The Annals of Probability
1995, Vol. 23, No. 4, 1939–1949

RANDOM WALKS ON THE GROUPS OF UPPER TRIANGULAR MATRICES

BY RICHARD STONG

Rice University

This paper gives sharp bounds on the eigenvalues of a natural random walk on the group of upper triangular $n \times n$ matrices over the field of characteristic p, an odd prime, with 1's on the diagonal. In particular, this includes the finite Heisenberg groups as a special case. As a consequence we get bounds on the time required to achieve randomness for these walks. Some of the steps are done using the geometric bounds on the eigenvalues of Diaconis and Stroock. However, the crucial step is done using more subtle and idiosyncratic techniques. We bound the eigenvalues inductively over a sequence of subspaces.

Introduction. There has been a great deal of work done in the past few years on bounding the eigenvalues of Markov chains by geometric techniques [2, 3, 5]. These techniques use the minimax characterization of the second largest eigenvalue to reduce the problem to studying the Dirichlet form of the random walk. Using the Cauchy–Schwarz inequality in a pattern guided by geometric data (a choice of paths) produces geometric bounds on the second eigenvalue. While this technique has been very successful, there are nevertheless many Markov chains for which the bounds attained are not sharp. In this paper we consider one such example. We will show that by using the Cauchy–Schwarz inequality in a very different fashion we can obtain bounds of the right order. Since the results of [2] and [3] allow one to bound the eigenvalues of a Markov chain by comparing it to a Markov chain with known eigenvalues, it may be possible to use the results of this paper to bound the eigenvalues of related Markov chains. Also it is reasonable to suppose that other Markov chains for which the geometric bounds fail to be sharp could also be approached by a different use of Cauchy–Schwarz. The exact form of this approach would presumably depend heavily on the particular problem.

Let p be an odd prime and let $G_n \subset \mathrm{GL}(n, p)$ be the group of upper triangular $n \times n$ matrices over the finite field $\mathbf{Z}/p\mathbf{Z}$ with 1's on the diagonal. The special cases $n = 3$ are often referred to as the finite Heisenberg groups. The upper triangular matrices are one of the simplest families of non-Abelian nilpotent groups. For this reason they are an obvious first example to consider to understand random walks on nilpotent groups. These groups have a natural symmetric generating set. Let E_k, $1 \le k \le n - 1$ be the $n \times n$ matrix with 1's on the diagonal and in the $(k, k + 1)$ entry and 0's elsewhere.

Received January 1993; revised January 1995.
AMS 1991 *subject classification*. 60J10.
Key words and phrases. Eigenvalues, random walk.

Then $S = (E_1^{\pm 1}, \dots, E_{n-1}^{\pm 1}\}$ is a generating set for G_n. This generating set gives a random walk on G. That is, start at the identity matrix and at each step multiply on the left by a (uniformly) randomly chosen statement of S. Let P_n be the transition probability matrix for this random walk on G_n. Let $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1} \geq -1$ be the eigenvalues of P_n, where $N = |G_n| = p^{n(n-1)/2}$. Let U be the uniform distribution on G_n.

This random walk in the case $n = 3$ was first discussed by Zack [9]. Her goal was to study this example and to understand how one could use a non-Abelian group to produce fast random number generators. The extension to the case of general n was proposed to the author by Professor Diaconis as a model for random additions. We can define it instead as: start with the identity matrix and at each step randomly add a row to the row above it or subtract that row from the row above it. The family of examples G_n is also interesting when compared with recent results of Diaconis and Saloff-Coste [4]. They considered convergence of random walks on finite groups with "moderate rate of growth" conditions, such as nilpotent groups. Suppose G is a nilpotent group with a symmetric set of n generators (containing the identity), class number l and diameter γ. Let P be the transition probability matrix for the associated random walk on G. In terms of eigenvalues, Diaconis and Saloff-Coste show ([3], Corollary 5.3) that there are constants $B = B(l, n)$ and $\eta = \eta(l, n)$ such that

