05/20/03

A SUPER-CLASS WALK ON UPPER-TRIANGULAR MATRICES

Ery Arias-Castro
Department of Statistics
Stanford University

Persi Diaconis
Depts. of Mathematics & Statistics
Stanford University

Richard Stanley
Department of Mathematics
M.I.T.

ABSTRACT

Let G be the group of $n \times n$ upper-triangular matrices with elements in a finite field and ones on the diagonal. This paper applies the character theory of Andre, Carter and Yan to analyze a natural random walk based on adding or subtracting a random row from the row above.

1. Introduction

For a prime p, let $G_n(p) = G$ be the group of unipotent upper-triangular matrices with elements in the finite field $\mathbb{F}_p$. This group has generating set

$$S = \{I \pm E_{i,i+1}\} \; 1 \le i \le n-1. \tag{1.0}$$

A natural random walk may be performed, beginning at the identity, each time choosing one of the $2(n-1)$ generators at random, and multiplying. More formally, define a probability measure on $G_n(p)$ by

$$Q_0(g) = \begin{cases} 1/2(n-1) & \text{if } g = I \pm E_{i,i+1} \quad 1 \le i \le n-1 \\ 0 & \text{otherwise .} \end{cases} \tag{1.1}$$

Let $Q_0^{*2}(g) = \Sigma_h Q_0(h) Q_0(gh^{-1})$, $Q_0^{*k}(g) = Q_0 * Q_0^{*(k-1)}(g)$. These convolution powers give the chance that the walk is at g after k steps. Denote the uniform distribution by

$$\pi(g) = \frac{1}{p^{n(n-1)/2}} \; . \tag{1.2}$$

If p is an odd prime, $Q_0^{*k}(g) \to \pi(g)$ as $k \to \infty$. To study the speed of convergence let

$$\|Q_0^{*k} - \pi\| = \max_A |Q_0^{*k}(A) - \pi(A)| = \frac{1}{2} \sum_g |Q_0^{*k}(g) - \pi(g)| \; . \tag{1.3}$$

Given $\epsilon > 0$, how large must k be so $\|Q_0^{*k} - U\| < \epsilon$? Partial results due to Zack, Diaconis, Saloff-Coste, Stong and Pak are described at the end of this introduction. There are good answers if n is fixed and p is large but the general problem is open.

The present paper develops an approach to the problem using character theory as described in Diaconis and Saloff-Coste [1993], Diaconis [2003]. This involves bounding the rate of convergence of a random walk driven by a probability measure that is constant on the union of the conjugacy classes containing the generating set. Then, a comparison theorem is used to bound the original walk. The character theory of $G_n(p)$ is a well known nightmare. In recent work, Carlos Andre, Roger Carter and Ning Yan have developed a theory based on certain unions of conjugacy classes (here called super-classes) and sums of irreducible characters (here called super-characters). The present paper gives a sharp analysis of the conjugacy class walk and gives partial results for the original walk.

Here is one of our main results. The conjugacy class containing $I + aE_{i,i+1}$ consists of upper triangular matrices with a in position $(i, i+1)$, arbitrary field elements $\alpha_1, \alpha_2, \ldots, \alpha_{i-1}$ in column $i+1$ above this a, arbitrary field elements $\beta_1, \beta_2, \ldots, \beta_{n-(i+1)}$ in row i to the right of the a. In the block bounded by these α_j, β_k, the (j,k) entry is $a^{-1}\alpha_j\beta_k$.
Call this class $C_i(a), 1 \le i \le n-1$. Thus $|C_i(a)| = p^{n-2}$.
For p an odd prime, define

$$Q(g) = \begin{cases} 1/[2(n-1)p^{n-2}] & \text{if } g \in C_i(\pm 1) \quad 1 \le i \le n-1 \\ 0 & \text{otherwise} \end{cases} \tag{1.4}$$

For $p = 2$, define

$$Q(g) = \begin{cases} 1/n & \text{if } g = id \\ 1/[n2^{n-2}] & \text{if } g \in C_i(1)\ 1 \le i \le n-1 \\ 0 & \text{otherwise.} \end{cases} \tag{1.4}$$

Theorem 1. For the random walk (1.4) on the group of unipotent upper-triangular matrices $G_n(p)$, there are universal constants γ_i so that for all $n \ge 2$ and all k,

$$\gamma_1 e^{-\gamma_2 k/(p^2 n \log n)} \le \|Q^{*k} - \pi\| \le \gamma_3 e^{-\gamma_4 k/(p^2 n \log n)}. \tag{1.5}$$

Remarks.

