

Numerical Results for the Metropolis Algorithm

Persi Diaconis and J. W. Neuberger

CONTENTS

- 1. Introduction
 - 2. The Example
 - 3. Discrete Approximations to T
 - 4. Results of Computations
- References

This paper deals with the spectrum of an operator associated with a special kind of random walk. The operator is related to the Metropolis algorithm, an important tool of large-scale scientific computing. The spectrum of this operator has both discrete and continuous parts. There is an interesting challenge due to the fact that any finite-dimensional approximation has only eigenvalues. Patterns are presented which give an idea of the full spectrum of this operator.

1. INTRODUCTION

The Metropolis algorithm [Metropolis et al. 53] is a mainstay of scientific computing. Indeed it appears first on a list of the “Top Ten Algorithms” [Sullivan 00]. It gives a method for sampling from probability distributions on high-dimensional spaces when these distributions are only known up to a normalizing constant. For background and references to extensive applications in physics, chemistry, biology, and statistics, see [Diaconis and Saloff-Coste 98] and [Billera and Diaconis 01].

One basic question is how long the algorithm should be run to have done its job properly, since the Metropolis algorithm involves iterations of a stochastic kernel (Markov chain). This question is illuminated by knowledge of the spectrum. There has been some study of the spectral gap for the relevant operator. This is surveyed in [Diaconis and Saloff-Coste 98]; also see [Billera and Diaconis 01] and [Mengersen and Tweedie 96]. As far as we know, the higher eigenvalues and eigenfunctions have *not* been investigated.

This paper presents the first analysis of the full spectrum (and associated eigenfunctions) for the Metropolis algorithm on a continuous state space. The example is the simplest possible but, as will emerge, the spectrum is full of mysterious possibilities.

2. THE EXAMPLE

We begin by explaining the example as it would be used by a practitioner. Our problem is to choose a point from

2000 AMS Subject Classification: Primary 65C05, 65F15; Secondary 47A10

Keywords: Metropolis algorithm, random walk, continuous spectrum

the uniform distribution (Lebesgue measure) on $[-a, a]$ with $a > 1$. The Metropolis algorithm will be based on a proposed distribution which is a simple random walk. Thus from x , choose a point y uniformly in $[x - 1, x + 1]$. If this point remains in $[-a, a]$, the chain moves to y . If the point is outside $[-a, a]$, the chain stays at x . If this simple procedure is repeated, the associated point becomes uniformly distributed on $[-a, a]$. Technically, the Metropolis algorithm generates a reversible Markov chain with the uniform distribution on $[-a, a]$ as stationary distribution. Using methods explained in [Tierney 94], this chain can be shown to be Harris-recurrent and so converges exponentially fast, in total variation distance, to its stationary distribution.

To give quantitative estimates of the rate of convergence, it is useful to introduce a stochastic operator $T : H \rightarrow H$ where H is $L^2([-a, a])$. For $f \in H$, the operator is defined as

$$(Tf)(t) = \begin{cases} (Tf)(t) = \frac{1}{2} \int_{t-1}^{t+1} f, & -a + 1 \leq t \leq a - 1 \\ (Tf)(t) = \frac{1}{2} \int_{-a}^{t+1} f + \frac{1-a-t}{2} f(t), & -a \leq t < -a + 1 \\ (Tf)(t) = \frac{1}{2} \int_{t-1}^a f + \frac{1+t-a}{2} f(t), & a - 1 < t \leq a \end{cases}$$

for almost all $t \in [-a, a]$.

It should be clear that T is not compact since

$$Tf = \rho f + Sf,$$

where

$$\rho(t) = \begin{cases} 0, & t \in [-a + 1, a - 1], \\ \frac{1-a-t}{2}, & t \in [-a, -a + 1], \\ \frac{1+t-a}{2}, & t \in [a - 1, a], \end{cases}$$

and

$$(Sf)(t) = \frac{1}{2} \int_{\max(t-1, -a)}^{\min(t+1, a)} f, \quad t \in [-a, a].$$

The transformation S is compact but the operation of multiplication by ρ is not and hence the sum T of these two transformations is not compact.

There has been some previous study of the spectral gap of such random walk Metropolis algorithms by [Kienetz 00], [Miclo and Roberto 00], and [Yuen 00]. Each of the authors gives techniques that allow a proof of the following bound.