$$1 - \frac{B}{\gamma^2} \leq \lambda_1 \leq 1 - \frac{\eta}{\gamma^2}$$

(and a similar bound for λ_{N-1}). For families of nilpotent groups with n and l fixed this shows that the first eigenvalue goes like $1 - C/\gamma^2$. (Diaconis and Saloff-Coste also give impressive convergence results that, for families of nilpotent groups with n and l, fixed order γ^2 steps are necessary and sufficient to achieve randomness.) For the upper triangular groups G_n, both n and l are unbounded and the results of [4] do not apply. In fact, by a clever argument of Ellenberg [7] there are positive constants c and C such that the diameter γ of G_n satisfies

$$c(np + n^2 \log p) \leq \gamma \leq C(np + n^2 \log p).$$

Thus from the bounds proved in this paper we see that the conclusions of [4] do not hold in this example if n is allowed to increase.

The goal of this paper will be to prove the following theorem.

THEOREM 1. *There are constants c_1, c_2 and c_3 such that*

$$1 - \frac{c_1}{p^2 n} \geq \lambda_1 \geq 1 - \frac{c_2}{p^2 n}$$

and

$$\lambda_{N-1} \geq -1 + \frac{c_3}{p^2}.$$

Therefore, there are constants c and C such that if $k = cp^2n^3 \log p + p^2ns$, *then*

$$\|P_n^k - U\| < e^{-Cs}.$$

The bound on the convergence rate in Theorem 1 uses the crudest result for bounding convergence in terms of eigenvalues and is therefore probably not sharp. At present the correct rate is not known. It is easy to see that a time of order at least n^2 is necessary. For example, we have to wait about n steps for the first occurrence of the matrix E_{n-1}. This will be the first time the $(n-1, n)$ entry will be nonzero. We have to wait about n steps for the next occurrence of the matrix E_{n-2}. This is the first time the $(n-2, n)$ entry might be nonzero. Continuing in this way, in order for the $(1, n)$ entry of our matrix to be nonzero, we must have waited at least n^2 steps. At present this crude lower bound is about as good as any lower bound the author is aware of. It is possible that one could refine the bounds below to control all the eigenvalues and obtain a better upper bound, but this seems very hard.

Geometric bounds. The lower bound on λ_{N-1} claimed in Theorem 1 is a straightforward application of the geometric bounds of Diaconis and Stroock [6]. In addition, these techniques give an upper bound on λ_1 which is not sharp. The limits on this method are discussed below. These results are summarized by the following lemma.

LEMMA 1. *There is a constant* c_3 *such that* $\lambda_{N-1} \geq -1 + c_3/p^2$.

PROOF. For the lower bound on λ_{N-1} we apply Corollary 2 of [6], page 41. To do so we must choose for every $A \in G_n$ a loop of odd length based at A (or more generally, for every A a collection of loops with weights such that the weights add to 1 [6], Remark 2.2, page 46). Then we have

$$\lambda_{N-1} \geq -1 + \frac{2}{d_* \sigma_* b_*}$$

where d_* is the maximum degree of a vertex, σ_* is the length of the longest cycle and b_* is the maximum over all directed edges of the sum of the weights of cycles that contain that edge. For every A choose the $2n-2$ cycles [each with weight $1/(2n-2)$] of the form

$$A, E_k^{\pm 1}A, E_k^{\pm 2}A, \dots, E_k^{\pm p-1}A, E_k^{\pm p}A = A.$$

For this choice of loops, $d_* = 2n-2$, $\sigma_* = p$ and $b_* = p/(2n-2)$. The last of these follows since every edge is in the same number p of cycles each of which has weight $1/(2n-2)$. Therefore, Corollary 2 of [6] gives

$$\lambda_{N-1} \geq -1 + \frac{2}{d_* \sigma_* b_*} = -1 + \frac{2}{p^2}.$$

For comparison consider applying the methods of [6] to upper bound λ_1. For the upper bound on λ_1, we need to choose for each pair of elements $x, y \in G$ a path from x to y. Then Corollary 1 of [6] states that