1. The natural analog of the walks (1.1) and (1.4) over the finite field $\mathbb{F}_q$ use generators $\{I + a_j E_{i,i+1}\}$ and $C_i(a_j)$ where a_j are an additive basis for $\mathbb{F}_q$ over $\mathbb{F}_p$. If $q = p^u$, then (1.5) holds with $p^2(n \log n)$ replaced by $p^2(nu \log(nu))$. See Section 3B.
2. The walk (1.4) is easy to implement as a series of 'rank one steps'. To choose an element of the conjugacy class $C_i(a)$ uniformly, form a random vector V by choosing field elements $V_1, V_2, \ldots, V_{i-1}$ uniformly in $\mathbb{F}_p$, setting $V_i = a$ and $V_j = 0$ for $j > i$. Form a random vector W by setting $W_k = 0$, $1 \le k \le i$, $W_{i+1} = 1$, $W_j = a^{-1}U_j$ with U_j-chosen uniformly in $\mathbb{F}_p$, $i + 2 \le j \le n$. The matrix $I + VW^T$ is uniformly distributed in $C_i(a)$.

Section Two below reviews the super-class theory needed. As new results, it derives the basic upper bound lemma, proves that super-class functions form a commutative, semi-simple algebra indexed by set partitions and derives a closed formula for the value of a super-character on a super-class with no restrictions on n and q. Theorem 1 is proved in Section Three in a stronger norm than (1.3). This is needed for comparison theorems. Section Four gives a character-free proof of Theorem 1 using a new form of stopping time arguments which may be of independent interest. Section Five gives our analysis of the original walk (1.1) by comparison. The main novelty in the present paper is showing that super-class theory can be used to solve problems usually solved by character theory.

Literature Review. For background on random walk on finite groups see Diaconis [1988], Saloff-Coste [1997], [2003]. The comparison approach is developed in Diaconis and Saloff-Coste [1993] with recent developments surveyed in Diaconis [2002]. There have been previous applications of comparison theory in the symmetric group and for finite groups of Lie type. The present paper is the first serious incursion into p-groups.

When $n = 3$, the random walk (1.1) on the Heisenberg Group was studied by Zack [1984]. For fixed $n \ge 3$ and large p, sharp rates of convergence are given in joint work with Saloff-Coste [1993A, 1994A,B]. Roughly, order p^2 steps are necessary and sufficient for convergence. The solution was achieved by three quite different routes. In [1994B], geometric volume growth arguments are used. In [1994A], the walk is realized as a projection of a walk on the free nilpotent group. Decay bounds of Hebisch and Saloff-Coste along with Harnack inequalities are used. The implicit constants depend badly on n. They are of order e^{n^2}.

Perhaps the earliest large n-results follow from work of Ellenberg [1993]. If γ is the diameter of $G_n(p)$ in the generating set S of (1.0) he shows there are explicit constants c, C such that

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

From this, standard bounds (see e.g. Diaconis and Saloff-Coste [1993A]) show that there are constants α, β such that

$$\|Q_0^{*k} - \pi\| \leq p^{n(n-1)/2} \left(1 - \frac{2}{n\gamma^2}\right)^k .$$

Thus, for p fixed and n large, order n^7 steps suffice.

