Journal of Computer and System Sciences **57**, 20–36 (1998)
Article No. SS981576

What Do We Know about the Metropolis Algorithm?*

P. Diaconis

Department of Statistics, Stanford University, Stanford, California 94305

and

L. Saloff-Coste

CNRS, Statistique et Probabilités, Université Paul Sabatier, Toulouse, 31062 cedex, France; and Department of Mathematics, Cornell University, Ithaca, New York 14853

Received January 1998

The Metropolis algorithm is a widely used procedure for sampling from a specified distribution on a large finite set. We survey what is rigorously known about running times. This includes work from statistical physics, computer science, probability, and statistics. Some new results (Propositions 6.1–6.5) are given as an illustration of the geometric theory of Markov chains. © 1998 Academic Press

1. INTRODUCTION

Let $\mathfrak{X}$ be a finite set and $\pi(x) > 0$ a probability distribution on $\mathfrak{X}$. The Metropolis algorithm is a procedure for drawing samples from $\mathfrak{X}$. It was introduced by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller [34]. The algorithm requires the user to specify a connected, aperiodic Markov chain $K(x, y)$ on $\mathfrak{X}$. This chain need not be symmetric but it must have $K(x, y) > 0$ if and only if $K(y, x) > 0$. The chain K is modified by auxiliary coin tossing to a new chain M with stationary distribution π. In other words, if the chain is currently at x, one chooses y from $K(x, y)$. Let the acceptance ratio be defined by

$$A(x, y) = \frac{\pi(y)\, K(y, x)}{\pi(x)\, K(x, y)}. \tag{1.1}$$

If $A(x, y) \geqslant 1$, the chain moves to y. If $A(x, y) < 1$, flip a coin with probability of heads $A(x, y)$. If the coin comes up heads, the chain moves to y. If the coin comes up tails, the chain stays at x. Formally,

$$M(x, y) = \begin{cases} K(x, y) & \text{if} \quad A(x, y) \geqslant 1;\ y \neq x \\ K(x, y)\, A(x, y) & \text{if} \quad A(x, y) < 1 \\ K(x, y) + \displaystyle\sum_{z : A(x, z) < 1} K(x, z)(1 - A(x, z)) & \\ \quad \text{if} \quad y = x. & \end{cases} \tag{1.2}$$

* Both authors research partially supported by NSF Grant DMS 9504379 and Nato Grant CRG 950686.

The following lemma says that the new chain has π as its stationary distribution:

LEMMA 1.1. *The chain $M(x, y)$ at (1.2) is an irreducible, aperiodic Markov chain on $\mathfrak{X}$ with*

$$\pi(x)\, M(x, y) = \pi(y)\, M(y, x) \qquad \text{for all} \quad x, y. \tag{1.3}$$

In particular, for all x, y

$$\lim_{n \to \infty} M^n(x, y) = \pi(y). \tag{1.4}$$

Proof. Equation (1.3) is easily verified directly: if $A(x, y) > 1$, $\pi(x)\, M(x, y) = \pi(x)\, K(x, y)$, and $A(y, x) < 1$, so

$$\pi(y)\, K(y, x)\, A(y, x) = \pi(x)\, M(x, y).$$

The same conclusion holds if $A(x, y) = 1$ and if $A(x, y) < 1$. The chain is clearly connected and is aperiodic by assumption. Now, the basic convergence theorem for Markov chains (see e.g., Karlin and Taylor [26]) implies the result.

Remark. In applications, $\mathfrak{X}$ is often a huge set and the stationary distribution π is given as $\pi(x) \propto e^{-H(x)}$ with $H(x)$ easy to calculate. The unspecified normalizing constant is usually impossible to compute. Note that this constant cancels out of the ratios $A(x, y)$ so that the chain $M(x, y)$ is easy to run.

The limit result (1.4) is unsatisfactory. In applied work, one needs to know how large n should be to have $M^n(x, y)$ suitably close to $\pi(y)$. One standard quantification of "close to stationarity" is the total variation distance:

$$\|M_x^n - \pi\| = \max_{A \subset \mathfrak{X}} |M^n(x, A) - \pi(A)| \qquad \text{with} \quad \pi(A) = \sum_{y \in A} \pi(y).$$

0022-0000/98 $25.00
Copyright © 1998 by Academic Press
All rights of reproduction in any form reserved.

If this distance is small, then the chance that the chain is in a set A is close to $\pi(A)$, uniformly. The techniques described below give fairly sharp bounds on convergence in terms of the size $|\mathfrak{X}|$ and the geometry of a graph with vertex set $\mathfrak{X}$ and an edge from x to y if $M(x, y) > 0$.

Section 2 describes a collection of examples where very sharp results are known. These include a chain on the symmetric group drawn from our joint work with Phil Hanlon and a variety of birth and death chains drawn from thesis work of Eric Belsley and Jeff Silver. There are also results for independence sampling base chains drawn from work by Jun Liu. All of these chains are special, having a high degree of symmetry. They give a collection of examples where the correct answer is known, so different bounds can be compared with the truth.

Section 3 describes work on statistical physics models widely used in image analysis. These include the Ising model and many variations. In low dimensions, away from "critical temperatures" and "phase transitions" the results show that order $n^2 \log n$ steps are necessary and suffice where n is the number of lattice sites. In phase transition regions, the running time can be exponential in n. The main work here is due to Martinelli, Schonnman, Stroock, Zegarlinski, and their co-authors.

Section 4 gives an overview of the geometric theory. This consists of Poincaré, Cheeger, Nash, Sobolev, and log–Sobelev inequalities.

Section 5 describes work on sampling from log concave densities on convex sets. This work has been developed in computer science by Frieze, Kannan, Lovász, Simonovits, and their co-authors in connection with the celebrated problem of computing the volume of a convex set. It uses Cheeger's inequality and work transferring information between discrete and continuous problems.

Section 6 describes some new work which allows sharp bounds for Metropolis chains on low-dimensional grids. This work is presented as an introduction to the geometric theory of Markov chains developed in [8–11]. It gives matching upper and lower bounds (up to good constants) for problems like sampling from

$$\pi(i) \propto i(n-i) \qquad \text{or} \qquad \pi(i) \propto e^{-(i-n/2)^2}$$

on $\{1, 2, ..., n-1\}$, with the base chain being a reflecting random walk.

The final section attempts to survey other literature, extensions to general state spaces, and some of the many improvements on the Metropolis algorithm which (currently) seem beyond rigorous analysis.

Real applications of the Metropolis algorithm are widespread. If the reader needs convincing, we recommend the three discussion papers in the *Journal of the Royal Statistical Society Ser. B* **55**, No. 3 (1993); these give many illustrations and pointers to the huge applied literature.

We have made no attempt to cover closely related work on annealing or the Gibbs sampler (Glauber dynamics). We have attempted to give a reasonably complete picture of what is rigorously known about the Metropolis algorithm.

2. EXAMPLES

This section reports work on examples where symmetry allows careful analysis.

2.1. A Walk on the Symmetric Group

Let S_n denote the permutations of n items. In psychophysical experiments (e.g., "rank these sounds for loudness"), taste-testing, and preference studies, a variety of nonuniform distributions on S_n are used. One family, the Mallows model through the metric d, has form

$$\pi_\theta(\sigma) = \theta^{d(\sigma, \sigma_0)}/Z, \tag{2.1}$$

with $d(\cdot, \cdot)$ a metric on S_n and σ_0 a centering permutation. We take $0 < \theta \leqslant 1$ and $Z = Z(\theta, \sigma_0)$ a normalizing constant. Thus, if $\theta = 1$, π_θ is the uniform distribution. If $\theta < 1$, the distribution peaks at the permutation σ_0 and falls off exponentially as σ moves away from σ_0. A variety of metrics are in use. For example,

$$d(\sigma, \sigma_0) = \text{minimum number of transpositions required to bring } \sigma \text{ to } \sigma_0 \text{ (Cayley distance).} \tag{2.2a}$$

$$d(\sigma, \sigma_0) = \sum |\sigma(i) - \sigma_0(i)| \quad \text{(Spearman's footrule).} \tag{2.2b}$$

Detailed discussion can be found in Diaconis [6, Chap. 6], Critchlow [4], or Fligner and Verducci [14].

For n large (e.g., $n = 52$) the normalizing constant is impossible to calculate and samples from π_θ would routinely be drawn using the Metropolis algorithm from the base chain of random transpositions. Thus, if the chain is currently at σ, the chain proceeds by choosing i, j at random in $\{1, 2, ..., n\}$ and transposing, forming $\sigma' = (i, j)\,\sigma$. If $d(\sigma', \sigma_0) \leqslant d(\sigma, \sigma_0)$, the chain moves to σ'. If $d(\sigma', \sigma_0) > d(\sigma, \sigma_0)$ a coin is flipped with probability $\theta^{d(\sigma', \sigma_0) - d(\sigma, \sigma_0)}$. If this comes out heads, the chain moves to σ'. Otherwise, the chain stays at σ.

The running time of this chain for the Cayley distance was analyzed in [7]. The following result shows that order $n \log n$ steps are necessary and suffice for convergence.

THEOREM 2.1. *For fixed* $0 < \theta < 1$, *let* M^k *be the kth power of the Metropolis chain* (2.1), *starting at the identity, with the Cayley metric* (2.2a). *Suppose*

$$k = an \log n + cn \qquad \textit{with} \quad a = \frac{1}{2\theta} + \frac{1}{4\theta}\left(\frac{1}{\theta} - \theta\right); \qquad c > 0.$$

Then

$$\|M^k - \pi_\theta\| \leqslant f(\theta, c)$$

for $f(\theta, c)$ *an explicit function, independent of* n, *with* $f(\theta, c) \searrow 0$ *as* $c \nearrow \infty$.

Conversely, if $k = \frac{1}{2} n \log n - cn$,

$$\|M^k - \pi_\theta\| \nearrow 1 \qquad as \quad c \nearrow \infty.$$

Remarks. 1. The upper bound requires $k = an \log n + cn$ while the lower bound shows $\frac{1}{2} n \log n - cn$ steps are not enough. We conjecture that the lower bound can be improved to showing that $an \log n - cn$ steps are needed. This would prove that this chain has a sharp cutoff in its convergence behavior.

2. The proof of Theorem 2.1 depends in crucial ways on the choice of d as Cayley's metric. It uses delicate estimates of all eigenvalues and eigenvectors, available through symmetric function theory.

3. We conjecture that order $n \log n$ steps are necessary and suffice for any reasonable metric (e.g., (2.2b)). At present, the best that can be rigorously proved is that order $n!$ steps suffice and order $n \log n$ steps are necessary.