$$\lambda_1 \leq 1 - \frac{2|E|}{d_*^2 \gamma_* b},$$

where $|E|$ denotes the number of edges, d_* denotes the maximum degree, γ_* denotes the diameter and b is the maximum number of paths going over any directed edge. Let γ be the average length of the paths chosen. Then one necessarily has $b \geq N^2\gamma/|E|$. Therefore, the upper bound above is bounded below by $1 - 2(\gamma_* \gamma)^{-1}$. Suppose A and B are in G_n. If BA^{-1} has nonzero $(1, n)$ entry (which occurs with probability $1 - p^{-1}$), then one must use each of the E_i at least once to get from A to B. Hence the path from A to B must have length at least $(n - 1)$. Therefore, $\gamma_* \geq n - 1$ and $\gamma > (p - 1)(n - 1)/p$. [A little extra work will raise these bounds. The underlying graph is regular of degree $2(n - 1)$ and the graph contains $p^{n(n-1)/2}$ vertices. Therefore, $\gamma_* > \bar{\gamma} > Cn^2/\log n$. The exact diameter is not immediately clear.] Hence the upper bound which the methods of [6] will give is bounded below by $1 - 2p/((p - 1)(n - 1)^2)$. Thus as a consequence of the results below, geometric bounds cannot give a sharp upper bound on λ_1. □

Lower bounds on the second largest eigenvalue. For the bounds on λ_1 consider the following construction of an orthonormal basis for $L^2(G_n)$. The subgroup G_n has an interesting property. In addition to being a group, it is also an affine subspace of the vector space of all $n \times n$ matrices over $\mathbf{Z}/p\mathbf{Z}$. Specifically it is the translate by I of the subspace U of strictly upper triangular matrices. Therefore, left multiplication of $G_n = I + U$ on itself gives an affine left action of G_n on U and a right action of G_n on the dual space U^*. Identify the space U^* with the strictly lower triangular matrices by taking the pairing of $M \in U^*$ and $A \in U$ to be given by $(M, A) = \operatorname{tr}(MA) = \Sigma_{i<j} M_{ji} A_{ij}$. We can restrict this pairing to a map $U^* \times G_n \to \mathbf{Z}/p\mathbf{Z}$ given by the same formulas. With these identifications the action of $A \in G_n$ on $M \in U^*$ (which we will denote $M \cdot A$) produces the matrix MA truncated so as to be strictly lower triangular. (It would probably be more natural to regard U^* as all matrices modulo the upper triangular matrices. This would remove the need to truncate.)

Let $\omega = \exp\{2\pi i/p\}$ and for any $M \in U^*$ define the function $\varphi_M \in L^2(G_n)$ by

$$\varphi_M(A) = \omega^{(M, A)}/N.$$

Note that $\varphi_0(A) = 1/N$ is the stationary distribution π for the Markov chain and that the φ_M $(M \in U^*)$ form an orthonormal basis for $L^2(G_n)$ with respect to the natural inner product $\langle f, g \rangle = N\Sigma_{A \in G_n} \overline{f(A)} g(A)$. Furthermore, the functions φ_M behave very well under the action of the Markov chain on

$L^2(G_n)$. Specifically, since $(M, E_k A) = M_{k+1\,k} + (M \cdot E_k, A)$, we have

$$(1) \qquad P_n \varphi_M = \frac{1}{2(n-1)} \sum_{k=1}^{n-1} \left\{ \omega^{M_{k+1\,k}} \varphi_{M \cdot E_k} + \omega^{-M_{k+1\,k}} \varphi_{M \cdot E_k^{-1}} \right\}.$$

There are two immediate consequences of (1). The first is the desired lower bound on λ_1 and the fact that the lower bound on λ_N above is sharp.