Richard Stong [1995] has given sharp estimates of the second eigenvalue of the walk (1.1). He showed there are universal constants c_i such that the second eigenvalue λ_1 satisfies $1 - \frac{c_1}{p^2 n} \leq \lambda_1 \leq 1 - \frac{c_2}{p^2 n}$. He also showed that the smallest eigenvalue satisfies

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

Using these, he shows that if $k = c_4 p^2 n^3 \log p + p^2 n\theta$ then

$$\|Q_0^k - \pi\| < e^{-c_5\theta}.$$

Stong also shows that at least order n^2 steps are needed

Pak [2000] treats the case of n large, with steps $I + aE_{i,i+1}$ for a chosen uniformly. Using an elegant stopping time argument he shows that order $n^{2.5}$ steps are necessary and suffice for this case. The arguments are extended to nilpotent groups in Atashkevich and Pak [2001]. Coppersmith and Pak [????] showed that order n^2 steps suffice provided $p \gg n$.

To conclude this survey we note that the parallel walk on the generating class of transpositions in the symmetric group S_n had many applications through projections to quotient walks on subgroups. The subgroup $S_k \times S_{n-k}$ yields the Bernoulli-Laplace Model of diffusion. The subgroup $S_n w_r S_2$ yields a walk on perfect matchings, the walk projected onto conjugacy classes gives an analysis of coagulation-fragmentation process appearing in chemistry. These and many further applications are surveyed in Diaconis [2003]. For the walk on upper-triangular matrices, the projection onto the Frattini quotients gives the basic product walk on $\mathbb{F}_p^{n-1}$ analyzed in Diaconis and Saloff-Coste [1993]. The group $G_n(q)$ is a semi-direct product of $G_{n-1}(q)$ and $\mathbb{F}_q^{n-1}$ with $\mathbb{F}_q^{n-1}$ seen as all matrices in $G_n(q)$ which are zero except in the last column and $G_{n-1}(q)$ seen as all matrices in $G_n(q)$ which are zero in the last column. The quotient walk on $G_n(q)/G_{n-1}(q)$ is an example of a facilitated kinematics model where a site can turn on or off only if its left most neighbor is on. See Aldous and Diaconis [2002] or Ritort and Sollich [2002] for extensive references. At this writing we do not have a simple interpretation of the projection of the walk (1.4) on super-classes but we presume it will give a natural walk on set partitions.

2. Background Throughout, $q = p^u$ for a prime p. The group $G_n(q)$ of $n \times n$ matrices which are upper triangular with ones on the diagonal is the Sylow p-subgroup of the general linear group $GL_n(\mathbb{F}_q)$. Throughout, we write G for $G_n(q)$. As is well known, G has center $Z(G)$ consisting of matrices in G which are zero in all coordinates except $(1, n)$. The commutator G' equals the Frattini subgroup $\Phi(G)$ which consists of matrices in G which are zero along the super-diagonal. It follows that the matrices $(I \pm E_{i,i+1})$ $1 \leq i \leq n-1$ form a minimal generating set for $G_n(p)$ and that there are q^{n-1} distinct linear characters.

The character theory and conjugacy classes of G have been a persistent thorn in the side of group theorists. They are not known for $n \geq 7$ and considered unknowable. Indeed, Poljak [1966] shows that a nice description of the conjugacy classes leads to a nice description of wild quivers. Presumably, this does not exist. The difficulty of applying the orbit method to G is reviewed by Kirilov in [1995, 1999]. Further study is in Issacs [1995] who shows that the degree of a nonlinear character is a power of q. Thompson [2003] studies the apparently difficult problem of proving that the number of conjugacy classes is a polynomial in q.

In a series of papers [1995A,B, 1996, 1996], Carlos Andre has developed what Roger Carter calls super-class and super-character theory. Super-classes are certain unions of conjugacy classes and super-characters are sums of irreducible characters. These have nice duality and orthogonality properties and a very useful super-character formula.

We follow an elegant elementary approach of Ning Yan [2001]. This does not have the restrictions of earlier work that $p > n$. It also contains all that we need to analyze the random walks of interest.

In Section A, super-classes are defined. The algebra $\mathcal{A}$ of super-class functions is introduced. Section B defines super-characters and gives their dimension and intertwining numbers. Section C gives the Andre-Carter-Yan Character formula. Section D shows these objects are naturally associated to Bell numbers and set partitions. Section E derives a Plancherel formula and the basic upper bound lemma needed to prove Theorem 1.