The operator T has eigenvalue 1 with multiplicity 1. The rest of the spectral measure is concentrated on $[-1 + \frac{c}{a^2}, 1 - \frac{c}{a^2}]$, for an explicit positive constant c . In [Diaconis et al. 03], it is shown that the spectrum near $1 - \frac{c}{a^2}$ is

discrete. Of course, since T is a compact perturbation of a multiplication operator, there is also a continuous spectrum.

3. DISCRETE APPROXIMATIONS TO T

We discretize the operator T as follows.

Pick a positive integer $a > 1$ and choose a positive integer $n > 1$. Consider the evenly spaced grid G_n of mesh $1/n$ imposed on $[-a, a]$. This grid has $2an + 1$ points. For n a positive integer, define $T_{a,n} : R^{2na+1} \rightarrow R^{2na+1}$ as follows: for $x = (x_{-an}, x_{-an+1}, \dots, x_{an-a}, x_{an})$

$$(T_{a,n})_k x = \frac{1}{2n+1} [(2n - (\min(an, k+n) - \max(-an, k-n)))x_k + \sum_{j=\max(-an, k-n)}^{\min(an, k+n)} x_j],$$

$$k = -an, \dots, an.$$

We are particularly interested in how the spectra of $T_{a,n}$, as $n \rightarrow \infty$, points to facts concerning the spectrum of T . A little background on this issue is in order.

In [Neuberger and Renka 99] we calculated approximations to eigenfunctions and eigenvectors for the Neumann Laplacian L on a fractal region (Rooms and Passages region; see [Edmunds and Evans 87]). The region was designed so that L , restricted to the orthogonal complement of the constant vectors, had continuous spectra in addition to eigenvalues (L^{-1} is not compact). It is a particular challenge to calculate such continuous (essential) spectra, since any finite dimensional approximation to the underlying region carries with it an operator that has only eigenvalues, continuous spectra being impossible in finite dimensions. The paper by Neuberger and Renka [Neuberger and Renka 99], and the present one, both seek to contribute to an understanding of computational phenomena when such spectra is present. The underlying idea is that finite dimensional spectra on increasingly finer approximate regions in a sense point to continuous spectra. It is hoped that the present case study of such an instance will be useful in the eventual formulation of a theory for numerically indicating continuous spectra.

We use the power method, with deflation ([Wilkinson 65]), to calculate eigenvalues and eigenvectors of numerical approximations to T . For our graphical report, we pick the interval $[-20, 20]$ for our calculations, i.e., we choose $a = 20$ in the above. The code we have written belongs to a family of codes starting with [Neuberger

and Noid 87] and including [Lapidus et al. 96] and [Neuberger and Renka 99]. These last three papers concern eigenvalues for $L = -\Delta + V$, V a potential function. These papers use the inverse power method for such a transformation L on regions of various shapes in various Euclidean spaces, i.e., the power method applied to L^{-1} . In the present case, T already has somewhat the form of a Green's function so that the power method is indicated (rather than the inverse power method)—hence the similarity of the present code with codes for these last three papers.

4. RESULTS OF COMPUTATIONS

The first 200 eigenvalues and eigenvectors for $a = 20$, $n = 20, 40, 80$ were calculated. Figures 1, 2, and 3 show plots of these 200 eigenvalues of $T_{a,n}$ for each of these cases. The scale on the horizontal axis is eigenvalue number, the first eigenvalue being the largest. The next 23 or so eigenvectors follow the familiar pattern of those for a vibrating string, i.e., each one after the first contains one-half of a cycle more than the preceding one. After the twenty-fourth or so eigenvalue has been calculated, patterns shift dramatically. This corresponds to the break in pattern at eigenvalue 24 in each of Figures 1, 2, and 3. In each of the cases $n = 20, 40, 80$, following the twenty-fourth eigenvalue for awhile, the period remains nearly the same as in eigenvector 24 but with some blow-up evidenced at both endpoints of $[-a, a]$ (see detail for the right endpoint in Figure 8—the left endpoint is similar). It is noted that the blow-up near the right endpoint of $[-a, a]$ shifts to the left as eigenvalue number increases (see Figure 9 for detail). It is not shown in the figures that from the twenty-fourth eigenvector up to the seventy-sixth eigenvalue one observes (still for the case $n = 40$) that wavelengths decrease much more slowly than in the 1–24 series of eigenvectors. They appear to

