \text{for } l = \frac{n + 1}{2} \left( \frac{1}{2} \log n + c \right), c > 0.$$

In particular, the result obtained by using  $\alpha \geq 1/(n + 1)$  shows that, as  $n$  tends to infinity,  $\frac{1}{4}n \log n$  steps are sufficient to reach stationarity. This is known to be sharp (see [9]).

We now present a typical application of the log-Sobolev technique. The above chain  $K$  is the simple random walk on the natural graph structure of  $\mathbb{Z}_2^n$  with a loop at each vertex. The edge set  $\mathcal{A} \subset \mathbb{Z}_2^n \times \mathbb{Z}_2^n$  of this graph is the set of all  $(x, y)$  such that  $x$  and  $y$  differ at most by one coordinate. Consider a new graph structure on  $\mathbb{Z}_2^n$  which is obtained by erasing and adding a few edges according to the following rules.

1. When erasing edges, for each original square,  $x, x + e_i, x - e_i, x + e_j, x - e_j$ , at most one edge is erased, and for each edge left in place, there are at most  $C$  squares containing this edge and an erased edge. For simplicity, no loop is erased.
2. When adding edges, the degree  $\delta'(x)$  of any vertex  $x$  must stay bounded by  $Cn$ .

Let  $\mathcal{A}' \subset \mathbb{Z}_2^n \times \mathbb{Z}_2^n$  be the symmetric edge set of this new graph. The simple random walk on  $(\mathbb{Z}_2^n, \mathcal{A}')$  has kernel  $K'(x, y) = 1/\delta'(x)$  if  $(x, y) \in \mathcal{A}'$  and  $K(x, y) = 0$  otherwise. Its reversible measure is  $\pi'(x) = \delta'(x)/|\mathcal{A}'|$ . Because of the above simple rules, it is an easy matter to compare  $\pi'$  with the uniform distribution  $1/2^n$  and to compare the Dirichlet forms  $\mathcal{E}, \mathcal{E}'$  of the chains  $K, K'$ ; see [10, 12]. Thus, using Lemma 3.3 and the known value of  $\alpha$ , we get

$$\alpha' \geq \frac{1}{C_1 n}$$

for a constant  $C_1 > 0$  depending on  $C$  but not on  $n$ . Then, applying Theorem 3.7, we find that the perturbed chain  $K'$  satisfies

$$\|h_x^l - 1\|_2 \leq e^{-c} \quad \text{for } l = C_1 n (C_2 \log n + C_3 + c), c > 0,$$

for universal constants  $C_2, C_3$ .

Let us illustrate this technique by looking at a simple case of the above example. Assume  $\mathcal{A}'$  is obtained from the edge set  $\mathcal{A}$  of the hypercube with loops by erasing the loops at  $(0, \dots, 0)$  and  $(1, \dots, 1)$  and adding an extra edge joining these two vertices. This gives a regular graph. Let  $K'$  be the kernel of the simple random walk on this graph and let  $\mathcal{E}'$  be the corresponding Dirichlet form. It is obvious that  $K'$  is reversible with respect to the uniform measure  $\pi(x) = 2^{-n}$  and that  $\mathcal{E} \leq \mathcal{E}'$ . It follows that  $\alpha' \geq 1/(n + 1)$ . In this

case one can also show that  $\beta_{\min} \geq -1 + 1/(n + 1)$  for  $n \geq 3$ . Hence Corollary 3.8 yields

$$\|k_x^l - 1\|_2 \leq (1 + 2e^2)^{1/2} e^{-c} \quad \text{for } l = \frac{n + 1}{2} \left( \frac{1}{2} \log \log 2^n + c \right) + 1, c > 0,$$

for  $n \geq 3$ . We do not know any other way to prove this result.

See also [17] for a different example of comparison involving the hypercube and log-Sobolev constants.

4.2. *The finite circle  $\mathbb{Z}_m$ .* Consider the simple random walk on  $\mathbb{Z}_m$  with  $m \geq 4$ . It has kernel  $K(x, x \pm 1) = 1/2$  and uniform stationary measure  $1/m$ . The eigenvalues of  $I - K$  are

$$1 - \cos \frac{2\pi i}{m}, \quad 0 \leq i \leq m - 1.$$

For this example we show that the log-Sobolev constant is of the same order as the spectral gap. We have

$$\|H_t - E\|_{2 \rightarrow \infty}^2 = \sum_1^{m-1} \exp(-2t\lambda_j) \leq 2 \sum_{1 \leq i \leq m/2} \exp(-2t[1 - \cos(2\pi i/m)]).$$

For  $0 \leq x \leq \pi/2$ ,  $1 - \cos x \geq 2x^2/5$ . For  $\pi/2 < x \leq \pi$ ,  $1 - \cos x \geq 1$ . These inequalities yield

$$\|H_t - E\|_{2 \rightarrow \infty}^2 \leq 2 \left( \sum_{1 \leq i \leq m/4} \exp\left(-\frac{16\pi^2 t}{5m^2} i^2\right) + \frac{m + 1}{4} e^{-2t} \right).$$

Now (see [10], Section 4.B, for details),

$$\sum_2^{m/4} \exp\left(-\frac{16\pi^2 t}{5m^2} i^2\right) \leq \int_1^\infty \exp\left(-\frac{16\pi^2 t}{5m^2} u^2\right) du \leq \frac{\sqrt{5} m}{8\sqrt{\pi t}} \exp\left(-\frac{16\pi^2 t}{5m^2}\right).$$

Hence,

$$\|H_t - E\|_{2 \rightarrow \infty}^2 \leq 2 \left( 1 + \frac{\sqrt{5} m}{8\sqrt{\pi t}} \right) \exp\left(-\frac{16\pi^2 t}{5m^2}\right) + \frac{m + 1}{2} e^{-2t}.$$

For  $t_0 = 5m^2/16\pi^2$  and  $m \geq 5$ , we get

$$\begin{aligned} \|H_{t_0} - E\|_{2 \rightarrow \infty}^2 &\leq 2 \left( 1 + \frac{\sqrt{\pi}}{2} \right) e^{-1} + \frac{m + 1}{2} \exp\left(-\frac{5m^2}{8\pi^2}\right) \\ &\leq 2 \left( 2 + \frac{\sqrt{\pi}}{2} \right) e^{-1} \leq e. \end{aligned}$$

Using this in Theorem 3.10 gives

$$\frac{2\pi^2}{m^2} \geq \frac{\lambda}{2} \geq \alpha \geq \frac{8\pi^2}{25m^2} \geq \frac{2\lambda}{25}$$

for  $m \geq 5$  and one can check that this also holds for  $m = 4$ . The exact value of  $\alpha$  is not known for  $m \geq 4$ . For  $m = 3$ ,  $\alpha$  is computed in the Appendix and is equal to  $1/[2 \log 2]$ .

4.3. *Random transpositions.* Let  $\mathcal{X} = S_n$  be the symmetric group. Consider the chain with kernel  $K(\theta, \sigma) = 2/[n(n - 1)]$  if  $\theta^{-1}\sigma$  is a transposition and  $K(\theta, \sigma) = 0$  otherwise. This chain has invariant measure  $\pi \equiv 1/(n!)$ . It is studied in detail in [9]. There, it is proved that the spectral gap is  $\lambda(K) = 2/(n - 1)$ .

Let  $H_t = e^{-t(I-K)}$  be the corresponding semigroup. Using the information and the techniques presented in [9], pages 40–43, one can show that

$$(4.2) \quad \|H_t - E_\pi\|_{2 \rightarrow \infty} = \|h_t^{\text{id}} - 1\|_2 \leq 1$$

for  $t = n \log n$ . By Theorem 3.10, this implies that the log-Sobolev constant of this chain satisfies

$$\frac{1}{3n \log n} \leq \alpha \leq \frac{1}{n - 1}.$$

This result is used in [14] to study random walk on very sparse contingency tables. It would be very interesting to compute  $\alpha$  exactly or to significantly improve upon the bounds stated above.