A. Super-Classes. Let $\mathcal{U}_n(q)$ denote the set of upper triangular matrices with zeros on the diagonal. The product group $G \times G$ acts on $\mathcal{U}_n(q)$ by left/right multiplication. Let Ψ index the orbits of this action. The orbits indexed by Ψ will be called *transition orbits* below. Yan [Th 3.1] shows that each transition orbit contains a unique element with at most one non-zero entry in each row and each column. If D denotes the positions of the non-zero entries and $\phi : D \to \mathbb{F}_q^*$ denotes the entries, Ψ may be represented by pairs (D, ϕ). For example, when $n = 3$, there are five possible choices of D shown in Figure 1 below

Figure 1

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & \blacksquare & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \blacksquare \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & \blacksquare \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & \blacksquare & 0 \\ 0 & 0 & \blacksquare \\ 0 & 0 & 0 \end{pmatrix}$$

In Section D below we show that the number of allowable configurations D is the Bell number $B(n)$. Here $B(1) = 1, B(2) = 2, B(3) = 5, B(4) = 15, B(5) = 52, \ldots$ is the number of set partitions of n.

Figure 1 also shows two combinatorial features of D that figure prominently in later developments. The *Dimension Index* $d(D)$ denotes the sum of the vertical distances from the boxes in D to the super-diagonal $\{(i,i+1)\}_{1\le i\le n-1}$. Thus if all of the boxes in D are on the super-diagonal $d(D)=0$. The *Intertwining Index* $i(D)$ counts the number of pairs of boxes in D, that is, $(i,j),(k,\ell)$ in D, with $1\le i<k<j<\ell\le n$ so that the 'corner' (k,j) is above the diagonal. Pictorially

$$\begin{array}{cc} (i,j) & \\ & \\ (k,j) & (k,\ell) \end{array}$$

The $n=3$ example above was close to trivial. Here is another with $n=5$. On the left, $d(D)=2; i(D)=0$; on the right, $d(D)=3; i(D)=1$.

$$\begin{pmatrix} 0 & 0 & \blacksquare & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & \blacksquare & 0 & 0 \\ 0 & 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

As will emerge in Section B, the super-characters are also indexed by pairs (D,ϕ). The associated super-character has dimension of $q^{d(D)}$ and intertwining number $q^{i(D)}$. We will call the positions in D "boxes" below.

Following Kirilov [1995] and Yan [2001] we may map transition orbits in $\mathcal{U}_n(q)$ into the group G by adding the identity to each matrix in the orbit. These will be called *super-classes* and labeled $C(D,\phi)$. Subtracting the identity from each element of $C(D,\phi)$ gives the transition class $K(D,\phi)$. It is clear that $C(D,\phi)$ is a union of conjugacy classes. As an example, the super-class corresponding to transition orbit for a single box consists of matrices in G with a fixed, non-zero field element a where the box is; arbitrary field elements α_i directly above the box, arbitrary field elements β_j directly to the right of the box. In the rectangle above and to the right of the box it has element $a^{-1}\alpha_i\beta_j$. Note that the super-class with one box containing a in position $(i,i+1)$ contains the generator $I+aE_{i,i+1}$. Clearly, the size of the super-class corresponding to one box is $q^{s(D)}$ with $s(D)$ equal to the number of places above and to the right of the box. Yan shows that any transition class is a sum of the "elementary" transition-classes it contains:

$$K(D,\phi)=\sum_{d\in D}K(d,\phi),$$

and further, each $x\in K(D,d)$ can be written in exactly $r(D)$ ways as such a sum, where

(2.0) $$r(D)=\#\ \{(i,j),(k,\ell)\in D: i<k,\ j<\ell\}.$$

Define the super-class functions $\mathcal{A}$ via

(2.1) $$\mathcal{A}=\{f:G\to\mathbb{C}\text{ with } f \text{ constant on super-classes}\}.$$

