Journal of Theoretical Probability, Vol. 9, No. 2, 1996

Nash Inequalities for Finite Markov Chains

P. Diaconis[1] and L. Saloff-Coste[2]

Received November 4, 1994; revised August 6, 1995

This paper develops bounds on the rate of decay of powers of Markov kernels on finite state spaces. These are combined with eigenvalue estimates to give good bounds on the rate of convergence to stationarity for finite Markov chains whose underlying graph has moderate volume growth. Roughly, for such chains, order (diameter)^2 steps are necessary and suffice to reach stationarity. We consider local Poincaré inequalities and use them to prove Nash inequalities. These are bounds on ℓ_2-norms in terms of Dirichlet forms and ℓ_1-norms which yield decay rates for iterates of the kernel. This method is adapted from arguments developed by a number of authors in the context of partial differential equations and, later, in the study of random walks on infinite graphs. The main results do not require reversibility.

KEY WORDS: Markov chains; Dirichlet forms; infinite graphs; Nash inequalities.

1. INTRODUCTION

We begin with an example of a natural problem which is (partially) solved by using present techniques. Let C be the lattice points inside a compact convex set in $\mathbb{R}^d$. Assume that two points in C can be connected by a lattice path within C. A random walk proceeds by uniformly choosing one of the $2d$ possible neighbors of $x \in C$. If the neighbor is inside C, the walk moves to the chosen point. If the neighbor is outside C, the walk stays at x. This gives a Markov kernel:

$$K(x, y) = \begin{cases} 1/2d & \text{for } \; x \neq y \text{ neighboring points in } C \\ g(x)/2d & \text{for } \; x = y \in C \end{cases} \tag{1.1}$$

[1] Harvard University, Department of Mathematics, Cambridge, MA., 02138.
[2] CNRS, Université Paul Sabatier, Laboratoire Statistique et Probabilités, 31062 Toulouse Cedex, France.

where $g(x)$ is the number of neighbors of x that do not belong to C. This is a reversible Markov chain on C with a uniform stationary distribution. Such walks arise in the age-old problem of computing the volume of a convex set (Dyer and Frieze[18]; Lovasz and Simonovits[27]). They also arise in statistical computing (Diaconis and Sturmfels[17]). The following theorem is proved in Section 6.

Theorem 1.1. Let C be a connected set of lattice points inside a compact convex set S in $\mathbb{R}^2$ and let U be the uniform distribution on C. Let $\gamma \geqslant 1$ be the Euclidean diameter of S. There are universal constants $a_1, a_2 > 0$ such that, for any $x \in C$, the walk in Eq. (1.1) satisfies

$$\|K^n(x, \cdot) - U(\cdot)\|_{\mathrm{TV}} \leqslant a_1 e^{-a_2 c}, \qquad \text{for} \quad n = c\gamma^2, \quad c > 0$$

Further, there are universal constants $a_3, a_4 > 0$ such that

$$\sup_x \|K^n(x, \cdot) - U(\cdot)\|_{\mathrm{TV}} \geqslant a_3 e^{-a_4 c}, \qquad \text{for} \quad n = c\gamma^2, \quad c > 0$$

Roughly, Theorem 1.1 says that order γ^2 steps are necessary and suffice to approach the uniform distribution in total variation distance. For c large, this distance is exponentially close to zero whereas, for c small, the distance is bounded away from zero. The result stated in Theorem 1.1 does not hold in this generality in higher dimensions.

We now turn to our general setting and explain the main techniques of this paper. Let X be a finite set, $K(x, y)$ a Markov kernel on X with stationary distribution $\pi(x)$. Write $K^n_x(y) = K^n(x, y) = \sum_{z \in X} K^{n-1}(x, z) K(z, y)$. Our goal is to bound the rate of convergence of K^n to π for irreducible aperiodic Markov chains. Define total variation distance by

$$\|K^n_x - \pi\|_{\mathrm{TV}} = \max_{A \subset X} |K^n(x, A) - \pi(A)| = \tfrac{1}{2} \sum_{y \in X} |K^n(x, y) - \pi(y)| \tag{1.2}$$

All our bounds on variation distance proceed by bounding the ℓ_1-norm in Eq. (1.2) by the ℓ_2-norm:

$$2 \|K^n_x - \pi\|_{\mathrm{TV}} \leqslant \|(K^n_x/\pi) - 1\|_2 \tag{1.3}$$

Here, ℓ_2 has inner product $\langle f, g \rangle = \sum_{x \in X} f(x) g(x) \pi(x)$ for real functions f, g. The ℓ_2-norm is divided into parts as follows

$$\max_x \|(K^n_x/\pi) - 1\|_2 = \|K^n - \pi\|_{2 \to \infty} \leqslant \|K^{n_1}\|_{2 \to \infty} \|K^{n_2} - \pi\|_{2 \to 2} \tag{1.4}$$

where $n_1 + n_2 = n$ and K and π denote also the operators corresponding to the kernels $K(x, y)$ and $\pi(y)$ (the end of the introduction has further details). The norms on the right-hand side can be represented as

$$\|K^n - \pi\|_{2\to 2} = \sup\{ \|(K^n - \pi) f\|_2 : f \in \ell_2(\pi), \|f\|_2 = 1\} \stackrel{\text{def}}{=} \mu(n) \quad (1.5)$$

$$\|K^n\|_{2\to\infty} = \sup_x \|K^n_x/\pi\|_2 \stackrel{\text{def}}{=} D(n) \quad (1.6)$$

The quantity $\mu(n)$ of Eq. (1.5) can be estimated by eigenvalue techniques. The decay rate $D(n)$ of Eq. (1.6) can be estimated by the Nash inequalities developed next. As an illustration, the bounds above give

Lemma 1.1. Assume the notation of Eqs. (1.2)–(1.6). Let $\pi_* = \min_x \pi(x)$. Then, for any $n \geqslant 1$,

$$2 \|K^n_x - \pi\|_{\text{TV}} \leqslant \|(K^n_x/\pi) - 1\|_2 \leqslant \min_{n_1+n_2=n} D(n_1)\, \mu(n_2) \leqslant \pi_*^{-1/2} \mu^n(1)$$

The final bound in Lemma 1.1 has been frequently used to give non-asymptotic bounds for reversible Markov chains (c.f., Sinclair and Jerrum[32]). It bounds total variation by a power of a second eigenvalue. It correspond to the choice $n_1 = 0$, $n_2 = n$ and uses the easy facts that $D(0) = \pi_*^{-1/2}$, $\mu(n) \leqslant \mu(1)^n$. To see the improvement possible, consider the chain in Eq. (1.1), in dimension $d = 2$. As shown in Section 6, for this example, $\mu(1) \leqslant 1 - a/\gamma^2$ for a universal $a > 0$, and $\pi_*^{-1/2} = |C|^{1/2} \leqslant \gamma$. Thus the final bound requires n large enough to make $\gamma(1 - a/\gamma^2)^n$ small. This needs n of order $\gamma^2 \log \gamma$ which is off by a factor of $\log \gamma$. This factor can be picked up by using the decay rate $D(n)$. In Section 6, it is shown that $D(n) \leqslant A(\gamma/\sqrt{n})^2$ for $1 \leqslant n \leqslant \gamma^2$ and a universal constant A. Using this with $n_1 = \gamma^2$, $n_2 = c\gamma^2$, $c > 0$, the middle bound of Lemma 1.1 proves that n of order γ^2 suffices. Section 6 proves a complementary lower bound showing that order γ^2 steps are actually needed. For a continuous time version of Lemma 1.1, see Lemma 2.3.

Bounds for the $2 \to 2$ norm are developed in Section 2. These involve various symmetrizations suggested by Lawler and Sokal[26]; Mihail[28]; and Fill,[1] which allow eigenvalues to be used for nonreversible Markov chains. We discuss bounds in discrete and continuous time.

One main theoretical contribution comes in Section 3 which develops Nash-type inequalities as a tool for bounding the $2 \to \infty$ norm. In the context of partial differential equations, Nash[29] introduced an inequality which, transplanted to our setting, reads

$$\|f\|_2^{2+1/D} \leqslant B\left\{\mathscr{E}(f, f)+\frac{1}{N}\|f\|_2^2\right\}\|f\|_1^{1/D} \tag{1.7}$$

where $\mathscr{E}$ is the underlying Dirichlet form defined in Section 2. Adapting developments of the ideas of Nash along the line of Carlen *et al.*[2]; Coulhon and Saloff-Coste[4,5]; and Varopoulos,[35–37] we will show that Eq. (1.7) is equivalent to decay-rate estimates of the type

$$D(n)=\sup_x \|K_x^n/\pi\|_2 \leqslant C/n^D, \qquad 1 \leqslant n \leqslant N \tag{1.8}$$

where D and N are the same in Eqs. (1.7) and (1.8). Theorem 3.1 states one form of the bounds on total variation distance achieved by combining the $2 \to 2$ and $2 \to \infty$ bounds. Section 3 also makes the connection with Sobolev inequalities.

Section 4 describes how Nash inequalities can be used together with comparison arguments to study certain chains. This is illustrated for random walk on a box in $\mathbb{Z}^d$.

Section 5 contains our second main contribution. It shows how local Poincaré inequalities can be combined with the notion of moderate growth of the graph underlying the chain to prove Nash inequalities. This technique is adapted from an original idea introduced by Robinson,[1] in the context of Lie groups and further developed in Ref. 6. Path techniques along the lines of Jerrum and Sinclair[25]; see also Refs. 9 and 16, are introduced to prove these Poincaré inequalities. A volume growth condition that we call moderate growth plays a crucial role here. It generalizes a notion introduced for groups in Ref. 11.

Section 6 proves Theorem 1.1 and discusses the difficulties of extending the result to convex sets in higher dimensions.

Section 7 shows how Nash inequalities extend to time inhomogeneous problems, e.g., random walk on groups with time varying step distribution.

To conclude, we collect together frequently used notation. The kernel $K(x, y)$ satisfies $K(x, y) \geqslant 0$, $\sum_y K(x, y)=1$ for each fixed $x \in X$. The stationary distribution satisfies $\pi(x) \geqslant 0$, $\sum_y \pi(y)=1$ and $\sum_x \pi(x)\, K(x, y)=\pi(y)$ for all $y \in X$. The space $\ell_p(\pi)$ has norm $\|f\|_p=(\sum |f(x)|^p\, \pi(x))^{1/p}$, $1 \leqslant p<\infty$, and $\|f\|_\infty=\sup_x |f(x)|$. If $Q\colon \ell_p \to \ell_{p'}$ is a linear map, we write $\|Q\|_{p \to p'}$ for the smallest number b such that $\|Qf\|_{p'} \leqslant b\, \|f\|_p$ for all $f \in \ell_p$. The kernel $K(x, y)$ defines an operator (also denoted by K) which acts on ℓ_p by

$$Kf(x)=\sum_y K(x, y)\, f(y)$$

Jensen's inequality shows that K is a contraction on ℓ_p for $1 \leqslant p \leqslant \infty$, so $\|K\|_{p \to p} \leqslant 1$. We often regard the stationary distribution π as a map π from ℓ_p to ℓ_p which takes f into the constant function $\pi f(x) = \sum f(y)\, \pi(y)$.

Sinclair[33] gives a book length treatment of related material with a thorough review of the literature. The present paper builds on Refs. 9–11, and 16, which may be consulted for background and examples.

2. EIGENVALUE BOUNDS

This section gives bounds on $\|K^n - \pi\|_{2 \to 2}$ of Eq. (1.5) by using the eigenvalues of the multiplicative symmetrization of the operator K, namely K^*K, where K^* is the adjoint of K on $\ell_2(\pi)$ with kernel

$$K^*(x, y) = K(y, x)\, \pi(y)/\pi(x) \tag{2.1}$$

An easy calculation shows that K^*K is reversible with respect to π and so self-adjoint on $\ell_2(\pi)$. It follows that K^*K has nonnegative real eigenvalues

$$\beta_0 = 1 \geqslant \beta_1 \geqslant \beta_2 \geqslant \cdots \geqslant \beta_{|X| - 1} \geqslant 0$$

Define

$$\mu = \mu(K) = \sqrt{\beta_1(K^*K)} \tag{2.2}$$

The minimax characterization for eigenvalues of reversible chains gives

$$1 - \mu^2 = \min\{\mathscr{E}_*(f, f)\colon \pi(f) = 0,\ \|f\|_2 = 1\} \tag{2.3}$$

where the Dirichlet form $\mathscr{E}_*$ is given by any of the expressions

$$\begin{aligned} \mathscr{E}_*(f, f) &= \langle (I - K^*K) f, f \rangle = \tfrac{1}{2} \sum_{x, y} |f(x) - f(y)|^2\, K^*K(x, y)\, \pi(x) \\ &= \|f\|_2^2 - \|Kf\|_2^2 \end{aligned} \tag{2.4}$$

See Ref. 16, for background and references. The use of singular values to analyze non-selfadjoint operators is classical. See e.g., Gohberg and Krein.[22] The following lemma originates with Elena Mihail.[28] It was isolated, developed, and applied in nontrivial problems by Jim Fill.[21] We give a short proof for completeness.

Lemma 2.1. Let K, π be a Markov chain on a finite state space X. Then

$$\|K - \pi\|_{2 \to 2} = \mu$$

$$\|K^n - \pi\|_{2 \to 2} \leqslant \mu^n \qquad \text{for any} \quad n \geqslant 1$$

where μ is defined at Eq. (2.2).

Proof. Observe that $(K-\pi)f = K(f-\pi f)$ and $\|f-\pi f\|_2^2 = \|f\|_2^2 - (\pi f)^2$. Thus,

$$\|K-\pi\|_{2\to 2} = \max\{\|Kf\|_2 : \pi f = 0, \|f\|_2 = 1\} = \mu$$

where the last equality follows from Eqs. (2.3) and (2.4).

Example 2a. Take $X = S_n$, the symmetric group. Using cycle notation, let the probability Q be defined by

$$Q(\mathrm{id}) = Q((1,2)) = Q((1,2,\ldots,n)) = 1/3 \tag{2.5}$$

The associated kernel is $K(\sigma,\eta) = Q(\eta\sigma^{-1})$. This corresponds to a card-shuffle random walk which proceeds by choosing the identity, transpose top two, or the n-cycle top to bottom, each with equal probability. The invariant measure for this walk is the uniform distribution $\pi(\sigma) = 1/n!$. The walk is connected and aperiodic but not symmetric. Define the reversed walk $Q^*(\mathrm{id}) = Q^*((1,2)) = Q^*((n, n-1,\ldots,1)) = 1/3$. The multiplicative symmetrization K^*K corresponds to the probability

$$Q^* \star Q(\mathrm{id}) = 1/3, \qquad Q^* \star Q((1,2)) = 2/9$$

$$Q^* \star Q((1,2,\ldots,n)^{\pm 1}) = Q^* \star Q([(1,2)(1,2,\ldots,n)]^{\pm 1}) = 1/9$$

Using comparison techniques of Diaconis and Saloff-Coste,[10] Section 3, it is straightforward to show

$$\mu^2(Q) = \beta_1(Q^*Q) \leqslant 1 - 1/(41n^3)$$

One problem with the symmetrization K^*K, is that it can destroy connectivity. As an example, modify Eq. (2.5) by setting $Q((1,2)) = Q((1,\ldots,n)) = 1/2$. The multiplicative symmetrization corresponds to the probability $Q^* \star Q(\mathrm{id}) = 1/2$, $Q^* \star Q([(1,2)(1,\ldots,n)]^{\pm 1}) = 1/4$. The support of $Q^* \star Q$ does not generate the symmetric group but only a cyclic subgroup of size $n-1$. Here, symmetrizing the measure $Q^2 = Q \star Q$ leads to a successful analysis. We will return to this example at the end of this section.

Example 2b. Consider the walk on $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ defined by $X_{n+1} = 2X_n + \varepsilon_n$. Here, p is a prime, ε_i are idd taking values 0, ± 1 with probability 1/3 each. Chung *et al.*,[3] show that this walk gets random in order $[\log p][\log\log p]$ steps. The multiplicative symmetrization takes order p^2 steps to get random and $\mu^2 = 1 - c/p^2 + O(1/p^4)$ for a fixed constant c.

We turn next to continuous time and the additive symmetrization. Let the Markov semigroup associated with K be defined by

$$H_t = e^{-t(I-K)} = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} K^n, \qquad H_0 = I \tag{2.6}$$

We consider the Dirichlet form associated with K and defined by

$$\mathscr{E}(f, f) = \langle (I-K) f, f \rangle = \tfrac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 K(x, y)\, \pi(x) \tag{2.7}$$

for any **real valued** function f.