4. The paper [7] with Hanlon gives several other special cases, where such careful analysis can be carried out: Metropolis algorithms on the hypercube and families of matchings. In these cases, the Metropolis chains (as at (2.1)) give a one-parameter family of deformations of transition matrix of the base chain having interesting special functions as eigenfunctions. Ross and Xu [38], have made a fascinating connection between some of these twisted walks and convolution of hypergroups.

5. Belsley [1] and Silver [41] have carried out a delicate analysis of related cases: changing the base chain of random walk on a path to a negative binomial distribution. Their results are described further in Section 6. Again, the eigenvalues are available in closed form. One sees here a new phenomenon worth further investigation: the Metropolis algorithm gives a one-parameter family of deformations (the parameter is θ in Theorem 2.1) with eigenvalues and eigenvectors that deform to interesting special functions.

2.2. Independence Base Chains

Let π be a probability on the finite set $\mathfrak{X}$. Consider as the base chain repeated independent samples from a fixed probability $p(x)$ on $\mathfrak{X}$. Thus $K(x, y) = p(y)$ for all x. Jun Liu [28] has explicitly diagonalized the Metropolis chain in this case. To describe his results, let $w(x) = \pi(x)/p(x)$. The chain can be written

$$M(x, y) = \begin{cases} p(y) \min\left\{1, \dfrac{w(y)}{w(x)}\right\} & \text{if} \quad y \neq x, \\ p(x) + \sum_z p(z) \max\left\{0, 1 - \dfrac{w(z)}{w(x)}\right\} & \text{if} \quad y = x. \end{cases} \tag{2.3}$$

Such a chain arises naturally when comparing the widely used schemes of importance and rejection sampling with the Metropolis algorithm. In these schemes an independent sample is drawn from p. In importance sampling, averages of functions with respect to π are estimated by weighting the sample value x by $w(x)$. In rejection sampling, sample values x are kept in the sample with probability $w(x)$ and thrown away otherwise. These are close cousins to the Metropolis algorithm.

To describe Liu's results, let the states be numbered (without loss) so that $w(x_1) \geqslant w(x_2) \geqslant \cdots \geqslant w(x_{|\mathfrak{X}|})$. Write $w(i) = \omega(x_i)$, $\pi(i) = \pi(x_i)$, etc. Let

$$S_\pi(k) = \pi(k) + \cdots + \pi(|\mathfrak{X}|),$$
$$S_p(k) = p(k) + \cdots + p(|\mathfrak{X}|).$$

THEOREM 2.2 (Liu). *The Metropolis chain* (2.3) *has eigenvalues* $1 = \beta_0 > \beta_1 \geqslant \cdots \geqslant \beta_{|\mathfrak{X}|-1} \geqslant 0$ *with*

$$\beta_j = \sum_{i \geqslant j} \pi(i) \left(\frac{1}{w(i)} - \frac{1}{w(j)}\right)$$
$$= S_p(j) - \frac{S_\pi(j)}{w(j)} = S_p(j) - \frac{p(j)}{\pi(j)} S_\pi(j).$$

In particular, $\beta_1 = 1 - [p(1)/\pi(1)]$. *Furthermore, the variation distance for the chain started at* x *is bounded by*

$$4 \|M_x^k - \pi\|^2 \leqslant \sum_{j=1}^{x-1} \frac{\pi(j)}{S_\pi(j)\, S_\pi(j+1)} \beta_j^{2k} + \frac{S_\pi(x+1)}{S_\pi(x)\, \pi(x)} \beta_x^{2k}. \tag{2.4}$$

For the chain started at x,

$$4 \|M_x^k - \pi\|^2 \leqslant \frac{\beta_1^{2k}}{\pi(x)}. \tag{2.5}$$

Proof. With hindsight, it is quite straightforward to verify the result discovered by Liu: with states numbered as above, an eigenvector corresponding to eigenvalue β_k is

$$(0, ..., 0, S_\pi(k+1), -\pi(k), ..., -\pi(k)),$$

where there are $(k-1)$ zero entries. For reversible Markov chains, the Cauchy–Schwarz inequality and the spectral theorem give

$$4\,\|M_x^k-\pi\|^2 \leqslant \left\| \frac{M_x^k}{\pi} - 1 \right\|_2^2 = \sum_{j=1}^{|\mathfrak{X}|-1} \beta_j^{2k} f_j^2(x) \leqslant \frac{\beta_*^{2k}}{\pi(x)} \tag{2.6}$$

with f_j an orthonormal basis of right eigenfunctions for the matrix M and $B_* = \max\{\beta_1, |\beta_{|\mathfrak{X}_*|-1}|\}$. See, e.g., [11, Proposition 3]. Normalizing the eigenfunctions and straightforward computation give (2.4) while (2.5) follows from the rightmost inequality of (2.6).

Here is a simple example for comparison with later examples: take

$$\mathfrak{X} = \{0, 1, 2, ..., n-1\}, \qquad \pi(j) = \theta^j/Z \qquad \text{with} \quad Z = \frac{1-\theta^n}{1-\theta}$$

and $0<\theta<1$ fixed. Take the base chain uniform on $\mathfrak{X}$: $p(j)=1/n$. Thus the states are naturally ordered and $w(j)=n\pi(j)$. From the theorem,

$$\beta_1 = 1 - \frac{1}{n}\frac{1-\theta^n}{1-\theta}$$

and the upper bound (2.5) gives

$$4\,\|M_{n-1}^k-\pi\|^2 \leqslant \frac{1}{(1-\theta)\,\theta^{n-1}}\left(1-\frac{1}{n}\right)^{2k}.$$

This shows that k of order n^2 steps suffice to achieve stationarity. (The extra n is needed to kill off the factor θ^{-n}.) Use of all the eigenvalues, as at (2.4), shows that order n steps actually suffice for any starting state. It is clear that at least n steps are necessary; even if the chain starts at 0, the most likely state, it takes order n steps to have a good chance of moving once. See the following remarks.

Liu uses the results above to compare importance sampling, rejection sampling, and the Metropolis algorithm for estimating expected values like

$$\sum_{x\in\mathfrak{X}} h(x)\,\pi(x).$$

Using the criterion of mean square error, he concludes, roughly speaking, that the Metropolis algorithm and rejection methods have essentially the same efficiency, but importance sampling can show big gains. Of course, this application of the Metropolis algorithm is far from the original motivation; importance sampling assumes we can compute, or at least approximate, normalizing constants while the Metropolis algorithm can proceed without them.

Remark. Jim Fill has graciously passed on the following observation on Theorem 2.2. Things work out especially neatly if the chain is started in state 1 (in the numbering above). From the diagonalization given, one sees that

$$M^k(1,1) = \pi(1) + (1-\pi(1))\,\beta_1^k,$$
$$M^k(1,j) = \pi(j)(1-\beta_1^k), \qquad 2 \leqslant j.$$

Thus

$$\|M_1^k - \pi\| = (1-\pi(1))\,\beta_1^k.$$

3. MODELS FROM STATISTICAL PHYSICS

Statistical physics has introduced a variety of models which are also used to analyze spatial data and model images in vision and image reconstruction. In this description, we restrict attention to binary spatial patterns in a portion of a lattice. For simplicity, we also restrict attention to the Ising model. The references cited apply to much more general situations.

Thus let Λ be a finite connected subset of the lattice $\mathbb{Z}^2$. Let

$$\mathfrak{X} = \{x\colon \Lambda \to \mathbb{Z}_2\}.$$

We think of $\mathbb{Z}_2 = \{\pm 1\}$ and $\mathfrak{X}$ as the set of two-colorings of the sites in Λ. If $\{\pm 1\}$ is replaced by $\{0, 1\}$, we may think of an element of $\mathfrak{X}$ as a picture. Let s be a two-coloring of the boundary of Λ (points in $\mathbb{Z}^2-\Lambda$ at distance 1 from points in Λ). This is a specified set of boundary conditions.

The Ising model is a probability distribution on $\mathfrak{X}$ specified by

$$\pi(x) \propto e^{\beta(\sum_{(i,j)} x_i x_j + h \sum_i x_i)}, \tag{3.1}$$

where the first sum is over neighboring pairs in $\mathbb{Z}^2$ with one or both of i, j in Λ and the second sum is over i in Λ. Here $\beta>0$ is called inverse temperature and h, $-\infty<h<\infty$, is called the external field strength. With β, h, s fixed, (3.1) is a well-specified probability measure on $\mathfrak{X}$. In applications, Λ is usually a square grid of size, e.g., 64×64 or 128×128, and it is clearly impossible to calculate the normalizing constant implicit in (3.1).

The Metropolis algorithm gives an easy way to generate from π; as base chain, let us take the following: pick i in Λ at random (uniformly) and change x_i to $-x_i$. This gives a connected chain on $\mathfrak{X}$. Call this random single site updating. This chain is periodic, but the Metropolis algorithm clearly has some holding probability so the chain $M^n(x, y)$ converges to $\pi(y)$.

There is a huge rigorous literature on properties of the stationary distribution π as a function of β, h, and s. McCoy and Wu [32] or Simon [42] give a careful extended discussion. We will not review this here but merely mention that there are regions of the β, h plane where the behavior of the boundary conditions s matter (phase transitions occur) and regions where the behavior does not matter. Phase transitions occur for $h=0$ and $\beta > \beta_c$ and not otherwise. As will be described, the Metropolis algorithm converges rapidly for (β, h) away from the phase transition values (order roughly $|\Lambda| \log |\Lambda|$ steps suffice). It is believed to take order $|\Lambda|^p$ steps (with some $p > 2$) to converge for $h=0$, $\beta = \beta_c$. It takes an exponential number of steps to converge for $h=0$, $\beta > \beta_c$. The behavior of the constants involved as (β, h) approach the critical values is currently under active study. Schonmann [39, 40] gives a review of this fascinating subject.

To state a precise result an annoying periodicity problem must be dealt with. Let

$$\hat{M}(x, y) = \tfrac{1}{2}(I + M(x, y)) \tag{3.2}$$

be a modified Metropolis chain.

THEOREM 3.1 (Martinelli–Olivieri–Schonmann, [31]). *Let Λ be a square grid in $\mathbb{Z}^2$ with $|\Lambda| = n$. Then, for β, h not on the segment $h=0$, $\beta \geqslant \beta_c$, and any boundary condition s, the Metropolis chain* (3.2) *for π defined at* (3.1) *based on random single site updating satisfies*

$$\|\hat{M}_x^k - \pi\| \leqslant A e^{-Bk/(n \log n)}$$

with A, B explicit functions of β, h which do not depend on n, s, or the starting state x.