4.4. *Moderate growth.* Random walk on a path (Example 3.6) and the circle  $\mathbb{Z}_m$  are the simplest examples of chains having moderate growth and satisfying a local Poincaré inequality in the sense of [10], Section 5. We briefly recall these definitions. For simplicity, consider a simple random walk on a finite graph  $(\mathcal{X}, \mathcal{A})$ , where  $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$  is a symmetric set of edges. This is the chain with kernel  $K(x, y) = 1/N(x)$  if  $(x, y) \in \mathcal{A}$  and  $K(x, y) = 0$  otherwise, where  $N(x)$  is the number of  $y$  such that  $(x, y) \in \mathcal{A}$ . This chain has stationary distribution  $\pi(x) = N(x)/|\mathcal{A}|$ . It has Dirichlet form

$$\mathcal{E}(f, f) = \frac{1}{2|\mathcal{A}|} \sum_{(x, y) \in \mathcal{A}} |f(x) - f(y)|^2.$$

Using the graph distance on  $\mathcal{X}$ , we consider the ball  $B(x, r)$  of radius  $r$  around  $x$ . We set  $V(x, r) = \pi(B(x, r))$  and  $f_r(x) = V(x, r)^{-1} \sum_{B(x, r)} f(y)$ . Let  $\gamma$  be the diameter of the graph.

The graph  $(\mathcal{X}, \mathcal{A})$  has  $(A, d)$  moderate growth if and only if

$$V(x, r) \geq \frac{1}{A} \left( \frac{r + 1}{\gamma} \right)^d \quad \text{for all } x \in \mathcal{X} \text{ and all } r \leq \gamma.$$

The graph  $(\mathcal{X}, \mathcal{A})$  satisfies a local Poincaré inequality with constant  $a > 0$  if and only if

$$\|f - f_r\|_2^2 \leq ar^2 \mathcal{E}(f, f) \quad \text{for all } r \leq \gamma \text{ and any function } f.$$

The following theorem asserts that a graph satisfying these two hypotheses has a spectral gap and a log-Sobolev constant of roughly the same order, both comparable to  $(\text{diameter})^{-2}$ .

**THEOREM 4.1.** *Let  $(\mathcal{X}, \mathcal{A})$  be a finite graph having  $(A, d)$  moderate growth and satisfying a local Poincaré inequality with constant  $a > 0$ . Let  $\gamma$  be the diameter of  $(\mathcal{X}, \mathcal{A})$ . There are constants  $c_i > 0$  depending only on  $A, d$  such that the following statements hold.*

(i) *The spectral gap of the simple random walk on  $(\mathcal{X}, \mathcal{A})$  satisfies  $1/a\gamma^2 \leq \lambda \leq c_1/\gamma^2$ .*

(ii) *The log-Sobolev constant  $\alpha$  of the simple random walk on  $(\mathcal{X}, \mathcal{A})$  satisfies  $c_2/a\gamma^2 \leq \alpha \leq c_3/\gamma^2$ . Here, one can take  $c_2^{-1} = 2(2 + \log[(e(1 + d)A)^{1/2}(2 + d)^{d/4}])$ .*

**PROOF.** The upper bound on  $\lambda$  follows readily for the hypothesis that a local Poincaré inequality is satisfied (take  $r = \gamma$ ). The upper bound on  $\alpha$  follows by Lemma 3.1. The lower bound on  $\lambda$  follows from the hypothesis of moderate growth; see [10, 15]. The lower bound on  $\alpha$  is a consequence of Theorem 3.10 and [10], Theorem 5.8.  $\square$

Examples of moderate growth are random walks on nilpotent groups of bounded class with a bounded number of generators, for example, the Heisenberg group mod  $p$ ; see [10, 13, 15, 16].

**4.5. Random graphs and expanders.** For simplicity, consider an  $r$ -regular finite graph  $(\mathcal{X}, \mathcal{A})$ , where  $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$  is a symmetric set of edges. Since each vertex has exactly  $r$  neighbors, the simple random walk on  $(\mathcal{X}, \mathcal{A})$  has kernel  $K(x, y) = 1/r$  if  $(x, y) \in \mathcal{A}$  and  $K(x, y) = 0$  otherwise. The stationary measure  $\pi$  is the uniform distribution  $\pi = 1/|\mathcal{X}|$ .

**LEMMA 4.2.** *The log-Sobolev constant of any finite  $r$ -regular graph  $(\mathcal{X}, \mathcal{A})$  satisfies*

$$\alpha \leq \log r \frac{(4 + \log \log |\mathcal{X}|)}{2 \log [3|\mathcal{X}|/4]}.$$

For the  $d$ -dimensional cube or random transpositions, this bound is worse than the one given by  $\alpha \leq \lambda/2$ . Nevertheless, Lemma 4.2 shows that for any fixed  $r$ , the log-Sobolev constant of an  $r$ -regular graph tends to zero with the size of the vertex set.

**PROOF.** Consider the chain  $\bar{K} = \frac{1}{2}(I + K)$ . On the one hand, by a straightforward counting argument, we have  $\|\bar{k}_n^x - 1\|_2 \geq 1/2$  if  $n$  satisfies  $|\mathcal{X}|^{-1}r^n \leq 3/4$ . On the other hand, applying Theorem 3.7 and Corollary 2.2 to  $\bar{K}$ , which has no negative eigenvalues, we find that

$$\|\bar{k}_n^x - 1\|_2 \leq e^{1-c} \quad \text{for } n \geq \frac{1}{4\bar{\alpha}}(\log \log |\mathcal{X}| + 2c),$$



where  $\bar{\alpha} = \alpha/2$  is the log-Sobolev constant of  $\bar{K}$ . Hence  $(\log[3|\mathcal{X}|/4])/\log r \leq (4 + \log \log |\mathcal{X}|)/2\alpha$ . This yields the desired result.  $\square$

REMARKS. (i) One motivation for Lemma 4.2 comes from the general theory of log-Sobolev inequalities. Recall that the inequality  $\alpha \leq \lambda/2$  (stated here as Lemma 3.1) is valid in full generality for Markov semigroups admitting a reversible probability measure. In the few examples where both  $\lambda$  and  $\alpha$  are explicitly known, they satisfy  $\alpha = \lambda/2$ . This is the case for the symmetric two-point space, the Orstein-Uhlenbeck process on  $\mathbb{R}^n$  or the standard diffusion on the  $n$ -sphere; see [24, 25, 5], for instance. A diffusion where  $\alpha = \lambda/4$  is given in [30]. Not surprisingly, the question of bounding  $\alpha$  from below in terms of  $\lambda$  is often raised in the literature. Lemma 4.2, together with known deep estimates on the spectral gap of a certain family of graphs, yields a host of examples where  $\alpha \ll \lambda$ . Some are described below. From this point of view, the most striking of these results, which might well be the easiest to prove, asserts that a generic  $r$ -regular graph has  $\alpha \ll \lambda$ .

(ii) Lemma 4.2 extends easily to any finite reversible Markov chain  $(K, \pi)$ . Let  $\hat{r} = \sup_x \#\{y \neq x: K(x, y) > 0\}$  and  $\hat{\pi} = \max_x \{\pi(x)\}$ . Then

$$\alpha(K) \leq \frac{(\log \hat{r})(4 + \log \log [1/\hat{\pi}])}{\log(3/4\hat{\pi})}.$$

Bounds in terms of the diameter can also be derived.