For complex valued function, set

$$\mathscr{E}_+(f, f) = \langle (I - \tfrac{1}{2}[K + K^*]) f, f \rangle$$

One has $\langle K^* f, f \rangle = \langle f, Kf \rangle = \overline{\langle Kf, f \rangle}$ and this yields

$$\mathscr{E}_+(f, f) = \operatorname{Re}(\langle (I-K) f, f \rangle)$$

Since we work here with real valued funtions only, there is no distinction to be made between $\mathscr{E}$ and $\mathscr{E}_+$. The proof of the next statement shows how the Dirichlet form $\mathscr{E}$ appears naturally when studying the semigroup H_t.

Lemma 2.2. The semigroup H_t satisfies

$$\|H_t - \pi\|_{2 \to 2} \leqslant e^{-t\lambda}$$

where

$$\lambda = \min\{\mathscr{E}(f, f) \colon \pi(f) = 0, \|f\|_2 = 1\} \tag{2.8}$$

Moreover, λ is the largest number α such that $\|H_t - \pi\|_{2 \to 2} \leqslant e^{-\alpha t}$ for all $t > 0$. In particular $1 - \lambda \leqslant \mu$, where μ is defined at Eq. (2.2).

Proof. Observe that $(d/dt) H_t = -(I-K) H_t$. This implies

$$\frac{d}{dt} \|H_t f\|_2^2 = -\langle (2I - [K + K^*])\, H_t f, H_t f \rangle = -2\mathscr{E}(H_t f, H_t f) \tag{2.9}$$

For f such that $\pi(f) = 0$, the definition of λ and Eq. (2.9) give

$$\frac{d}{dt} \|H_t f\|_2^2 \leqslant -2\lambda \, \|H_t f\|_2^2$$

If also $\|f\|_2=1$, integrating from 0 to t gives $\|H_t f\|_2^2 \leqslant e^{-2t\lambda}$ which implies the first bound claimed.

To show that λ is optimal, again let f satisfy $\pi(f)=0$, $\|f\|_2=1$. If $\|H_t-\pi\|_{2\to 2} \leqslant e^{-t\alpha}$, we get $\|(H_t-\pi)f\|_2^2 = \|H_t f\|_2^2 \leqslant e^{-2t\alpha}$ or $\|H_t f\|_2^2 - 1 \leqslant e^{-2t\alpha}-1$. Thus,

$$\frac{d}{dt}\|H_t f\|_2^2\bigg|_{t=0} = -2\mathscr{E}(f,f) \leqslant -2\alpha$$

and so $\alpha \leqslant \lambda$. Finally, writing

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

and using Lemma 2.1, we get $\|H_t-\pi\|_{2\to 2} \leqslant e^{-t(1-\mu)}$ and so $1-\mu \leqslant \lambda$.

It will be useful to have the following continuous-time version of Lemma 1.1.

Lemma 2.3. Let K be a Markov chain on a finite set X with invariant probability measure π. Let H_t be defined by Eq. (2.6). Then

$$2\,\|H_t^x - \pi\|_{\mathrm{TV}} \leqslant \|(H_t^x/\pi) - 1\|_2 \leqslant \min_{t_1+t_2=t}\{D_+(t_1)\,e^{-\lambda t_2}\} \leqslant \pi_*^{-1/2} e^{-\lambda t}$$

where λ is defined at Eq. (2.8), $\pi_* = \min_x \pi(x)$, and $D_+(t) = \|H_t\|_{2\to +\infty}$.

We now discuss the discrete time applications of the additive symmetrization $\frac{1}{2}[K+K^*]$ of K. This is irreducible and aperiodic if K is. The following examples show that this symmetrization can change the rates in unpredictable ways.

Example 2c. Consider the deterministic Markov chain on $\mathbb{Z}_p$ which always moves one step clockwise. If started at 0, it never gets random! The multiplicative symmetrization is the chain that stays still at its starting point. The additive symmetrization is the usual random walk on $\mathbb{Z}_p$.

Example 2d. Define a Markov chain on $\mathbb{Z}_2^d$ by $X_0=0$, $X_{n+1} = SX_n + \varepsilon_{n+1}$ with ε_1 idd taking values $(0,\ldots,0,0)$ and $(0,\ldots,0,1)$ with equal probability, and $S(x_1, x_2,\ldots, x_d) = (x_2,\ldots, x_d, x_1)$. This chain is exactly uniform after d steps. It is not hard to show that the additive symmetrization takes at least order d^2 steps to get random. The multiplicative symmetrization isn't connected.

Example 2e. In the other direction, let p be an odd integer and consider the random walk on $\mathbb{Z}_p$ which takes steps 1 or $\lfloor \sqrt{p} \rfloor$ with equal probability. This walk takes order p^2 steps to get random (it is, up to an affine transformation, the walk taking value ± 1 with equal probability). The additive symmetrization takes order p steps to get random (see e.g. Ref. 11, Section 1). Here, the symmetrization speeds things up.

Despite these examples, both symmetrizations are often useful (e.g., see the discussion in Fill's paper[21]). In general, the additive symmetrization is easier to use and gives bounds for the first eigenvalue of the process in continuous time. The multiplicative symmetrization is crucial for the Nash inequalities of Section 3. We find it instructive to look carefully at the following simple example of a nonreversible chain with a nonobvious stationary stationary distribution which can be analyzed in detail by the techniques developed in this paper.

Example 2f. Let $n \geqslant 3$ be an odd integer. Consider the usual nearest neighbor walk on an n-point path with an extra directed edge from 1 to n (Fig. 1).

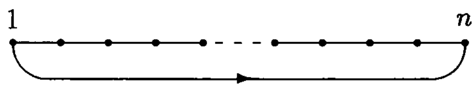

Fig 1. The neighbor walk on an n-point path with an extra directed edge from 1 to n.

This gives a nonreversible Markov chain

$$\begin{aligned}
K(i, i+1) = K(i, i-1) &= \tfrac{1}{2} \qquad \text{if} \quad 2 \leqslant i \leqslant n-1 \\
K(1, 2) = K(1, n) &= \tfrac{1}{2} \\
K(n, n-1) &= 1 \\
K(i, j) &= 0 \qquad \text{otherwise}
\end{aligned} \tag{2.10}$$

While the usual random walk on a path has essentially a uniform stationary distribution, adding the extra edge gives

Lemma 2.4. The Markov chain K at Eq. (2.10) has stationary distribution

$$\pi(i) = \frac{2i}{n^2}, \qquad \text{if} \quad 1 \leqslant i \leqslant n-1, \quad \pi(n) = \frac{1}{n}$$

Proof. The equations for stationarity become

$$\frac{\pi(2)}{2}=\pi(1), \qquad \frac{\pi(1)}{2}+\frac{\pi(3)}{2}=\pi(2),...$$

$$..., \frac{\pi(n-3)}{2}+\frac{\pi(n-1)}{2}=\pi(n-2), \qquad \frac{\pi(n-2)}{2}+\pi(n)=\pi(n-1)$$

Solving inductively for $\pi(i)$ in terms of $\pi(1)$ gives $\pi(i)=i\pi(1)$ for $1 \leqslant i \leqslant n-1$ and $\pi(n)=n\pi(1)/2$. Then, use $\pi(1)+\cdots+\pi(n)=1$ to solve for $\pi(1)$.

Let us compute the additive and multiplicative symmetrizations. First, recall that $K^*(x, y)=K(y, x)\,\pi(y)/\pi(x)$; thus

$$K^*(i, i\pm 1)=\frac{1}{2}\pm\frac{1}{2i} \qquad \text{for} \quad 2 \leqslant i \leqslant n-1$$

$K^*(1, 2)=1$, $K^*(n, 1)=1/n$, $K^*(n, n-1)=1-1/n$, and $K^*(x, y)=0$ otherwise.

The additive symmetrization $K_+=\frac{1}{2}(K+K^*)$ becomes

$$K_+(i, i\pm 1)=\frac{1}{2}\pm\frac{1}{4i} \qquad \text{for} \quad 2 \leqslant i \leqslant n-1$$

$K_+(1, 2)=3/4$, $K_+(1, n)=1/4$, $K_+(n, 1)=1/(2n)$, $K_+(n, n-1)=1-1/(2n)$, and $K_+(x, y)=0$ otherwise.

The multiplicative symmetrization has kernel

$$P(x, y)=[K^*K](x, y)=\sum_z K^*(x, z)\,K(z, y)$$

Thus

$$P(i, i+2)=\frac{1}{4}+\frac{1}{4i} \qquad \text{for} \quad 1 \leqslant i \leqslant n-2$$

$$P(i, i-2)=\frac{1}{4}-\frac{1}{4i} \qquad \text{for} \quad 3 \leqslant i \leqslant n-1$$

$$P(i, i)=\frac{1}{2} \qquad \text{for} \quad 2 \leqslant i \leqslant n-2$$

$$P(1, 1)=\frac{1}{2}, \qquad P(2, n)=\frac{1}{8}, \quad P(n-1, n-1)=\frac{3}{4}+\frac{1}{4(n-1)}$$

$$P(n, n-2)=\frac{1}{2}-\frac{1}{2n}, \qquad P(n, n)=\frac{1}{2}, \quad P(n, 2)=\frac{1}{2n}$$

Using techniques as in Refs. 16 and 32, the quantities μ, λ defined at Eqs. (2.2) and (2.8) can be bounded as $\mu \leqslant 1 - c/n^2$ and $\lambda \geqslant c/n^2$ for a universal constant c. The geometric techniques developed in Section 5 allow us to show that the chain K defined at Eq. (2.10) is closed to equilibrium after order n^2 steps. See Lemmas 5.5 and 5.6 for details. □

We conclude this section with two results on higher eigenvalues suggested to us by Jim Fill.

Lemma 2.5. Let K be a Markov chain on a finite set X. Let K^*K have eigenvalues β_i, in nonincreasing order, and set $\mu_i^* = \sqrt{\beta_i}$. Let $\frac{1}{2}[K + K^*]$ have eigenvalues μ_i^+, in nondecreasing order. Then

$$\mu_i^+ \leqslant \mu_i^* \qquad \text{for all } i$$

If in addition we assume that $\min_x K(x, x) \geqslant \varepsilon > 0$, then

$$\mathscr{E}_* \geqslant 2\varepsilon \mathscr{E} \tag{2.11}$$

and so

$$\mu_i^* \leqslant 1 - \varepsilon(1 - \mu_i^+)$$

Proof. The first inequality follows from a classical inequality for singular values. See e.g., Marshall and Olkin,[30] or Horn and Johnson,[24] (p. 150). In the other direction, the identity

$$I - K^*K = 2\varepsilon(I - \tfrac{1}{2}[K + K^*]) + (1 - \varepsilon)^2[I - (1 - \varepsilon)^{-2}(K^* - \varepsilon I)(K - \varepsilon I)]$$

yields Eq. (2.11) (see also Ref. 21). Then, the minimax characterization implies

$$\mu_i^* \leqslant \sqrt{1 - 2\varepsilon(1 - \mu_i^+)} \leqslant 1 - \varepsilon(1 - \mu_i^+) \qquad \square$$

Lemma 2.6. With the same notation as in Lemma 2.5,

$$\sum_x \|(K_x^n/\pi) - 1\|_2^2\, \pi(x) \leqslant \sum_1^{|X|-1} \beta_i^n$$

Proof. Let $\beta_i(n)$ be the eigenvalues of $K^{*n}K^n$, and observe that these are the same as the eigenvalues of K^nK^{*n}. Let (ψ_i), $i = 0,..., |X| - 1$, be the corresponding basis of orthonormal eigenfunctions for K^nK^{*n} (these eigenfunctions also depend on n). Elementary linear algebra shows that

$$\|(K_x^n/\pi) - 1\|_2^2 = \sum_1^{|X|-1} \beta_i(n)\, |\psi_i(x)|^2$$

Thus,

$$\sum_x \|(K_x^n/\pi) - 1\|_2^2 \, \pi(x) = \sum_1^{|X|-1} \beta_i(n)$$

Now, a classical inequality (see e.g., Horn and Johnson,[24] p. 190) gives

$$\sum_1^{|X|-1} \beta_i(n) \leqslant \sum_1^{|X|-1} \beta_i^n \qquad \square$$

Remark 2.1. Lemma 2.6 gives bound on the average ℓ_2-norm (and so the average ℓ_1-norm) in terms of the eigenvalues of K^*K. For random walk on groups, the ℓ_p-norms involved here do not depend on the starting points so that Lemma 2.6 gives bounds on $\|(K_x^n/\pi) - 1\|_2$ and on $\|K_x^n - \pi\|_{\mathrm{TV}}$. As an example, consider the probability Q on the symmetric group S_n defined at Eq. (2.5). Here, $\pi = U$ is the uniform distribution. As shown earlier, $\mu = \sqrt{\beta_1} \leqslant 1 - c/n^2$ for an explicit constant $c > 0$. The final bound o Lemma 1.1 shows

$$2 \, \|Q^m - U\|_{\mathrm{TV}} \leqslant (n!)^{1/2} (1 - c/n^3)^m$$

Thus, m of order $n^4 \log n$ suffices to make variation distance small. This can be improved to order $n^3 \log n$ (which is presumably the right answer) using Lemma 2.6. For a proof, bound the ℓ_2 rate of convergence of Q^*Q using the comparison technique of Ref. 10. The same argument applies to the slightly different measure considered at the end of Example 2a.

3. NASH INEQUALITIES AND DECAY BOUNDS

The main result of this section shows that a Nash inequality implies a bound on the decay of K^n. Following this a converse: for reversible chains, decay bounds are equivalent to Nash inequalities. Similar results are given for continuous time. At the end of the section we give some history of these techniques in partial differential equations and probability theory. We also sketch their connection to Sobolev inequalites. Methods for proving Nash inequalities are given in Sections 4 and 5.

Theorem 3.1. Let $K(x, y)$ be a Markov kernel of a finite set X. With the notation as in Eqs. (1.2)–(1.6), (2.1)–(2.4), assume that the Nash inequality

$$\|f\|_2^{2+1/D} \leqslant C\left(\mathscr{E}_*(f, f) + \frac{1}{N} \|f\|_2^2\right) \|f\|_1^{1/D} \tag{3.1}$$

holds for some constants $C, D>0$, $N\geqslant 1$ and all functions f. Then,

$$\|K^n\|_{2\to\infty}=D(n)\leqslant(4CB/(n+1))^D \qquad \text{for} \quad 0\leqslant n\leqslant N \tag{3.2}$$

with $B=B(D,N)=(1+1/N)(1+\lceil 4D\rceil)$. Moreover, if K is irreducible and aperiodic, then for any $x\in X$ and any $\theta>0$,

$$\|(K_x^n/\pi)-1\|_2\leqslant e^{-\theta} \qquad \text{for} \quad n\geqslant N+\frac{1}{1-\mu}\left[D\log(4C(1+\lceil 4D\rceil)/N)+\theta\right] \tag{3.3}$$

This yields exactly the same bound for $2\,\|K_x^n-\pi\|_{\mathrm{TV}}$. When K is reversible (i.e., $K=K^*$) the factor $4C$ can be replaced by C in Eqs. (3.2) and (3.3).

To set up the proof of Theorem 3.1, fix a function f with $\|f\|_1=1$. Set $t(n)=\|K^nf\|_2^2$ and notice that $t(n)\leqslant t(n-1)$ for $n\geqslant 1$. The following argument, the heart of the proof, works for any nonincreasing sequence $t(n)$.

Lemma 3.1. Suppose $t(n)$, $0\leqslant n\leqslant N$, is a nonincreasing sequence of nonnegative real numbers that satisfies

$$t(n)^{1+1/(2D)}\leqslant C(t(n)-t(n+1)+t(n)/N) \qquad \text{for} \quad 0\leqslant n\leqslant N-1 \tag{3.4}$$

with C, $D>0$. Then

$$t(n)\leqslant(CB/(n+1))^{2D} \qquad \text{for} \quad 0\leqslant n\leqslant N$$

where $B=B(D,N)=(1+1/N)(1+\lceil 4D\rceil)$.

Proof. Note that $t(0)^{1/(2D)}\leqslant C(1+1/N)$. Thus, for any integer b and any integer $n<b$,

$$t(n)\leqslant t(0)\leqslant\left(C\left(1+\frac{1}{N}\right)\frac{b}{n+1}\right)^{2D}$$

Regard b as fixed and let $n_0\geqslant b$ be the first integer less than or equal to N (if any) such that

$$t(n_0)>\left(C\left(1+\frac{1}{N}\right)\frac{b}{n_0+1}\right)^{2D} \tag{3.5}$$

Then

$$\left(C\left(1+\frac{1}{N}\right)\frac{b}{n_0+1}\right)^{2D}<t(n_0)\leqslant t(n_0-1)\leqslant\left(C\left(1+\frac{1}{N}\right)\frac{b}{n_0}\right)^{2D}$$

Using this in Eq. (3.4) gives

$$t(n_0) \leqslant \left(1+\frac{1}{N}\right) t(n_0-1) - \frac{t(n_0-1)^{1+1/(2D)}}{C}$$
$$\leqslant \left(C\left(1+\frac{1}{N}\right)\frac{b}{n_0+1}\right)^{2D}\left[\left(1+\frac{1}{N}\right)\left\{\left(1+\frac{1}{n_0}\right)^{2D}-\frac{b}{n_0+1}\right\}\right]$$

We now argue that, if b is chosen so that $b=1+\lceil 4D\rceil$, the factor in square brackets is at most 1 for all $0 \leqslant n_0 \leqslant N$. This contradicts Eq. (3.5) and shows that

$$t(n) \leqslant \left(C\left(1+\frac{1}{N}\right)\frac{1+\lceil 4D\rceil}{n+2}\right)^{2D}$$

for all $0 \leqslant n \leqslant N$ which is the desired conclusion.