Thus $f \in \mathcal{A}$ if and only if $f(g) = f(g')$ whenever $g - I = h_1(g' - I)h_2$. We show below that $\mathcal{A}$ is a commutative, semi-simple sub-algebra of the class functions on G under convolution

$$f_1 * f_2(g) = \sum_{h \in G} f_1(h) f_2(gh^{-1}). \tag{2.2}$$

B. Super-Characters. Let $\mathcal{U}_n^*(q)$ be the space of linear maps from $\mathcal{U}_n(q)$ to $\mathbb{F}_q$. The group G acts on the left and right of $\mathcal{U}_n^*(q)$ via

$$g * \lambda(m) = \lambda(mg), \;\; \lambda * g(m) = \lambda(gm), \;\; g \in G, \;\; \lambda \in \mathcal{U}_n^*(q), \;\; m \in \mathcal{U}_n(q).$$

The orbits of the product group $G \times G$ on $\mathcal{U}_n^*(q)$ are called *cotransition orbits* and indexed by Ψ^*. Fix a non-trivial homomorphism $\theta : \mathbb{F}_q$ to $\mathbb{C}^*$. For $\lambda \in \mathcal{U}_n^*(q)$, define $v_\lambda : G \to \mathbb{C}^*$ by

$$v_\lambda(g) = \theta[\lambda(g - I)]$$

Yan [2001, sec. 2] shows that $\{v_\lambda : \lambda \in \mathcal{U}_n^*(q)\}$ is an orthonormal basis of $\mathbb{C}[G]$ with the usual inner product $\langle f_1 | f_2 \rangle = \frac{1}{|G|} \sum_g f_1(g) \overline{f_2(g)}$.

By direct computation,

$$g v_\lambda(\cdot) = v_\lambda(g) v_{g\lambda}(\cdot).$$

It follows that if L is a left orbit of G acting on $\mathcal{U}_n^*$, the linear span of $\{v_\lambda\}_{\lambda \in L}$ is a submodule of $\mathbb{C}[G]$. Let χ_λ be the character of this representation for any $\lambda \in L$. Yan [2001, R.2] shows that if λ and λ' are in the same *right* orbit of G acting on $\mathcal{U}_n^*$ then $\chi_\lambda = \chi_{\lambda'}$. The characters $\{\chi_\lambda\}_{\lambda \in \Psi^*}$ are called super-characters. Yan [2001, 2.6] shows that the super-characters are in fact super-class functions, that they are orthogonal and

$$\langle \chi_{D,\phi} | \chi_{D',\phi'} \rangle = \begin{cases} 0 & \text{if } (D,\phi) \neq (D',\phi') \\ q^{i(d)} & \text{if } (D,\phi) = (D',\phi') \end{cases} \tag{2.3}$$

Here, the set Ψ^* is identified with Ψ and the labeling of (D, ϕ) pairs will be used.

One further useful fact Yan [2001, 2.4]: let χ_G be the character of the regular representation of G; its decomposition into super-characters is

$$\chi_G(\cdot) = \sum_{D,\phi} \frac{|\psi(D,\phi)|}{\chi_{D,\phi}(1)} \chi_{D,\phi}(\cdot), \tag{2.4}$$

where $\chi_{D,\phi}(1) = q^{d(D)}$ is the character degree and $|\psi(D,\phi)| = q^{2d(D) - i(D)}$ is the size of the $G \times G$ orbit in $\mathcal{U}_n^*$ indexed by (D, ϕ). The sum is over all cotransition orbits.

These facts allow us to prove an apparently new result.

Proposition 1. The space $\mathcal{A}$ of super-class functions defined at (2.1) is a commutative semi-simple algebra.

Proof. We will show that $\mathcal{A}$ is closed under convolution. It is thus a sub-algebra of the class functions on G and so commutative. Further, it has a basis of orthogonal idempotents, the super-characters, so it is semi-simple.