EXAMPLE 4.1 (Quotients of a group with Kazhdan’s property  $T$ ). Let  $G$  be an infinite, finitely generated group with a finite symmetric set  $S$  of generators. Consider the finite groups  $\mathcal{X} = G/\Gamma$  that are quotient of  $G$  by a normal subgroup  $\Gamma$ . The set  $S$  can also be considered as a generating set in  $\mathcal{X}$  and thus determines a symmetric ergodic walk on  $\mathcal{X}$ . Margulis [34] and Alon and Milman [3] showed that if  $G$  has Kazhdan’s property  $T$ , then there exists a constant  $\tau = \tau(G) > 0$  such that  $\lambda = \lambda(\mathcal{X}, S)$  satisfies  $\lambda \geq \tau$ . By Lemma 4.2,  $\alpha \ll \lambda$  for these graphs. The group  $SL_3(\mathbb{Z})$  is an example of a group with property  $T$ . For  $S$  one can take

$$\left( \begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{matrix} \right)^{\pm 1}, \left( \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \right)^{\pm 1}.$$

Thus the Cayley graphs  $(SL_3(\mathbb{Z}_m), S)$ ,  $m = 2, 3, \dots$ , have

$$\inf_m \lambda(SL_3(\mathbb{Z}_m), S) > 0 \quad \text{and} \quad \alpha(SL_3(\mathbb{Z}_m), S) \leq \frac{c \log \log m}{\log m}.$$

Inequality (3.9) gives a lower bound of order  $1/\log m$  in this case.

This generalizes to  $SL_n(\mathbb{Z})$  for any fixed  $n \geq 3$ . The group  $SL_2(\mathbb{Z})$  does not have property  $T$ . For all of this, see [33].

EXAMPLE 4.2 [ $SL_2(\mathbb{Z}_p)$ ]. Let  $\mathcal{X} = SL_2(\mathbb{Z}_p)$ , where  $p$  is a prime, with generating set

$$S = \left\{ \left( \begin{matrix} 1 & 1 \\ 0 & 1 \end{matrix} \right)^{\pm 1}, \left( \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} \right)^{\pm 1} \right\}.$$

It can be shown (but this is rather difficult) that there exists  $\varepsilon > 0$  such that  $\lambda(\mathrm{SL}_2(\mathbb{Z}_p), S) \geq \varepsilon$ ; see [33]. Lemma 4.2 yields  $\alpha(\mathrm{SL}_2(\mathbb{Z}_p), S) \leq c(\log \log p)/\log p$ . Thus  $\alpha \ll \lambda$ . Here, inequality (3.9) gives a lower bound of order  $1/\log p$ .

It is worth mentioning the following simple example which is obtained as a quotient of these Cayley graphs: set  $\mathcal{Y} = \{0, 1, \dots, p-1, \infty\}$ , where  $p$  is a prime, and connect  $x \in \mathcal{Y}$  to  $x+1$ ,  $x-1$  and  $-1/x$ . This is a family of cubic graphs with  $\lambda \geq \varepsilon$  uniformly in  $p$ ; see [33], where an explicit value of  $\varepsilon$  is also given. Again, we have

$$\alpha \leq \frac{(\log 3)(4 + \log \log(p+1))}{2 \log(3(p+1)/4)}.$$

**EXAMPLE 4.3 (Random regular graphs).** Fix an integer  $r \geq 3$ . A model for random  $r$ -regular graphs on  $n$  vertices was introduced by Bollobas [6]; see also [33, 42]. Theorem 4.2 in [2] states that, for this model and as  $n$  tends to infinity, a random  $r$ -regular graph  $\mathcal{G}$  on  $n$  vertices satisfies  $\lambda(\mathcal{G}) \geq \varepsilon(r) > 0$  with probability  $1 - o(1)$  as  $n$  tends to infinity. Lemma 4.2 shows that  $\alpha(\mathcal{G}) \rightarrow 0$  as  $n$  tends to infinity. Thus, for fixed  $r$ , a generic  $r$ -regular graph has  $\alpha \ll \lambda$ .

**4.6. Exclusion process.** Simple exclusion is a well studied process; see Liggett [31] for an overview. Here, we consider the simple case where  $r$  particles are hopping around on the vertices of a finite graph. We refer the reader to [23, 12] for motivation and more details. In particular, we will mainly keep the notation introduced in [12].

Thus, let  $\mathcal{G}_0 = (\mathcal{X}_0, \mathcal{A}_0)$  be a finite graph with vertex set  $\mathcal{X}_0$  having cardinality  $|\mathcal{X}_0| = n$  and symmetric edge set  $\mathcal{A}_0 \subset \mathcal{X}_0 \times \mathcal{X}_0$ . Let  $d(x) = \#\{y \in \mathcal{X}_0 : (x, y) \in \mathcal{A}_0\}$  be the edge degree of  $x$  in  $\mathcal{G}_0$  and let  $d_0 = \max\{d(x) : x \in \mathcal{X}_0\}$ .

For any fixed  $r \leq n$ , the exclusion process is defined as a Markov chain with state space the  $r$  sets of  $\mathcal{X}_0$ . Informally, if the current state is the  $r$  set  $A$ , pick an element in  $A$  with probability proportional to its degree, pick a neighboring site at random and move the element to the neighboring site provided this site is unoccupied. If the site is occupied, the chain stays at  $A$ . Formally, let  $X = X_r$  be the set of the  $r$  sets of  $\mathcal{X}_0$  and  $A_1, A_2$  be  $r$  sets. Define  $K(A_1, A_2)$  as follows:

$$(4.3) \quad \begin{aligned} &\text{If } |A_1 \cap A_2| \leq r-2, K(A_1, A_2) = 0. \\ &\text{If } |A_1 \cap A_2| = r-1 \text{ and } A_1 = A \cup \{a_1\}, A_2 = A \cup \{a_2\} \\ &\text{with } (a_1, a_2) \notin \mathcal{A}_0, K(A_1, A_2) = 0. \\ &\text{If } |A_1 \cap A_2| = r-1 \text{ and } A_1 = A \cup \{a_1\}, A_2 = A \cup \{a_2\} \\ &\text{with } (a_1, a_2) \in \mathcal{A}_0, K(A_1, A_2) = 1/\sum_{a \in A_1} d(a). \\ &\text{If } A_1 = A_2, K(A_1, A_1) = \sum_{a \in A_1} d_a^*(A_1)/\sum_{a \in A_1} d(a), \text{ where} \\ &d_a^*(A_1) = \#\{b \in A_1 : (a, b) \in \mathcal{A}_0\}. \end{aligned}$$

This is a reversible chain with stationary distribution

$$\pi(A) = \frac{n \sum_{a \in A} d(a)}{r \binom{n}{r} |\mathcal{A}_0|}.$$

Note that if  $\mathcal{E}_0$  is  $d_0$ -regular [i.e.,  $d(x) \equiv d_0$ ], then  $K$  is symmetric and  $\pi$  is uniform on  $r$  sets.

In [12, 37] the exclusion process on a given graph is studied by comparison with a similar but simpler process  $\tilde{K}$  known as the Bernoulli–Laplace model of diffusion. This is also a process on the  $r$  sets of  $\mathcal{E}_0$ : if its current state is the  $r$  set  $A$ , pick an element in  $A$  at random, pick an element in  $\mathcal{E}_0 \setminus A$  at random, and switch the two elements. Formally, for two  $r$  sets  $A_1, A_2$ , define  $\tilde{K}$  as follows:

$$(4.4) \quad \begin{aligned} &\text{If } |A_1 \cap A_2| \leq r - 2 \quad \text{or} \quad A_1 = A_2, \quad \tilde{K}(A_1, A_2) = 0. \\ &\text{If } |A_1 \cap A_2| = r - 1, \quad \tilde{K}(A_1, A_2) = 1/r(n - r). \end{aligned}$$

This chain is reversible with uniform stationary distribution  $\tilde{\pi} \equiv 1/\binom{n}{r}$ . The following result is proved in [12], Section 3.