LEMMA 2. *There are constants c and c_2 such that $\lambda_1 \geq 1 - c_2/p^2 n$ and $\lambda_{N-1} \leq 1 - c/p^2$.*

PROOF. Suppose we choose M in (1) to be supported in the first off-diagonal. Then $M \cdot E_k^{\pm 1} = M$. Therefore, (1) simplifies to

$$P_n \varphi_M = \frac{1}{n-1} \sum_{k=1}^{n-1} \cos\left(\frac{2\pi M_{k+1\,k}}{p} \right) \varphi_M.$$

For the bound on λ_1 take one of the $M_{k+1\,k}$ to be ± 1 and the others 0. This gives $2(n-1)$ eigenvectors with eigenvalue $(n - 2 + \cos(2\pi/p))/(n-1)$. Therefore,

$$\lambda_1 \geq (n - 2 + \cos(2\pi/p))/(n-1) \geq 1 - c_2 p^{-2} n^{-1},$$

which is the desired lower bound on λ_1. For the bound on λ_{N-1} take all of the $M_{k+1\,k}$ to be $(p \pm 1)/2$. This produces 2^{n-1} eigenvectors with eigenvalue $-\cos(\pi/p)$. Therefore,

$$\lambda_{N-1} \leq -\cos(\pi/p) \leq 1 - cp^{-2}. \qquad \square$$

The upper bound on $\boldsymbol{\lambda_1}$. This leaves only the upper bound on λ_1. This will be done using essentially the same variational description of the second eigenvalue as was used for the geometric bounds on [6]. Recall that

$$\lambda_1 = \max\left\{ \langle \psi, P_n \psi \rangle / \langle \psi, \psi \rangle : \psi \in L^2(G_n) \setminus \{0\}, \langle \psi, \varphi_0 \rangle = 0 \right\}$$

and in fact the maximum is attained exactly on the nonzero vectors of the eigenspace corresponding to the second largest eigenvalue. We will bound λ_1 by bounding $\langle \psi, P_n \psi \rangle / \langle \psi, \psi \rangle$. Toward this end, suppose we have a decomposition of $L^2(G_n)$ into orthogonal P_n invariant subspaces $V_1, V_2, \dots, V_m$. Then for each i we may choose a basis of V_i consisting of eigenvectors of P_n. Therefore, the maximum must be attained for a vector in one of the V_i. That is, it suffices to bound $\langle \psi, P_n \psi \rangle / \langle \psi, \psi \rangle$ on each of the V_i separately. This observation may be summarized by the following lemma.

LEMMA 3. *Let P: $L^2(G) \to L^2(G)$ be the transition probability matrix for a reversible Markov chain. Let $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1} \geq -1$ be the eigenvalues of P, where $N = |G|$. Let $\varphi_0 \in L^2(G)$ denote an eigenvector of P with eigenvalue 1 (i.e., the constant function.) Suppose we have an orthogonal decomposition $L^2(G) = V_1 \oplus V_2 \oplus \cdots \oplus V_m$ of $L^2(G)$ into P-invariant subspaces. Then $\lambda_1 = \max_i \max\{\langle \psi, P_n \psi \rangle / \langle \psi, \psi \rangle : \psi \in V_i \setminus \{0\}, \langle \psi, \varphi_0 \rangle = 0\}$.*

This leads us to the second important feature of (1). The formula for $P_n\varphi_M$ involves only $\varphi_{M'}$ for M' in the same orbit of the G_n action on U^* as M. Therefore, we have split $L^2(G_n)$ into a large number of P_n invariant subspaces. Let $[M]$ be the orbit of M under the action of G_n. This orbit may be described very explicitly. Elements of it have the same first column as M. Their second column is the second column of M plus some multiple of the first column of M (truncated). Their third column is the third column of M plus some multiples of the (truncated) first and second columns of M, and so forth. Let $V_{[M]} = \operatorname{span}\{\varphi_{M'}: M' \in [M]\}$. Then the $V_{[M]}$ give a decomposition of $L^2(G_n)$ into orthogonal P_n invariant subspaces.