We set $\alpha = 2D$, $b = 1+\lceil 2\alpha\rceil$ and proceed to prove the inequality

$$\left[\left(1+\frac{1}{N}\right)\left\{\left(1+\frac{1}{n}\right)^{\alpha}-\frac{b}{n+1}\right\}\right] \leqslant 1 \qquad \text{for} \quad b \leqslant n \leqslant N \tag{3.6}$$

in stages. Once this is done, we will have proved Lemma 3.1.

Claim 1. For positive integers k, n and any real positive α such that $k-1<\alpha\leqslant k$,

$$\left(1+\frac{1}{n}\right)^{\alpha} \leqslant 1+\frac{\alpha}{n}+\cdots+\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!\,n^k}$$

Proof. The remainder is nonpositive.

Claim 2. Let $k-1<\alpha$. Set $b=1+\lceil 2\alpha\rceil$. Then, for $n \geqslant b$,

$$\frac{\alpha}{n}+\cdots+\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!\,n^k}-\frac{b}{n+1} \leqslant -\frac{1}{n+1}$$

Proof. Consider that α and $n \geqslant b$ are fixed and set

$$\Delta(k)=\frac{\alpha}{n}+\cdots+\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!\,n^k}-\frac{b-1}{n+1}$$

We wish to show that $\Delta(k) \leqslant 0$ for all k such that $k-1<\alpha$. Observe that

$$\Delta(1)=\frac{\alpha}{n}-\frac{b-1}{n+1} \leqslant \frac{\alpha}{n}-\frac{2\alpha}{n+1} = -\frac{\alpha(n-1)}{n(n+1)}$$

This proves the claim for $k=1$ and also shows that,

$$\frac{\alpha}{n}-\frac{b-1}{n+1}\leqslant -\frac{\alpha(\alpha-1)}{n(n+1)}$$

since $n>\alpha>0$. Proceeding by induction on k, assume that $1\leqslant k-1<\alpha$ and that we have shown $\Delta(k-1)\leqslant 0$ and

$$\Delta(k-1)\leqslant \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{(k-1)!n^{k-1}(n+1)}$$

Then

$$\begin{aligned}\Delta(k)&=\Delta(k-1)+\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!n^k}\\&\leqslant\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!n^k}-\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{(k-1)!n^{k-1}(n+1)}\\&=-\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!n^k}\left(\frac{kn}{n+1}-1\right)\end{aligned}$$

This shows that $\Delta(k)\leqslant 0$ if $1\leqslant k-1<\alpha$ and that

$$\Delta(k)\leqslant\frac{\alpha(\alpha-1)\cdots(\alpha-k)}{(k)!n^k(n+1)}$$

To obtain this last inequality, we have used the fact that $k(n+1)\geqslant 2n+1\geqslant\alpha+n+1$. This completes the induction and proves Claim 2.

Claim 3. Fix $\alpha>0$. For $b=1+\lceil 2\alpha\rceil$ and $b\leqslant n\leqslant N$,

$$\left(1+\frac{1}{N}\right)\left\{\left(1+\frac{1}{n}\right)^{\alpha}-\frac{b}{n+1}\right\}\leqslant 1$$

Proof. From Claims 1 and 2,

$$\left(1+\frac{1}{n}\right)^{\alpha}-\frac{b}{n+1}\leqslant 1-\frac{1}{n+1}\leqslant 1-\frac{1}{N+1}$$

and $(1+1/N)(1-1/(N+1))=1$.

Claim 3 with $\alpha=2D$ proves Eq. (3.6) and thus finishes the proof of Lemma 3.1.

Proof of Theorem 3.1. As said before, take $t(n) = \|K^n f\|_2^2$ with $\|f\|_1 = 1$. The assumption of Eq. (3.1), i.e., the Nash inequality, applied to $K^n f$ shows that Eq. (3.4) holds. Thus, Lemma 3.1 yields

$$\|K^n\|_{1 \to 2} \leqslant (C(1 + 1/N)(1 + \lceil 4D \rceil)/(n+1))^D \qquad \text{for} \quad 0 \leqslant n \leqslant N$$

By duality, this gives

$$\begin{aligned} \|K^{*n}\|_{2 \to \infty} &= \|K^n\|_{1 \to 2} \\ &\leqslant (C(1 + 1/N)(1 + \lceil 4D \rceil)/(n+1))^D \qquad \text{for} \quad 0 \leqslant n \leqslant N \qquad (3.7) \end{aligned}$$

if $K = K^*$, this ends the proof of the bound in Eq. (3.2). If $K \neq K^*$, we need a further argument which is adapted from Ref. 4. Fix f with $\|f\|_1 = 1$ and set

$$M(f) = \max_{0 \leqslant n \leqslant N} \{(n+1)^{2D} \|K^{*n} f\|_\infty\}$$

We want to get a bound on $M(f)$. For any integer $0 \leqslant n \leqslant N$, write $n = n_1 + n_2$ where $n_1 = \lfloor n/2 \rfloor$ and observe that

$$\begin{aligned} \|K^{*n} f\|_\infty &\leqslant (CB/(n_1 + 1))^D \|K^{*n_2} f\|_2 \\ &\leqslant (CB/(n_1 + 1))^D \|K^{*n_2} f\|_\infty^{1/2} \\ &\leqslant (4CB/(n+1)^2)^D M(f)^{1/2} \end{aligned}$$

Here, $B = B(D, N) = (1 + 1/N)(1 + \lceil 4D \rceil)$. The first inequality follows from Eq. (3.7), the second from

$$\|K^{*n_2} f\|_2 \leqslant \|K^{*n_2} f\|_\infty^{1/2} \|K^{*n_2} f\|_1^{1/2} \leqslant \|K^{*n_2} f\|_\infty^{1/2}$$

and the third follows from the definition of $M(f)$ and the fact that $(n_1 + 1)(n_2 + 1) \geqslant (n+1)^2/4$. This yields $M(f) \leqslant (4CB)^D M(f)^{1/2}$ or $M(f) \leqslant (4CB)^{2D}$. It follows that

$$\|K^{*n}\|_{1 \to \infty} \leqslant (4CB/(n+1))^{2D} \qquad \text{for} \quad 0 \leqslant n \leqslant N$$

By duality,

$$\|K^n\|_{1 \to \infty} \leqslant (4CB/(n+1))^{2D} \qquad \text{for} \quad 0 \leqslant n \leqslant N$$

Now, by the Riesz-Thorin interpolation theorem (see Ref. 34, p. 179), this gives

$$\|K^n\|_{2 \to \infty} \leqslant (4CB/(n+1))^D \qquad \text{for} \quad 0 \leqslant n \leqslant N$$

which is the first inequality stated in Theorem 3.1. The bound for the ℓ_2 norm in Eqs. (3.3) follows from (3.2) and Lemma 1.1.

Remark 3.1.

1. For Q a symmetric probability distribution on a finite group G, the decay rate satisfies $D(n)^2 = |G|\, Q^{2n}(\mathrm{id}) \geqslant 1$. For example random walk on $\mathbb{Z}_p$, p odd, $D(n)^2 = p\binom{2n}{n}/2^{2n} \sim p/\sqrt{\pi n}$ for $1 \leqslant n \leqslant p/2$, as n becomes large.
2. When there is a natural dimension d (e.g., random walk on a square box in $\mathbb{Z}^d$ has dimension d), the decay rate has exponent $D = d/4$.
3. We have chosen to symmetrize K to K^*K. The Dirichlet form associated with the other symmetrization on KK^* is

$$\mathscr{E}_{\#}(f, f) = \|f\|_2^2 - \|K^*f\|_2^2$$

This proof works for this symmetrization also. It shows that the Nash inequality

$$\|f\|_2^{2+1/D} \leqslant C\left\{\mathscr{E}_{\#}(f, f) + \frac{1}{N}\|f\|_2^2\right\} \|f\|_1^{1/D}$$

implies

$$D(n) \leqslant (CB/(n+1))^D \qquad \text{for} \quad 0 \leqslant n \leqslant N$$

where B is as in Theorem 3.1. This is better than Eq. (3.2) by a factor of 4^D. There is often no practical difference between bounding $\mathscr{E}_{\#}$ and $\mathscr{E}_*$, so this may offer a useful improvement.

4. There is a universal inequality between $\mathscr{E}$ defined at Eq. (2.7) and $\mathscr{E}_*$, namely, $\mathscr{E}_*(f, f) \leqslant 2\mathscr{E}(f, f)$ for any real function f. To see this, observe that $0 \leqslant \|f - Kf\|_2^2 = \|f\|_2^2 - \langle Kf, f\rangle - \langle f, Kf\rangle + \|Kf\|_2^2$. For any real function f, this gives $\|f\|_2^2 - \langle Kf, f\rangle \geqslant \langle Kf, f\rangle - \|Kf\|_2^2$. Thus,

$$\mathscr{E}_*(f, f) = \|f\|_2^2 - \langle Kf, f\rangle + \langle Kf, f\rangle - \|Kf\|_2^2 \leqslant 2\mathscr{E}$$

The same inequality holds with $\mathscr{E}_{\#}$ instead of $\mathscr{E}_*$. There is no converse inequality, even for reversible chains. Any reversible chain with -1 as an eigenvalue provides a counter example. Thus, direct use of Theorem 3.1 for reversible Markov chains requires working with $\mathscr{E}_*$ which may be more complex than $\mathscr{E}$. The following result allows direct use of $\mathscr{E}$.

Corollary 3.1. Let K, π be a reversible Markov chain on a finite set X. With notation as in Eqs. (1.2)–(1.6), (2.1)–(2.4), assume that the Nash inequality

$$\|f\|_2^{2+1/D} \leqslant C\left\{\mathscr{E}(f,f)+\frac{1}{N}\|f\|_2^2\right\}\|f\|_1^{1/D}$$

holds for some constants C, $D>0$, $N\geqslant 1$, and all functions f. Then

$$\|K^n\|_{2\to\infty}=D(n)\leqslant\sqrt{2}\,[2CB/(n+1)]^D \qquad \text{for} \quad 0\leqslant n\leqslant 2N$$

with $B=(1+(2N)^{-1})(1+\lceil 4D\rceil)$.

Proof. Let $K_+=\frac{1}{2}(I+K)$ have $\mathscr{E}^+$ and $\mathscr{E}_*^+$ as corresponding Dirichlet forms. These satisfy

$$\mathscr{E}_*^+\geqslant\mathscr{E}^+=\tfrac{1}{2}\mathscr{E}$$

The hypothesis and Theorem 3.1 yield

$$\|K_+^n\|_{2\to\infty}\leqslant(2CB/(n+1))^D \qquad \text{for} \quad 0\leqslant n\leqslant 2N$$

Fix a function $f\geqslant 0$ and note that $\|K^nf\|_2$ is nonincreasing in n. Write

$$\begin{aligned}\|K_+^nf\|_2^2&=\left\langle\left(\frac{I+K}{2}\right)^{2n}f,f\right\rangle=\frac{1}{2^{2n}}\sum_0^{2n}\binom{2n}{i}\langle K^if,f\rangle\\&\geqslant\frac{1}{2^{2n}}\sum_0^{n}\binom{2n}{2i}\langle K^{2i}f,f\rangle\\&=\frac{1}{2^{2n}}\sum_0^{n}\binom{2n}{2i}\|K^if\|_2^2\geqslant\frac{1}{2}\|K^nf\|_2^2\end{aligned}$$

This completes the proof, for it shows that

$$\|K^n\|_{2\to\infty}=\|K^n\|_{1\to2}\leqslant\sqrt{2}\,\|K_+^n\|_{1\to2}=\sqrt{2}\,\|K_+^n\|_{2\to\infty}$$

Here, we have used the fact that

$$\|K^n\|_{p\to q}=\sup_{\|f\|_p\leqslant 1}\|K^nf\|_q=\sup_{\substack{\|f\|_p\leqslant 1\\ f\geqslant 0}}\|K^nf\|_q \qquad \square$$

Remark 3.2. Example 2c shows that the hypothesis that K is reversible cannot be comitted in this corollary. Indeed, the form $\mathscr{E}$ corresponding to this example is the same as the form of the usual symmetric random walk on the circle. Thus, $\mathscr{E}$ satisfies a Nash inequality with C and N of order p^2 and $D=1/4$. However, $D(n)\equiv\sqrt{p}$ for all n.

The next result gives a converse to Theorem 3.1. It shows that, for reversible Markov chains, polynomial decay of the kernel implies a Nash inequality. This is a direct adaptation of arguments from Carlen *et al.*[2] It is quite useful for comparison of different chains. See Section 4 for an example.

Theorem 3.2. Let K, π be a reversible Markov chain on a finite space χ. If there are C, $D>0$, $N\geqslant 1$ such that

$$\|K^n\|_{2\to\infty}=D(n)\leqslant C(n+1)^{-D} \qquad \text{for} \quad 0\leqslant n\leqslant N$$

then $\mathscr{E}_*$ satisfies the Nash inequality

$$\|f\|_2^{2+1/D}\leqslant C'\left\{\mathscr{E}_*(f,f)+\frac{1}{N}\|f\|_2^2\right\}\|f\|_1^{1/D}$$

with $C'=(1+1/2D)[((1+2D)^{1/2}C]^{1/D}$.

Proof. Since K is self-adjoint, the hypothesis and duality imply that

$$\|K^n\|_{1\to 2}\leqslant C(n+1)^{-D} \qquad \text{for} \quad 0\leqslant n\leqslant N$$

Hence we have, for $0\leqslant n\leqslant N$ and any $f\not\equiv 0$,

$$\begin{aligned}\|f\|_2^2&=\sum_{i=0}^{n-1}(\|K^if\|_2^2+\|K^{i+1}f\|_2^2)+\|K^nf\|_2^2\\&=\sum_{i=0}^{n-1}\|(I-K^2)^{1/2}K^if\|_2^2+\|K^nf\|_2^2\\&\leqslant n\mathscr{E}_*(f,f)+C^2(n+1)^{-2D}\|f\|_1^2\end{aligned}$$

Here, $(I-K^2)^{1/2}$ is the symmetric square root given by spectral theory. This gives

$$\|f\|_2^2\leqslant n\left\{\mathscr{E}_*(f,f)+\frac{1}{N}\|f\|_2^2\right\}+C^2(n+1)^{-2D}\|f\|_1^2 \qquad \text{for any integer } n$$

Now choose n as the integer with

$$\begin{aligned}\left(\frac{2DC^2\|f\|_1^2}{\mathscr{E}_*(f,f)+N^{-1}\|f\|_2^2}\right)^{1/(1+2D)}-1&<n\\&\leqslant\left(\frac{2DC^2\|f\|_1^2}{\mathscr{E}_*(f,f)+N^{-1}\|f\|_2^2}\right)^{1/(1+2D)}\end{aligned}$$

It follows that

$$\|f\|_2^2 \leqslant ((2D)^{-2D/(1+2D)} + (2D)^{1/(1+2D)}) \\ \times \left\{ \mathscr{E}_*(f,f) + \frac{1}{N}\|f\|_2^2 \right\}^{2D/(1+2D)} (C\|f\|_1)^{2/(1+2D)}$$

Raising this to the power $(1+2D)/2D$ gives the stated conclusion. □

Remark 3.3. A little calculus shows that the constant $C' = (1 + 1/2D)[(1+2D)^{1/2}C]^{1/D}$ in Theorem 3.2 can be bounded by $2^{1+1/2D}C^{1/D}$.

2. Theorem 3.2 really needs reversibility. Example 2d gives a chain K that has nontrivial decay $D(n) \leqslant C(n+1)^{-D}$ for $0 \leqslant n \leqslant N$, with some constants C, D, N. However, the multiplicative symmetrization $K^*K = Q$ does not decay at all. This and Corollary 3.1 show that $\mathscr{E}_*$ does not satisfy a useful Nash inequality.

3. Theorem 3.2, Theorem 3.1 and the inequality $\mathscr{E}_* \leqslant 2\mathscr{E}$ show that, if $Q = K^*K$ satisfies $\|Q^n\|_{2\to\infty} \leqslant C(n+1)^{-D}$ for $0 \leqslant n \leqslant N$, then $\|K^n\|_{2\to\infty} \leqslant C'(n+1)^{-D}$ for $0 \leqslant n \leqslant N$ with $C' = B(D)C$ where $B(D)$ depends only on D.