**THEOREM 4.3.** *The Dirichlet forms  $\mathcal{E}, \tilde{\mathcal{E}}$  of the chains  $K, \tilde{K}$  defined in (4.3) and (4.4) satisfy*

$$\tilde{\mathcal{E}} \leq \frac{|\mathcal{A}_0| \Delta_0}{n(n - r)} \mathcal{E}$$

with

$$\Delta_0 = \max_{e_0 \in \mathcal{E}_0} \left\{ \sum_{\gamma_{x,y} \ni e_0} |\gamma_{x,y}| \right\},$$

where, for each  $(x, y) \in \mathcal{E}_0 \times \mathcal{E}_0$ , a path  $\gamma_{x,y}$  of length  $|\gamma_{x,y}|$  has been chosen in  $\mathcal{E}_0$  and the sum is over all  $(x, y)$  such that the edge  $e_0$  is an edge used in  $\gamma_{x,y}$ .

One of the main features of the above result is that  $\Delta_0$  is a quantity that depends only on the underlying graph  $\mathcal{E}_0$ . Now, a lot is known about the Bernoulli–Laplace chain  $\tilde{K}$ . It has been studied by Diaconis and Shahshahani [20], where they show that  $\tilde{\lambda} = n/r(n - r)$ . Further, using eigenvalues and eigenfunctions as in [20], one can show that

$$\|\tilde{H}_t - E_{\tilde{\pi}}\|_{2 \rightarrow \infty} = \sup_{x \in \mathcal{X}} \|\tilde{h}_t^x - 1\|_2 \leq e \quad \text{for } t = \frac{r(n - r)}{n} \log n.$$

By Theorem 3.10, this yields the following lemma.

**LEMMA 4.4.** *The log-Sobolev constant  $\tilde{\alpha}$  of the Bernoulli–Laplace chain (4.4) satisfies*

$$\frac{n}{2r(n - r)} \geq \tilde{\alpha} > \frac{n}{3r(n - r) \log n}.$$

Lemmas 3.3 and 4.4, Theorem 4.3 and a direct comparison of  $\pi, \tilde{\pi}$  yield the following theorem.

**THEOREM 4.5.** *The chain  $K$  of the exclusion process (4.3) satisfies*

$$\lambda \geq \frac{n}{r d_r \Delta_0} \quad \text{and} \quad \alpha \geq \frac{n}{3r d_r \Delta_0 \log n},$$

where

$$d_r = \max_{A \in X_r} \left\{ \frac{1}{r} \sum_{a \in A} d(a) \right\} \leq d_0.$$

The estimate of  $\lambda$  is one of the main results of [12] and the bound on  $\alpha$  was announced there. Theorem 3.7 can now be used to improve upon the convergence results stated in Section 6 of [12].

**THEOREM 4.6.** *The exclusion process at (4.3) satisfies*

$$\|h_t^x - 1\|_2 \leq e^{1-c} \quad \text{for } t \geq \frac{r d_r \Delta_0}{n} \left( \frac{3}{4} (\log n) \left( \log \log \left( \frac{n d_r}{|\mathcal{S}_0|} \binom{n}{r} \right) \right) + c \right).$$

In many cases, [12] also provides estimates on  $\beta_{\min}$  and one can apply Corollary 3.8 to get convergence results for the corresponding discrete time exclusion process. Examples follow.

**EXAMPLE 4.4.** As a first example, consider the process of  $\lfloor n/2 \rfloor$  particles around the finite circle  $\mathbb{Z}_n$  with its standard graph structure. This is Example 1 in Section 5 of [12], which gives  $\Delta_0 \leq n(n+2)^2/24$ ,  $d_0 = 2$ . Hence, Theorem 4.6 gives

$$\|h_t^x - 1\|_2 \leq e^{1-c} \quad \text{for } t \geq \frac{n(n+2)^2}{24} ((\log n)^2 + c).$$

By Corollary 3.8 and the results of [12], Section 5, a similar estimate holds as well for the discrete time chain. Thus, using the log-Sobolev constant, we can assert approximate randomness for this process after order  $n^3(\log n)^2$  steps. Using only the spectral gap as in [12] gives order  $n^4$ .

**REMARKS.** (i) In the important case where  $\mathcal{S}_0$  is a path with  $n$  points or a square grid or a finite square box in  $d$  dimensions of side length  $n$ , Lu and Yau [32, 50] have been able to show that the log-Sobolev constant of the exclusion process of  $r$  particles is bounded by  $\alpha \geq c/\lfloor rn^2 \rfloor$  for a constant  $c$  independent of  $r$  and  $n$ . That is,  $\alpha \approx \lambda$  in this case. Their proof is much more involved than the argument used above. What their argument gives for more complicated underlying graph structures  $\mathcal{S}_0$  is not clear to us. It would be interesting to decide what is the exact order of the log-Sobolev constant  $\tilde{\alpha}$  of the Bernoulli-Laplace model considered in Lemma 4.4. If we knew that  $\tilde{\alpha}$  is

in fact of order  $n/[r(n - r)]$ , the comparison argument used above would give an alternative proof of Lu and Yau’s result and extend it to other underlying graph structure.

(ii) There is an interesting nonreversible variant of the above process where a continuous time exclusion process of  $r$  particles on  $\mathbb{Z}_n$  is constructed from the deterministic walk that moves to the nearest right neighbor at each step; see [23] for details. This nonreversible process has the same (symmetrized) Dirichlet form as the above reversible process, and our analysis applies as well to this nonreversible case.

EXAMPLE 4.5. Consider now the exclusion process of  $d$  particles on the standard graph of the  $d$ -dimensional cube  $\mathbb{Z}_2^d$ . For this case, one has  $\Delta_0 = (d + 1)2^{d-2}$ ,  $d_0 = d$ , and Theorem 4.6 gives

$$\|h_t^x - 1\|_2 \leq e^{1-c} \quad \text{for } t \geq \frac{d^2(d + 1)}{4}(2d \log d + c).$$

In other words, we find here that approximate equilibrium is reached for  $t$  of order  $d^4 \log d$ , whereas Theorem 1 in Section 5 of [12] asserts approximate randomness for  $t$  of order  $d^6$ .

EXAMPLE 4.6. Consider the problem of picking (say)  $d^2$  permutations at random, without replacement, in the symmetric group  $S_d$ . A possible way of doing that is to run the exclusion process (4.3) on the Cayley graph of  $S_d$  with the transpositions as set of generators. This is Example 4 in Section 5 in [12] and we have  $\Delta_0 \leq d!$ ,  $d_0 = d(d - 1)/2$ . We find that

$$\|h_t^x - 1\|_2 \leq e^{1-c} \quad \text{for } t \geq \frac{d^4}{2}(3d(\log d)^2 + c).$$

Thus,  $t$  of order  $d^5(\log d)^2$  is enough for approximate randomness of this process. Using only the spectral gap as in [12], one would ask for  $t$  of order  $d^7 \log d$ .

4.7. *The Metropolis algorithm in a box.* Let  $\mathcal{X} = C(n, d)$  be a discrete box of side length  $n$  in  $d$ -dimensions. The extreme points of  $C(n, d)$  are the  $2^d$  vectors with coordinates 0 or  $n$ .

The usual nearest-neighbor walk in  $C(n, d)$  has stationary distribution proportional to the degree  $\delta(x)$  of the vertex  $x \in C(n, d)$ . This varies between  $d$  and  $2d$  and so is not uniform. The Metropolis algorithm is a widely used method for changing the transition probabilities to have a given stationary distribution. Here, this is a Markov chain on the points in  $C(n, d)$  with transitions  $P(x, y) = 0$  unless  $x = y$  or  $x$  and  $y$  differ by  $\pm 1$  in a single

coordinate, in which case  $P(x, y)$  is given by

$$(4.5) \quad P(x, y) = \begin{cases} \frac{1}{d(x)}, & \text{if } \delta(x) \geq \delta(y) \text{ and } x \neq y, \\ \frac{1}{\delta(y)}, & \text{if } \delta(x) < \delta(y), \\ \frac{1}{\delta(x)} \sum_{\substack{z: \delta(x) < \delta(z) \\ z \sim x}} \left( \frac{1 - \delta(x)}{\delta(z)} \right), & \text{if } x = y. \end{cases}$$

The chain  $P$  is a reversible, aperiodic, irreducible Markov chain on  $C(n, d)$  with uniform stationary distribution  $\pi(x) = 1/(n + 1)^d$ . As far as we can say, the eigenvalues and eigenvectors of this chain are not explicitly known. Our aim is to prove the following sharp result.