We now turn to the problem of bounding $\langle \psi, P_n\psi\rangle / \langle \psi, \psi\rangle$ on each of these subspaces. Every element $M' \in [M]$ has the same first column as M; hence, restricting to the final $(n-1)$ rows and columns gives an embedding i: $V_{[M]} \to L^2(G_{n-1})$. Under this embedding the action of $E_k \in G_n$, $2 \le k \le n-1$, gets taken to the action of $E_{k-1} \in G_{n-1}$. Therefore, we have

$$P_n|_{V_{[M]}} \simeq \frac{n-2}{n-1} P_{n-1}|_{i(V_{[M]})} + \frac{1}{n-1} T, \tag{2}$$

where T is a random walk on a p-cycle coming from the contribution of E_1. More explicitly, if the first column of M is $(0 \quad a \quad b_2 \quad b_3 \quad \cdots \quad b_n)^T$, let D be the $(n-1) \times (n-1)$ matrix whose first column is $(0 \quad b_2 \quad b_3 \quad \cdots \quad b_n)^T$ and all of whose other entries are zeroes. Then T is given by the formula

$$T\varphi_{M'} = \tfrac{1}{2}(\omega^a \varphi_{M'+D} + \omega^{-a}\varphi_{M'-D}).$$

This identification will allow us to inductively give upper bounds for the first eigenvalue on each of these invariant subspaces.

CASE 1 $[\varphi_0 \notin i(V_{[M]})]$. In this case, $i(V_{[M]}) \subset (\varphi_0)^\perp$ and hence $\langle \varphi, P_{n-1}\varphi\rangle \le \lambda_1(P_{n-1})\|\varphi\|^2$ for every $\varphi \in i(V_{[M]})$. The embedding i rescales the inner products $\langle \cdot, \cdot \rangle$ by a constant (coming from the difference in the normalizations); hence, from (2) and the trivial bound $\langle \psi, T\psi\rangle \le \|\psi\|^2$, we obtain

$$\langle \psi, P_n\psi\rangle \le ((n-2)\lambda_1(P_{n-1}) + 1)\|\psi\|^2/(n-1)$$

for every $\psi \in V_{[M]}$. Thus any eigenvector $\psi \in V_{[M]}$ has eigenvalue λ satisfying

$$\begin{aligned}\lambda &\le \frac{n-2}{n-1}\lambda_1(P_{n-1}) + \frac{1}{n-1}\\ &\le \frac{n-2}{n-1}\left(1 - \frac{c_1}{p^2(n-1)}\right) + \frac{1}{n-1}\\ &\le 1 - \frac{c_1}{p^2 n}.\end{aligned}$$

CASE 2 $[\varphi_0 \in i(V_{[M]})]$. This is by far the harder case and it will be divided into two subcases depending upon whether a parameter b_n defined below is zero (easy) or nonzero (hard). First note that the condition $\varphi_0 \in i(V_{[M]})$ is equivalent to the statement that $[M]$ contains a matrix, say M itself, whose only nonzero entries are in the first column. Suppose M has first column $(0 \quad a \quad b_2 \quad b_3 \quad \cdots \quad b_n)^T$ and all other entries zero. If $b_n = 0$, then the entire bottom row of M is zero and hence every element of $[M]$ has all zeroes in the bottom row. Hence restricting to the first $n-1$ rows and columns gives an embedding $j\colon V_{[M]} \to L^2(G_{n-1})$. For this embedding,

$$P_n|_{V_{[M]}} \simeq \frac{n-2}{n-1} P_{n-1}|_{j(V_{[M]})} + \frac{1}{n-1}$$

and hence repeating the argument above shows that any eigenvector in $V_{[M]}$ has eigenvalue λ satisfying $\lambda \le 1 - c_1 p^{-2} n^{-1}$.