We now briefly discuss the use of Nash inequalities in continuous time. The links between decay rates and Nash inequalities for continuous time semigroups have been actively studied. Nash[29] introduced the inequalities named after him to study divergence form, uniformly elliptic, second order differential operators with measurable coefficients in $\mathbb{R}^d$. He showed how a Nash inequality implies a polynomial decay in time of the corresponding heat kernel. We refer the reader to Nash[29] and to the more recent works of Varopoulos and his students[4,5,35,36] and of Carlen *et al.*[2] Fabes[19] gives a nice survey of the use of Nash inequality in PDE. Chapter 2 of Ref. 37 contains information about this and related techniques. The proofs of the following results are similar to, although easier than, those earlier.

Theorem 3.3. Let K be an irreducible Markov kernel on a finite set $\mathcal{X}$ with invariant probability measure π. Let $H_t = e^{-t(I-K)}$ be the corresponding semigroup. With the notation as in Eqs. (1.2)–(1.6), (2.1)–(2.7), assume that the Nash inequality

$$\|f\|_2^{2+1/D} \leqslant C\left\{ \mathscr{E}(f,f) + \frac{1}{T}\|f\|_2^2 \right\} \|f\|_1^{1/D}$$

holds for some constants C, D, $T > 0$ and all functions f. Then,

$$\|H_t\|_{2\to\infty} \leqslant e(DC/t)^D \qquad \text{for} \quad 0 \leqslant t \leqslant T \tag{3.8}$$

Moreover,

$$2\,\|H_t(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}} \leqslant \|(H_t(x,\cdot)/\pi(\cdot))-1\|_2 \leqslant e^{1-c} \tag{3.9}$$

for

$$t \geqslant T+\frac{1}{\lambda}\left(D\log\left(\frac{DC}{T}\right)+c\right) \qquad \text{with} \quad c>0$$

Conversly, for self-adjoint K, if there exist positive constants $C, D, T>0$ such that

$$\|H_t\|_{2\to\infty} \leqslant Ct^{-D} \qquad \text{for} \quad 0\leqslant t\leqslant T$$

then, for all functions f,

$$\|f\|_2^{2+1/D} \leqslant C'\left\{\mathscr{E}(f,f)+\frac{1}{2T}\|f\|_2^2\right\}\|f\|_1^{1/D}$$

with $C'=2(1+1/(2D))((1+2D)^{1/2}C)^{1/D} \leqslant 2^{2+1/(2D)}C^{1/D}$.

We close this section with some comments about Nash inequalities and how they compare to Sobolev inequalities. In terms of Dirichlet forms, a Sobolev inequality is an inequality of the type

$$\|f\|_q^2 \leqslant C_S\left\{\mathscr{E}(f,f)+\frac{1}{T}\|f\|_2^2\right\} \tag{3.10}$$

where $q>2$, and C_S, $T>0$. These functional inequalities were introduced by Sobolev in the thirties and have played an important role in analysis, PDE and geometry ever since. For instance, they provide the basic compactness properties for the study of solutions of elliptic PDE. Varopoulos[36] (see also Ref. 37, Chs. 6 and 7) introduced these Nash-Sobolev techniques in the study of random walk on finitely generated groups where they proved to be very effective.

If we set $q=4D/(2D-1)$ with $D>1/2$, and use Hölder's inequality

$$\|f\|_2^{2+1/D} \leqslant \|f\|_{4D/(2D-1)}^2\,\|f\|_1^{1/D}$$

we see that the Sobolev inequality in Eq. (3.10) implies the Nash inequality

$$\|f\|_2^{2+1/D} \leqslant C_N\left\{\mathscr{E}(f,f)+\frac{1}{T}\|f\|_2^2\right\}\|f\|_1^{1/D} \tag{3.11}$$

with $C_N=C_S$. Lest out notation anger a classical analyst, we hasten to add that, classically, $4D=d$, the natural dimension. The value of the parameter

q in terms of $d>2$ is then $q=2d/(d-2)$. We stick to D to match notation with the rest of this paper.

Less obvious is the converse, that Eq. (3.11) implies Eq. (3.10) with $C_S \leqslant ADC_N$ for a universal constant A. One way to see this is to use the equivalence of each of these two inequalities with the polynomial decay of the corresponding semigroup. See e.g., Ref. 37, Ch. 2. A more direct proof in a very general setting is given in Ref. 1. We have chosen to work with Nash inequalities because they appear to be more convenient in the present setting.

4. RANDOM WALK IN A BOX

This section analyzes an example of the Metropolis algorithm on the lattice points inside a box of side n in d-dimensions. This should be an easy problem but it is not (at present). In outline, the argument proceeds by comparison with an auxiliary product chain. A detailed eigenanalysis leads to sharp decay rates for the product chain. This implies a Nash inequality for this auxiliary chain and, by comparison, a Nash inequality for the original chain. This finally gives decay rates for the original chain.

4.1. The Metropolis Algorithm in a Box

Let $C(n, d)$ 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 method for changing the transition probabilities to have a given stationary distribution. In this section, we analyse the Metropolis algorithm for the uniform stationary distribution. 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 ± 1 in a single coordinate, in which case $P(x, y)$ is given by:

$$P(x, y)=\begin{cases} 1/\delta(x) & \text{if } \delta(x) \geqslant \delta(y) \text{ and } x \neq y \\ 1/\delta(y) & \text{if } \delta(x)<\delta(y) \\ (1/\delta(x)) & \sum_{\substack{z:\, \delta(x)<\delta(z) \\ z \sim x}} (1-(\delta(x)/\delta(z))) \text{ if } x=y \end{cases} \tag{4.1}$$

Here $z \sim x$ stands for z, x neighbors on the grid. As an example, in 2-dimensions, the walk becomes the weighted nearest neighbor walk on $C(n, 2)$ with loops added and weighted as in Fig. 2).

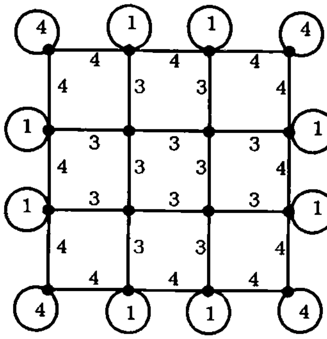

Fig. 2. In two-dimensions, the walk becomes the weighted nearest neighbor walk on $C(n, 2)$ with loops added and weighted.

For general n and d, P is a reversible, aperiodic, irreducible Markov chain on $C(n, d)$ with uniform stationary distribution

$$\pi(x) = \frac{1}{(n+1)^d} \tag{4.2}$$

The analysis to follow gives

Theorem 4.1. Let $d \geqslant 2$. For P and π defined at Eqs. (4.1) and (4.2),

$$\|P_x^k - \pi\|_{\mathrm{TV}} \leqslant a_1 e^{-a_2 c} \qquad \text{for} \quad k = n^2 d(d \log d + c), \quad c > 0$$

where $a_1, a_2 > 0$ are universal constants. For $k = cn^2 d \log d$,

$$2 \sup_x \|P_x^k - \pi\|_{\mathrm{TV}} \geqslant a_3 e^{-a_4 c} \qquad \text{for} \quad c > 0$$

with $a_3, a_4 > 0$ universal constants.

4.2. Analysis for the Comparison Chain

We build a product chain with known eigenvalues and eigenvectors. Consider the Markov kernel W on $\{0, 1, ..., n\}$ defined by

$$\begin{aligned} W(x, x+1) &= 1/2 \qquad \text{for} \quad x \in \{0, ..., n-1\} \\ W(x, x-1) &= 1/2 \qquad \text{for} \quad x \in \{1, ..., n\} \\ W(0, 0) &= W(n, n) = 1/2 \end{aligned} \tag{4.3}$$

This is a symmetric kernel with uniform stationary distribution. Feller,[20] (p. 436) gives its eigenvalues and eigenfunctions as

$$\beta_0 = 1, \qquad \psi_0(x) \equiv 1$$

$$\beta_j = \cos\frac{\pi j}{n+1}, \qquad \psi_j(x) = \sqrt{2}\cos(\pi j(x-1/2)/(n+1)) \qquad \text{for} \quad j = 1,..., n$$

We need an analysis of this process in continuous time. Let $V_t = e^{-t(I-W)}$ and write

$$\begin{aligned} V_t(x, y) &= \frac{1}{n+1}\left(1 + \sum_{j=1}^{n} \psi_j(x)\,\psi_j(y)\, e^{-t(1-\cos(\pi j/(n+1)))}\right) \\ &\leqslant \frac{1}{n+1}\left(1 + 2\sum_{j=1}^{n} e^{-2tj^2/(n+1)^2}\right) \\ &\leqslant \frac{1}{n+1}\left(1 + 2e^{-2t/(n+1)^2}(1 + \sqrt{(n+1)^2/2t})\right) \end{aligned}$$

To obtain the last inequality, use

$$\sum_{2}^{n} e^{-2tj^2/(n+1)^2} \leqslant \int_1^\infty e^{-2ts^2/(n+1)^2}\, ds = \frac{n+1}{\sqrt{2t}} \int_{\sqrt{2t}/n+1}^\infty e^{-u^2}\, du$$

and

$$\frac{2}{\sqrt{\pi}} \int_z^\infty e^{-u^2}\, du = \frac{2e^{-z^2}}{\sqrt{\pi}} \int_z^\infty e^{-(u-z)^2 - 2(u-z)z}\, du \leqslant e^{-z^2}$$

Next, consider the kernel $\tilde{P}$ on $C(n, d)$ which proceed by choosing one of the d coordinates at random and changing that coordinate using W from Eq. (4.3). Thus,

$$\tilde{P} = \frac{1}{d}\sum_{i=1}^{d} I \otimes \ldots \otimes I \otimes \underset{i-1}{W \otimes I} \ldots \otimes I \tag{4.4}$$

In fact, the transition kernel Eq. (4.4) is exactly given by Eq. (1.1) for this case. As an example, when $d = 2$, the walk becomes the weighted nearest neighbor walk on $C(n, 2)$ with loops and weighted as in Fig. 3.

For general n, d, $\tilde{P}$ is a reversible Markov chain with uniform distribution $\tilde{\pi} = \pi$ on $C(n, d)$ as stationary distribution. The eigenvalues of such a

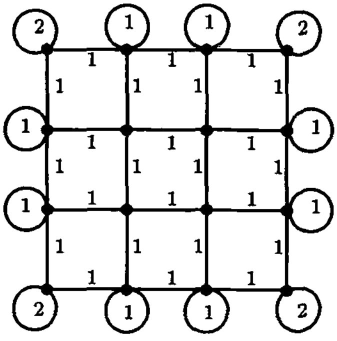

Fig. 3. When $d=2$, the walk becomes the weighted nearest neighbor on $C(n, 2)$ with loops.

product chain are easy to derive in terms of the eigenvalues of W (see Ref. 10, Sec. 6). Here

$$\tilde{\beta}_1 = 1 - \frac{1}{d}\left(1 - \cos\frac{\pi}{n+1}\right) \leqslant 1 - \frac{2}{d(n+1)^2}$$

The semigroup $\tilde{H}_t = e^{-t(I-\tilde{P})}$ has the property that the coordinates evolve independently of each other, each according to V_t defined earlier. It follows that

$$\begin{aligned}\|\tilde{H}_t\|_{2\to\infty}^2 &= \max_{x,y}\{(n+1)^d V_{2t/d}(x_1, y_1) \ldots V_{2t/d}(x_d, y_d)\}\\ &\leqslant (1 + 2e^{-4t/d(n+1)^2}(1 + \sqrt{d(n+1)^2/4t}))^d\\ &\leqslant (4d(n+1)^2/t)^{d/2} \qquad \text{for} \quad t \leqslant d(n+1)^2/16\end{aligned}$$

For this last inequality, let $f(u) = u + 2ue^{-u^2} + 2e^{-u^2}$ and check that $f(u) \leqslant 4$ for $0 \leqslant u \leqslant 1/2$.

From this, the converse in Theorem 3.3 yields the Nash inequality

$$\|f\|_2^{2+4/d} \leqslant 64d(n+1)^2\left(\tilde{\mathscr{E}}(f, f) + \frac{8}{d(n+1)^2}\|f\|_2^2\right)\|f\|_1^{4/d} \tag{4.5}$$

From Eq. (4.5), we have a Nash inequality for the Dirichlet form associated to $\tilde{P}$. Corollary 3.1 gives decay rates for the discrete time chain $\tilde{P}^k$. Using these and the second eigenvalue of $\tilde{P}^2$ in Lemma 1.1 implies that the upper bound of Theorem 4.1 holds with P replaced by $\tilde{P}$. However, for $\tilde{P}$, it can shown that order $n^2 d \log d$ steps suffice to reach approximate equilibrium by a direct elementary comparison with the continuous time chain. It is not hard to prove the lower bound of Theorem 4.1 for $\tilde{P}$ or P.

Remark 4.1. 1. The walk $\tilde{P}$ defined at Eq. (4.4) is exactly of the form defined at Eq. (1.1). The Euclidean diameter of the convex hull of $C(n, d)$ is $\gamma_{n,d} = n\sqrt{d}$. The argument outlined before shows that order $\gamma^2_{n,d} \log d$ steps are necessary and suffice for $\tilde{P}$ to reach equilibrium. For the walk P, Theorem 4.1 only says that order $\gamma^2_{n,d} \log d$ steps are necessary and that order $d\gamma^2_{n,d} \log d$ suffice. This last result can be improved to order "$\gamma^2_{n,d} \log d$ steps suffice" by using the present technique **and** logarithmic Sobolev inequalities; see Ref. 13 for details. Direct use of geomtric techniques lead to far cruder bounds when d is large. See Example 5b.

2. Bounding rates of convergence of $\tilde{P}$ required an excursion in continuous time. This is also true for the product walks in Section 6 of Ref. 5.

4.3. Proof of Theorem 4.1

We argue by comparing P and $\tilde{P}$. Observe that $\pi = \tilde{\pi}$. For $x, y \in C(n, d)$, and $x \neq y$,

$$\tilde{P}(x, y) \leqslant P(x, y)$$

Thus, $\tilde{\mathscr{E}} \leqslant \mathscr{E}$. This implies $\beta_1 \leqslant \tilde{\beta}_1$. For the lowest eigenvalue, $\beta_{|X|-1}$, use of Corollary 2, p. 41 of Ref. 16, gives

$$\beta_{|X|-1} \geqslant -1 + \frac{1}{dn^2}$$

To see this, use the loops at the boundary of $C(n, d)$ and paths along the first coordinate direction. All of this shows

$$\mu(1) \leqslant 1 - \frac{1}{d(n+1)^2} \tag{4.6}$$

Further, the Nash inequality Eq. (4.5) implies

$$\|f\|_2^{2+4/d} \leqslant 64d(n+1)^2 \left(\mathscr{E}(f, f) + \frac{8}{d(n+1)^2} \|f\|_2^2 \right) \|f\|_1^{4/d} \tag{4.7}$$

Now, Corollary 3.1 gives

$$\|P^k\|_{2\to\infty} \leqslant \sqrt{2}\,[2CB/(k+1)]^{d/4}, \qquad \text{for} \quad 0 \leqslant k \leqslant d(n+1)^2/4$$

with $C = 64d(n+1)^2$ and $B = (1 + 4/d(n+1)^2)(1+d)$. Thus,

$$\|P^k\|_{2\to\infty} \leqslant \sqrt{2}\,[256(1+d)]^{d/4}, \qquad \text{for} \quad k = d(n+1)^2/4 \tag{4.8}$$

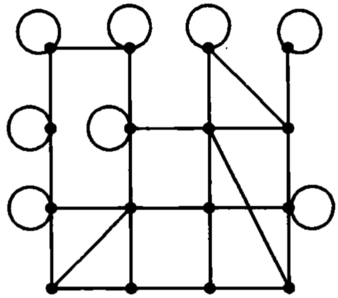

Fig. 4. An example when $d=2$, $n=3$, and $M=3$.

Finally, using Eqs. (4.6) and (4.8) in Lemma 1.1 gives the upper bound in Theorem 4.1. The proof of the lower bound is straightforward and omitted.

The techniques used to prove Theorm 4.1 give results for a variety of graph structures on $C(n, d)$ obtained by adding and subtracting edges. For example, diagonally adjacent points could be connected. To be specific, fix $M \geqslant 1$ and consider adding and erasing edges according to the following rules:

- When erasing edges, for each basic unit cube at most one edge is erased, and for each edge left in, there are at most M cubes containing this edge and an erased edge.
- When adding edges, the degree of any vertex $x \in C(n, d)$ stays bounded by Md.

An example with $d=2$, $n=3$, and $M=3$ is shown in Fig. 4.
This defines a new Markov chain $\hat{P}$ on $C(n, d)$, with stationary distribution $\hat{\pi}$ proportional to the degree of each vertex. To avoid parity problems, we work in continuous time with $\hat{H}_t = e^{-t(I-\hat{P})}$.