**THEOREM 4.7.** *The semigroup  $H_t = e^{-t(I-P)}$  associated with the Metropolis chain  $P$  on  $C_{n,d}$  satisfies*

$$\|h_t^x - 1\|_2 \leq e^{1-c} \quad \text{for } t = d(n + 1)^2(\log(d) + 1/8 + c/2),$$

with  $c > 0$ . Similarly,

$$\|p_x^l - 1\|_2 \leq (1 + e^2)^{1/2} e^{-c} \quad \text{for } l \geq d(n + 1)^2(\log(d) + 1 + c).$$

Let us first describe the ideas of the proof. There are several variants of the chain  $P$  that are product chains in the sense of Section 2.4. For these variants, statements similar to Theorem 4.7 follow directly from Theorems 2.9 and 2.10. Now, the reversible measure and Dirichlet form of  $P$  are easily comparable to the reversible measure and Dirichlet form of any of these variants. In [10], we used this line of reasoning and Nash inequalities to show that the chain  $P$  is close to equilibrium for  $l$  of order  $d^2 n^2 \log d$ . Here, we are going to use the same idea, the results of [10], and the machinery of log-Sobolev inequality to improve upon the results of [10] and show that  $l$  of order  $dn^2 \log d$  suffices. Easy arguments show that this is best possible.

We have chosen to present this analysis for the Metropolis chain (4.5), but it is worth noting that the arguments developed below clearly work for many other natural chains on  $C_{n,d}$ .

**PROOF.** Consider the Markov kernel  $K$  on  $\{0, 1, \dots, n\}$  defined by

$$\begin{aligned} K(x, x + 1) &= 1/2 \quad \text{for } x \in \{0, \dots, n - 1\}, \\ K(x, x - 1) &= 1/2 \quad \text{for } x \in \{1, \dots, n\}, \\ K(0, 0) &= K(n, n) = 1/2. \end{aligned}$$

This is a symmetric kernel with uniform stationary distribution. It has known eigenvalues and eigenfunctions and we proved in Example 3.6 that its log-Sobolev constant on  $\alpha(K)$  is bounded by  $\alpha(K) \geq 1/[2(n + 1)^2]$ .

Next, consider the kernel  $\tilde{P}$  on  $C(n, d)$  which proceeds by choosing one of the  $d$  coordinates at random and changing that coordinate using  $K$  above. Thus,

$$\tilde{P} = \frac{1}{d} \sum_{i=1}^d I \otimes \cdots \otimes I \otimes \underbrace{K \otimes I \otimes \cdots \otimes I}_{i-1}.$$

It follows from Lemma 3.2 that this chain has spectral gap

$$(4.6) \quad \tilde{\lambda} = \frac{1}{d} \left( 1 - \cos \frac{\pi}{n+1} \right) \geq \frac{2}{d(n+1)^2}$$

and log-Sobolev constant bounded by

$$(4.7) \quad \tilde{\alpha} \geq \frac{1}{2d(n+1)^2}.$$

Further, there is an obvious comparison between the chains  $P$  and  $\tilde{P}$ . They have the same stationary distribution  $\pi(x) = 1/(n+1)^d$  and satisfy  $\tilde{P}(x, y) \leq P(x, y)$  for  $x \neq y$ . Hence, their Dirichlet forms satisfy

$$\tilde{\mathcal{E}} \leq \mathcal{E}.$$

This, (4.6), (4.7) and Lemma 3.3 imply that the spectral gap and log-Sobolev constant of  $P$  are bounded by

$$(4.8) \quad \lambda \geq \frac{2}{d(n+1)^2}, \quad \alpha \geq \frac{1}{2d(n+1)^2}.$$

Further, it is proved in [10] that the least eigenvalue  $\beta_{\min}$  of  $P$  satisfies

$$(4.9) \quad \beta_{\min} \geq -1 + \frac{1}{dn^2}$$

and that the chain  $P$  satisfies the Nash inequality

$$\|f\|_2^{2+4/d} \leq 64d(n+1)^2 \left( \mathcal{E}(f, f) + \frac{8}{d(n+1)^2} \|f\|_2^2 \right) \|f\|_1^{4/d}.$$

For  $H_t = e^{-t(I-P)}$ , this Nash inequality implies

$$(4.10) \quad \sup_x \|h_t^x\|_2 = \|H_t\|_{2 \rightarrow \infty} \leq \left( \frac{16d^2(n+1)^2}{t} \right)^{d/4} \quad \text{for } 0 \leq t \leq \frac{d(n+1)^2}{8}.$$

See [10], Theorem 3.5, and the references given therein.

To prove the first statement in Theorem 4.7, we follow the line of reasoning used in the proof of Theorem 3.7. First, observe that  $(H_t - E)h_u^x(y) = (H_t^* - E)h_u^x(y) = h_{t+u}^x(y) - 1$  because  $P$  is reversible. For  $s > 0$ , define  $q(s) = 1 + e^{4as}$ ,  $1/q(s) + 1/q'(s) = 1$  and note that Theorem 3.5 and duality

imply  $\|H_s\|_{q'(s) \rightarrow 2} \leq 1$ . Second, write  $t = c + s + u$  and

$$\begin{aligned} \|h_t^x - 1\|_2 &= \|(H_{c+s} - E)h_u^x\|_2 \leq \|H_s h_u^x\|_2 \|H_c - E\|_{2 \rightarrow 2} \\ &\leq \|h_u^x\|_{q'(s)} \|H_2\|_{q'(s) \rightarrow 2} e^{-\lambda c} = \|h_u^x\|_{q'(s)} e^{-\lambda c} \\ &\leq \|h_u^x\|_2^{2/q(s)} e^{-\lambda c}, \end{aligned}$$

where the last inequality follows from the Hölder inequality  $\|f\|_{q'} \leq \|f\|_1^{1-2/q} \|f\|_2^{2/q}$  which holds for all  $f$  and all  $q \geq 2$ ,  $1/q + 1/q' = 1$ . Now, we pick  $u = d(n + 1)^2/8$  so that (4.10) implies  $\|h_u^x\|_2 \leq e(2^7 d)^{d/4}$ . This yields

$$\|h_t^x - 1\|_2 \leq \left[ e(2^7 d)^{d/4} \right]^{2/q(s)} e^{-\lambda c}.$$

Next, we pick  $s$  so that  $[e(2^7 d)^{d/4}]^{2/q(s)} \leq e$ . Since  $q(s) = 1 + e^{4\alpha s}$ ,  $2^7 \leq e^5$ , it is enough to choose

$$s \geq \frac{1}{4\alpha} \log \left( 1 + \frac{5d}{2} + \frac{d \log d}{2} \right).$$

for instance,  $s = (1/(2\alpha)) \log d$  works. For this choice of  $s$  and  $u$ , we have

$$\|H_t - E\|_{2 \rightarrow \infty} \leq e^{1-\lambda c},$$

where

$$t = \frac{d(n + 1)^2}{8} + \frac{1}{2\alpha} \log(d) + c.$$

Using the bounds (4.8), this implies

$$\|H_t - E\|_{2 \rightarrow \infty} \leq e^{1-c} \quad \text{for } t = d(n + 1)^2(\log(d) + 1/8 + c/2).$$

This is the first assertion of Theorem 4.7. The second assertion then follows from Corollary 2.2 and (4.9).

### APPENDIX

This appendix gives the exact value of the log-Sobolev constant of the chain  $K(x, y) = \pi(y)$  on a finite set  $\mathcal{X}$ , where  $\pi$  is a given probability distribution which satisfies  $\pi(y) > 0$ . It also determines the log-Sobolev constant for all chains on a two-point space.