Remarks. 1. A very similar result was proved earlier by Stroock and Zegarlinski [46]. Their result holds for somewhat fewer values of β, h (e.g., $|h| \geqslant 4$ is required) but is stronger in holding uniformly for all Λ (not just square grids). They also give results which hold for larger dimensions while the techniques of Martinelli–Olivieri–Schonmann lean heavily on the assumption of $\mathbb{Z}^2$. A detailed comparison is in Frigessi, Martinelli, and Stander [17].

2. For β, h on the phase transition segment, things change drastically. Results of Martinelli [30] and Thomas [47] combine to show that the chain $\hat{M}$ takes order $e^{Bn^{1/2}}$ steps to converge. Again, B is a function of β, h, and now s; indeed in the critical segment the stationary distribution π depends strongly on the boundary conditions which now do not wash away for large grids.

The proofs of the theorems above depend on detailed study of the stationary distribution π and build on years of work by the statistical physics community. There is not much hope of carrying them over in any straightforward way to other high-dimensional uses of the Metropolis algorithm such as the permutation distributions of Section 2. There is one very useful ingredient which is clearly broadly useful, the log–Sobolev inequality. The next section gives a brief description of this emerging technique.

4. GEOMETRIC TECHNIQUES

A hierarchy of technical tools have emerged for studying powers of Markov chains. At present, these go well beyond bounds on eigenvalues. The geometric tools are named after cousins from differential geometry and differential equations: inequalities of

Poincaré, Cheeger, Sobolev, Nash, log–Sobolev.

It is beyond the scope of this paper to give a thorough introduction to these; we give a brief outline and pointers to good expositions. Basic references are [8–11] with Sinclair [44] a useful recent book.

For simplicity, we work in the context of reversible Markov chains although one of the exciting breakthroughs (see Fill [13] and [9, 10]) is that much can be pushed through in the nonreversible case.

Let $\mathfrak{X}$ be a finite set, $K(x, y)$ an irreducible, aperiodic Markov matrix on $\mathfrak{X}$. Let π be the stationary distribution and suppose (π, K) is reversible (so that $\pi(x) K(x, y) = \pi(y) K(y, x)$). Define an inner product on real functions from $\mathfrak{X}$ by $\langle f | g \rangle = \sum f(x) g(x) \pi(x)$. Then reversibility is equivalent to saying the operator K which tales f to $Kf(x) = \sum K(x, y) f(y)$ is self-adjoint on $\ell^2(\pi)$ (so $\langle Kf | g \rangle = \langle f | Kg \rangle$). This implies that K has real eigenvalues,

$$1 = \beta_0 > \beta_1 \geqslant \cdots \geqslant \beta_{|\mathfrak{X}|-1} > -1,$$

and an orthonormal basis of real eigenvectors f_i (so $Kf_i = \beta_i f_i$).

One aim is to bound the total variation distance between $K^n(x, \cdot)$ and $\pi(\cdot)$. This is accomplished by using the Cauchy–Schwarz inequality to bound

$$4 \|K_x^n - \pi\|^2 \leqslant \left\| \frac{K_x^n}{\pi} - 1 \right\|_2^2 = \sum_{i=1}^{|\mathfrak{X}|-1} f_i^2(x) \beta_i^{2n} \leqslant \frac{1}{\pi(x)} \beta_*^{2n} \tag{4.1}$$

with $\beta_* = \max(\beta_1, |\beta_{|\mathfrak{X}|-1}|)$. This final bound is clearly proved by Jerrum and Sinclair [23]. See also [8, Section 6].

Thus, one can get bounds on rates of convergence using eigenvalues. Next, one needs to get bounds on eigenvalues.

This can be accomplished by using the minimax characterization. This involves the quadratic form

$$\mathscr{E}(f|f) = \langle (I-K)\, f|f \rangle$$

$$= \tfrac{1}{2} \sum_{x,\, y} (f(x) - f(y))^2\, \pi(x)\, K(x,\, y).$$

Then

$$1 - \beta_1 = \min_f \frac{\mathscr{E}(f|f)}{\mathrm{var}(f)} \tag{4.2}$$

with

$$\mathrm{var}(f) = \sum (f(x) - \bar{f})^2\, \pi(x), \qquad \bar{f} = \sum f(x)\, \pi(x).$$

Because of (4.2), bounds of $\mathrm{var}(f)$ in terms of $\mathscr{E}(f|f)$

$$\mathrm{var}(f) \leqslant A\mathscr{E}(f|f),$$

or, equivalently, bounds on the $\ell^2(\pi)$ norm on functions with $\bar{f} = 0$:

$$\|f\|_2^2 \leqslant A\mathscr{E}(f|f).$$

give bounds $\beta_1 \leqslant 1 - 1/A$. Such bounds are called *Poincaré inequalities*. An illustration of these techniques is given in Section 6 below.

In [11] a simple technique for proving a Poincaré inequality is given using paths γ_{xy} from x to y in a graph with vertex set $\mathfrak{X}$ and an edge from z to w if $K(z, w) > 0$. These paths had been suggested earlier by work of Jerrum and Sinclair [23] to bound eigenvalues using conductance (see our discussion of Cheeger's inequality below). It emerged that whenever paths were available, their direct use in Poincaré inequalities was preferable to their use via conductance. For example, Jerrum and Sinclair's pioneering work on approximation of the permanent used paths and conductance to give a bound for the second eigenvalue of the underlying chain

$$\beta_1 \leqslant 1 - \frac{c}{n^{12}}.$$

Using just their calculations and replacing conductance by Poincaré, [11] shows

$$\beta_1 \leqslant 1 - \frac{c}{n^7}.$$

Sinclair [43] then went through several other arguments (problems of generating graphs with given degree, dimer problems) and obtained substantial improvements in every case.

Cheeger's inequality bounds eigenvalues by considering *conductance*, defined as

$$h = \min_{\pi(A) \leqslant 1/2} \frac{\sum_{x \in A,\, y \in A^c} \pi(x)\, K(x,\, y)}{\pi(A)}. \tag{4.3}$$

Bounds on eigenvalues are obtained via

$$1 - 2h \leqslant \beta_1 \leqslant 1 - \frac{h^2}{2}. \tag{4.4}$$

These ideas were introduced into combinatorial work by Alon and his coworkers for building expander graphs. There, the quantity h is of interest. One builds graphs where group theory can be used to bound β_1 and this gives bounds on h. The idea of getting bounds on β_1 by getting bounds on h directly is standard in differential geometry. It was introduced in probabilistic contexts by Lawler and Sokal [27] and independently by Jerrum and Sinclair [22, 23].

An interesting class of problems where graphs can be embedded in Euclidean space and then tools from continuous geometry (Payne–Weinberger inequalities) can be used to give direct bounds on h has been intensively studied in computer science by Dyer, Frieze, Kannan and Lovász, Simonovits. This leads to remarkable bounds for problems like approximating the volume of convex sets. These seem unobtainable by other methods at present writing.

Roughly, these bounds proceed by taking a fine mesh (the underlying graph) in an ambient Euclidean space. Then, the eigenvalues of the graph Laplacian are shown to be close to the known eigenvalues of the combinatorial Laplacian.

A superb survey was given by Kannan [25]. A recent very interesting effort along these lines is given by work of Chung, Graham, and Yau. See Chung's book [2] and the references therein.

Cheeger and Poincaré inequalities are fairly basic tools in modern geometry. More refined results are obtainable by using Nash, Sobolev, and log–Sobolev inequalities to which we now turn. Details for the following can be found in [9, 10].

While Poincaré inequalities bound the $\ell^2(\pi)$ norm in terms of the quadratic form, Nash inequalities ask for more, a bound on a power of the ℓ^2 norm. In terms of the form, this appears as

$$\|f\|_2^{2+1/D} \leqslant B \left\{ \mathscr{E}(f|f) + \frac{1}{N} \|f\|_2^2 \right\} \|f\|_1^{1/D}. \tag{4.5}$$

In (4.5), B, D, and N are constants which enter into any conclusions.

In [9] it is shown that (4.5) is *equivalent* to the conclusion that powers of the kernel M^n decay like C/n^D for $1 \leqslant n \leqslant N$. This gives crude bounds: "the Nth power is roughly flat" from which one can then use eigenvalue bounds. When applicable, Nash inequalities allow elimination of the $\pi(x)^{-1}$ term in the upper bound (4.1).

One of the main accomplishments of [9] was a useful set of conditions which imply Nash inequalities. These involve the graph structure of the state space $\mathfrak{X}$ with an edge from x to y if $K(x, y) > 0$. This allows a distance $d(x, y) =$ length of shortest path from x to y. Let $B(x, r) = \{z: d(x, z) \leqslant r\}$ be the closed ball around x with radius r. Define the volume of $B(x, r)$ as

$$V(x, r) = \sum_{z \in B(x, r)} \pi(z).$$

The first geometrical definition is moderate growth; roughly, this says that the volume $V(x, r)$ is bounded below by r^d at each x.

DEFINITION 4.1. Fix $A, d \geqslant 1$. A reversible Markov chain (K, π) has (A, d) moderate growth if

$$V(x, r) \geqslant \frac{1}{A}\left(\frac{r+1}{\gamma}\right)^d \qquad \text{for all} \quad x \in \mathfrak{X}$$
$$\text{and integers} \quad r \in \{0, 1, 2, ..., \gamma\},$$

where γ is the diameter of the graph.

The second geometrical definition is a local version of the Poincaré inequalities described above. For any real function f and integer r, set

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

DEFINITION 4.2. Let K, π be a reversible Markov chain with Dirichlet form $\mathscr{E}$. Say (K, π) satisfies a local Poincaré inequality if there is $a > 0$ such that for any real function f and all positive integers r,

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

As motivation, note that when $r = \gamma$, the bound becomes

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

which is a Poincaré inequality.

Section 6 contains several classes of examples of Metropolis chains where moderate growth and local Poincaré inequalities can be effectively demonstrated. One of the main results of [9] shows that local Poincaré and moderate growth imply Nash inequalities and that these in turn lead to good upper and lower bounds on convergence. For simplicity, we state one result in continuous time; let $H_t^x(y) = e^{-t(I-K)}(x, y)$. This represents the chance of going from x to y in time t if the steps occur at the jumps of a Poisson process of rate 1.