Corollary 4.1. For each fixed $M \geqslant 1$, any Markov chain $\hat{H}_t$ as previously satisfies

$$\|\hat{H}_t - \hat{\pi}\|_{\mathrm{TV}} \leqslant a_1 e^{-a_2 c} \qquad \text{for} \quad t \geqslant dn^2(d \log d + c), \qquad c > 0$$

where a_1, a_2 do not depend on d or n.

5. GEOMETRIC THEORY

5.1. Introduction

The section introduces volume growth conditions and path arguments as a way of proving Nash inequalities and bounding rates of convergence to stationarity. Let $K(x, y)$ be a Markov chain on a finite set X. Assume that K is irreducible with stationary distribution π.

The geometric arguments developed in this section use an underlying graph structure on the finite set X which is given by a set $\mathscr{A}$ of oriented edges. This graph structure must be **compatible** with the chain K. A precise general definition will be given in Section 5.2.

In this introduction, we describe only the simplest and most useful way to associate a graph structure with K. Namely, let $\mathscr{A}_K$ be the set of pairs (x, y) such that either $K(x, y)>0$ or $K(y, x)>0$. This defines a symmetric graph structure on X. Let $d(x, y)$ be the shortest path distance corresponding to the edge-set $\mathscr{A}_K$ and let $B(x, r)=\{z: d(x, z) \leqslant r\}$ be the closed ball around x with radius r. We define the volume of $B(x, r)$ by setting $V(x, r)=\sum_{z \in B(x, r)} \pi(z)$.

Definition 5.1. For $A, d \geqslant 1$, the finite Markov chain (K, π) has (A, d)**-moderate growth** if

$$V(x, r) \geqslant \frac{1}{A}\left(\frac{r+1}{\gamma}\right)^d \qquad \text{for all } x \in X \text{ and integers } r \in \{0, 1, \ldots, \gamma\} \tag{5.1}$$

where γ is the diameter of the graph $(X, \mathscr{A}_K)$.

Moderate growth is a variation of polynomial growth which has proved effective in studying random walk on groups. A thorough exposition and many examples appear in Ref. 11. Several other example are given in Section 5.2.

The second geometric notion needed is that of local Poincaré inequalities. For any real function f and integer r, set

$$f_r(x)=\frac{1}{V(x, r)} \sum_{y \in B(x, r)} f(y)\, \pi(y)$$

Definition 5.2. Let K, π be a Markov chain on the finite set X. Let $\mathscr{E}$ be the Dirichlet form of Eq. (2.7) associated with K. We say that (K, π) satisfies a **local Poincaré inequality** if there exists $a>0$ such that, for any real function f and integer r,

$$\|f-f_r\|_2^2 \leqslant a r^2 \mathscr{E}(f, f) \tag{5.2}$$

As motivation, when $r=\gamma$ and K, π is reversible, the bound becomes

$$\|f-f_\gamma\|_2^2=\operatorname{Var}_\pi(f) \leqslant a \gamma^2 \mathscr{E}(f, f)$$

and the minimax characterization of the second largest eigenvalue gives $\beta_1 \leqslant 1 - 1/(a\gamma^2)$. Even when K, π is not reversible, Eq. (5.2) yields the inequality $\lambda \geqslant 1/(a\gamma^2)$ for the quantity λ defined at Eq. (2.8). Thus, specializing a **local** Poincaré inequality to the case $r = \gamma$ (= the diameter) yields a Poincaré inequality as considered in Ref. 16. Such Poincaré inequalities have proved classically useful. In Ref. 16, geometric path techniques were shown to yield Poincaré inequalities. Section 5.2 shows how paths yield useful **local** Poincaré inequalities. For instance, it will be shown that random walk on any finite group satisfies Eq. (5.2).

The following result will be proved in greater generality in Section 5.3. It shows that, for chains with moderate growth satisfying a local Poincaré inequality, order γ^2 steps are necessary and sufficient to guarantee converge. For simplicity, we state the result in continuous time. Several discrete analogs are given in Section 5.3.

Theorem 5.1. Let K, π be a Markov chain of a finite set X. Assume that (K, π) has moderate growth (5.1) and satisfies a local Poincaré inequality of Eq. (5.2). Then, the continuous time semigroup $H_t = e^{-t(I-K)}$ defined at Eq. (2.6) satisfies, for all $t > 0$,

$$2 \|H_t^x - \pi\|_{TV} \leqslant a_1 e^{-t/(a\gamma^2)}$$

with $a_1 = (e^5(1+d)A)^{1/2}(d/4)^{d/4}$.

Conversely, if K, π is reversible, there are constants $a_2, a_3 > 0$ depending only on A, a, d from Eqs. (5.1) and (5.2) such that

$$\sup_x \|H_t^x - \pi\|_{TV} \geqslant a_2 e^{-a_3 t/\gamma^2} \qquad \text{for} \quad t > 0$$

Remark 5.1. Of course, any irreducible chain on a finite set satisfies Eqs. (5.1) and (5.2) for some A, a, d. Unfortunately, the constants a_i, $i = 1, 2, 3$, depend exponentialy on these parameters, especially on d. The bounds work well for chains with small values of A, a, d. As shown later, there are many natural families of graphs with bounded parameters.

Section 5.2 introduces path techniques and uses these to treat simple examples. It also treats random walk on groups, showing how Theorem 5.1 implies the main result of Ref. 11.

Section 5.3 shows that moderate growth and local Poincaré inequalities imply a Nash inequality. It contains the proof of Theorem 5.1 and develops its discrete time analogs.

5.2. Path Techniques and Examples

As stated in Section 5.1 introduction, the geometric techniques to be develop later use an underlying graph structure. Of course, this graph structure must be related to the chain K we want to analyze. However, there is some freedom in the way the graph and the chain have to be related. We think that it is useful and instructive to present the argument in some generality.

The first notion we need to introduce is the notion of **compatibility** between a graph $(X, \mathscr{A})$, symmetric or not, and a Dirichlet form $\mathscr{E}_Q$ associated with a nonnegative kernel Q by

$$\mathscr{E}_Q(f, f) = \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 \, Q(x, y)$$

Definition 5.3. Let $\mathscr{E}_Q$ be the Dirichlet form on X associated with a nonnegative kernel Q. Let $\mathscr{A} \subset X \times X$ be a set of oriented edges. We say that Q (or $\mathscr{E}_Q$) and $\mathscr{A}$ are compatible if $Q(x, y) > 0$ for all $(x, y) \in \mathscr{A}$.

In most applications, we are given a chain K, π on X and we construct the edge-set $\mathscr{A}$ from K. There are several interesting possible choices for doing that.

- The most obvious choice is to define a graph with vertex set X and an edge from x to y is $K(x, y) > 0$. This graph may well not be symmetric if K, π is not reversible. This graph is compatible with $Q(x, y) = K(x, y) \, \pi(x)$ which corresponds to the Dirichlet from $\mathscr{E}(f, f) = \langle (I - K) f, f \rangle$.
- Another possible choice is to put an edge from x to y if either $K(x, y) > 0$ or $K(y, x) > 0$. This corresponds to the edge-set $\mathscr{A}_K$ introduced in Section 5.1. This graph is symmetric by construction. It is compatible with

$$Q(x, y) = \tfrac{1}{2}(K(x, y) \, \pi(x) + K(y, x) \, \pi(y))$$

  which corresponds to the same Dirichlet form $\mathscr{E}(f, f) = \langle (I - K) f, f \rangle$ as before. This is the construction that we will use in most applications.
- We can also build the edge set $\mathscr{A}$ as the set of pairs (x, y) such that

$$Q(x, y) = \sum_z K(z, x) \, K(z, y) \, \pi(z) > 0$$

  This always gives a symmetric graph that corresponds to the Dirichlet form $\mathscr{E}_*$.
- Finally, in some cases, we may want to use only a part of the obvious edges previously considered. For instance, we may want to fix $\varepsilon > 0$ and define $\mathscr{A}$ as the set of pairs (x, y) such that $K(x, y) \, \pi(x) \geqslant \varepsilon$. Indeed, using pairs (x, y) with very small weight $K(x, y) \, \pi(x)$ produces bad factors in the bounds developed later.

Any fixed oriented graph structure $(X, \mathscr{A})$ defines a shortest path distance $d(x, y)$ between x and y. This distance need not be symmetric if $\mathscr{A}$ is not symmetric. Let $B(x, r) = \{y: d(x, y) \leqslant r\}$ be the closed ball around x with radius r. By definition, this ball has volume

$$V(x, r) = \sum_{y \in B(x,r)} \pi(y) \qquad (5.3)$$

where π is a given fixed probability measure on X. We also set

$$f_r(x) = \frac{1}{V(x, r)} \sum_{y \in B(x,r)} f(y)\, \pi(y) \qquad (5.4)$$

Let γ be the diameter of the graph, so $B(x, \gamma) = X$ for any $x \in X$. We now extend to the present general setting the notion of moderate growth and local Poincaré inequality introduced in Section 5.1.

Definition 5.4. Fix an edge set $\mathscr{A}$ and a probability measure π on X. For A, $d \geqslant 1$, $(X, \mathscr{A}, \pi)$ has (A, d)-**moderate growth** if

$$V(x, r) \geqslant \frac{1}{A}\left(\frac{r+1}{\gamma}\right)^d \qquad \text{for all } x \in X \text{ and integers } r \in \{0, 1,..., \gamma\} \qquad (5.5)$$

Definition 5.5. Fix an edge set $\mathscr{A}$ and a probability measure π on X. Let $\mathscr{E}_Q$ be a Dirichlet form compatible with $\mathscr{A}$. We say that $(\mathscr{A}, \pi, \mathscr{E}_Q)$ satisfies a **local Poincaré inequality** if there exists $a > 0$ such that, for any real function f and integer r,

$$\|f - f_r\|_2^2 \leqslant a r^2 \mathscr{E}_Q(f, f) \qquad (5.6)$$

To relate these definitions to the ones given in Section 5.1, observe that we say that a Markov chain (K, π) has **moderate growth** as in Eq. (5.1) or that (K, π) satisfies a **local Poincaré inequality** as in Eq. (5.2) if, respectively, Eqs. (5.5) or (5.6) are satisfied by $(\mathscr{A}_K, \pi, \mathscr{E}_Q)$ where $Q(x, y) = \frac{1}{2}(K(x, y)\, \pi(x) + K(y, x)\, \pi(y))$, $\mathscr{E}_Q(f, f) = \mathscr{E}(f, f) = \langle (I - K) f, f \rangle$ and $\mathscr{A}_K = \{(x, y): Q(x, y) > 0\}$.

We can now give a first result concerning local Poincaré inequalities. Throughout, if $e = (z, w)$ is an edge of the graph, we set $e_+ = w$ and $e_- = z$ and $Q(e) = Q(z, w)$. We assume that $(X, \mathscr{A})$ is a connected graph and let γ be its diameter.

For each pair of points $x, y \in X$, choose a path $\gamma_{x,y}$ joining x to y in $(X, \mathscr{A})$. Usually, these paths are geodesics but this is not necessary. The following result is a local version of results in Ref. 16. The quality of the bounds given next depends on the choice of paths.

Lemma 5.1. Let $\mathcal{X}$, $\mathcal{A}$, π be a finite graph equipped with a probability measure π. Let $\mathcal{E}_Q$ be a Dirichlet form and assume that $\mathcal{A}$ and Q are compatible. Let $\gamma_{x,y}$ be a path from x to y in the underlying graph $(\mathcal{X}, \mathcal{A})$. Then, with notation as in Eqs. (5.3) and (5.4), for each integer $0 \leqslant r \leqslant \gamma$,

$$\|f-f_r\|_2^2 \leqslant \eta(r)\,\mathcal{E}_Q(f,f)$$

with

$$\eta(r) = \max_{e \in \mathcal{A}} \left\{ \frac{2}{Q(e)} \sum_{\substack{\gamma_{x,y} \ni e \\ d(x,y) \leqslant r}} |\gamma_{x,y}| \frac{\pi(x)\,\pi(y)}{V(x,r)} \right\} \tag{5.7}$$

Proof. For any $x \in \mathcal{X}$,

$$\begin{aligned}
|f(x)-f_r(x)|^2 &\leqslant \frac{1}{V(x,r)} \sum_{y \in B(x,r)} |f(x)-f(y)|^2\,\pi(y) \\
&= \frac{1}{V(x,r)} \sum_{y \in B(x,r)} \left| \sum_{e \in \gamma_{x,y}} f(e_+)-f(e_-) \right|^2 \pi(y) \\
&\leqslant \frac{1}{V(x,r)} \sum_{y \in B(x,r)} |\gamma_{x,y}|\,\pi(y) \sum_{e \in \gamma_{x,y}} |f(e_+)-f(e_-)|^2
\end{aligned}$$

Multiply both sides by $\pi(x)$ and sum in x. Bring the sum over directed edges e outside to get

$$\|f-f_r\|_2^2 \leqslant \sum_e |f(e_+)-f(e_-)|^2 \frac{Q(e)}{Q(e)} \sum_{\substack{\gamma_{x,y} \ni e \\ d(x,y) \leqslant r}} |\gamma_{x,y}| \frac{\pi(x)\,\pi(y)}{V(x,r)} \leqslant \eta(r)\,\mathcal{E}_Q(f,f)$$

We proceed to some examples. In each case, we use path techniques to show that local Poincaré inequalities are satisfied. We also determine moderate growth so that Theorem 5.1 is in force and shows that order γ^2 steps are necessary and suffice to achieve randomness.

Example 5a. The n-point path. Take $\mathcal{X} = \{1,..., n\}$, $K(x, x+1) = 1/2$ for $1 \leqslant x \leqslant n-1$, $K(x, x-1) = 1/2$ for $2 \leqslant x \leqslant n$, $K(1,1) = K(n,n) = 1/2$. This is nearest neighbor random walk on a path with holding probabilities on the ends. We use the obvious underlying graph. The chain has stationary distribution $\pi(x) = 1/n$ and is reversible, aperiodic, and irreducible. Take $\gamma_{x,y}$ as the unique geodesic path from x to y. Then $\gamma = n-1$. Further, moderate growth with $A = n/(n-1)$, $d = 1$, follows from

$$\frac{r+1}{n} \leqslant V(x,r) \leqslant \frac{2r+1}{n}$$

for any x and $r \leqslant \gamma$.

Finally, for any edge e, $Q(e)=1/(2n)$. Using these ingredients, the quantity $\eta(r)$ of Lemma 5.1 satisfies

$$\eta(r)\leqslant 4n\frac{r}{n^2}\frac{n}{r+1}N(r), \qquad N(r)=\max_{e\in\mathscr{A}}|\{\gamma_{x,y}\ni e\colon |\gamma_{x,y}|\leqslant r\}$$

Clearly, $N(r)\leqslant r(r+1)/2$. Combining bounds, $\eta(r)\leqslant 2r^2$, so

$$\|f-f_r\|_2^2\leqslant 2r^2\mathscr{E}(f,f)$$

Thus, Theorem 5.1 yields $2\,\|H_t^x-\pi\|_{\mathrm{TV}}\leqslant a_1e^{-a_2t/n^2}$ for $t>0$ with explicit constants a_1, a_2.

Example 5b. Random walk on a box. It is instructive to see how these bounds work out for the example of the discrete box studied in Section 4. In summary, they work well in bounded dimension but are exponentially off in high dimensions. Again, we use the obvious underlying graph.

With notation as in Section 4, consider the chain in Eq. (4.1) on the box $C(n,d)$. Paths between $x, y\in C(n,d)$ are chosen inductively (in d) as follows. For $d=2$, there is a unique shortest path making at most one 90 degrees counterclockwise turn. For $d=3$, given, $x, y\in C(n,d)$, project y to the plane $\{(z_1,z_2,z_3)\colon z_3=x_3\}$. Connect x to this projected point using the two-dimensional paths and then connect the third coodinates. Continuing inductively defines paths from x to y in any dimension. The resulting paths have at most $d-1$ "turns." Note that the diameter is $\gamma=nd$.

Let e be a directed edge in $C(n,d)$. We will bound the number of paths of length at most r that use that edge. For any path $\gamma_{x,y}$ using this edge, suppose the edge appears after turn $i-1$ and before turn i, $i\in\{1,\dots,d\}$. There at most r^i starting points x and at most $(2r)^{d+1-i}$ ending points y. Hence, bounding $|\gamma_{x,y}|$ by $|\gamma_{x,y}|\leqslant r$ yields

$$\sum_{\substack{\gamma_{x,y}\ni e\\ |\gamma_{x,y}|\leqslant r}}|\gamma_{x,y}|\leqslant(2r)^{d+2}$$

The quantity $V(x,r)$ can be bounded between

$$\frac{1}{d^d d!}\left(\frac{r+1}{n+1}\right)^d\leqslant V(x,r)\leqslant(2d)^d\left(\frac{r+1}{n+1}\right)^d \qquad \text{for}\quad 0\leqslant r\leqslant\gamma \tag{5.8}$$

The upper bound is needed for x in the center of the box and the lower bound for x in a corner.