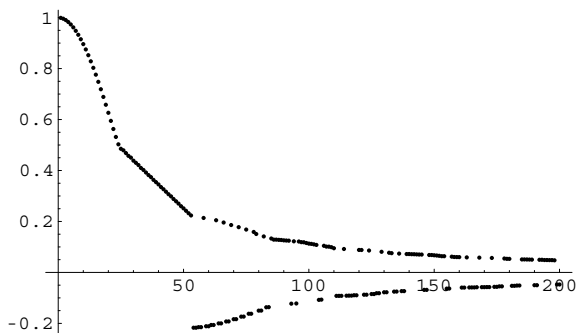


FIGURE 1. Eigenvalues 1–200, $a = 20$, $n = 20$.

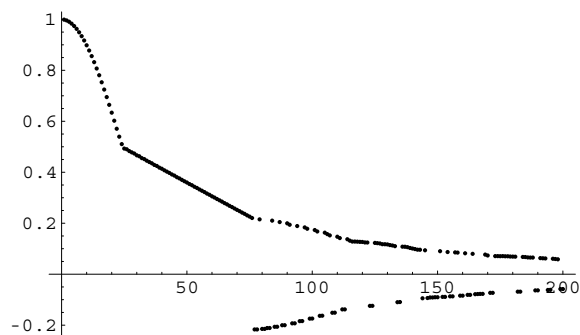


FIGURE 2. Eigenvalues 1–200, $a = 20$, $n = 40$.

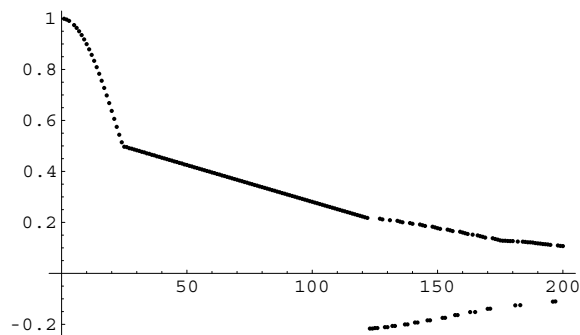


FIGURE 3. Eigenvalues 1–200, $a = 20$, $n = 80$.

come in groups of six or seven eigenvectors each, there being nearly a common wavelength in each group.

In each of the Figures 1, 2, and 3, following the twenty-fourth eigenvalue, there is what appears to be a nearly straight line of eigenvalues, persisting up to eigenvalue 50 for $n = 20$, eigenvalue 77 for $n = 40$, and eigenvalue 122 for $n = 80$. Call these lines L_{20} , L_{40} , and L_{80} , respectively. It is striking that the apparent slope of L_{40} appears to be $\frac{1}{2}$ of that of L_{20} , that of L_{80} to be $\frac{1}{2}$ of that of L_{40} . This phenomenon persists up to the highest values of n for which calculations have been made: $n = 200, 400, 800$. We would like very much to be able to follow up on a referee's suggestion to make mathematics supporting this observation of slopes falling in half as n doubles. We have to leave this as a problem for readers. Behavior of eigenvectors with eigenvalues in the range approximately $[\frac{1}{23}, \frac{1}{5}]$ points to continuous spectrum in this interval, possibly the entire interval.

For $n = 40$, a striking change occurs at eigenvalue 77 in the appearance of a negative eigenvalue (eigenvalues are ordered according to decreasing magnitude—the same order in which they are calculated by our FORTRAN code). Note the break in the pattern at eigenvalue 77 in Figure 2 as well as the form of the seventy-seventh eigenvector given in Figure 10. As seen in Figure 2, after a negative eigenvalue appears, there are still

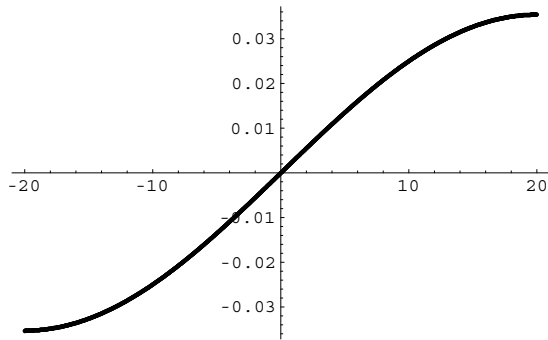


FIGURE 4. Eigenvector 1, Eigenvalue = .99894, $a = 20$, $n = 40$.