Finally suppose $b_n \neq 0$. In this case, we will bound $\langle \psi, P_n \psi \rangle$ using the Cauchy–Schwarz inequality in a somewhat unusual way. To motivate this odd construction, notice that the argument above fails if $b_n \neq 0$ because $\varphi_0 \in i(V_{[M]})$ and it is an eigenfunction of P_{n-1} with eigenvalue 1. Thus we cannot simply bound the contribution of T by a constant. Fortunately φ_0 is not close to an eigenvector for T. We will show that T forces enough mixing to push the eigenvalues down. View P_n as an average of $n-1$ pieces, one for each pair $E_k^{\pm 1}$. The last pair $(E_{n-1} + E_{n-1}^{-1})/2$ acts on any φ_M as a multiplication by $\cos(2\pi M_{n\,n-1}/p)$. The other pairs either act as multiplication by a similar factor or by permuting p of these basis vectors around in a p-cycle (with some phases thrown in). That is, P_n behaves like a "random walk" on a graph whose vertices are $[M]$ and whose edges are a union of p-cycles. (This fails to be a true random walk since there are some phases associated to the edges and some probability of the walk disappearing.) The Cauchy–Schwarz inequality says that if we alter P_n by replacing any of these p-cycles by the identity on these p elements, then $\langle \psi, P\psi \rangle$ will only increase. Such replacements strategically carried out will "decouple" P_n on $V_{[M]}$.

Suppose $M' \in [M]$ has its leftmost r columns nonzero and its rightmost $n-r$ columns all zeroes. Then P_n is the average of r terms that permute a p-cycle and $n-r-1$ terms that act as the identity. We wish to use Cauchy–Schwarz to remove the first $r-2$ of these p-cycles. Any other $M' \in [M]$ has a leftmost nonzero column (say, the rth column) which has a column of all zeroes to the left of it. In this case we wish to use Cauchy–Schwarz to remove all the p-cycles corresponding to the nonzero columns among the first $r-1$ columns of M'. These desires are consistent. If we wish to remove a p-cycle because of any element of a cycle, then we wish to remove it for every element of that cycle. Let P' be the Markov chain that results from doing so. Then as remarked above, $\langle \psi, P_n \psi \rangle \le \langle \psi, P'\psi \rangle$ for all $\psi \in V_{[M]}$.

The Markov chain P' in turn has several invariant subspaces and as above it suffices to bound $\langle \psi, P'\psi \rangle$ on each of them separately. For the first set of these invariant subspaces, suppose M' has its rth column the first nonzero column after a column of all zeroes. Let $W = \operatorname{span}(\varphi_{M''}\colon M''$ has the same first r columns as $M'\}$. Then W is invariant under P' since the terms in P_n which alter the first r columns have been removed in P'. Restriction to the last $n-r+1$ rows and columns gives an embedding $j\colon W \to L^2(G_{n-r+1})$ and

$$P'\Big|P_W = \frac{r-1}{n-1} + \frac{n-r}{n-1}P_{n-r+1}\Big|_{j(W)}.$$

Since every element of W has nonzero rth column, $\varphi_0 \notin j(W)$. Hence $j(W) \subset (\varphi_0)^{\perp}$ and as above we have

$$\langle \psi, P'\psi \rangle \leq ((r-1) + (n-r)\lambda_1(P_{n-r+1}))\|\psi\|^2/(n-1),$$

for all $\psi \in W$. Since by induction $\lambda_1(P_{n-r+1}) \leq 1 - c_1 p^{-2}(n-r)^{-1}$, this gives $\langle \psi, P'\psi \rangle \leq (1 - c_1 p^{-2}(n-1)^{-1})\|\psi\|^2$, as desired.

The elements of $[M]$ not covered above are those that correspond to matrices which have the leftmost r columns nonzero and the remaining $n-r$ columns all zeroes for some r. The corresponding φ's give an invariant subspace on which P' has an interesting form. It is a "random walk" on the graph shown in Figure 1.