Using these ingredients, $\eta(r)$ of Eq. (5.7) can be bounded by

$$\eta(r) \leqslant \frac{d(n+1)^d}{(n+1)^{2d}} d^d d! \left(\frac{n+1}{r+1}\right)^d \sum_{\substack{\gamma_{x,y} \ni e \\ |\gamma_{x,y}| \leqslant r}} |\gamma_{x,y}| \leqslant (2d)^{d+1} d! r^2$$

This gives a local Poincaré inequality. Moderate growth follows from Eq. (5.8). These ingredients and Theorem 5.1 show that, at fixed d, order γ^2 steps are necessary and suffice to achieve randomness. Evidently, the result becomes useless for large d. Section 4 gives much sharper results for this case.

For some Markov chains, there are many geodesic paths and it is natural and effective to average over paths. See Refs. 9, 16, 21, and 33 for examples. The bounds above fit well with such averages. We proceed to details.

Let $L(x, y)$ be the set of all geodesic paths connecting x to y. Set

$$L_r(x) = \bigcup_{y \in B(x,r)} L(x, y), \qquad L_r = \bigcup_x L_r(x), \qquad L = \bigcup_{x, y \in X} L(x, y)$$

For $\ell \in L$, let $|\ell|$ denote its length. A function $\omega: L \to [0, 1]$ is a **flow** if

$$\sum_{\ell \in L(x,y)} \omega(\ell) = \pi(x)\, \pi(y)$$

Lemma 5.1 used a trivial flow $\omega(\gamma_{x,y}) = \pi(x)\,\pi(y)$, $\omega(\ell) = 0$ for other paths in $L(x, y)$. The argument for Lemma 5.1 gives

Lemma 5.2. Let the edge-set $\mathscr{A}$ and the Dirichlet form $\mathscr{E}_Q$ with kernel Q be compatible. Fix a flow ω. Then, for all integers r and all real functions f,

$$\|f - f_r\|_2^2 \leqslant \eta_\omega(r)\, \mathscr{E}_Q(f, f)$$

where

$$\eta_\omega(r) = \max_{e \in \mathscr{A}} \left\{ \frac{2}{Q(e)} \sum_{x \in X} \frac{1}{V(x, r)} \sum_{\substack{\ell \in L_r(x) \\ \ell \ni e}} |\ell|\, \omega(\ell) \right\}$$

The following corollary makes use of a nontrivial flow to derive local Poincaré inequalities for graphs with symmetry. It is used in Example 5d. Say that a one to one map $\Phi: X \to X$ preserves Q and π if $Q(x, y) = Q(\Phi(x), \Phi(y))$ for all $(x, y) \in \mathscr{A}$, and $\pi(x) = \pi(\Phi(x))$ for all $x \in X$. Note that we consider only the pairs (x, y) in $\mathscr{A}$.

Corollary 5.1. Let the edge $\mathscr{A}$ and the Dirichlet form $\mathscr{E}_Q$ with kernel Q be compatible. Let Γ be a group of permutations of X which preserves Q and π. Suppose that the set of oriented edges $\mathscr{A}$ partitions as

$$\mathscr{A} = \bigcup_1^\sigma \mathscr{A}_i$$

with Γ operating transitively on each $\mathscr{A}_i$. Then, for each integer $r \geqslant 0$,

$$\|f - f_r\|_2^2 \leqslant \max_{1 \leqslant i \leqslant \sigma} \left\{ \frac{2r^2}{|\mathscr{A}_i| \; Q(e_i)} \right\} \mathscr{E}_Q(f, f)$$

Here, e_i is any element in $\mathscr{A}_i$.

Proof. Define a flow ω supported on geodesic paths with

$$\omega(\ell) = \frac{\pi(x)\,\pi(y)}{\# L(x, y)} \qquad \text{for} \quad \ell \in L(x, y) \tag{5.9}$$

By hypothesis,

$$\sum_{x \in X} \frac{1}{V(x, r)} \sum_{\substack{\ell \in L_r(x) \\ \ell \ni e}} |\ell| \, \omega(\ell)$$

does not depend on the edge $e \in \mathscr{A}_i$. Averaging over all $e \in \mathscr{A}_i$, gives

$$\max_{e \in \mathscr{A}_i} \left\{ \sum_{x \in X} \frac{1}{V(x, r)} \sum_{\substack{\ell \in L_r(x) \\ \ell \ni e}} |\ell| \, \omega(\ell) \right\}$$
$$\leqslant \frac{1}{|\mathscr{A}_i|} \sum_{x \in X} \frac{1}{V(x, r)} \sum_{y \in B(x, r)} d^2(x, y)\, \pi(x)\, \pi(y) \leqslant \frac{r^2}{|\mathscr{A}_i|}$$

To see the first inequality, consider the inner sum on the left also averaged over $e \in \mathscr{A}_i$. Only paths $\ell \in L(x, y)$ for $y \in B(x, r)$ can appear. Such a path can appear for at most $|\ell| = d(x, y)$ edges. Summing first over $\ell \in L(x, y)$ and using Eq. (5.9) yields the second inequality and completes the proof. □

Example 5c. Random walks on groups. Let G be a finite group and S a set of generators. Define a random walk on G as

$$K(x, y) = \begin{cases} 1/|S| & \text{if} \quad yx^{-1} \in S \\ 0 & \text{if} \quad yx^{-1} \notin S \end{cases} \tag{5.10}$$

The stationary distribution is the uniform probability $\pi(x) = 1/|G|$. The next lemma shows that the random walk in Eq. (5.10) satisfies the local Poincaré inequality.

Lemma 5.3. Let G be a finite group with generating set S. Consider the graph with edge set $\mathscr{A} = \{(x, sx): x \in G, s \in S \cup S^{-1}\}$. Then the random walk Eq. (5.10) with Dirichlet form $\mathscr{E}(f, f) = \langle (I-K) f, f \rangle$ satisfies

$$\|f - f_r\|_2^2 \leqslant 2\,|S|\, r^2 \mathscr{E}(f, f) \qquad \text{for all integer } r, \qquad \text{if } S = S^{-1}$$

and

$$\|f - f_r\|_2^2 \leqslant 4\,|S|\, r^2 \mathscr{E}(f, f) \qquad \text{for all integers } r$$

if S is not symmetric.

Proof. Set $Q(x, y) = (1/2\,|G|)(K(x, y) + K(y, x))$. Then $\mathscr{E}(f, f) = \frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 Q(x, y)$ and Q is compatible with $\mathscr{A}$. Any element g of the group G acts as an automorphism of the chain by right multiplication since $Q(xg, yg) = Q(x, y)$. Each G-edge-orbit has size at least $|G|$. If S is symmetric, we have $Q(e) = 1/(|S|\,|G|)$. If S is not symmetric, $1/(2\,|S|\,|G|) \leqslant Q(e) \leqslant 1/(|S|\,|G|)$. In both cases, Corollary 5.1 yields Lemma 5.3. ☐

Remark 5.2. Assume for simplicity that $S = S^{-1}$. For $y \in G$, write $y = z_1 z_2 \cdots z_k$ with $z_i \in S$ and $k = |y| = d(\mathrm{id}, y)$. Let $W(z, y) = |\{i: z_i = z\}| \leqslant |y|$. Arguing as in Ref. 10, the bound above can be refined to

$$\|f - f_r\|_2^2 \leqslant 2\,|S|\, m(r)\, \mathscr{E}(f, f)$$

with

$$m(r) = \max_{z \in S} \left\{ \frac{1}{|B(r)|} \sum_{w \in B(r)} |w|\; W(z, w) \right\} \leqslant r^2$$

2. In Ref. 11, random walks on groups with moderate growth were studied. Lemma 5.3 shows that such groups also satisfy local Poincaré inequalities. Thus, Theorem 5.1 gives a different proof of the main results of Ref. 11. That paper contains many examples which may serve as motivation for the present paper. In fact, Theorem 5.1 and its discrete time variations extend the results of Ref. 11 to the nonsymmetric case. Some further examples appear later.

3. The factor $|S|$ that appears in Lemma 5.3 can sometimes be eliminated by symmetry. As an example, take $G=\mathbb{Z}_m^d$ with generating set

$$S=\{\pm e_1,..., \pm e_d\}$$

with $e_i=(0,..., 0, 1, 0,..., 0)$ the 1 being in ith position. For this example, using obvious paths and Remark 5.2, condition 1 gives $\|f-f_r\|_2^2 \leqslant 2r^2\mathscr{E}(f, f)$. This improves upon Lemma 5.3 by a factor of $2\,|S|$. Here, an edge $(x, x\pm e_i)$ can be moved to $(y, y\pm e_j)$ by first transposing the ith and jth coordinates and then translating. This shows that the automorphism group of the chain acts transitively on oriented edges. Thus, Corollary 5.1 also yields $\|f-f_r\|_2^2 \leqslant 2r^2\mathscr{E}(f, f)$. The next example contains further illustration.

Example 5d. Random walk generated by conjugacy classes. Let G be a finite group and consider the random walk defined at Eq. (5.10). We show that if the set S is invariant under conjugation, i.e., $t^{-1}St=S$ for all $t\in G$, then the bounds of Lemma 5.3 can be improved. The random transpositions walk of Refs. 8 and 15, or the Hildebrand's random transvections walk,[23] are examples of natural walks which are constant on conjugacy classes. If S is invariant under conjugation, write

$$S=C_1\cup\cdots\cup C_\sigma \tag{5.11}$$

with C_i disjoint conjugacy classes. We set $I=\{i\in\{1,..., \sigma\}\colon C_i\neq\{id\}\}$.

Lemma 5.4. Let G a finite group with generating set S. Consider the graph with edge set $\mathscr{A}=\{(x, sx)\colon x\in G, s\in S\cup S^{-1}\}$. For the random walk in Eq. (5.10), with S satisfying Eq. (5.11),

$$\|f-f_r\|_2^2 \leqslant \frac{4\,|S|\,r^2}{C_*}\,\mathscr{E}(f, f)$$

with $C_*=\min_{i\in I}|C_i\cup C_i^{-1}|$. If we further assume that S is symmetric, then

$$\|f-f_r\|_2^2 \leqslant \frac{2\,|S|\,r^2}{C_*}\,\mathscr{E}(f, f)$$

Proof. An oriented edge is a pair (x, sx) with $x\in G$, $s\in S^{\pm 1}$. The group G acts on edges both on the right $((x, sx)\,g=(xg, sxg))$ and on the left $(g(x, sx)=(g^{-1}x, g^{-1}sgg^{-1}x))$. These actions preserve $Q(x, y)=(1/2\,|G|)(K(x, y)+K(y, x))$. Moreover, the map $x\to x^{-1}$ also preserves Q because

$$(x^{-1}, x^{-1}s^{-1})=(x^{-1}, u^{-1}x^{-1}) \qquad \text{where} \quad u=x^{-1}sx$$

These actions generate a group Γ. It is easy to see that a Γ-edge-orbit is precisely given by

$$\mathscr{A}_i = \{(x, sx) : x \in G, \quad s \in C_i \cup C_i^{-1}\}, \quad 2 \leqslant i \leqslant \sigma$$

Thus $|\mathscr{A}_i| = |G|\,|C_i \cup C_i^{-1}|$. Now the result follows from Corollary 5.1.

It is interesting to specialize Lemma 5.4 to the case where S is symmetric and $r = \gamma =$ the diameter. In this case, the argument yields easily

$$\|f - f_\gamma\|_2^2 = \mathrm{Var}(f) \leqslant \frac{|S|\,\gamma^2}{C_*}\,\mathscr{E}(f, f)$$

which improves by a factor of 2 the bound of Lemma 5.4. In particular, this implies the eigenvalue bound

$$\beta_1 \leqslant 1 - \frac{C_*}{|S|\,\gamma^2}$$

for the random walk on G associated with the generating set S.

Remark 5.3. Usually, walks that are constant on conjugacy classes are analysed by using the fact that the eigenvalues can be expressed in terms of the characters χ of the group. Lemma 5.4 offers the opportunity of turning this around, giving bounds on the ratio $\mathrm{Re}(\chi(s)/\chi(\mathrm{id}))$ in terms of the diameter of the group generated by the conjugates of s. This is developed in Ref. 14.

5.3. Moderate Growth and Local Poincaré Imply Nash Inequalities

This section proves Theorem 5.1. The first result leans on an idea of Robinson[31] which is also used in Ref. 6.

Theorem 5.2. Let X be a finite set equipped with a probability measure π and an edge set $\mathscr{A}$. Let $\mathscr{E}_Q$ be a Dirichlet form with kernel Q compatible with $\mathscr{A}$. Suppose that, for some integer R, there are reals M, $d \geqslant 1$ and $a > 0$ such that

$$V(x, r) \geqslant \frac{(r+1)^d}{M} \quad \text{and} \quad \|f - f_r\|_2^2 \leqslant a r^2 \mathscr{E}_Q(f, f)$$

for all $x \in X$, $r \in [0, R]$, and all functions f. Then

$$\|f\|_2^{2+4/d} \leqslant C\left[\mathscr{E}_Q(f, f) + \frac{1}{aR^2}\|f\|_2^2\right]\|f\|_1^{4/d}$$

with $C = (1 + 1/d)^2 (1 + d)^{2/d} M^{2/d} a$.

Proof. Write

$$\|f\|_2^2 = \langle f, f\rangle = \langle f - f_r, f\rangle + \langle f_r, f\rangle$$

Then, for any integer $r \in [0, R]$,

$$\langle f - f_r, f\rangle \leqslant \|f - f_r\|_2 \, \|f\|_2 \leqslant a^{1/2} r \mathscr{E}_Q(f, f)^{1/2} \, \|f\|_2$$

and

$$\langle f_r, f\rangle \leqslant \|f_r\|_\infty \, \|f\|_1 \leqslant M(r+1)^{-d} \, \|f\|_1^2$$

This gives, for all integers $r \in [0, \infty[$,

$$\|f\|_2^2 \leqslant a^{1/2} r [\mathscr{E}_Q(f, f) + (aR^2)^{-1} \, \|f\|_2^2]^{1/2} \, \|f\|_2 + M(r+1)^{-d} \, \|f\|_1^2$$

Thus, for all reals $s \in [0, \infty[$,

$$\|f\|_2^2 \leqslant a^{1/2} s [\mathscr{E}_Q(f, f) + (aR^2)^{-1} \, \|f\|_2^2]^{1/2} \, \|f\|_2 + M s^{-d} \, \|f\|_1^2$$

Minimizing the right-hand side in s gives

$$\|f\|_2^2 \leqslant (d^{-d/(d+1)} + d^{1/(d+1)}) [a(\mathscr{E}_Q(f, f) + (aR^2)^{-1} \, \|f\|_2^2) \, \|f\|_2^2]^{d/2(d+1)} [M \, \|f\|_1^2]^{1/(d+1)}$$

The result now follows from routine simplifications. □

Remark 5.4. 1. The factor $(1 + 1/d)^2 (1 + d)^{2/d}$ is bounded by 16 for all $d \geqslant 1$.

2. This argument works as well if we replace the local Poincaré inequality with an inequality of the form

$$\|f - f_r\|_2^2 \leqslant a r^\alpha \mathscr{E}_Q(f, f)$$

for some $\alpha > 0$. This and the volume growth hypothesis $V(x, r) \geqslant (r+1)^d/M$ yields the Nash inequality

$$\|f\|_2^{2 + 2\alpha/d} \leqslant C \left[\mathscr{E}_Q(f, f) + \frac{1}{aR^\alpha} \, \|f\|_2^2 \right] \|f\|_1^{2\alpha/d}$$

with $C = (1 + \alpha/(2d))^2 (1 + 2d/\alpha)^{\alpha/d} M^{\alpha/d} a$. Moreover, the volume $V(x, r)$ and the mean value $f_r(x)$ do not need to be defined in terms of balls: any family of sets depending on the parameter r could be used instead. However, it is not clear that this extra generality is of any real use.

Proof of Theorem 5.1. We can now give the proof of the upper bound in Theorem 5.1. Here, we work with $\mathscr{E}_Q(f, f) = \mathscr{E}(f, f) = \langle (I-K\rangle f, f\rangle$ and the symmetric edge-set associated with $Q(x, y) = \frac{1}{2}(K(x, y)\,\pi(x) + K(y, x)\,\pi(y))$.

Under the assumption of Eqs. (5.1) and (5.2), the quantity M in Theorem 5.2 can be taken as $A\gamma^d$ with $R = \gamma$. Now, the Nash inequality in Theorem 5.2 together with Theorem 3.3 give the decay bound

$$\|H_t\|_{2\to\infty} \leqslant e(dC/4t)^{d/4}$$

for $t \in [0, a\gamma^2]$ with $C = (1 + 1/d)^2 (1+d)^{2/d} A^{2/d} \gamma^2 a$.