THEOREM 4.3. *Assume a Markov chain (K, π) satisfies moderate growth and local Poincaré. Then, for all $t > 0$ and all x*

$$2\, \|H_t^x - \pi\| \leqslant a_1 e^{-t/a\gamma^2}$$

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

Remark. 1. Conversely (see [9]), there are constants a_2, a_3 depending on A, a, d of the definitions (4.1)(4.2), such that for all $t > 0$,

$$\sup_x \|H_t^x - \pi\| \geqslant a_2 e^{-a_3 t/\gamma^2}.$$

2. Any irreducible chain satisfies moderate growth and local Poincaré for some A, a, d. These constants enter exponentially into a_1, a_2, a_3 above. As shown in Section 5, it all fits together and gives good bounds for the classes of examples.

Sobolev inequalities are essentially equivalent to Nash inequalities. They ask for bounds of form

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

where $q > 2$ and C, T are constants. See [9] for the equivalence of Sobolev and Nash inequalities. See Chung and Yau [3] for a development of Sobolev inequalities on graphs.

Log–Sobolev inequalities give a tool that is not plagued by the curse of dimension. Indeed, these inequalities were invented by analysts in trying to get results in infinite dimensions. A splendid introduction and survey to the continuous work is in Gross [19]. The volume this is contained in has further useful articles. A careful account of log–Sobolev inequalities for finite problems is in [10], from which the present account is drawn.

We say a chain K satisfies a *log–Sobolev inequality* if

$$c\mathscr{L}(f) \leqslant \mathscr{E}(f|f)$$

for some $c > 0$ and all f with

$$\mathscr{L}(f) = \sum_x f^2(x) \log\left(\frac{f^2(x)}{\|f\|_2^2}\right) \pi(x).$$

The best constant $\alpha = \min(\mathscr{E}(f|f))/\mathscr{L}(f)$ is called the log–Sobolev constant.

If such an inequality is available, then the continuous time chain satisfies

$$2\,\|H_t^x - \pi\|^2 \leqslant \left(\log \frac{1}{\pi_*}\right) e^{-4\alpha t}$$

with $\pi_* = \min_x \pi(x)$. This is often a considerable improvement over (4.1).

Going from Poincaré/Cheeger inequalities to Nash/Sobolev inequalities necessitates more sophisticated use of available information; paths must be used locally and additional information such as polynomial growth of the underlying graph must be incorporated.

Good log–Sobolev inequalities are yet more difficult to prove. In fact, essentially the only nontrivial finite cases where this value is known is for Markov chains on a 2-point space, and the chain on the complete graph with all rows of k equal to π. See [10] for details. The situation is not all bad; the log–Sobolev inequality for the direct product of two Markov chains follows easily from this inequality for the factors. This gives the log–Sobolev inequality for the hypercube $\mathbb{Z}_2^d$. Further, log–Sobolev inequalities with poor constants can be extremely useful. Many mathematicians are working hard on these problems and there is much progress.

All of the proofs for the Metropolis algorithm for Ising models cited in Section 3 use log–Sobolev inequalities. In the language of Theorem 3.1, the authors show that both the log–Sobolev constant α and the spectral gap $\lambda = 1 - \beta_1$ are bounded below by const/n.

For another example, the running example in [10] is an analysis of the rate of convergence of the Metropolis algorithm for simulating from the binomial distribution on $\{0, 1, 2, ..., n\}$ from the base chain of nearest neighbor random walk. It is shown that order $n \log n$ steps are necessary and suffice for convergence to stationarity.

5. SAMPLING FROM LOG CONCAVE DENSITIES AND VOLUME APPROXIMATION

Let K be a compact, convex set in Euclidean space $\mathbb{R}^d$. Let f be a log concave probability density on K. For example, K might be the standard orthant where all coordinates are positive and f might be a standard normal density restricted to K. Consider the problem of sampling from f. This problem has been intensively studied in recent years in close connection with the problem of approximations to the volume of K. A comprehensive survey is given by Kannan [25]. We focus here on the parts of the work having to do with the Metropolis algorithm.

5.1. Discrete Algorithms

Frieze, Kannan, and Polson [15] discretized the problem, dividing $\mathbb{R}^d$ into hypercubes of size δ and running the Metropolis algorithm on a graph with vertices as the centers of cubes intersecting K, with an edge between vertices if the cubes are adjacent. The weight at center x is the average of f over the cube containing x.

They assume an approximation $\bar{f}(x)$ (defined only on the cube centers) is available which satisfies the approximation and continuity requirements for some $\alpha > 0$,

$$(1+\alpha)^{-1}\,\bar{f}(x) \leqslant \bar{f}(y) \leqslant (1+\alpha)\,\bar{f}(x) \quad \text{for adjacent points } x,\ y, \tag{5.1}$$

$$(1+\alpha)^{-1}\,\delta^d \bar{f}(x) \leqslant \int_{c(x)} f(z)\,dz \leqslant (1+\alpha)\,\delta^d \bar{f}(x), \tag{5.2}$$

$$(1+\alpha)\,\delta^{d-1} \bar{f}(x) \leqslant \int_{c(x)\cap c(y)} f(z)\,dz \leqslant (1+\alpha)\,\delta^{d-1} \bar{f}(x) \tag{5.3}$$

for $c(x)$, $c(y)$ cubes having $c(x) \cap c(y)$ of dimension $d-1$.

With these assumptions, it is sufficient to analyze the Metropolis algorithm with weight $\bar{f}(x)$ at x.

We state here a special case of their result, where $K = B(R)$, the Euclidean ball of radius R centered at 0, and where f satisfies the following assumption on its support. Consider the half line $L_u = \{ru: r \in \mathbb{R}^+\}$ with $u \in \mathbb{R}^d$. Let $h(r) = r^{d-1} f(ru)$ be defined for $r > 0$. This is a log concave function of r if f is log concave. The following assumption says that the tails of f are at the boundary of $B(R)$. Let $R = r(u) = |L_u \cap K|$. Let $R_1 = |L_u \cap T|$, where T is the union of all cubes that intersect K. It is assumed that $L_u \cap T$ is an interval, that $R_1 < 2R$, and, with $s = R_1 - R$,

$$h(r') \leqslant K_1 h(r) \text{ for } R - s \leqslant r \leqslant r' \leqslant r + s \text{ for some } K_1 \geqslant 1. \tag{5.4}$$

With these assumptions, the following result can be stated.

THEOREM 5.1 (Frieze, Kannan, and Polson). *Let f be a log concave probability density which is positive on $\mathbb{R}^d$ and satisfies* (5.1)–(5.4). *Let $M(x, y)$ be the Metropolis algorithm on the centers of cubes of side δ which intersect the ball $B(R)$. Assume $\delta \leqslant R$. Then*

$$\|M_x^k - \pi\|_{TV} \leqslant \bar{f}(x)^{-1/2}\,(1-\lambda)^k,$$

where

$$\lambda^{-1} = \max\left\{(1+\alpha)^3\,(K_0(K_1+1)+K_2 + 2\sqrt{K_0K_1K_2})\,\frac{n}{\delta}, \frac{\gamma^2}{6\delta^2}\right\} \sim \frac{d}{2}\left(\frac{\gamma}{\delta}\right)^2$$

for γ, *the Euclidean diameter of the set of cubes involved* (*the greatest distance between two such cubes*),

$$K_0 = \frac{\gamma}{4\delta}\,(\gamma + 2\sqrt{d}\,\delta),$$

K_1 *is from* (5.4), *and* $K_2 = 2K_1(s/\delta + \sqrt{d})$. *Here* s *is from* (5.4). *The final approximation holds as* $\alpha \searrow 0$, $K_1 \sim 1$, $K_2 \ll K_0$, *and* $\sqrt{n}/\gamma \to 0$.

Remarks. 1. This result is remarkable even in fixed dimensions for a Gaussian density. Then it basically says that a natural algorithm converges exponentially fast, in a useful sense, that is, with reasonable constants.

2. In high dimensions, observe that the constants do *not* get bad.

3. The above is a special case of the arguments. The restriction to balls or the restriction (5.4) are not required. The final result is more complicated to state.

4. In the end, the argument rests heavily on properties of convex sets in Euclidean space. It does not seem easy to adapt the tools involved to more general graphs. One interesting technical development seems broadly useful: a technique is introduced for dealing with a small "bad" set of the stated space, where, e.g., $\pi(x)$ is very small. This should not affect things, since basically the chain does not visit small sets. However, the usual conductance approach involves an infimum over all sets. A different, useful approach for treating a small bad set appears in Lovász and Simonovits [29].

5.2. Continuous Algorithms

Lovász and Simonovits have introduced a series of techniques for analyzing a Metropolis algorithm for sampling from a log concave density f on a compact convex set. A Convenient recent reference is [29]. Their work analyzes the following natural algorithm: suppose the chain is at x. Pick y from the uniform distribution on a ball of radius δ centered at x. If y is not in K, the walk stays at x. If y is in K, and $f(y)/f(x) \geqslant 1$, the walk moves to y. For $f(y)/f(x) < 1$, the usual Metropolis coin flip is executed. The chain moves to y or stays at x depending on the outcome.

This walk is analyzed without discretization. Following Lawler and Sokal [27] they work with the tools of conductance in general state spaces. This must prove useful. The heart of the argument is the same set of ideas about convex geometry in Euclidean spaces that are used by Frieze, Kannan and Polson. These have evolved from the original work of finding polynomial algorithms for volume computation due to Dyer, Frieze, and Kannan [12].

One main focus of [29] is getting good bounds on the complexity of volume computation (they get an order $n^7(\log n)^3$ algorithm). The Metropolis algorithm enters as a tool: for a convex body K, let $\varphi(x)$ be the smallest number t for which $x \in tK$. Set $f(x) = e^{-\varphi(x)}$. Then $\mathrm{Vol}(K) = 1/(n!)\int_{\mathbb{R}^n} f(x)\,dx$. Further, sampling from f gives an algorithm for approximating $\mathrm{Vol}(K)$.

Meyn and Tweedie [35] and Mengersen and Tweedie [33] have begun work on extending the tools of Harris recurrence to get useful quantitative results. They develop the theory for abstract spaces but do try a simple example of the Metropolis algorithm for sampling from the normal distribution on $\mathbb{R}$, the base chain being discrete time steps from a different normal. Rosenthal [37] develops results for general spaces that seem to give fairly good bounds.

6. LOW-DIMENSIONAL EXAMPLES

This section treats low-dimensional examples, probability distributions on a low-dimensional grid with nearest neighbor random walk "Metropolized" to the given stationary distribution.