For each (D,ϕ), let $S(D,\phi)$ be the labels of the irreducible characters of G contained in $\chi_{D,\phi}$. By orthogonality of $\chi_{D,\phi}$, the $S(D,\phi)$ are disjoint. From (2.4), every irreducible character appears in a unique $S(D,\phi)$. Since each irreducible character χ_s appears in the regular character $\chi_s(1)$ times, (2.4) yields that the multiplicity of χ_s in the appropriate $\chi_{D,\phi}$ is $q^{i(D)-d(D)}\chi_s(1)$. Thus

$$\chi_{D,\phi}(\cdot) = q^{i(D)-d(D)} \sum_{s\in S(D,\phi)} \chi_s(1)\chi_s(\cdot). \tag{2.5}$$

It is classical that for two irreducible characters

$$\chi_s * \chi_t = \delta_{st}\frac{|G|}{\chi_s(1)}\cdot\chi_s(\cdot)$$

See e.g. Isaacs [1976, 2.13]. Thus, $\chi_{D,\phi} * \chi_{D',\phi'}$ is zero unless $(D,\phi)=(D',\phi')$ and then

$$\chi_{D,\phi} * \chi_{D,\phi}(\cdot) = q^{2(i(D)-d(D))} \sum_{s\in S(D,\phi)} \chi_s^2(1)\frac{|G|}{\chi_s(1)}\chi_s(\cdot) = q^{i(D)-d(D)}|G|\chi_{D,\phi}(\cdot) \quad \square \tag{2.6}$$

We may identify the super-characters as tensor products of certain induced characters studied by Lehrer [1974]. First, for $1\le i\le n-1$ identify the i-dimensional group $G_i(q)$ with a subgroup of $G_n(q)$ having non-zero entries only in the left $i\times i$ upper corner. This subgroup has a normal complement $H_i(q)$ consisting of matrices in $G_n(q)$ having the identity in the left $i\times i$ left upper corner. Thus $G_n(q)$ is the semi-direct product of $G_i(q)$ with $H_i(q)$. It follows that any character of $G_i(q)$ extends trivially to $G_n(q)$. Let $G_{i,k}$, $1\le k\le i-1$ be the subgroup of G_i consisting of matrices that are zero in the kth row (to the right of the diagonal). For χ a character of $\mathbb{F}_q^*$, let $\chi^{i,k}$ be the character $(0,\ldots,0,\chi,0,\ldots,0)$ (zeros except in the kth place) of the Abelian subgroup of G_i consisting of matrices which are zero except in the last column. Using the trivial character of G_{i-1} gives a character ??? of G_i. Inducing this up to G_n gives precisely $\chi_{D,\phi}$ with $D=\{(k,i)\}$ and ϕ uniquely associated to χ. Further, for any D, $\chi_{D,\phi}=\bigotimes_{d\in D}\chi_{d,\phi}$. Details may be found in Lehrer [1974] and Yan [2001].

C. The Character Formula. There is a remarkable closed form formula for the value of a super-character on a super-class. Andre [1996] gave such a result for p sufficiently large compared to n. Using tools developed by Yan, we are able to show that Andre's formula holds for all values of n and p.

Theorem 2. On the group $G_n(q)$ of upper-triangular matrices, with ones on the diagonal and entries in $\mathbb{F}_q$, the value of the super-character $\chi_{D,\phi}$ on the super-class $C(D',\phi')$ equals

$$\begin{cases} q^{p(D,D')} \quad \theta\left(\displaystyle\prod_{(i,j)\in D\cap D'} \phi(i,j)\phi'(i,j)\right) & \text{if} \quad D\subseteq R(D') \\ \\ 0 & \text{Otherwise} \end{cases} \tag{2.7}$$

where $R(D')$ is the complement in $\{1 \leq i < j \leq n\}$ of the positions directly above and to the right of positions in D' (thus $D' \subseteq R(D)$) and $p(D, D')$ is the number of positions directly below positions in D which are also in $R(D')$. Finally, θ is an isomorphism from $\mathbb{F}_q$ (additively) to $\mathbb{C}$.

Remarks and Examples.