For $t = a\gamma^2 + s =^{\text{def}} t_1 + t_2$, use of Lemma 2.3 along with $\lambda \geqslant 1/(a\gamma^2)$ gives

$$\|(H_t^x/\pi) - 1\|_2 \leqslant e^3(1+d)A)^{1/2} (d/4)^{d/4} \exp\left(-\frac{s}{a\gamma^2}\right)$$

which yields the desired result. More precisely, we proved

Theorem 5.3. Let K, π be a Markov chain on a finite set X. Assume that (K, π) has moderate growth in Eq. (5.2) and satisfies a local Poincaré inequality in Eq. (5.3). Then, the continuous time semigroup at Eq. (2.6) satisfies, for all $t > 0$,

$$\|(H_t^x/\pi) - 1\|_2 \leqslant a_1 e^{-s/(a\gamma^2)}$$

for all $t \geqslant a\gamma^2 + s$ with $s > 0$. Here, $a_1 = (e^3(1+d)A)^{1/2} (d/4)^{d/4}$.

For the proof of the lower bound of Theorem 5.1 for reversible chains, we refer the reader to the arguments developed in Refs. 11 and 12. These arguments are elementary and can easily be adapted to the present setting.

We now describe discrete time results that are the analogs of Theorem 5.1.

Theorem 5.4. Let K, π be a Markov chain on a finite set X. Assume that (K, π) has moderate growth in Eq. (5.1) and satisfies a local Poincaré inequality Eq. (5.2). Assume further that $\inf_X K(x, x) = \varepsilon > 0$. Then,

$$\|(K_x^n/\pi) - 1\|_2 \leqslant a_1 e^{-m/(a\gamma^2)} \qquad \text{for} \quad n = (2\varepsilon)^{-1} a\gamma^2 + m + 1$$

with $m \geqslant 0$ and $a_1 = (e(1+d)A)^{1/2} (2+d)^{d/4}$.

Proof. Use Theorem 5.2, the comparison in Eq. (2.11) between $\mathscr{E}$ and $\mathscr{E}_*$, Theorem 3.1 and Lemma 1.1. In fact, we used the form $\mathscr{E}_\#$ instead of $\mathscr{E}_*$ as in Remark 3.1 following Theorem 3.1. □

Theorem 5.5. Let K, π be a reversible Markov chain on a finite set X. Assume that (K, π) has moderate groth (5.1) and satisfies a local Poincaré inequality in Eq. (5.2). Assume further that the least eigenvalue $\beta_{|X|-1}$ satisfies $\beta_{|X|-1} \geqslant -1 + 1/a\gamma^2$. Then

$$\|(K_x^n/\pi) - 1\|_2 \leqslant a_1 e^{-m/(a\gamma^2)} \qquad \text{for} \quad n = 2a\gamma^2 + m + 1$$

with $m \geqslant 0$ and $a_1 = (2e(1+d)A)^{1/2}(2+d)^{d/4}$. Moreover, there exists $a_2, a_3 > 0$ such that, for

$$a_2 e^{-a_3(n/\gamma^2)} \leqslant \sup_x \|K_x^n - \pi\|_{\text{TV}} \leqslant a_1 e^{-n/a\gamma^2}$$

Proof. For the upper bound, use Theorem 5.2, Corollary 3.1, and Lemma 1.1. For the lower bound, adapt the arguments given in Refs. 11 and 12.

Theorem 5.6. Let K, π be a Markov chain on a finite set X. Let $\mathscr{A}_*$ be the set of edge (x, y) such that $Q_*(x, y) = \sum_z K(z, x)\, K(z, y)\, \pi(z) > 0$. Assume that $(X, \mathscr{A}_*, \pi)$ has moderate growth as in Eq. (5.5) and that $(\mathscr{A}_*, \pi, \mathscr{E}_*)$ satisfies a local Poincaré inequality as in Eq. (5.6). Then

$$\|(K_x^n/\pi) - 1\|_2 \leqslant a_1 e^{-m/(a\gamma^2)} \qquad \text{for} \quad n = a\gamma^2 + m + 1$$

with $m \geqslant 0$ and $a_1 = (e(1+d)A(^{1/2}(4(2+d))^{d/4}$.

Proof. Use Theorem 5.2, Theorem 3.1, and Lemma 1.1.

We close this section by showing how these results apply to the non-reversible chain in Eq. (2.10) of Example 2f. Recall that K is the nearest neighbor random walk on a n point path with an extra directed edge from 1 to n. By Lemma 2.4, this chain has stationary distribution π given by

$$\pi(i) = (2i)/n^2 \qquad \text{for} \quad 1 \leqslant i \leqslant n-1, \quad \pi(n) = 1/n \tag{5.12}$$

Lemma 5.5. The n point path with the probability measure π at (5.12) has 4-2-moderate growth.

Proof. Consider first $V(1, r)$. We have

$$V(1, r) = \sum_1^{r+1} \frac{2i}{n^2} = \frac{(r+1)(r+2)}{n^2} \geqslant \frac{1}{4}\left(\frac{r+1}{n-1}\right)^2$$

Now, the diameter of the n point path is $n-1$, so this is just what is needed. Further, $V(i, r) \geqslant V(1, r)$ for all $i = 1,..., n$ and all r.

Lemma 5.6. The kernel $Q(x, y) = K(x, y)\,\pi(x)$ associated with the chain in Eq. (2.10) is compatible with the graph structure of the n point path and the corresponding Dirichlet form $\mathscr{E}(f, f) = \langle (I-K) f, f\rangle$ satisfies the local Poincaré inequality

$$\|f - f_r\|_2 \leqslant 24 r^2 \mathscr{E}(f, f)$$

Proof. First consider the edge $(i, i+1)$. Then $Q(i, i+1) = i/n^2$ for $1 \leqslant i \leqslant n-1$. We must bound, for all i, r,

$$\eta(i, r) = \frac{2}{Q(i, i+1)} \sum_{\substack{|k-j| \leqslant r, j \leqslant i \\ k \geqslant i+1}} |k-j| \frac{\pi(j)\,\pi(k)}{V(j, r)}$$

We may bound $|k-j|$ by r. Consider two cases: $i < 2r$, $i \geqslant 2r$.

Case 1. $i < 2r$. Then, for $j \leqslant i$,

$$V(j, r) \geqslant \frac{1}{n^2} \sum_{\ell \leqslant r} \ell = \frac{r(r+1)}{2n^2} \geqslant \frac{r^2}{2n^2}$$

Using this, we get

$$\begin{aligned} \eta(i, r) &\leqslant \frac{16}{ir} \sum_{\substack{j \leqslant i \\ i+1 \leqslant k \leqslant i+r}} jk = \frac{16}{ir} \frac{i(i+1)}{2} \left(\frac{(i+r+1)(i+r)}{2} - \frac{i(i+1)}{2} \right) \\ &= 4 \frac{i+1}{r} (2ir + r + r^2) \leqslant 24 r^2 \end{aligned}$$

Case 2. $i \geqslant 2r$. Then, for $j \leqslant i$,

$$V(j,r) \geqslant V(i-r, r) \geqslant \frac{2}{n^2} \sum_{\ell = i-2r}^{i-r} \ell \geqslant \frac{2}{n^2} \left(\frac{(i-r)^2}{2} - \frac{(i-2r)^2}{2} \right) = \frac{4ir + 5r^2}{n^2}$$

Using this, we obtain

$$\begin{aligned} \eta(i, r) &\leqslant \frac{8r}{4i(ir + r^2)} \sum_{\substack{i-r \leqslant j \leqslant i \\ i+1 \leqslant k \leqslant i+r}} jk \\ &= \frac{r}{2i(ir + r^2)} ([(2i+r)(r+1)][(2r+1)i + r(r+1)]) \\ &= \frac{r}{2i(ir + r^2)} (4(i+r)r)(6ri) \leqslant 12 r^2 \end{aligned}$$

The edges $(i, i-1)$ can be treated similarly. Lemma 5.6 follows.

The two last lemmas show that Theorem 5.1 applies to the present chain and shows that the continuous time chain is close to equilibrium after a time of order n^2.

It is interesting to see what has to be done to reach the same conclusion for the discrete time chain. In Example 2f, we computed the kernel $P = K^*K$. The natural graph corresponding to this kernel is a different n point path: starting from 2 it goes to 4, 6,..., $n-1$, 1, and then 3, ..., n. The stationary distribution is given (of course) by Eq. (5.12). Working as before, one can argue moderate groth and a local Poincaré inequality. Thus, Theorem 5.11 applies. It shows that order n^2 steps are sufficient to reach stationary in discrete time as well.

6. CONVEX SETS IN TWO-DIMENSIONS

Let C be a connected set of lattice points inside a compact convex set $S \subset \mathbb{R}^2$. Let U be the uniform distribution, and let K be the Markov kernel defined at Eq. (1.1). Theorem 1.1, stated in the introduction, asserts that order γ_e^2 steps are necessary and suffice for K^n to reach approximate equilibrium in total variation. Here, γ_e stands for the maximum of the Euclidean distance between two points of C. We shall see shortly that this Euclidean diameter is comparable with the diameter γ of C for the graph distance induced by $\mathbb{Z}^2$. This section proves Theorem 1.1 as a corollary of Theorem 5.5 by showing that the Markov chain in Eq. (1.1) has moderate growth with $d = 2$ for a constant A independent of the convex set $S \subset \mathbb{R}^2$ and satisfies a local Poincaré inequality with a constant a independent of S. From now on, we fix an orthonormal basis (e_1, e_2) of $\mathbb{R}^2$ and identify $\mathbb{Z}^2$ with $\{n_1 e_1 + n_1 e_2 \colon n_1, n_2 \in \mathbb{Z}\}$.

Lemma 6.1. Let $C = S \cap \mathbb{Z}^2$ be a set of lattice points inside a convex set $S \subset \mathbb{R}^2$. Assume that the graph induced by $\mathbb{Z}^2$ on C is connected. For $x \neq y \in C$, let $D(x, y)$ be the straight line passing through x and y. Then, for any $x, y \in C$, there exists a graph geodesic path $\gamma_{x,y} \in \mathbb{Z}^2$ such that

1. The path $\gamma_{x,y}$ stays in C.
2. Each edge of $\gamma_{x,y}$ belongs to a unit square that intersects $D(x, y)$.

It follows that

$$\gamma_e \leqslant \gamma \leqslant \sqrt{2}\, \gamma_e$$

where γ is the graph diameter of C and γ_e is the euclidean diameter of the convex hull of C in $\mathbb{R}^2$. Moreover, the graph distance between two points in C is the same as the graph distance between these points in $\mathbb{Z}^2$.

Proof. Start at $x=x_0$ and construct $x_1,\ldots,x_k=y$ inductively as follows. Let v denote a unit vector for $D(x,y)$, pointing from x to y. Without loss of generality, we can assume that $v\cdot e_1>0$ and $v\cdot e_2>0$. Suppose that x_i has been constructed such that $x_0=x, x_1,\ldots,x_i$ is the beginning of a geodesic path from x to y staying in C, and that each of the edges (x_j, x_{j+1}), $j=0,\ldots,i-1$, belongs to some unit square that intersects $D(x,y)$. Look at $w_1=x_i+e_1$, $w_2=x_i+e_2$. Consider two cases:

1. If w_1, w_2 are on the same side of $D(x,y)$, one of the halflines $[x_i, e_1[$, $[x_i, e_2[$ intersects $D(x,y)$ between x and y, say $[x_i, e_1[$. Then, by convexity, w_1 is in C and we set $x_{i+1}=w_1$. It is easy to check that (x_i, x_{i+1}) belongs to a squate intersecting $D(x,y)$.
2. If w_1, w_2 are separated by $D(x,y)$, let W be the straight line defined by w_1, w_2. This line cuts $D(x,y)$ between w_1 and w_2. Moreover, we know that there is a lattice path ℓ in C that goes from x to y. This path must intersect W somwhere, and it can not be in the segment $]w_1, w_2[$. Thus, we have two points $\ell\cap W\notin\,]w_1,w_2[$ and $D(x,y)\cap W\in[w_1,w_2]$ that are in S. By convexity, it follows that one of the points w_1 or w_2 belongs to C. If only one of them belongs to C pick that one to be x_{i+1}. If both belong to C, pick the one closest to y in Euclidean distance.

In any case, we have succeeded in constructing x_{i+1} with the required properties. Observe that each step decreases the Euclidean distance from y by at least a fixed amount. It is thus clear that this process ends and produces a geodesic path from x to y.

Remark 6.1. The result given by Lemma 6.1 (in dimension 2) is simply wrong in higher dimensions. Consider, in $\mathbb{Z}^3=\{n_1e_1+n_2e_2+n_3e_3: n_1, n_2, n_3\in\mathbb{Z}\}$, the set

$$C=\{(n,0,0): 0\leqslant n\leqslant N\}\cup\{(0,1,0)\}\cup\{(n,1,1): 0\leqslant n\leqslant N\}$$

and let S be the convex hull of C. The Euclidean distance between $(N,0,0)$ and $(N,1,1)$ is $\sqrt{2}$ whereas the graph distance in C between these two points is $2(N+1)$. Using this type of construction, it is possible to produce a set $C=\mathbb{Z}^3\cap S$ with S a convex set in $\mathbb{R}^3$ such that C is connected with graph-diameter γ and for which equilibrium is **not** reached after order γ^2 steps (e.g., two 2-dimensional squares of side a attached by a path of length b with a and b arbitrary).

We now want to describe some properties of the number of lattice points in C that are at (lattice) distance less than or equal to r from a fixed point $x \in C$. Let $N(r)$ denote the number of lattice points that are at distance less than or equal to r from the origin in $\mathbb{Z}^2$ (here we are using the graph distance d in $\mathbb{Z}^2$). We have

$$N(r) = 1 + 2r + 2r^2$$

Now, in C, equipped with the lattice graph structure, consider the ball $B(x, r)$ of radius r around x. By Lemma 6.1, we have

$$B(x, r) = \{z \in \mathbb{Z}^2 \colon d(x, z) \leqslant r\} \cap C$$

Let $N(x, r) = |B(x, r)|$ be the number of points in $B(x, r)$.

Proposition 6.1. Let $C = S \cap \mathbb{Z}^2$ be a set of lattice points inside a convex set $S \subset \mathbb{R}^2$ and assume that the graph induced by $\mathbb{Z}^2$ on C is connected. Then we have

$$\forall x \in C, \qquad \forall 0 \leqslant s \leqslant r, \qquad \frac{N(x, r)}{N(x, s)} \leqslant 81 \frac{N(r)}{N(s)}$$

Specializing to $r = \gamma$, it follows that the Markov chain K defined at Eq. (1.1) has $(A, 2)$-moderate groth for some universal constant A.

Proof. For $r \geqslant 0$, let $N'(x, r)$ be the number of points in C at distance exactly r from x and let $N'(r)$ be the number of points in $\mathbb{Z}^2$ at distance r from $(0, 0)$ (i.e., $N'(0) = 1$, $N'(r) = 4r$ if $r \geqslant 1$). The points of C at distance r from x lie on the boundary of a Euclidean square (see Fig. 5). Call the boundary of this square $\delta(x, r)$.

Let $\ell(x, r)$ be the length of the part of $\delta(x, r)$ that is inside the convex set S and let $\ell(r) = 4\sqrt{2}\, r$ be the total length of $\delta(x, r)$. An elementary argument involving dilation and convexity—as in Fig. 6—shows that

$$\forall 0 \leqslant s \leqslant r, \qquad \frac{\ell(x, r)}{\ell(r)} \leqslant \frac{\ell(x, s)}{\ell(s)} \tag{6.1}$$

Since $N'(r) = \ell(r)/\sqrt{2}$ and

$$N'(x, r) \leqslant 4 + \ell(x, r)/\sqrt{2}, \qquad \ell(x, r)/\sqrt{2} \leqslant 4 + N'(x, r)$$

we deduce from Eq. (6.1) that

$$\forall 0 \leqslant s \leqslant r, \qquad \frac{N'(x, r)}{N'(r)} \leqslant \frac{N'(x, s) + 8}{N'(s)}$$

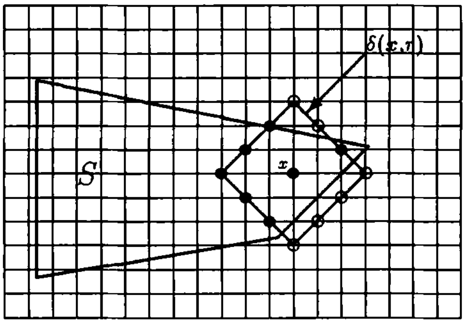

Fig. 5. The points of C at distance r from x lie on the boundary of a Euclidean square.

It follows that

$$\forall 0 \leqslant s \leqslant r, \qquad \frac{N'(x,r)}{N'(r)} \leqslant 9 \frac{N'(x,s)}{N'(s)} \tag{6.2}$$

since $N'(x,r) \geqslant 1$ implies $N'(x,s) \geqslant 1$ for all $0 \leqslant s \leqslant r$.