Recall that a nearest neighbor walk on a grid of side length n takes order n^2 steps to reach stationarity in any fixed dimension. If the target distribution has an exponential (or faster) fall-off from a central peak, our analysis shows that the Metropolis chain reaches stationarity in order n steps. This is the fastest possible; the chain has to travel order n steps to go between opposite corners of the grid. For some distributions with polynomial fall-off from a certain peak, the analysis shows that a polynomial number of steps suffice to reach stationarity. Examples are given to show how these polynomials vary.

The analysis is described in some detail as an illustration of geometric methods described in Section 4 above. In the exponential case, one novelty is the use of different weights in the Cauchy–Schwarz inequality. This suggestion of Alan Sokal is shown to give improved results. In the polynomial case, the Nash inequalities of [9] are the driving tool.

6.1. Exponential Fall-off

To fix ideas, consider a one-dimensional grid $\mathfrak{X} = \{0, 1, 2, ..., n-1\}$. Let the base chain be the nearest neighbor random walk with holding $\frac{1}{2}$ at both ends. Represent the stationary distribution as

$$\pi(i) = z(a)\,a^{h(i)}, \qquad 0 < a < 1, \tag{6.1}$$

with $z(a)$ the normalizing constant. Assume

$$h(i+1)-h(i) \geqslant c \geqslant 1, \qquad 0 \leqslant i \leqslant n-2. \tag{6.2}$$

Thus $\pi(i)$ falls off at least exponentially from 0. Examples are $h(i)=i^b$, for $b \geqslant 1$. For π defined by (6.1) the Metropolis chain becomes

$$\left.\begin{aligned} M(i,i) &= \frac{1}{2} - \frac{a^{h(i+1)-h(i)}}{2} \\ M(i,i+1) &= \frac{a^{h(i+1)-h(i)}}{2} \\ M(i,i-1) &= \frac{1}{2} \end{aligned}\right\} \quad \text{for} \quad 1 \leqslant i \leqslant n-2,$$

$$M(0,0) = 1 - \frac{a^{h(1)-h(0)}}{2}, \qquad M(0,1) = \frac{a^{h(1)-h(0)}}{2}$$

$$M(n-1,n-2) = M(n-1,n-1) = \frac{1}{2}. \tag{6.3}$$

The main result is the following bound for the second eigenvalue of the chain.

PROPOSITION 6.1. *Assume* (6.1)–(6.3). *Then the second eigenvalue of the chain satisfies*

$$\beta_1 \leqslant 1 - \frac{(1-a^{c/2})^2}{2}.$$

Remark. Thus, the eigenvalue is bounded away from 1 uniformly in the size of the state space. In remarks following the proof, this will be used to show that order n steps are necessary and suffice for total variation convergence.

Proof. The argument uses the path techniques of [11] in a novel way. We have

$$1-\beta_1 = \min_f \frac{\mathscr{E}(f|f)}{\operatorname{var}(f)}$$

with the min taken over nonconstant f,

$$\operatorname{var}(f) = \tfrac{1}{2} \sum_{x,y} (f(x)-f(y))^2\, \pi(x)\, \pi(y),$$

and the Dirichlet form

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

for $Q(x,y) = \pi(x)\, M(x,y)$. Choose paths: for $x<y$, $\gamma_{xy} = (x, x+1, x+2, \ldots, y)$. The same path is used backwards to connect y to x. Then,

$$\begin{aligned} 2\operatorname{var}(f) &= \sum_{x,y} |f(x)-f(y)|^2\, \pi(x)\, \pi(y) \\ &= \sum_{x,y} \Bigg(\sum_{e \in \gamma_{xy}} [f(e^+)-f(e^-)] \Bigg)^2 \pi(x)\, \pi(y). \end{aligned} \tag{6.4}$$

The inner sum will be bounded by the Cauchy–Schwarz inequality. Usually, this is done with weights taken as 1 which gives a factor of $|\gamma_{xy}|$. The novelty here is to use weights depending on the stationary distribution. For the edge e, the weights are chosen as $Q(e)^\theta$. Subsequent calculations show that any fixed θ in $(0, \frac{1}{2})$ will do, e.g., $\theta = \frac{1}{4}$. To bring this out, we keep θ as a parameter. Multiply and divide $f(e^+)-f(e^-)$ in (6.4) by $Q(e)^\theta$. Writing $|\gamma_{xy}|_\theta = \sum_{e \in \gamma_{xy}} Q(e)^{-2\theta}$, we have

$$\begin{aligned} 2\operatorname{var}(f) &\leqslant \sum_{x,y} |\gamma_{xy}|_\theta \sum_{e \in \gamma_{xy}} Q(e)^{2\theta} \\ &\qquad \times (f(e^+)-f(e^-))^2\, \pi(x)\, \pi(y) \\ &= \sum_e (f(e^+)-f(e^-))^2\, Q(e)\, Q(e)^{2\theta-1} \\ &\qquad \times \sum_{\gamma_{xy} \ni e} \pi(x)\, \pi(y)\, |\gamma_{x,y}|_\theta \\ &\leqslant 2A\mathscr{E}(f|f) \end{aligned}$$

with

$$A = \max_e Q(e)^{2\theta-1} \sum_{\gamma_{xy} \ni e} \pi(x)\, \pi(y)\, |\gamma_{xy}|_\theta.$$

To bound A, observe first that $Q(i,i+1) = Q(i+1,i) = \pi(i+1)/2$. Next, the dominant term in $|\gamma_{xy}|_\theta$ is $Q(y-1,y)^{-2\theta}$ (for $x<y$). Pull this out and bound the ratio with the other terms using (6.2):

$$|\gamma_{x,y}|_\theta \leqslant \frac{(\pi(y)/2)^{-2\theta}}{1-a^{2c\theta}}.$$

Suppose that $e=(i,i+1)$. The quantity to be bounded is

$$2^{2\theta}(1-a^{2c\theta})^{-1}\, Q(e)^{2\theta-1} \sum_{\substack{0 \leqslant j \leqslant i \\ i+1 \leqslant k \leqslant n}} \pi(j)\, \pi(k)^{1-2\theta}.$$

The sum in k is bounded above by

$$\frac{\pi(i+1)^{1-2\theta}}{1-a^{c(1-2\theta)}}.$$

The sum in j is bounded above by 1. Combining bounds, we have

$$A \leqslant \frac{2}{(1-a^{c(1-2\theta)})(1-a^{2c\theta})};$$

choosing $\theta = \frac{1}{4}$ gives the bound announced.

Remarks. 1. In Proposition 6.1 the stationary distribution was chosen to have its maximum at 0. The same argument works if the maximum is taken on at any point in $\mathfrak{X}$. Thus $h(i)$ decreases up to x_0 and increases past x_0; the analog of (6.2) is assumed.

2. The easiest upper bound for total variation using the eigenvalue bound of Proposition 6.1 is as follows. First, bound the smallest eigenvalue $\beta_{n-1} \geqslant -1 + 2 \min M(i, i) \geqslant -1 + 2(\frac{1}{2} - a^c/2) = -a^c$. Thus,

$$\beta_* = \max(\beta_1, |\beta_{n-1}|) \leqslant \max\left(1 - \frac{(1-a^{c/2})^2}{2}, a^c\right).$$

Now, the upper bound at (4.1) gives for any $x \in \mathfrak{X}$

$$2\, \|M_x^k - \pi\| \leqslant \pi(x)^{-1/2}\, \beta_*^k.$$

This is correct (up to constants) when $h(x) = x$; it says order n steps are necessary and suffice to reach stationarity for any starting position. If h grows faster, e.g., $h(x) = x^2$, the bound shows that for a walk starting at $x = 0$, order n steps suffice. For walks starting at $n-1$ the bound shows order $h(n-1)$ steps are sufficient. This is off. The following argument shows how to conclude that for total variation convergence order n steps suffice for any starting position, provided h satisfies (6.2). Consider the walk started at $n-1$. This essentially stays still or goes left. It is straightforward to show that the chance that the walk hits 0 in the first $3n$ steps is exponentially close to 1, with constants depending only on $(1-a)$. Once the walk hits zero, the argument above shows it is close to stationarity in at most order n further steps. This shows order n steps suffice for variation distance convergence. We do not know how many steps are required to make the ℓ^2 norm small.

3. The argument for Proposition 4.1 is fairly robust and will handle many variations. It does depend on the roughly unimodal nature of π. (See Section 6.3 below for more on this.) There are techniques in Deuschel and Mazza [5] and Ingrassia [21] for bounding essentially arbitrary π. While these bounds are sharp (in the sense that there are examples where they cannot be improved in nice examples such as those of Proposition 6.1) they can be very far off, suggesting that exponentially many steps are needed. Much remains to be done in giving useful tools for natural examples.

4. Consider the special case when $h(i) = i$. Then Belsley [1] and Silver [41] have essentially given a complete analysis of the Metropolis chain (6.3). We state some of their results here to compare with what comes out of general theory.

To begin with, the Metropolis chain reduces to biased reflecting random walk on $\{0, ..., n-1\}$. The eigenvalues and eigenvectors are classically known. For large n, $\beta_1 \sim 1 - (1+a)/2 + \sqrt{a}$. Proposition 6.1 gives the bound $\beta_1 \leqslant 1 - (1 - a^{1/2})^2/2$. This is actually equal to the value of β_1, up to terms of order $1/n^2$.

The bound (4.1) gives two upper bounds for the total variation distance; one using all the eigenvectors and eigenvalues and the second using just the second largest eigenvalue. Figure 1 shows the actual bounds as a function of the number of steps k for $a = 0.3$, $a = 0.9$. In both cases $N = 50$ and the starting state was chosen as 50 also. Figure 1 also shows the exact total variation (dotted).

We see that the bound using the second eigenvalues is off by a factor of 5 or more when $a = 0.9$ and off by a factor of about 2 when $a = 0.3$. The bound using all the spectral data does better.

Belsley [1] has worked out sharp asymptotics for variation convergence in this case. He shows that for an explicit $b(a)$, $b(a)\, n + c(a)\, \sqrt{n}$ steps are necessary and suffice: if $c(a)$ is large and positive, the variation distance is close to zero. If $c(a)$ is large and negative, the variation distance is close to one.

5. The restriction $c \geqslant 1$ in (6.2) is made for simplicity. If $h(i+1) - h(i) \geqslant c > 0$, then $h(i+1)/c - h(i)/c \geqslant 1$ and the chain with a replaced by a^c and h replaced by h/c satisfies the conditions. This leads to the bound

$$\beta_1 \leqslant 1 - \frac{(1-a^{c^2/2})^2}{2} \qquad \text{for} \quad c > 0.$$

6. The argument goes through more or less as above for two-dimensional versions with $h(i, j)$, falling off at least linearly from a single peak. Here one chooses paths which move from x to y, first making the first coordinates equal, then the second coordinates equal, and so on. We hope to carry out a detailed analysis of the multimodal case on grids in low dimension.