The vertices are labeled by the remaining matrices in $[M]$. The top layer consists of M, the second layer consists of the matrices $M \cdot E_1^k$, $k \neq 0$. In general, the rth layer consists of those matrices with exactly the leftmost r columns nonzero which have the form $M \cdot E_1^{k_1} \cdot E_2^{k_2} \cdots E_{r-1}^{k_{r-1}}$ with all the k_i nonzero. Between each layer is a set of p-cycles as shown in Figure 1. The cycle between layer r and layer $r+1$ corresponds to multiplication by E_r. The random walk is described as follows. Each edge has weight $1/(2n-2)$. The vertex M has stationary probability $(n-2)/(n-1)$. The vertices in the middle layers have stationary probability $(n-3)/(n-1)$. The vertex in the bottom layer corresponding to the matrix M' has stationary probability $[(n-3) + \cos(2\pi M'_{nn-1}/p)]/(n-1)$. Since the holding probability for vertices in the bottom row is less than $(n-2)/(n-1)$, the total probability in this "random walk" is not conserved. We may regard this as saying that the vertices in the bottom row have some probability of absorbing the walk. Thus P' need not have an eigenvector with eigenvalue 1. Furthermore, to bound $\langle \psi, P'\psi \rangle$ for ψ in this subspace we can apply Cauchy–Schwarz to spread out

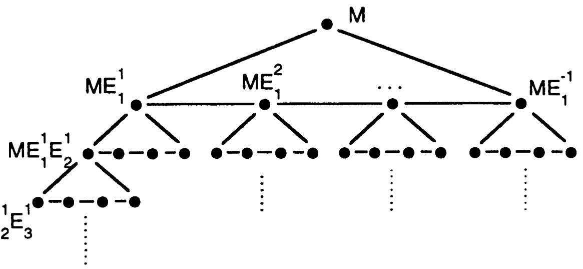

FIG. 1. *The graph sketched for $p = 5$.*

the absorbing by the bottom layer. For this we need the following easy lemma.

LEMMA 4. *There is a function* $f(\varepsilon) > 0$ (*depending on* p) *such that, for any* $0 < \varepsilon < 1.5p^{-2}$ *and any real numbers* $\phi_0, \phi_1, \dots, \phi_{p-1}$,

$$\phi_0\phi_1 + \phi_1\phi_2 + \cdots + \phi_{p-2}\phi_{p-1} + \phi_{p-1}\phi_0 \le (1 - f(\varepsilon))\phi_0^2 + (1 + \varepsilon) \sum_{i=1}^{p-1} \phi_i^2.$$

Furthermore $(p-1)\varepsilon \ge f(\varepsilon) \ge (p-1)\varepsilon - (p+1)^3\varepsilon^2/6$.

We will postpone the (elementary) proof of this lemma until the end of this paper. To complete the upper bounds, fix an ε to be determined later. In the bottom layer the holding probabilities are bounded above by $[n - 3 + \cos(2\pi/p)]/(n-1)$. Applying the lemma, we can remove the bottom layer of p-cycles, which leaves all the vertices with stationary probability of at most

$$\frac{n-3}{n-1} + \frac{1}{n-1}\cos\left(\frac{2\pi}{p}\right) + \frac{1}{n-1}(1+\varepsilon)$$

and leaves the layering above them with stationary probabilities of at most

$$\frac{n-3}{n-1} + \frac{1}{n-1}(1 - f(\varepsilon)).$$

Applying the lemma again removes the next layer of p-cycles and leaves that layer with stationary probabilities of at most

$$\frac{n-3}{n-1} + \frac{1}{n-1}(1 - f(\varepsilon)) + \frac{1}{n-1}(1+\varepsilon).$$

This procedure can be repeated to remove all the layers of p-cycles from the graph. The vertices are left with stationary probabilities of at most $[n - 2 + \varepsilon + \cos(2\pi/p)]/(n-1)$ in the bottom layer, $(n - 1 - f(\varepsilon) + \varepsilon)/(n-1)$ in the middle layers and $(n - 1 - f(\varepsilon))/(n-1)$ for M. Thus we have just shown that we can upper bound $\langle \psi, P'\psi \rangle$ by a diagonal form with these values as the entries. In particular,

$$\langle \psi, P'\psi \rangle \le \max\{[n - 2 + \varepsilon + \cos(2\pi/p)]/(n-1), \\ (n - 1 - f(\varepsilon) + \varepsilon)/(n-1), (n - 1 - f(\varepsilon))/(n-1)\}.$$

If we choose $\varepsilon = cp^{-3}$, then $f(\varepsilon) = c'p^{-2}$ and we conclude that

$$\langle \psi, P'\psi \rangle \le \left(1 - \frac{c_1}{p^2 n}\right)\|\psi\|^2.$$

This finally completes the proof of the upper bound on λ_1. We have now bounded $\langle \psi, P_n\psi \rangle$ on every one of the orthogonal invariant subspaces defined above. Therefore, the proof on the upper bound on λ_1 is complete.