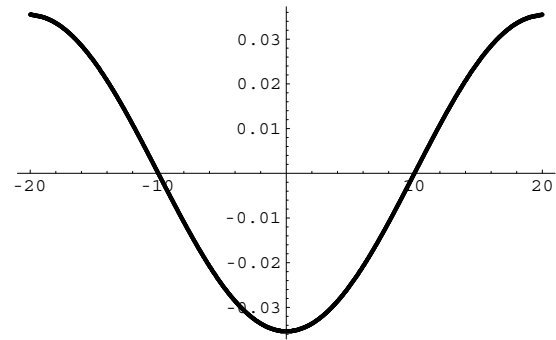


FIGURE 5. Eigenvector 2, Eigenvalue = .99579, $a = 20$, $n = 40$.

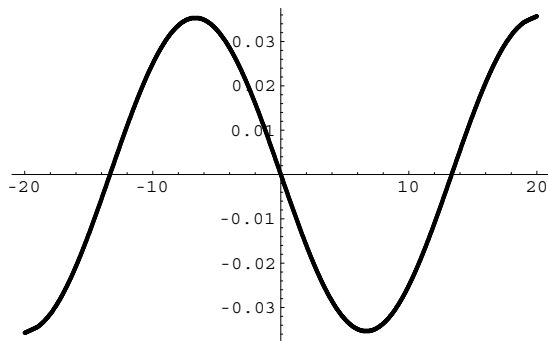


FIGURE 6. Eigenvector 3, Eigenvalue = .99056, $a = 20$, $n = 40$.

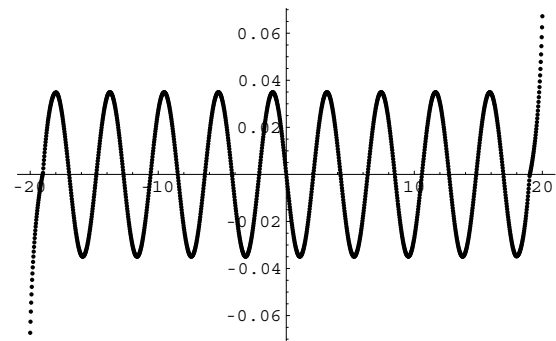


FIGURE 7. Eigenvector 19, Eigenvalue = .66466, $a = 20$, $n = 40$.

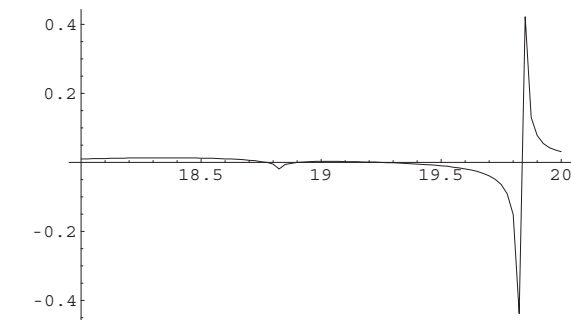


FIGURE 8. Eigenvector 40, Eigenvalue = .41369, $a = 20$, $n = 40$.

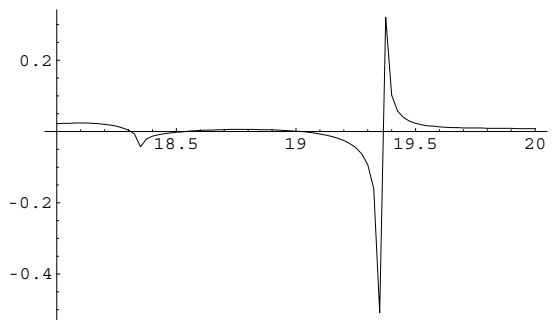


FIGURE 9. Eigenvector 98, Eigenvalue = .17825, $a = 20$, $n = 40$.