**THEOREM A.1.** *Let  $\pi$  be a probability measure on the finite set  $\mathcal{X}$ . Assume that  $\pi$  is positive and set  $\pi_* = \min_{\mathcal{X}} \pi$ . Consider the chain  $K(x, y) = \pi(y)$  which has invariant measure  $\pi$ . The log-Sobolev constant  $\alpha$  of this chain is given by*

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



In particular, for  $\pi \equiv 1/|\mathcal{X}|$ ,

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

We will first prove the special case where  $\mathcal{X} = \{0, 1\}$  is the two-point space. Then we will show that the above result follows from this special case.

**THEOREM A.2.** *Consider the chain on  $\{0, 1\}$  with matrix*

$$\begin{pmatrix} \theta & 1 - \theta \\ \theta & 1 - \theta \end{pmatrix} \text{ where } 0 < \theta \leq 1/2.$$

*This has invariant measure  $\pi(0) = \theta$ ,  $\pi(1) = 1 - \theta$ . The log-Sobolev constant of this chain is*

$$\alpha(\theta) = \frac{1 - 2\theta}{\log[(1 - \theta)/\theta]}.$$

*At  $\theta = 1/2$ , the function  $\theta \rightarrow (1 - 2\theta)/(\log[(1 - \theta)/\theta])$  must be replaced by its limit value, which is equal to  $1/2$ .*

Yoshida informed us that he and Higuchi [26] have independently discovered the value of the log-Sobolev constant of asymmetric two-point spaces.

**PROOF OF THEOREM A.2.** The proof is a tedious calculus exercise involving good guessing supported by numerical computations. The first step consists of picking nice coordinates. In order to do this, we follow the choice that leads to the easiest computation in the known symmetric case  $\theta = 1/2$ . Call  $x, y$  the values of a given function on  $\mathcal{X} = \{0, 1\}$  and set  $s = x - y$ , which will be our main variable. We can assume that  $x, y \geq 0$ . We want to compare

$$\begin{aligned} \mathcal{L} &= \theta x^2 \log x^2 + (1 - \theta)y^2 \log y^2 \\ &\quad - (\theta x^2 + (1 - \theta)y^2)\log(\theta x^2 + (1 - \theta)y^2) \end{aligned}$$

and

$$\mathcal{E} = \theta(1 - \theta)(x - y)^2.$$

By homogeneity, we can impose the condition

$$\theta x + (1 - \theta)y = 1,$$

which amounts to saying that our function has mean 1 under  $\pi$ . With this normalization, we can compute  $x$  and  $y$  as functions of  $s = x - y$ :

$$\begin{aligned} x &= 1 + (1 - \theta)s, \\ y &= 1 - \theta s. \end{aligned}$$

The parameter  $s$  varies between  $-(1 - \theta)^{-1}$  and  $\theta^{-1}$ . Also, we have  $\theta x^2 + (1 - \theta)y^2 = 1 + \theta(1 - \theta)s^2$ . Thus, our problem is to compare

$$\begin{aligned} l(s) &= \theta(1 + (1 - \theta)s)^2 \log(1 + (1 - \theta)s)^2 \\ &\quad + (1 - \theta)(1 - \theta s)^2 \log(1 - \theta s)^2 \\ &\quad - (1 + \theta(1 - \theta)s^2)\log(1 + \theta(1 - \theta)s^2) \end{aligned}$$

with

$$e(s) = \theta(1 - \theta)s^2$$

on  $[-(1 - \theta)^{-1}, \theta^{-1}]$ . Computing derivatives gives

$$e'(s) = 2\theta(1 - \theta)s, \quad e''(s) = 2\theta(1 - \theta),$$

and

$$l'(s) = 2\theta(1 - \theta)\left((1 + (1 - \theta)s)\log(1 + (1 - \theta)s)^2 - (1 - \theta s)\log(1 - \theta s)^2 - s\log(1 + \theta(1 - \theta)s^2)\right),$$

$$l''(s) = 4\theta(1 - \theta)\left(1 + (1 - \theta)\log(1 + (1 - \theta)s) + \theta\log(1 - \theta s) - \frac{1}{2}\log(1 + \theta(1 - \theta)s^2) - \frac{\theta(1 - \theta)s^2}{1 + \theta(1 - \theta)s^2}\right),$$

$$l'''(s) = \frac{4\theta(1 - \theta)b(s)}{(1 + \theta(1 - \theta)s^2)^2(1 + (1 - \theta)s)(1 - \theta s)},$$

where  $b(s) = (1 - 2\theta)(1 + 3\theta(1 - \theta)s^2) - 4\theta^2(1 - \theta)^2s^3$ . Finally, we compute

$$b'(s) = 6\theta(1 - \theta)s(1 - 2\theta - 2\theta(1 - \theta)s).$$

We want to find the smallest positive real  $a$  such that

$$l(s) \leq ae(s)$$

on  $[-(1 - \theta)^{-1}, \theta^{-1}]$ . Let us start with two simple observations.

(i) Since  $l(0), l'(0), e(0), e'(0)$  are equal to zero and  $l''(0) = 4\theta(1 - \theta)$ ,  $e''(0) = 2\theta(1 - \theta)$  we have  $a \geq 2$ .

(ii) Since  $l(1/\theta) - ae(1/\theta) = \theta^{-1}(\log(1/\theta) - a(1 - \theta))$ , we have  $a \geq 1/(1 - \theta)\log(1/\theta)$ .

Figure 2 shows the variation of  $l'' - ae''$ ,  $l' - ae'$  and  $l - ae$ . For  $a \geq 2$ ,  $l''(0) - ae''(0) = u_0 \leq 0$ . Further, we can assume that  $a$  is taken small enough so that  $u_1 > 0$ .

If  $u_3$  is negative it easily follows from the table in Figure 2 (center) and the fact that  $l(0) - ae(0) = 0$  that  $l - ae \leq 0$ . We can thus assume that  $a$  is small enough so that  $u_3$  is positive. Then  $l - ae$  varies as shown in the lower panel of Figure 2.

It follows from the study above that the equation  $l(s) - ae(s) = 0$  has zero, one or two nonzero solutions in  $[(1 - \theta)^{-1}, \theta^{-1}]$ , depending on the value of  $a$ . If  $a$  is too large, there are no nonzero solutions. When  $a$  is too small, either there is one nonzero solution  $s_*$  with  $l'(s_*) - ae'(s_*) > 0$  and  $l(s) - ae(s) > 0$  for  $s_* < s \leq \theta^{-1}$ , or there are two solution  $s_- < s_+$  with  $l'(s_-) - ae'(s_-) >$

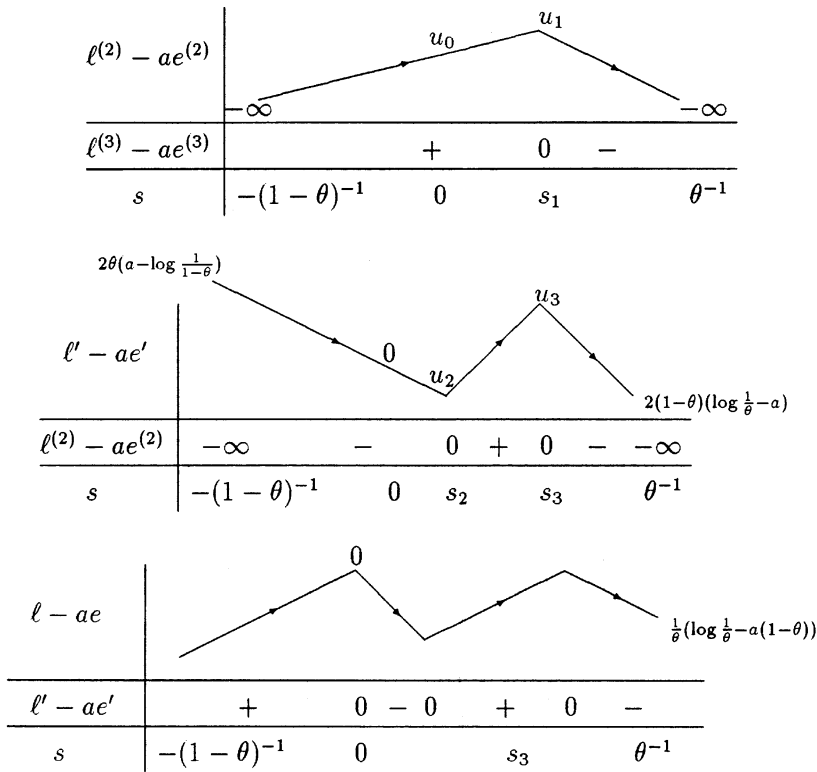


FIG. 2. Variation of  $l - ae$  and its derivatives.