The bound on the rate of convergence is now straightforward. Applying Proposition 3 of [6], page 41, gives

$$4\|P_n^k - U\|^2 < p^{n(n-1)/2}\left(1 - \frac{c_1}{p^2 n}\right)^{2k} \le \exp\left\{n(n-1)\frac{\log p}{2} - c_1 k p^{-2} n^{-1}\right\}.$$

Taking $c = 1/(2c_1)$ and $C = c_1/2$ gives the statement bound. This finally completes the proof of the theorem (except for the proof of Lemma 4 below). □

It should be emphasized that we have used one of the weakest results for bounding convergence in terms of eigenvalues. The other results use either knowledge of all the eigenvalues (the upper bound lemma of [1] and [5]) or knowledge of the eigenvalues and eigevectors (the L^1 upper bound lemma of [8]). An interesting extension of the work above would be to sharpen the analysis to gain some information about the number of eigenvectors with eigenvalues near λ_1 or bounds on the rate of convergence directly from the analysis.

PROOF OF LEMMA 4. We will do only the case $p = 2m + 1$ (since the case p even is not used above). Define a sequence of numbers a_k by $a_0 = 1$ and $a_{k+1}(2 + 2\varepsilon - a_k) = 1$. The Cauchy–Schwarz inequality gives $\phi_m \phi_{m+1} \le (\phi_m^2 + \phi_{m+1}^2)/2$ and

$$\phi_{m\pm k}\phi_{m\pm k+1} \le \left[(2 + 2\varepsilon - a_k)\phi_{m\pm k}^2 + a_{k+1}\phi_{m\pm k+1}^2\right]/2.$$

Summing these bounds up to $k = m$ gives

$$\phi_0\phi_1 + \phi_1\phi_2 + \cdots + \phi_{p-2}\phi_{p-1} + \phi_{p-1}\phi_0 \le a_m\phi_0^2 + (1 + \varepsilon)\sum_{i=1}^{p-1}\phi_i^2.$$

Therefore the lemma holds with $f(\varepsilon) = 1 - a_m$. To complete the proof one checks (by induction) that for $0 < \varepsilon < 1.5p^{-2}$ one has $1 - 2k\varepsilon \le a_k \le 1 - 2k\varepsilon + 4(k + 1)^3\varepsilon^2/3$. □

Acknowledgments. I would like to thank Persi Diaconis for suggesting this Markov chain and sharing his ideas on it and the referees for suggesting a number of improvements in the exposition.

REFERENCES

[1] DIACONIS, P. (1988). *Group Representations in Probability and Statistics.* IMS, Hayward, CA.

[2] DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison techniques for random walks on finite groups. *Ann. Probab.* **21** 2131–2156.

[3] DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730.

[4] DIACONIS, P. and SALOFF-COSTE, L. (1994). Moderate growth and random walks on finite groups. *Geom. Funct. Anal.* **4** 1–36.

[5] DIACONIS, P. and SHAHSHAHANI, M. (1979). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **46** 193–204.

[6] DIACONIS, P. and STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61.

[7] ELLENBERG, J. (1993). A sharp diameter bound for upper triangular matrices. Senior honors thesis, Dept. Mathematics, Harvard Univ.

[8] STONG, R. (1991). Choosing a random spanning subtree: a case study. *J. Theoret. Probab.* **4** 753–766.

[9] ZACK, M. (1992). Convergence to uniform on the finite Heisenberg group. Preprint.

DEPARTMENT OF MATHEMATICS
RICE UNIVERSITY
P.O. BOX 1892
HOUSTON, TEXAS 77251-1892