1. $D' = \emptyset$ corresponds to the identity, which forms a super-class by itself. Then, $R(D')$ is the full upper triangle, the product in (2.7) is one, and $p(D, D') = d(D)$ defined in Section 2A above. Thus
$$\dim \chi_{D,\phi} = \chi_{D,\phi}(1) = q^{d(D)}$$
2. The random walk Q of (1.4) is supported on the union of $2(n-1)$ super-classes
$$C_i(\pm 1) = C((i,\ i+1);\ \pm 1), \quad 1 \leq i \leq n-1.$$
For $D' = C_i(\pm 1)$, $R(D')$ consists of all positions in the upper-triangle which are not strictly above or strictly to the right of $(i, i+1)$. The product in (2.7) has a single term and $p(D, D')$ counts the distance from the entries in D down to the super-diagonal counting only positions in $R(D')$. Thus, if D_i is the set of positions in D in the rectangle strictly above and to the right of $(i, i+1)$
$$\frac{\chi_{D,\phi}(C_i(\pm 1))}{\chi_{D,\phi}(\emptyset)} = \begin{cases} q^{-|D_i|}\theta(\pm\phi(i, i+1)) & \text{if } D \subset R(\{(i, i+1)\}) \\ 0 & \text{Otherwise} \end{cases}$$

We make careful use of this in Section 3. We begin the proof of the theorem with a duality lemma. The super-characters of $G = G_n(q)$ are indexed by orbits of $G \times G$ on $\mathcal{U}_n^*(q)$ the set of $\mathbb{F}_q$-valued linear functions of $\mathcal{U}_n(q)$ taken as a vector space over $\mathbb{F}_q$. Yan shows these may also be indexed by Pairs (D, ϕ) as above. Call the set of orbits Φ^* with typical element $\psi(D, \phi)$.

Lemma 1. Fix $\lambda \in \psi(D, \phi)$ and $g \in C(D', \phi')$. Then,

$$\chi_{D,\phi}(g) = \frac{q^{d(D)}}{|\psi(D,\phi)|} \sum_{\lambda' \in \psi(D,\phi)} \theta(\lambda'(g - I)) = \frac{q^{d(D)}}{|C(D',\phi)|} \sum_{h \in C(D',\phi')} \theta(\lambda(h - I)). \tag{2.8}$$

Proof. The first equality in (2.8) is 2.5 of Yan [2001]. Write the first sum as

$$\begin{aligned}
\sum_{\lambda' \in \psi(D,\phi)} \theta(\lambda'(g - I)) &= \frac{1}{|G|^2} \sum_{s,t \in G} \sum_{\lambda' \in \psi(D,\phi)} \theta(s * \lambda' * t(g - I)) \\
&= \frac{1}{|G|^2} \sum_{\lambda' \in \psi(D,\phi)} \sum_{s,t \in G} \theta(s * \lambda' * t(g - I)) \\
&= \frac{|\psi(D,\phi)|}{|G|^2} \sum_{s,t \in G} \theta(s * \lambda * t(g - I)) \\
&= \frac{|\psi(D,\phi)|}{|G|^2} \sum_{s,t \in G} \theta(\lambda(t(g - I)s).
\end{aligned}$$

The last sum equals

$$\frac{|G|^2}{|C(D',\phi')|}\sum_{h\in C(D',\phi')}\theta(\lambda(h-I)).$$

Combining formulae gives the second equality in (2.8) $\square$

Proof of Theorem 2. Observe first that the claimed formula (2.7) is multiplicative: If $D=\{\delta_1,\delta_2,\ldots,\delta_r\}$ and the formula is known, then

$$\chi_{D,\phi}=\prod_{i=1}^{r}\chi_{\delta_i,\phi}.$$

Now, Yan [2001, th 6.1] has shown the super-characters $\chi_{D,\phi}$ is multiplicative. Thus it is enough to verify for any position δ

$$\chi_{\delta,\phi}(C(D',\phi'))=\begin{cases} q^{p(\delta,D')}\theta(\phi(\delta)\phi'(\delta)) & \text{if } \delta\in D' \\ q^{p(\delta,D')} & \text{if } \delta\in R(D')\setminus D' \\ 0 & \text{if } \delta\notin R(D') \end{cases}$$