Now, given $0 \leqslant s \leqslant r$, write

$$\begin{aligned}
\frac{N(x,r)}{N(x,s)} &= \frac{\sum_0^r N'(x,t)}{\sum_0^s N'(x,t)} = 1 + \frac{\sum_{s+1}^r N'(x,t)}{\sum_0^s N'(x,t)} \\
&= 1 + \frac{\sum_{s+1}^r N'(t)[N'(x,t)/N'(t)]}{\sum_0^s N'(t)[N'(x,t)/N'(t)]} \\
&\leqslant 1 + \frac{[9N'(x,s)/N'(s)][\sum_{s+1}^r N'(t)]}{[N'(x,s)/9N'(s)][\sum_0^s N'(t)]} \\
&\leqslant 1 + 81 \frac{\sum_{s+1}^r N'(t)}{\sum_0^s N'(t)} \leqslant 81 \frac{\sum_0^r N'(t)}{\sum_0^s N'(t)} \\
&\leqslant 81 \frac{N(r)}{N(s)}
\end{aligned}$$

Here, we have used Eq. (6.2) to obtain the first inequality. This ends the proof of Proposition 6.1. □

Proposition 6.2. Let $C = S \cap \mathbb{Z}^2$ be a set of lattice points inside a convex set $S \subset \mathbb{R}^2$ and assume that the graph induced by $\mathbb{Z}^2$ on C is connected.

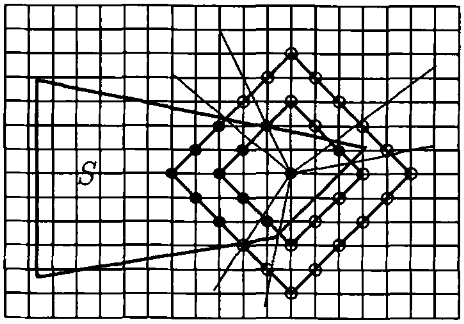

Fig. 6. An example of an elementary argument involving dilation and convexity.

Then the Markov chain defined at Eq. (1.1) satisfies the local Poincaré inequality

$$\forall r \geqslant 0,\ \forall f, \qquad \|f - f_r\|_2^2 \leqslant a r^2 \mathscr{E}(f, f)$$

for some universal constant a.

Proof. We use Lemma 5.1 with the paths $\gamma_{x,y}$ constructed in Lemma 6.1. Here, $\pi = U = 1/|C|$, $Q(e) = 1/4\,|C|$ for any edge e, and

$$\eta(r) = \max_e \left\{ 8 \sum_{\substack{\gamma_{x,y} \ni e \\ d(x,y) \leqslant r}} \frac{|\gamma_{x,y}|}{|B(x,r)|} \right\} \tag{6.3}$$

Fix an edge e, and let x, y be such that $e \in \gamma_{x,y}$ and $d(x, y) \leqslant r$. Without loss of generality, we can further assume that $d(x, e_-) \geqslant d(y, e_-)$ (if not, exchange the roles of x and y). Fix x and count how many y can qualify. Since $e \in \gamma_{x,y}$, the construction of $\gamma_{x,y}$ implies that the straight line $D(x, y)$ is at Euclidean distance at most $\sqrt{2}$ from e_-. This forces y to be in a Euclidean rectangle of length at most $r/2$ and width at most $4\sqrt{2}$. Thus the number of y that qualify for a given x is bounded by $8\sqrt{2}\,r$. See Fig. 7. It follows that the number of pairs (x, y) such that $d(x, y) \leqslant r$ and $e \in \gamma_{x,y}$ is bounded by

$$16\sqrt{2}\, r \max\{N(e_-, r),\ N(e_+, r\} \tag{6.4}$$

Now, for any x, $z \in C$ with $d(x, z) \leqslant r$, Proposition 6.1 implies

$$N(z, r) \leqslant N(x, 2r) \leqslant a_1 N(x, r)$$

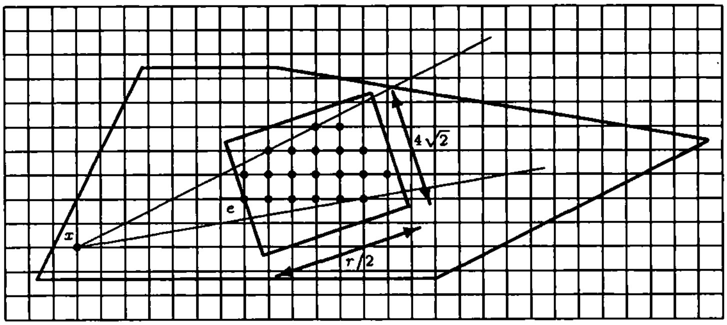

Fig. 7. The number of pairs (x, y) such that $d(x, y) \leqslant r$ and $e \in \gamma_{x,y}$ is bounded by $16\sqrt{2}\, r \max\{N(e_-, r), N(e_+, r)\}$.

for some universal constant a_1. Using this and Eq. (6.4) in Eq. (6.3) yields a universal constant a such that $\eta(r) \leqslant ar^2$, and this is the desired inequality.

In order to use the results of Section 5, we will have to check the condition $\beta_{|C|-1} \geqslant -1 + 1/(a\gamma^2)$ for the least eigenvalue of the chain (cf. Theorem 5.5).

For the chain in Eq. (1.1) we are dealing with, Proposition 2, p. 40, of Ref. 16 yields

$$\beta_{|C|-1} \geqslant -1 + \frac{1}{a\gamma^2}$$

The details are left to the reader (hint: use the loops at the boundary). Putting all these ingredients together, we get

Theorem 6.1. Let $C = S \cap \mathbb{Z}^2$ be a set of lattice points inside a convex set $S \subset \mathbb{R}^2$ and assume that the graph induced by $\mathbb{Z}^2$ on C is connected. Then the Markov chain defined at Eq. (1.1) satisfies

$$\|(K_x^n/U) - 1\|_2 \leqslant a_1 e^{-a_2 m} \qquad \text{for all} \quad n \geqslant (1+m)\,\gamma^2, \qquad m > 0$$

and

$$a_3 e^{-a_4 m} \leqslant \sup_x \|K_x^n - U\|_{\mathrm{TV}} \leqslant a_1 e^{-a_2 m} \qquad \text{for all} \quad n = \lceil m\gamma^2 \rceil, \qquad m > 0$$

Here, the a_i's are positive universal constants.

7. INHOMOGENEOUS CHAINS

Consider a sequence of Markov chains K_i on the same give finite state space X having a common stationary distribution π. Form the product

$$P_n = K_1 K_2 \cdots K_n \qquad (7.1)$$

There is very little theory to help understand the convergence of P_n to π given rates of convergence for the K_i's. The following result shows that, if the K_i have comparable second eigenvalues and satisfy comparable Nash inequalities, then, the product with variable factors P_n converges at least as rapidly as the power of a single factor.

Theorem 7.1. Let K_i, be a family of irreducible aperiodic Markov kernels on a finite state space $\mathcal{X}$ with common stationary distribution π. With notation as an Theorem 3.1, if the Nash inequalities

$$\|f\|_2^{2+1/D} \leqslant C\left(\mathscr{E}_{i,*}(f,f) + \frac{1}{N}\|f\|_2^2\right)\|f\|_1^{1/D}$$

hold for some constants $C, D > 0$, $N \geqslant 1$ and all $i = 1, 2,...$, then, P_n defined at Eq. (7.1) satisfies

$$\|P_n\|_{2\to\infty} \leqslant [4CB/(n+1)]^D \qquad \text{for} \quad 0 \leqslant n \leqslant N$$

with $B = B(D, N) = (1 + 1/N)(1 + \lceil 4D \rceil)$. If, moreover, $\mu_i(1) \leqslant \mu$ for all $i = 1, 2,...$, then,

$$2\,\|P_n^x - \pi\|_{\mathrm{TV}} \leqslant \|(P_n^x/\pi) - 1\|_2 \leqslant e^{-\theta}$$

for

$$n \geqslant N + \frac{1}{1-\mu}\,[D\log(4C(1 + \lceil 4D \rceil)/N) + \theta]$$

Proof. Observe first that $\|P_n - \pi\|_{2\to 2} \leqslant \prod_{i=1}^n \mu_i(1)$. This follows, as in Lemma 2.1, from the identity $P_n - \pi = \prod_{i=1}^n (K_i - \pi)$. The Nash inequalities imply the decay bound by using Lemma 3.1 and slight modifications of the argument for Theorem 3.1. We omit further details. □

Example 7a. Let p be an odd number. Let a_i, $i = 1, 2,...$, be a sequence of integers, all relatively prime to p. Define random walks on $\mathbb{Z}_p$ by

$$K_i(x, y) = \begin{cases} 1/2 & \text{if } \; y = x \pm a_i \\ 0 & \text{otherwise} \end{cases}$$

These are symmetric Markov kernels with uniform stationary distributions. For each i, $\mu_i(1) = \cos(2\pi/p)$ by simple Fourier analysis. Recall that $\mathscr{E}_{i,*} = \langle (I - K_i^2) f, f \rangle$ and observe that K_i^2 is supported by $\{\mathrm{id}, \pm 2a_i\}$. Since $2a_i$ is also relatively prime to p, $\{\mathrm{id}, 2a_i, -2a_i\}$ generates $\mathbb{Z}_p$. Using Lemma 5.1, it can be checked that

$$\|f-f_r\|_2^2 \leqslant 4r^2 \mathscr{E}_{i,*}(f,f)$$

Since $\mathbb{Z}_p$ has moderate growth (with $d=1$) for any $\pm 2a_i$ as generators, Theorem 5.3 implies that each K_i satisfies the same Nash inequality. Now, Theorem 7.1 shows that, for any choice of the a_i, $P_n = K_1 K_2 \dots K_n$ is close to uniform if n/p^2 is large.

In this example, special choices of a_i can drastically change the rate of convergence. If $a_i = 2^{i-1}$, $1 \leqslant i < +\infty$, we will show that $K_1 \dots K_n$ is close to uniform if $n/\log p$ is large. To see this, consider the random integer $\varepsilon_0 + \varepsilon_1 2 + \dots + \varepsilon_j 2^j$ with ε_ℓ independent random variables taking values 0 or 1 with probability 1/2. This has a uniform distribution on $\{0, 1,..., 2^{j+1}-1\}$. It is thus close to uniform, taken mod p, provided $2^j/p$ is large. The convolution $K_1 K_2 \dots K_j$ has this same distribution up to affine change of variables. It follows that $\|P_n - \pi\|_{\mathrm{TV}}$ is small for $n/\log p$ large.

Remark 7.1. 1. Many variations on this example are possible. The group $\mathbb{Z}_p$ can be replaced by any other finite p-group having moderate growth. Examples of such groups are given in Ref. 11. The reversible Markov chains K_i can be replaced by the similar but nonreversible chains supported on $\{\mathrm{id}, a_i\}$ instead of $\{a_i, -a_i\}$.

2. In the other direction, there are two random walks K_1, K_2 on the symmetric group S_n where K_1^2 is exactly uniform but $(K_1 K_2)^n$ is slow to converge. See Ref. 7, Section 6.

3. We do not know how to get similar bounds for such combinations if the rate for each individual has been established by other means (e.g., coupling).

ACKNOWLEDGMENTS

We thank Jim Fill for pointing out to us the utility of certains singular value inequalities in bounding rates of convergence. We also thank him for corrections resulting from his careful reading of the manuscript.

REFERENCES

1. Bakry, D., Coulhon, T., Ledoux, M., and Saloff-Coste, L. (1995). Sobolev inequalities in disguise. Université Paul Sabtier, Toulouse. *Stat. et. Prob.*, Preprint.
2. Carlen, E., Kusuoka, S., and Stroock, D. (1987). Upper bounds for symmetric Markov transition functions. *Ann. Inst. H. Poincaré, Prob. Stat.* **23**, 245–287.

3. Chung, F., Diaconis, P., and Graham, R. (1987). A random walk problem arising in random number generation. *Ann. Prob.* **15**, 1148–1165.
4. Coulhon, T., and Saloff-Coste, L. (1990a). Puissance d'un opérateur regularisant. *Ann. Inst. H. Poincaré, Prob. Stat.* **26**, 419–436.
5. Coulhon, T., and Saloff-Coste, L. (1990b). Marches aléatoires non-symétriques sur les groupes unimodulaires. *C. R. Acad. Sci. Paris*, Série I, **310**, 627–630.
6. Coulhon, T., and Saloff-Coste, L. (1993). Isopérimétrie sur les groupes et les variétés. *Rev. Mat. Iberoamericana*, **9**, 293–314.
7. Diaconis, P. (1982). *Applications of Non-Commutative Fourier Analysis to Probability Problems.* Springer L.N.M. **1362**, 51–100.
8. Diaconis, P. (1986). *Group Representations in Probability and Statistics.* IMS, Hayward.
9. Diaconis, P., and Saloff-Coste, L. (1933a). Comparison theorems for reversible Markov chains. *Ann. Appl. Prob.* **3**, 696–730.
10. Diaconis, P., and Saloff-Coste, L. (1993b). Comparison techniques for random walk on finite groups. *Ann. Prob.* **21**, 2131–2156.
11. Diaconis, P., and Saloff-Coste, L. (1994a). Moderate growth and random walk on finite groups. *Geom. and Funct. Anal.* **4**, 1–34.
12. Diaconis, P., and Saloff-Coste, L. (1995). An application of Harnack inequalities to random walk on nilpotent quotients. Actes du Colloque en l'honneur de J. P. Kahane. *J. Fourier Anal. Appl.* (special Kahane issue) 189–208.
13. Diaconis, P., and Saloff-Coste, L. (1992). Logarithmic Sobolev inequalities and finite Markov chains. *Annals Apol. Prob.* (to appear).
14. Diaconis, P., and Saloff-Coste, L. (1994c). Geometric bounds on character ratios and symmetric functions. Manuscript.
15. Diaconis, P., and Shashahani, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Geb.* **57**, 159–179.
16. Diaconis, P., and Stroock, D. (1991). Geometric bounds for eigenvalues for Markov chains. *Ann. Appl. Prob.* **1**, 36–61.
17. Diaconis, P., and Sturmfels, B. (1993). Algebraic algorithms for sampling from conditional distributions. *Ann. Stat.* (to appear).
18. Dyer, M., and Frieze, A. (1991). Computing the volume of convex bodies: a case where randomness proveably helps. In Bollobàs, B. (ed.), *Probabilistic Combinatorics and Its Applications, Proc. Symp. Appl. Math.* **44**, 123–170.
19. Fabes, E. (1993). Gaussian upper bounds on fundamental solutions of parabolic equations: the method of Nash. In *Dirichlet Forms*, L.N.M. Springer, p. 163.
20. Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. Vol. I*, Third Edition J. Wiley and Sons.
21. Fill, J. (1991). Eigenvalue bounds on convergence to stationarity for nonreveversible Markov chains with an application to the exclusion process. *Ann. Appl. Prob.* **1**, 62–87.
22. Gohberg, I., and Krein, I. (1969). Introduction to the theory of linear non-selfadjoint operators. Providence. *Amer. Math. Soc.*
23. Hildebrand, M. (1992). Generating random elements in $SL_n(F_q)$ by random transvections. *J. Alg. Combinatorics* **1**, 133–150.
24. Horn, R., and Johnson, C. (1990). *Topics in Matrix Analysis.* Cambridge University Press.
25. Jerrum, M., and Sinclair, A. (1989). Approximating the permanent. *SIAM J. Comp.* **18**, 1149–1178.
26. Lawler, G., and Sokal, A. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309**, 557–580.

27. Lovasz, L., and Simonovits, M. (1990). Random walks in a convex body and an improved volume algorithm. *Random Struct. and Algo.* **4**, 359–412.
28. Mihail, E. (1989). Combinatorial aspects of expanders. Ph.D. dissertation, Department Computer Sciences, Harvard University.
29. Nash, J. (1958). Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.* **80**, 931–954.
30. Marshall, A., and Olkin, I. (1979). *Inequalities: The Theory of Majorization and Its Applications.* Academic Press, New York.
31. Robinson, D. (1991). *Elliptic Operators on Lie Groups.* Oxford University Press.
32. Sinclair, A., and Jerrum, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.* **82**, 93–133.
33. Sinclair, A. (1993). *Algorithms for Random Generation and Counting: A Markov Chain Approach.* Birkhäuser, Boston.
34. Stein, E., and Weiss, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton University Press.
35. Varopoulos, N. (1984). Une généralisation du théorème de Hardy-Littlewood-Sobolev pour les espaces de Dirichlet. *C. R. Acad. Sc. Paris*, Série I, **299**, 651–654.
36. Varopoulos, N. (1985). Isoperimetric inequalities and Markov chains. *J. Funct. Anal.* **63**, 215–239.
37. Varopoulos, N., Saloff-Coste, L., and Coulhon, T. (1993). *Analysis and Geometry on Groups.* Cambridge University Press.