positive eigenvalues interspersed with the negative ones; indeed the ninety-eighth eigenvector (Figure 9) is typical. Eigenvectors corresponding to negative eigenvalues follow in a series. For $n = 40$, eigenvector 78 appears as an out-of-phase version of eigenvector 77—these two eigenvectors likely share a common eigenvalue. With Figure 15 depicting eigenvector 149 we see a splitting into two packets with nearly the same wavelength as in the

previous figure. For eigenvectors corresponding to negative eigenvalues, this pattern persists until eigenvalue 147 (Figure 14), where a much smaller wavelength appears. Eigenvector 148 is not shown but it is an out-of-phase version of eigenvector 147. Following eigenvector 147, a double packet appears in Figure 15. The wavelength in the double packet appears to be very close to that of eigenvector 147. Figure 17 shows detail of a blow-up at

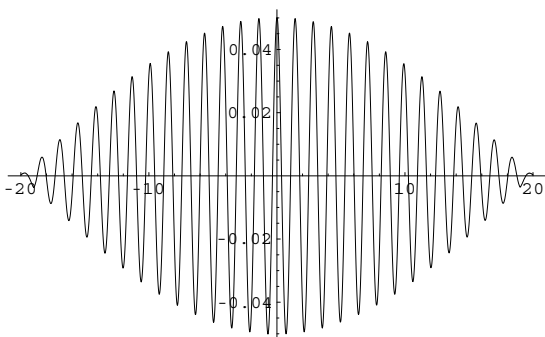


FIGURE 10. Eigenvector 77, Eigenvalue = $-.21665$, $a = 20$, $n = 40$.

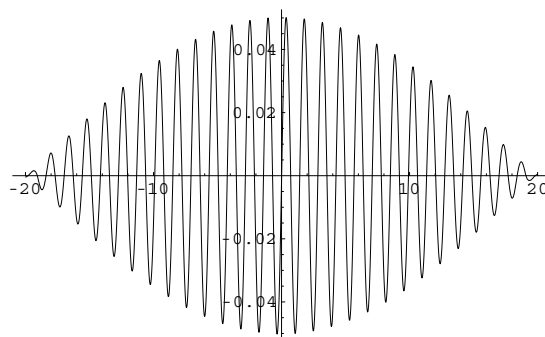


FIGURE 11. Eigenvector 78, Eigenvalue = $-.21665$, $a = 20$, $n = 40$.

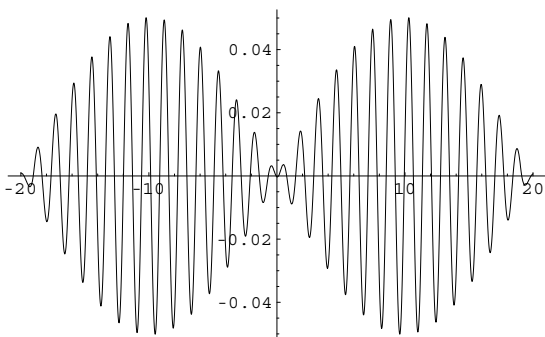


FIGURE 12. Eigenvector 80, Eigenvalue = $-.21455$, $a = 20$, $n = 40$.

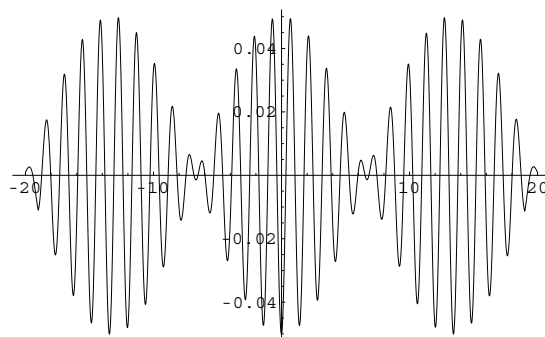


FIGURE 13. Eigenvector 82, Eigenvalue = $-.21111$, $a = 20$, $n = 40$.

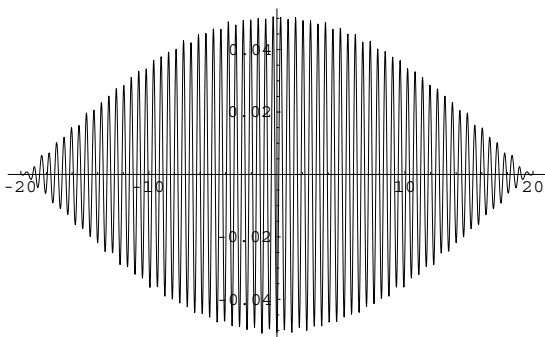


FIGURE 14. Eigenvector 147, Eigenvalue = $-.09130$, $a = 20$, $n = 40$.