7. For the one-dimensional case, it is worth pointing out that Cheeger's inequality can be used to give results similar to those in Propositon 6.1. Lawler and Sokal [27] do this when $h(i) = i$ and generalize to trees. See [11, Section 3] for further details. For higher-dimensional grids, we find paths much easier to work with.

8. In light of the results for sampling from log concave distributions in the continuous case (Section 5.2 above), it is natural to inquire how this type of condition works in

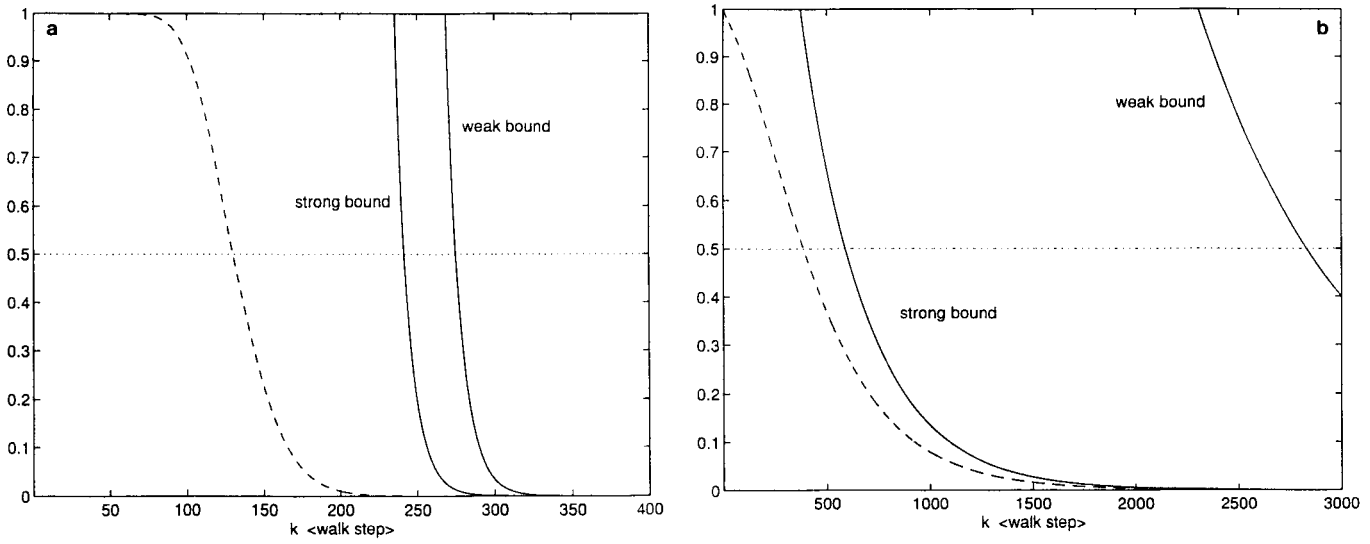

FIG. 1. (a) Total Variation Distance and Bounds Plot—$a = 0.3$, $N = 50$, $n = 50$; (b) Total Variation Distance and Bounds Plot—$a = 0.9$, $N = 50$, $n = 50$.

Proposition 6.1. While natural examples are easy to treat, the following shows that some care is needed. Consider the symmetric binomial distribution $\pi(i) = \binom{n-1}{i}/2^{n-1}$ on $\{0, 1, 2, ..., n-1\}$, with base chain reflecting random walk. The Metropolis chain is easily comparable to the classical Ehrenfest chain. The analysis shows the Metropolis chain has $c_1/n \leqslant 1 - \beta_1 \leqslant c_2/n$ for explicit constants c_1, c_2. The difference is this: the binomial falls off from its peak at $n/2$ exponentially, but at scale $\sqrt{n}$. It is (roughly) flat in a $\sqrt{n}$ neighborhood of $n/2$. The exponentials treated by Proposition 6.1 fall off exponentially at scale 1. A careful analysis carried out in [10] shows that order $n \log n$ steps are necessary and suffice for convergence in the binomial case.

6.2. *Polynomial Fall-off*

Consider $\mathfrak{X} = \{1, 2, ..., n\}$, with the base chain of nearest neighbor random walk with holding $\frac{1}{2}$ at both ends. We begin with a simple example. Take the stationary distribution

$$\pi(i) = zi, \qquad 1 \leqslant i \leqslant n, \qquad z^{-1} = n(n+1)/2. \tag{6.5}$$

Thus $\pi(i)$ rises linearly from 1. The Metropolis chain becomes

$$\left.\begin{aligned} &M(i, i-1) = (i-1)/(2i) \\ &M(i, i) = 1/(2i) \\ &M(i, i+1) = 1/2 \end{aligned}\right\} \quad \text{for} \quad 2 \leqslant i \leqslant n-1$$

$$M(1, 1) = M(1, 2) = 1/2$$

$$M(n, n-1) = (n-1)/(2n), \qquad M(n, n) = 1 - (n-1)/(2n). \tag{6.6}$$

The following result shows that the walk (6.6) reaches stationarity in order n^2 steps. This is the same rate as the base chain.

PROPOSITION 6.2. *There are explicit positive constants A, B, C, D such that the Metropolis chain* (6.6) *satisfies*

$$Ae^{-Bk/n^2} \leqslant \max_x \|M_x^k - \pi\| \leqslant Ce^{-Dk/n^2}$$

for all positive integer k, n.

Proof. We apply the geometric tools of [9]. Consider $\mathfrak{X}$ as a graph with an edge from i to $j+1$, $1 \leqslant i \leqslant n-1$. Write $|x-y|$ for the graph distance between x and y. Let $B(x, r) = \{y : |x-y| \leqslant r\}$ and $V(x, r) = \sum_{y \in B(x, r)} \pi(y)$. The diameter of $\mathfrak{X}$ is $\gamma = n-1$.

As in Section 4, a graph and stationary distribution have $(A; d)$ *moderate growth* if $V(x, r) \geqslant (1/a)((r+1)/\gamma)^d$ for all $x \in \mathfrak{X}$, and $r = \{0, 1, ..., \gamma\}$. An elementary verification shows that the Metropolis chain has (6,2) moderate growth.

For a real function f defined on $\mathfrak{X}$ and positive r, set

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

We will verify in Section 6.3 below that the chain satisfies a local Poincaré inequality:

$$\|f - f_i\|_2^2 \leqslant ar^2 \mathscr{E}(f|f) \qquad \text{with} \quad a = 4. \tag{6.7}$$

Finally, the smallest eigenvalue satisfies $\beta_- \geqslant -1 + 2\min(M(i, i)) \geqslant -1 + 1/n$.

For reversible chains satisfying moderate growth and local Poincaré inequalities, Theorem 4.3 shows that order (diameter)^2 steps are necessary and suffice for convergence. This result gives Proposition 6.2.

Remarks. 1. Very similar bounds can be obtained for stationary distributions of form $\pi(i) = zp(i)$, for positive polynomial p. Some of this is explored in Section 6.3 below.

2. Preliminary computations indicate that similar bounds hold for higher-dimensional grids when the stationary distribution has a unique maximum and polynomial decay. Order (diameter)^2 steps are necessary and sufficient to reach stationarity.

6.3. Some Variations and Extensions

In Sections 6.1 and 6.2 above we explored fairly well specified examples. In this section we give some general results for unimodal distributions on $\mathfrak{X} = \{1, 2, ..., n\}$. For distributions with a unique maximum, much of the above goes through. For U-shaped stationary distributions, some new behavior occurs. The results use the geometric arguments of Section 4. While we are stating them as results about the Metropolis algorithm, we note that all of the examples here are birth and death chains and our results may easily be re-said as giving an analysis of some general classes of birth and death chains.

The first few propositions are about stationary distributions with a unique local maximum.

PROPOSITION 6.3. *Let π have a unique local maximum at $k \in \{1, 2, ..., n\}$. Let M be the Metropolis chain for π based on nearest neighbor random walk. Then*

$$\beta_1(M) \leqslant 1 - \frac{1}{2n^2}, \tag{6.8}$$

(M, π) satisfies a local Poincaré inequality (Definition 4.2) with $a = 4$, so

$$\|f - f_r\|_2^2 \leqslant 4r^2 \mathscr{E}(f|f). \tag{6.9}$$

Proof. Choose paths γ_{xy} as in the proof of Proposition 6.1. Then, using the Poincaré inequality with weight 1, $\beta_1 \leqslant 1 - 1/A$ with

$$A = \max_e \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\, \pi(y)\, |\gamma_{xy}|.$$

Take the edge $e = (i, i+1)$. Consider the case $i + 1 \leqslant k$. Then

$$Q(e) = \tfrac{1}{2} \min\{\pi(i), \pi(i+1)\} = \tfrac{1}{2}\pi(i).$$

Now, bounding $|\gamma_{xy}| \leqslant n$,

$$\frac{1}{Q(i, i+1)} \sum_{\gamma_{xy} \ni (i, i+1)} \pi(x)\, \pi(y)\, |\gamma_{xy}| \leqslant 2n \left(\sum_{x \leqslant i} \frac{\pi(x)}{\pi(i)} \right) \left(\sum_{y \geqslant i+1} \pi(y) \right).$$

Since $\pi(x)/\pi(i) \leqslant 1$, for $x \leqslant i \leqslant k$ the first bracketed term above is at most n. The second bracketed term is at most 1; hence the bound. If $k \leqslant i$, the argument works as well. This proves (6.8).