$0, l'(s_+) - ae'(s_+) < 0$  and  $l(s) - ae(s) > 0$  in  $(s_-, s_+)$ . This shows that we are looking for the unique value of  $a$  for which the system of equations

$$(A.1) \quad \begin{aligned} l(s) - ae(s) &= 0, \\ l'(s) - ae'(s) &= 0 \end{aligned}$$

has a nonzero solution in  $[(1-\theta)^{-1}, \theta^{-1}]$ . Unfortunately, it is not clear how to compute this value explicitly from the system above.

Now, it is easy to plot  $s \rightarrow l(s) - ae(s)$  for a fixed value of  $\theta$  and different values of  $a$ . Experiments and good guesses led us to believe that (A.1) admits a solution if and only if

$$a = \frac{\log[(1-\theta)/\theta]}{1-2\theta}$$

and that the solution of (A.1) for that value of  $a$  was

$$s = \frac{1-2\theta}{2\theta(1-\theta)}.$$

Luckily enough, plugging these values in (A.1) shows that, indeed, they satisfy (A.1). This proves Theorem A.2.  $\square$

Above, we have chosen a specific two-state chain, but the result of Theorem 5.2 determines the log-Sobolev constant for all kernels on a two-point space.

**COROLLARY A.3.** *Let  $\mathcal{X} = \{0, 1\}$ . Let  $K$  be a Markov kernel on  $\mathcal{X}$  with stationary distribution  $\pi(x)$ . Assume that  $\pi(0) \leq \pi(1)$ . Then the log-Sobolev constant  $\alpha = \alpha(K)$  equals*

$$\frac{K(0, 1)[1 - 2\pi(0)]}{\pi(1)\log[\pi(1)/\pi(0)]}.$$

If  $\pi(1) \leq \pi(0)$ , reverse the roles of 0 and 1.

**PROOF.** The Dirichlet form  $\mathcal{E}_K$  equals  $K(0, 1)\pi(0)(f(0) - f(1))^2$ , whereas the Dirichlet form used in Theorem A.2 is  $\pi(0)\pi(1)(f(0) - f(1))^2$ . This proves the corollary.  $\square$

**REMARK.** Observe that  $K$  has spectral gap  $\lambda = K(0, 1) + K(1, 0)$ . It follows that the ratio  $\alpha(K)/\lambda(K) = (1 - 2\pi(0))/(\log[\pi(1)/\pi(0)]) \leq 1/2$  with strict inequality unless  $\pi(0) = 1/2$ . Thus, for all chains on a two-point space having  $\pi(0) \neq 1/2$ ,  $\alpha < \lambda/2$ , the two sides having different orders for  $\pi(0)$  small.

**PROOF OF THEOREM A.1.** The proof borrows many ideas from the work of Rothaus [38–40]. Recall that  $\alpha$  is defined by the variational formula

$$(A.2) \quad \alpha = \inf \left\{ \frac{\mathcal{E}(f, f)}{\mathcal{L}(f)} : \mathcal{L}(f) \neq 0 \right\}.$$

We know that we can restrict ourselves to nonnegative functions because  $\mathcal{E}(|f|, |f|) \leq \mathcal{E}(f, f)$ . Now, either there exists a nonconstant nonnegative minimizer (call it  $f_0$ ) or the minimum is attained at the constant function 1, where  $\mathcal{E}(1, 1) = \mathcal{L}(1) = 0$ . In this second case, the proof of Lemma 3.1 shows that we must have  $\alpha = \lambda/2$  since, for any function  $g \neq 0$  satisfying  $E_\pi(g) = 0$ ,

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{E}(1 + \varepsilon g, 1 + \varepsilon g)}{\mathcal{L}(1 + \varepsilon g)} = \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon^2 \mathcal{E}(g, g)}{2\varepsilon^2 \text{Var}_\pi(g)} \geq \frac{\lambda}{2}.$$

This reasoning is valid for any finite Markov chain. It implies that either  $\alpha = \lambda/2$  or there must exist a nonconstant nonnegative function  $f_0$  which minimizes (A.2). Further, it is not hard to show that any minimizer  $u$  of (A.2)

must satisfy

$$(A.3) \quad 2u \log u - 2u \log \|u\|_2 - \frac{1}{\alpha} (I - K)u = 0.$$

Let us now specialize to the case at hand, where

$$K(x, y) = \pi(y), \quad \mathcal{E}(f, f) = \text{Var}_\pi(f) = \|f\|_2^2 - (E_\pi(f))^2.$$

Equation (A.3) becomes

$$(A.4) \quad 2u \log u - 2u \log \|u\|_2 - \frac{1}{\alpha} u + \frac{1}{\alpha} E_\pi(u) = 0.$$

The function  $t \rightarrow t \log t$  is convex on  $[0, +\infty)$ . It follows that any straight line intersects the graph of  $t \rightarrow t \log t$  in at most two points. Hence any solution  $u$  of (A.4) takes at most two values. For the chain  $K(x, y) = \pi(y)$  with  $\inf \pi < 1/2$ , it is easy to rule out the possibility that  $\alpha = \lambda/2$ . Indeed,  $\lambda = 1$  and a well chosen test function shows that  $\alpha < 1/2$ . Thus, we can assume that there exists a nonconstant, nonnegative minimizer  $f_0$  for (A.2). As  $f_0$  must satisfy (A.4), it takes exactly two values  $x, y \geq 0$ . Let  $\theta$  be the probability that  $f_0$  takes the value  $x$ . Without loss of generality we can assume that  $0 < \theta \leq 1/2$ . We have reduced the problem to that of computing

$$\min_{\theta, x, y} \frac{\theta(1 - \theta)(x - y)^2}{\theta x^2 \log x^2 + (1 - \theta)y^2 \log y^2 - (\theta x^2 + (1 - \theta)y^2) \log(\theta x^2 + (1 - \theta)y^2)},$$

where  $\theta$  takes all the possible values  $\theta = \pi(A)$  with  $A \subset \mathcal{X}$ . In particular,  $\theta$  varies between  $\min_{\mathcal{X}} \pi = \pi_*$  and  $1/2$ . From Theorem A.2, we infer that the minimum is

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

This ends the proof of Theorem A.1.  $\square$

**COROLLARY A.4.** *For any finite Markov chain  $K$  with invariant measure  $\pi$ , the spectral gap  $\lambda(K)$  and the log-Sobolev constant  $\alpha(K)$  satisfy*

$$\alpha(K) \geq \frac{(1 - 2\pi_*)\lambda(K)}{\log[1/\pi_* - 1]}.$$

**PROOF.** The result of Theorem A.1 can be written

$$\frac{(1 - 2\pi_*)}{\log[1/\pi_* - 1]} \mathcal{L}_\pi(f) \leq \text{Var}_\pi(f).$$

The desired result follows since  $\lambda(K)\text{Var}_\pi(f) \leq \mathcal{E}_K(f, f)$ .  $\square$

**COROLLARY A.5.** *On a finite set  $\mathcal{X}$ , consider the Markov kernel  $K(x, y) = 1/(|\mathcal{X}| - 1)$  if  $x \neq y$  and  $K(x, x) = 0$ . This has stationary measure  $\pi \equiv 1/|\mathcal{X}|$ .*

The associated logarithmic Sobolev constant is

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

In particular, the simple random walk on  $\mathcal{X} = \mathbb{Z}_3$  has log-Sobolev constant  $\alpha = 1/(2 \log 2)$ .