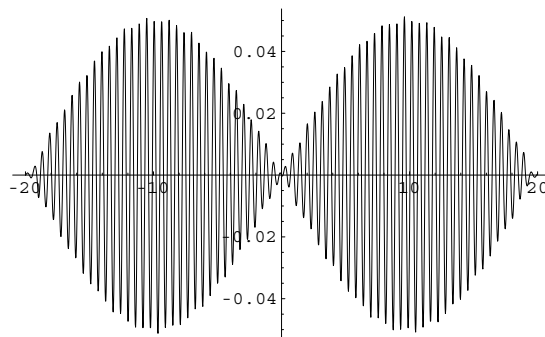


FIGURE 15. Eigenvector 149, Eigenvalue = $-.09041$, $a = 20$, $n = 40$.

the right end (the left end is similar) for the eigenvector in Figure 16. Figure 18 shows the start of a third series of eigenvectors for negative eigenvalues, again with a much smaller wavelength than in the previous series.

Above the twenty-fourth eigenvector (for $n = 40$ to be sure), all eigenvectors with positive eigenvalues have evidence of blow-up near the endpoints of $[-a, a]$. Furthermore, as noted, this blow-up shifts as eigenvalue number increases. We consider that both the presence of blow-ups and the shifting of positions of these blow-ups are

indicative of continuous spectra associated with eigenvalues in $[0, .5]$

Patterns reported above for $n = 40$ carry over completely to $a = 10, 30$. For $a = 20$, pictures of eigenvectors 1–24 are essentially independent of n and eigenvectors for negative eigenvalues have their counterparts for $a = 20, 30$.

For $a = 10, 20, 30$, we have, respectively, 12, 24, 36 eigenvalues not exceeding $.5$, independent of choice of n . The general pattern for various values of a is repeated

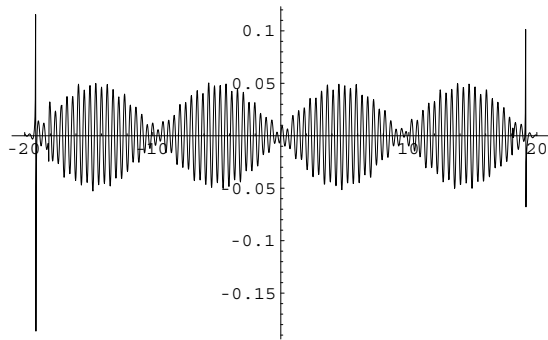


FIGURE 16. Eigenvector 185, Eigenvalue = .06739, $a = 20$, $n = 40$.

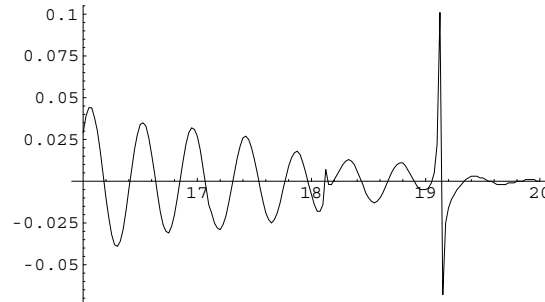


FIGURE 17. Eigenvector 185, Eigenvalue = .06739, $a = 20$, $n = 40$.

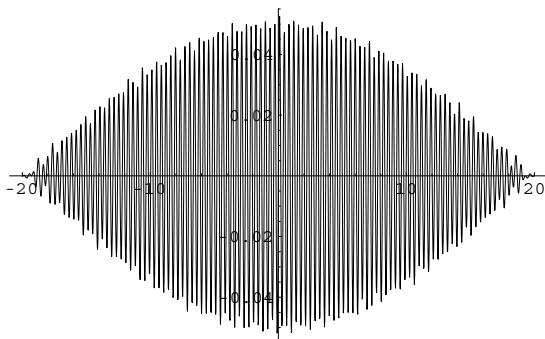


FIGURE 18. Eigenvector 200, Eigenvalue = -.05822, $a = 20$, $n = 40$.

but, the greater a is, the longer we must wait to see various modes found in our figures. We have made calculations up to $n = 800$ with no essentially new features to report.