To prove (6.9) we use the Poincaré argument locally as in Lemma 5.2 of [9]. This gives

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

with

$$\eta(r) = \max_e \frac{2}{Q(e)} \sum_{\substack{|\gamma_{xy}| \leqslant r \\ \gamma_{xy} \ni e}} \frac{\pi(x)\, \pi(y)\, |\gamma_{xy}|}{V(x, r)}.$$

We must thus show $\eta(r) \leqslant 4r^2$. We proceed as above. If $i + 1 \leqslant k$,

$$\frac{2}{Q(i, i+1)} \sum_{\substack{|\gamma_{xy}| \leqslant r \\ \gamma_{xy} \ni e}} |\gamma_{xy}| \frac{\pi(x)\, \pi(y)}{V(x, r)} \leqslant 4r \sum_{\substack{x \leqslant i \\ |x - i| \leqslant r}} \frac{\pi(x)}{\pi(i)} \sum_{\substack{y \geqslant i+1 \\ |y - x| \leqslant r}} \frac{\pi(y)}{V(x, r)}$$

If $k \leqslant i$,

$$\frac{2}{Q(i, i+1)} \sum_{\substack{|\gamma_{xy}| \leqslant r \\ \gamma_{xy} \ni e}} |\gamma_{xy}| \frac{\pi(x)\, \pi(y)}{V(x, r)} \leqslant 4r \sum_{\substack{y \geqslant i \\ |y - i + 1| \leqslant r}} \frac{\pi(y)}{\pi(i)} \sum_{\substack{x \leqslant i \\ |x - y| \leqslant r}} \frac{\pi(x)}{V(x, r)}.$$

The term $\pi(y)/\pi(i)$ is smaller than 1. To bound

$$S = \sum_{\substack{x \leqslant i \\ |x - y| \leqslant r}} \frac{\pi(x)}{V(x, r)} \qquad \text{with} \quad y \geqslant i + 1,$$

observe that $\{i-r, ..., i-1\} \subset B(x, r)$ because $i-r \leqslant x \leqslant i-1$. Hence $S \leqslant 1$,

$$\frac{2}{Q(i, i+1)} \sum_{\substack{|\gamma_{xy}| \leqslant r \\ \gamma_{xy} \ni e}} |\gamma_{xy}| \frac{\pi(x)\,\pi(y)}{V(x, r)} \leqslant 4r \sum_{\substack{y \geqslant i \\ |y-i+1| \leqslant r}} 1 \leqslant 4r^2,$$

and the proof is complete.

Proposition 6.3 gives a very general eigenvalue bound. Of course, this bound can be off, as the examples in Section 6.1 show. The examples in Section 6.2 show that this bound can be sharp in natural examples. Proposition 6.3 also gives local Poincaré inequalities quite generally. We turn next to conditions for moderate growth (Definition 4.2).

PROPOSITION 6.4. *Let π be any probability distribution on $\{1, 2, ..., n\}$. Let $\tilde{\pi}$ be the nondecreasing rearrangement of π. Assume that for some $d > 0$, $\tilde{\pi}(x)/x^d$ is decreasing in x. Then, π has $(4^{d+1}, d+1)$ moderate growth.*

Proof. The argument uses the following elementary fact. If f, g are positive functions on $\{1, 2, ..., n\}$ and f/g is decreasing, then F/G is also decreasing; here

$$F(y) = \sum_{x \leqslant y} f(x), \qquad G(y) = \sum_{x \leqslant y} g(x).$$

Clearly $V(x, r) \geqslant \tilde{V}(1, r) = \tilde{\pi}(1) + \cdots + \tilde{\pi}(r)$. Now the elementary fact and the hypothesis given show

$$\frac{\tilde{V}(1, r)}{N(r)} = \frac{\tilde{V}(1, r)}{\sum_{x \leqslant r} x^d} \qquad \text{is decreasing.}$$

Let

$$N(r) = \sum_{1 \leqslant x \leqslant r} x^d,$$

so

$$\frac{r^{d+1}}{d+1} \leqslant N(r) \leqslant \frac{(r+1)^{d+1}}{d+1}.$$

We have

$$\frac{\tilde{V}(1, r)}{N(r)} \geqslant \frac{\tilde{V}(1, n)}{N(n)} = \frac{1}{N(n)}$$

so

$$V(i, r) \geqslant \frac{N(r)}{N(n)} \geqslant \left(\frac{r}{n+1}\right)^{d+1} \geqslant \frac{1}{4^{d+1}} \left(\frac{r+1}{n}\right)^{d+1}.$$

The 4^{d+1} uses the crude bound $(r+1)/r \leqslant 2$, $(n+1)/n \leqslant 2$ for $r \geqslant 1$ (the case $r = 0$ can be checked directly).

Remark. As an example, suppose $\pi(i)$ is proportional to i^d. Then, Propositions 6.3 and 6.4 combine with Theorem 4.2 to show that order n^2 steps are necessary and suffice to achieve stationarity. The same conclusions hold if $\pi(i)$ is rearranged to take a unique local maximum in the middle of $\{1, 2, ..., n\}$.

It is natural to wonder what happens for multimodal distributions. We show that even a simple U-shaped distribution with both sides of the "U" linear can slow things down. To fix ideas, take $n = 2k+1$, $k \geqslant 2$, and set $\mathfrak{X} = \{0, ..., n\}$,

$$\pi(x) = \frac{|x - n/2| + 1/2}{c(n)}, \qquad c(n) = \frac{(n+1)(n+3)}{4}. \tag{6.10}$$

THEOREM 6.5. *The Metropolis chain M for π at* (6.10) *with base chain nearest neighbor random walk on $\mathfrak{X} = \{0, 1, ..., n\}$ satisfies*

$$c_1 e^{-c_2 t/(n^2 \log n)} \leqslant \max_x \|M_t^x - \pi\| \leqslant c_3 e^{-c_4 t/(n^2 \log n)}.$$

Proof. We begin by proving

$$\beta_1 \leqslant 1 - \frac{1}{(n+1)(n+3)[2 + \log((n-1)/2)]}. \tag{6.11}$$

To prove Eq. (6.11) use the Poincaré inequality with weights $Q(e)^{1/2}$ as in Section 6.1. This gives $\beta_1 \leqslant 1 - 1/A$ with

$$A = \max_{(i, i+1)} \sum_{\substack{x \leqslant i \\ y \geqslant i+1}} |\gamma_{xy}|_{1/2}\, \pi(x)\, \pi(y)$$

with

$$|\gamma_{xy}|_{1/2} = \sum_{e \in \gamma_{xy}} \frac{1}{Q(e)}.$$

Clearly

$$\begin{aligned} A &\leqslant \max_{x, y} |\gamma_{xy}|_{1/2} \leqslant 4c(n) \sum_{j=0}^{k} \left(\left|j + 1 - \frac{n}{2}\right| + \frac{1}{2}\right)^{-1} \\ &= 4c(n) \left(1 + \sum_{j=1}^{k} \frac{1}{j}\right) \\ &\leqslant 4c(n)(2 + \log[(n-1)/2]). \end{aligned}$$

We next show directly that the chain M, π satisfies a Nash inequality

$$\|f\|_2^{2(1+2/d)} \leqslant a_1 n^2 \left\{ \mathscr{E}(f|f) + \frac{a_2}{n^2} \|f\|_2^2 \right\} \|f\|_1^{4/d} \tag{6.12}$$

with a_1, a_2 universal constants and $d=2$.

To prove (6.12), break $\{0, ..., k, k+1, ..., n\}$ into Part I: $\{0, ..., k\}$ and Part II: $\{k+1, ..., n\}$. Call M_1, M_2 the kernels restricted to the two halves. For f a function on $\{0, ..., n\}$ call f_1, f_2 the restrictions to the two halves and $\pi_1 = 2\pi$, $\pi_2 = 2\pi$ the stationary distributions. Now

$$\|f_1\|_{2,\pi_1}^2 + \|f_2\|_{2,\pi_2}^2 = 2\,\|f\|_{2,\pi}^2, \tag{6.13}$$

$$\mathscr{E}_1(f_1, f_1) + \mathscr{E}_2(f_2, f_2) \leqslant 2\mathscr{E}(f|f). \tag{6.14}$$

From (6.13), (6.14) it is enough to prove (6.12) for M_1, M_2. But we know this because each of these chains has moderate growth and satisfies local Poincaré (Propositions 6.3 and 6.4). Then Theorem 5.7 of [9] shows (6.12) holds for each half and so for M.

Now, as explained in Section 4, the Nash inequality (6.12) implies that

$$\|M_t^x/\pi\|_2 \leqslant a_3 \qquad \text{for} \quad t \geqslant n^2$$

(Theorem 3.1 of [9]). This and the eigenvalue bound combine (Lemma 1.2 of [9]) to show that for $t = a_5 n^2 + a_6(n^2 \log n)\, c$, $c > 0$,

$$2\,\|M_t^x - \pi\| \leqslant \left\| \frac{M_t^x}{\pi} - 1 \right\|_2 \leqslant a_4 e^{-c}.$$

This gives the required upper bound.

For the lower bound, observe that the chain started at 1 takes order $n^2 \log n$ steps to hit k by an easy birth and death chain argument. Thus $\|M_t^x - \pi\| \geqslant \frac{1}{2}$ for $t < \varepsilon n^2 \log n$. Further details are omitted.

Remark. The eigenvalue estimate (6.11) is of the right order. To see this, use $1 - \beta_1 = \min(\mathscr{E}(f|f)/\mathrm{var}(f)) \leqslant \mathscr{E}(f|f)/\mathrm{var}(f)$ for any particular f. Choose

$$f(x) = \begin{cases} \log\left(\left|x - \dfrac{n}{2}\right| + \dfrac{1}{2}\right), & \text{if} \quad x \leqslant k, \\ -\log\left(\left|x - \dfrac{n}{2}\right| + \dfrac{1}{2}\right), & \text{if} \quad x \geqslant k+1. \end{cases}$$

Then straightforward calculus gives $E_\pi(f) = 0$,

$$\mathrm{var}(f) \geqslant \frac{3}{4c(n)} (k+1)^2 \left(\log \frac{k+1}{2}\right)^2,$$

$$\mathscr{E}(f|f) \leqslant \frac{2}{c(n)} \log \frac{n+1}{2}.$$

These give

$$\beta_1 \geqslant 1 - \frac{a_5}{n^2 \log n}.$$

7. FINAL REMARKS

The Metropolis algorithm is the most widely used way of changing the output of a Markov chain into a sampling mechanism with a given stationary distribution π. Hastings [20] determined a large class of such mechanisms. To explain his result, use the notation of Lemma 1.1. Let F be a function from $\mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $F(cu, cv) = cF(u, v)$, $F(u, v) = F(v, u)$, $F(u, v) \leqslant \min(u, v)$. For example $F(u, v) = uv/(u+v)$ or $F(u, v) = \min(u, v)$. Given a kernel $K(x, y)$ as in Lemma 1.1 define

$$M_F(x, y) = K(x, y)\, F(1, A(x, y)),$$

$$A(x, y) = \frac{\pi(y)\, K(y, x)}{\pi(x)\, K(x, y)}, \qquad x \neq y.$$

Then, it is straightforward to show that $M(x, y)$ is revresible with π as its stationary distribution.

The usual Metropolis chain has $F = \min(u, v)$. Jun Liu and Alan Sokal in personal communications have shown that $uv/(u+v)$ (called Barker dynamics) is the same thing as the Gibbs sampler when applied to the usual Ising model.