This readily follows from Theorem A.1. Observe that simple random walk on  $\mathbb{Z}_3$  has spectral gap  $\lambda = 1 - \cos 2\pi/3 = 3/2$ . Thus  $\lambda/2 = 0.75$  whereas  $\alpha \sim 0.72 < \lambda/2$  in this case.

**Acknowledgments.** We thank Nati Linial for encouraging us to compute the log-Sobolev constant for the complete graph and Susan Holmes for crucial help with the computer. This material was presented at an MSRI workshop during the summer of 1995. We thank Bob Osserman and the participants. We thank Jim Fill, Laurent Miclo and Nobuo Yoshida for their remarks.

## REFERENCES

- [1] ALDOUS, D. and DIACONIS, P. (1987). Strong uniform times and finite random walks. *Adv. in Appl. Math.* **8** 69–97.
- [2] ALON, N. (1987). Eigenvalues and expanders. *Combinatorica* **5** 83–96.
- [3] ALON, N. and MILMAN, V. (1985).  $\lambda_1$ , isoperimetric inequalities for graphs and superconcentrators. *J. Combin. Theory Ser. B* **38** 78–88.
- [4] BAKRY, D. and EMERY, M. (1985). Diffusions hypercontractive. *Séminaire de Probabilité XIX. Lecture Notes in Math.* **1123** 179–206. Springer, Berlin.
- [5] BAKRY, D. (1994). L'hypercontractivité et son utilisation en théorie des semigroups. *Ecole d'Été de Saint Flour 1992. Lecture Notes in Math.* **1581**. Springer, Berlin.
- [6] BOLLOBAS, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* **1** 311–316.
- [7] DAVIES, E. B. (1989). *Heat Kernels and Spectral Theory*. Cambridge Univ. Press.
- [8] DEUSCHEL, J.-D. and STROOCK, D. (1989). *Large Deviations*. Academic Press, New York.
- [9] DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, CA.
- [10] DIACONIS, P. and SALOFF-COSTE, L. (1996). Nash's inequalities for finite Markov chains. *Journal of Theoretical Probability* **9** 459–510.
- [11] DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison techniques for random walk on finite groups. *Ann. Probab.* **21** 2131–2156.
- [12] DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730.
- [13] DIACONIS, P. and SALOFF-COSTE, L. (1994). Moderate growth and random walk on finite groups. *Geometry and Functional Analysis* **4** 1–34.
- [14] DIACONIS, P. and SALOFF-COSTE, L. (1995). Random walks on contingency tables with fixed row and column sums. Unpublished manuscript.

- [15] DIACONIS, P. and SALOFF-COSTE, L. (1995). An application of Harnack inequalities to random walk on nilpotent quotients. *Journal of Fourier Analysis and Its Applications* (Kahane special issue) 189–207.
- [16] DIACONIS, P. and SALOFF-COSTE, L. (1995). Random walks on finite groups: a survey of analytic techniques. In *Probability Measures on Groups and Related Structures* **11** (H. Heyer, ed.) 46–75. World Scientific, Singapore.
- [17] DIACONIS, P. and SALOFF-COSTE, L. (1996). Random walks on generating sets of finite Abelian groups. *Probab. Theory Related Fields*. To appear.
- [18] DIACONIS, P. and SALOFF-COSTE, L. (1996). What do we know about the Metropolis algorithm. *J. Comput. System Sci.* To appear.
- [19] DIACONIS, P. and SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** 159–179.
- [20] DIACONIS, P. and SHAHSHAHANI, M. (1987). Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM J. Math. Anal.* **18** 208–218.
- [21] DIACONIS, P. and STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61.
- [22] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, 3rd ed., **1**. Wiley, New York.
- [23] FILL, J. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.
- [24] GROSS, L. (1976). Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** 1061–1083.
- [25] GROSS, L. (1993). Logarithmic Sobolev inequalities and contractivity properties of semigroups. *Lecture Notes in Math.* **1563**. Springer, Berlin.
- [26] HIGUCHI, Y. and YOSHIDA, N. (1995). Analytic conditions and phase transition for Ising models. Unpublished lecture notes (in Japanese).
- [27] HOLLEY, R. and STROOCK, D. (1987). Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys.* **46** 1159–1194.
- [28] HORN, P. and JOHNSON, C. (1985). *Matrix Analysis*. Cambridge Univ. Press.
- [29] HORN, P. and JOHNSON, C. (1990). *Topics in Matrix Analysis*. Cambridge Univ. Press.
- [30] KORZENIOWSKI, A. and STROOCK, D. (1985). An example in the theory of hypercontractive semigroups. *Proc. Amer. Math. Soc.* **94** 87–90.
- [31] LIGGETT, T. (1985). *Interacting Particle Systems*. Springer, New York.
- [32] LU, S. L. and YAU, H. T. (1993). Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Comm. Math. Phys.* **161** 399–433.
- [33] LUBOTZKY, A. (1994). *Discrete Groups, Expanding Graphs and Invariant Measures*. Birkhäuser, Boston.
- [34] MARGULIS, G. (1973). Explicit construction of concentrators. *Problemy Peredachi Informatsii* **9** 71–80. [English transl. *Problems of Information Transmission* (1975), 325–332.]
- [35] MICLO, L. (1995a). Sur les problèmes de sortie discrets inhomogènes. Preprint.
- [36] MICLO, L. (1995b). Remarques sur l’hypercontractivité et l’évolution de l’entropie pour des chaînes de Markov finies. Preprint.
- [37] QUASTEL, J. (1992). Diffusion of colour in the simple exclusion process. *Comm. Pure Appl. Math.* **45** 623–679.
- [38] ROTHHAUS, O. (1980). Logarithmic Sobolev inequalities and the spectrum of Sturm–Liouville operators. *J. Funct. Anal.* **39** 42–56.
- [39] ROTHHAUS, O. (1981). Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities. *J. Funct. Anal.* **42** 102–109.
- [40] ROTHHAUS, O. (1981). Logarithmic Sobolev inequalities and the spectrum of Schrödinger operators. *J. Funct. Anal.* **42** 110–120.
- [41] SALOFF-COSTE, L. (1994). Precise estimates on the rate at which certain diffusions tend to equilibrium. *Math. Z.* **217** 641–677.
- [42] SARNACK, P. (1990). *Some Applications of Modular Forms*. Cambridge Univ. Press.
- [43] SINCLAIR, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing* **1** 351–370.

- [44] SINCLAIR, A. (1993). *Algorithms for Random Generation and Counting: A Markov Chain Approach*. Birkhäuser, Boston.
- [45] STEIN, E. (1993). *Harmonic Analysis*. Princeton Univ. Press.
- [46] STEIN, E. and WEISS, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton Univ. Press.
- [47] STROOCK, D. (1993). Logarithmic Sobolev inequalities for Gibbs states. *Lecture Notes in Math.* **1563**. Springer, Berlin.
- [48] STROOCK, D. and ZEGARLINSKI, B. (1992). The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.* **149** 175–193.
- [49] SU, F. (1995). Ph.D. dissertation, Harvard Univ.
- [50] YAU, T. H. (1995). Logarithmic Sobolev inequality for zero range process. Report, Dept. Mathematics, New York Univ.

HARVARD UNIVERSITY  
DEPARTMENT OF MATHEMATICS  
CAMBRIDGE, MASSACHUSETTS 02138

CNRS, UNIVERSITÉ PAUL SABATIER  
STATISTIQUE ET PROBABILITÉS  
31062 TOULOUSE CEDEX  
FRANCE  
EMAIL: lsc.cict.fr