We summarize our speculations on how computational results in this note might be related to facts about essential spectra of T . It seems rather certain for the case $a = 20$, that T itself has about 24 positive eigenvalues (after the zeroth eigenvalue 1 with constant eigenfunction) ranging from nearly 1 down to around .5. It is conjectured that an interval approximately $[0, .5]$ contains essential spectra, perhaps filling approximately $[\cdot23, .5]$. As to the interval $[0, \cdot23]$, since Figures 1–3 indicate some symmetry between spectral distributions in $[0, \cdot23]$ and $[-\cdot23, 0]$, we speculate that both of these intervals contain eigenfunctions of the operator T . Since calculated eigenvectors for negative eigenvalues appear not to have blow-up, it may be that there is no essential spectrum below $\cdot23$ (or so). We hope that a full resolution of questions raised here will be forthcoming. We also hope that our calculations, together with a fuller mathematical ex-

planation, will contribute to a deeper understanding of the important Metropolis algorithm.

REFERENCES

- [Billera and Diaconis 01] L. Billera and P. Diaconis. “A Geometric Interpretation of the Metropolis Algorithm.” *Stat. Sci.* 20 (2001), 1–5.
- [Diaconis and Saloff-Coste 98] P. Diaconis and L. L. Saloff-Coste. “What Do We Know about the Metropolis Algorithm?” *Jour. Computer and Systems Sciences* 57 (1998), 20–36.
- [Diaconis et al. 03] P. Diaconis, J. Keller, and P. Lax. “On the Spectrum of a Simple Metropolis Algorithm. To appear, 2003.
- [Edmunds and Evans 87] D. E. Edmunds and W. D. Evans. *Spectral Theory and Differential Operators*. Oxford, UK: Oxford Sci. Pub., 1987.
- [Kienetz 00] J. Kienetz. “Convergence of Markov Chains via Analytic and Isoperimetric Inequalities.” PhD diss., Universität Bielefeld, 2000.
- [Lapidus et al. 96] M. L. Lapidus, J. W. Neuberger, R. J. Renka, and C. A. Griffith. “Snowflake Harmonics and Computer Graphics: Numerical Calculation of Spectra on a Fractal Drum.” *Int. J. Bifurcation and Chaos* 6 (1996), 1185–1210.
- [Mengersen and Tweedie 96] K. Mengersen and R. Tweedie. “Rates of Convergence of the Hastings and Metropolis Algorithms.” *Ann. Statist.* 24 (1996), 101–121.
- [Metropolis et al. 53] N. Metropolis, A. Rosenbluth, A. Rosenbluth, A. Teher, and E. Teher. “Equations of State Calculations by Fast Computing Machines.” *J. Chem. Phys.* 21 (1953), 1087–1092.

- [Miclo and Roberto 00] L. Miclo and C. Roberto. “Trous spectraux pour certains algorithmes de Metropolis sur R .” In *Séminaire de Probabilités, XXXIV*, pp. 336–352, Lecture Notes in Math. 1729. Berlin: Springer, 2000.
- [Neuberger and Noid 87] J. W. Neuberger and D. W. Noid. “Numerical Calculation of Eigenvalues for the Schrödinger Equation III.” *J. Comp. Chem.* 8 (1987), 459–461.
- [Neuberger and Renka 99] J. W. Neuberger and R. J. Renka. “Numerical Calculation of the Essential Spectrum of a Laplacian.” *Exp. Math.* 8 (1999), 301–308.
- [Sullivan 00] F. Sullivan. “Great Algorithms of 20th Century Scientific Computing.” *Computing in Science and Engineering* 2 (2000), special issue.
- [Tierney 94] L. Tierney. “Markov Chains for Exploring Posterior Distributions (with discussion).” *Ann. Statist.* 22 (1994), 1701–1762.
- [Wilkinson 65] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford, UK: Oxford Univ. Press, 1965.
- [Yuen 00] W. K. Yuen. “Applications of Geometric Bounds to the Convergence Rate of Markov Chains on R^n .” *Stoch. Proc. Appl.* 87 (2000), 1–23.

Persi Diaconis, Department of Mathematics, Stanford University, Palo Alto, CA 94305-4065

J. W. Neuberger, Department of Mathematics, University of North Texas, Denton, TX 76203-1430 (jwn@unt.edu)

Received November 4, 2003; accepted in revised form March 21, 2004.