It is natural to ask which of these procedures works best. Peskun [36] gives an elegant extremal characterization of the Metropolis algorithm in this class of chains. For $f\colon \mathfrak{X} \to \mathbb{R}$ a function of interest, the limiting variance of the usual estimate of the mean value of f is

$$\sigma^2(f) = \lim_{n \to \infty} n\, \mathrm{var} \left\{ \frac{1}{n} \sum_{i=1}^n f(X_i) \right\},$$

where $X_1, X_2, ...$ is a realization of the chain.

Consider two chains P_1, P_2 with the stationary distribution π. Call P_1 better than P_2 if $\sigma^2(f, P_1) \leqslant \sigma^2(f, P_2)$ for all f. Peskun [36] proves that the Metropolis algorithm is best in Hastings class of chains. His proof uses the following elegant theorem: Let P_1, P_2 be irreducible, reversible Markov chains with respect to π. If $P_1(x, y) \leqslant P_2(x, y)$ for

all $x \neq y$ then P_2 is better than P_1. This is a careful way of saying that an algorithm that holds less gets random faster. In contrast, note that [16] shows that if one chain is said to dominate a second if the first chain's second eigenvalue is smaller, then the Metropolis algorithm is not always best; things depend on temperature.

It is natural to compare the various dynamics in simple examples to see how their rates of convergence compare. In unpublished work, Jeff Silver has shown that any of Hastings variations can be analyzed from the base chain of simple random walk on an n-point path. To get an exponential stationary distribution the analysis of Section 6A goes through without criminal difficulties to give the bound

$$\beta_1(F) \leqslant 1 - \frac{(1-a^{1/2})^2}{2} F(1, a^{-1}).$$

We thus see that the convergence is (roughly) as quick for any of these chains, e.g., order n steps are necessary and sufficient for convergence. Of course, $F(1, a^{-1}) \leqslant 1$, so the straight Metropolis is fastest. Silver [41] extends these computations to a variety of other stationary distributionss on finite and infinite sets. He also shows some ways in which to beat the usual Metropolis chain.

Heuristically, one wants to choose the base chain K so that its stationary distribution is close to π. It is natural to try to estimate π, and change the base chain as information about π comes in. Gilks, Best, and Tan [18] is an early interesting effort in this direction. There is much to do here.

We have not attempted tic survey other, closely related algorithms for sampling from π. To begin with, for low-dimensional examples such as those of Section 6, there is a large body of competitive technology. In high dimensions, Glauber dynamics (known as the Gibbs sampler) is a closely related method that is beginning to have some useful finite sample convergence result. See Rosenthal [37] and the references cited there. There are many further ideas in the statistical physics literature. Sokal [45] gives a useful review of cluster algorithms, multigrid Monte Carlo and other techniques. Browsing through recent years of the *Journal of Statistical Physics* will reveal hundreds of other methods and variations.

All of these are fair game for careful mathematical analysis

ACKNOWLEDGMENTS

We thank Jeffrey Silver for several technical contributions to the present paper and Alan Sokal for extensive comments and technical contributions used in Section 6. We also thank Jiun Liu, Jim Fill, and Susan Holmes. An earlier version of this paper appeared in the 27*th STOC Meeting*. We thank Ravi Kannan and Allan Borodin for their help.

REFERENCES

1. E. Belsley, "Rates of Convergence of Markov Chains Related to Association Schemes," Ph.D. dissertation, Department of Mathematics, Harvard University, 1993.
2. F. Chung, "Spectral Graph Theory," Am. Math. Soc., Providence, RI, 1996.
3. F. Chung and S. T. Yau, Eigenvalues of graphs and Sobolev inequalities, *Combin. Probab. Comput.* **4** (1995), 11–25.
4. D. Critchlow, "Metric Methods for Analyzing Partially Ranked Data," Springer Lecture Notes in Statistics, Vol. B4, Springer-Verlag, Berlin, 1985.
5. J. D. Deuschel and C. Mazza, L^2 convergence of time non-homogeneous Markov processes. I. Spectral estimates, *Ann. Appl. Prob.* **4** (1994), 1012–1056.
6. P. Diaconis, "Group Representations in Probability and Statistics," Institute of Mathematical Statistics, Hayward, CA, 1988.
7. P. Diaconis and P. Hanlon, Eigenanalysis for some examples of the Metropolis algorithm, *Contemp. Math.* **138** (1992), 99–117.
8. P. Diaconis and L. Saloff-Coste, Comparison techniques for reversible Markov chains, *Ann. Appl. Probab.* **3** (1993), 696–730.
9. P. Diaconis and L. Saloff-Coste, Nash inequalities for finite Markov chains, *J. Theor. Probab.* **9** (1996), 459–510.
10. P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities and finite Markov chains, *Ann. Appl. Probab.* **6** (1996), 695–750.
11. P. Diaconis and D. Stroock, Geometric bounds for eigenvalues for Markov chains, *Ann. Appl. Probab.* **1** (1991), 36–61.
12. M. Dyer, A. Frieze, and R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies, *J. Assoc. Comput. Mach.* **38** (1991), 1–17.
13. J. Fill, Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process, *Ann. Appl. Probab.* **1** (1991), 62–87.
14. M. Fligner and J. Verducci, "Probability Models and Statistical Analysis for Ranking Data," Springer Lecture Notes in Statistics, Vol. 80, Springer-Verlag, New York, 1993.
15. A. Frieze, R. Kannan, and N. Polson, Sampling from log concave distributions, *Ann. Appl. Probab.* **4** (1994), 812–837.
16. A. Frigessi, C. Hwang, S. Sheu, and P. Di Stefano, Convergence rates of the Gibb sampler, the Metropolis algorithm and other single-site updating dynamics, *Jour. Roy. Stat. Soc. B* **55** (1993), 205–219.
17. A. Frigessi, F. Martinelli, and J. Stander, Computational complexity of Markov chain Monte Carlo methods, technical report, IAC, Rome, (1993).
18. W. Gilks, W. Best, and K. Tan, Adapted rejection Metropolis sampling within Gibbs sampling, *Jour. Roy. Stat. Soc. C* **44** (1995), 455–472.
19. L. Gross, Logarithmic Sobolev inequalities and contractivity properties of semigroups, Lecture Notes in Mathematics, Vol. 1563, pp. 54–82, Springer-Verlag, New York/Berlin, 1993.
20. W. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57** (1970), 97–109.
21. S. Ingrassia, On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds, *Ann. App. Probab.* **4** (1994), 347–389.
22. M. Jerrum and A. Sinclair, Approximating the permanent, *SIAM J. Comput.* **18** (1989), 19–78.
23. M. Jerrum and A. Sinclair, Approximate counting, uniform generation and rapidly mixing Markov chains, *Inform. and Comput.* **82** (1989), 93–133.
24. M. Jerrum, Large cliques elude the Metropolis process, *Random Struct. Algorithms* **3** (1992), 347–359.
25. R. Kannan, Markov chains and polynomial time algorithms, *in* "Proc. 35th FOCS," pp. 656–671, IEEE, Los Alamos, CA, 1994.
26. S. Karlin and H. Taylor, "A First Course in Stochastic Processes," 2nd ed., Academic Press, New York, 1975.

27. G. Lawler and A. Sokal, Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheegers inequality, *Trans. Amer. Math. Soc.* **309** (1988), 557–580.
28. J. Liu, Metropolized independent sampling and comparisons to rejection sampling and importance sampling, *Statist. and Comput.* **6** (1995), 113–119.
29. L. Lovász and M. Simonovits, Random walks in a convex body and an improved volume algorithm, *Random Struct. Algorithms* **4** (1993), 359–412.
30. F. Martinelli, On the two dimensional dynamical Ising Model in the phase coexistence region, *J. Statist. Phys.* **76** (1994), 1179–1246.
31. F. Martinelli, E. Olivieri, and R. Schonmann, For 2D lattice spin systems, weak mixing implies strong mixing, *Commun. Math. Phys.* **165** (1994), 33–47.
32. B. McCoy and T. Wu, "The Two-Dimensional Ising Model," Harvard University Press, Cambridge, MA, 1973.
33. K. Mengersen and R. Tweedie, Rates of convergence of the Hastings and Metropolis algorithms, *Ann. Statist.* **24** (1996), 101–121.
34. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.* **21** (1953), 1087–1092.
35. S. Meyn and R. Tweedie, Computable bounds for geometric convergence rates of Markov chains, *Ann. Appl. Probab.* **4** (1994), 981–1011.
36. P. Peskun, Optimum Monte Carlo sampling using Markov chains, *Biometrika* **60** (1973), 607–612.
37. J. Rosenthal, Minorization conditions and convergence rates for Markov chain Monte Carlo, *J. Am. Statist. Assoc.* **90** (1995), 558–566.
38. K. Ross and D. Xu, Hypergroup deformations and Markov chains, *J. Theor. Probab.* **7** (1994), 813–830.
39. R. Schonmann, Slow droplet-driven relaxation of stochastic Ising models in the vicinity of the phase coexistence region, *Commun. Math. Phys.* **161** (1994), 1–49.
40. R. Schonmann, Theorems and conjectures and the droplet-driven relaxation of stochastic Ising models, *in* "Probability and Phase Transition, Proceedings of the Nato Advanced Study Institute on Probability Theory of Spatial Disorder and Phase Transition, at the University of Cambridge, July 1993" (G. Grimmett, Ed.), pp. 265–301, 1995.
41. J. Silver, "Rates of Convergence for the Metropolis Algorithm and Its Variations," Ph.D. dissertation, Department of Mathematics, Harvard University, 1995.
42. B. Simon, "The Statistical Mechanics of Lattice Gases, I," Princeton University Press, Princeton, 1993.
43. A. Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flow, *Combin. Prob. Comput.* **1** (1992), 351–370.
44. A. Sinclair, "Algorithms for Random Generation and Counting: A Markov Chain Approach," Birkhauser, Boston, 1993.
45. A. Sokal, Monte Carlo methods in statistical mechanics: Foundations and new algorithms, Cours de troisième cycle de la physique en suisse romande, Available as lecture notes, Department of Physics, NYU, 1989.
46. D. Stroock and B. Zegarlinski, The logarithmic Sobolev inequality for spin system on a lattice, *Comm. Math. Physics* **149** (1992), 175–194.
47. L. Thomas, Bound on the mass gap for finite volume stochastic Ising models at low temperatures, *Comm. Math. Physics* **126** (1989), 1–11.