It will be convenient to use the correspondence $g\leftrightarrow g-I$ which takes $C(D',\phi')$ to $K(D',\phi')$. As explained in Section 2A above, every transition class $K(D',\phi')$ can be written as a sum of classes:

$$K(D',\phi')=\sum_{\delta'\in D'}K(\delta',\phi'),$$

with each $m\in K(D',\phi')$ expressible in exactly $r(D')$ ways – see (2.0). Thus

$$|K(D',\phi')|=\frac{1}{r(D')}\prod_{\delta'\in D'}|K(\delta',\phi')| \tag{2.9}$$

Using (2.8), for any $\lambda\in\psi^*(\delta,\phi)$

$$\chi_{\delta,\phi}(C(D',\phi'))=\frac{q^{d(\delta)}}{|K(D',\phi')|}\sum_{m\in K(D',\phi')}\theta(\lambda(m)).$$

Using the decomposition of m as a sum

$$\sum_{m\in K(D',\phi')}\theta(\lambda(m))=\frac{1}{r(D')}\prod_{\delta'\in D'}\sum_{m\in K(\delta',\phi')}\theta(\lambda(m)). \tag{2.10}$$

Using properties of trigonometric sums,

$$\sum_{m\in K(\delta',\phi')}\theta(\lambda(m))=\begin{cases} |K(\delta,\phi')|\ \theta(\phi(\delta)\phi'(\delta)) & \text{if } \delta=\delta' \\ |K(\delta',\phi')| & \text{if } \delta\in R(\delta')\setminus R_+(\delta') \\ |K(\delta',\phi')|\ q^{-1} & \text{if } \delta\in R_+(\delta') \\ 0 & \text{if } \delta\notin R(\delta') \end{cases}$$

We use the following notations.

$$\begin{pmatrix} 0 & 0 & \bullet & \blacklozenge & \blacklozenge \\ 0 & 0 & \blacksquare & \bullet & \bullet \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\blacksquare$ is in position δ', $\bullet$ are in positions belonging to $R(\delta')^C$ and $\blacklozenge$ are in positions belonging to $R_+(\delta')$. It follows that the sum (2.10) is

$$q^{-\sum_{\delta' \in D'} \mathbf{1}_{R_+(\delta')}(\delta)}\ \theta(\phi(\delta)\phi'(\delta))^{\mathbf{1}_{D'}(\delta)}\ \frac{1}{r(D')} \prod_{\delta' \in D'} |K(\delta', \phi')|.$$

The theorem folows from this, (2.8), (2.9) and the obvious fact

$$p(\delta, D') = d(\delta) - \sum_{\delta' \in D'} \mathbf{1}_{R_+(\delta')}(\delta) \qquad \square$$

D. Set Partitions and Bell Numbers. The algebra $\mathcal{A}$ of Proposition 1 has a close connection with set partitions and Bell numbers. Indeed, the allowable sets D correspond to set partitions of $[n]$ by declaring i and j to be in the same block if D contains (i, j). For example, when $n = 3$, the five subsets D displayed in Figure 1 correspond to $1/2/3, 12/3, 1/23, 13/2, 123$. Given a set partition, we associate D, a set of pairs (i, j) with $1 \le i < j \le n$, by beginning with 1 and adding a box $(1, j)$ to D for the smallest distinct entry j in the same block with one (if one is a singleton,no box is added). Then add a box $(2, j)$ if j is the smallest entry in the block with 2 (no box is added if there is no larger entry). Continue with $3, 4, \ldots, n-1$. As an example, $25/14/3$ corresponds to

$$\begin{pmatrix} 0 & 0 & 0 & \blacksquare & 0 \\ 0 & 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Under this correspondence, partitions with b blocks map to patterns with $n - b$ boxes.

There is an extensive enumerative theory of set partions, see e.g. Fristed [1987] or Pitman [2003] for authoritative surveys. We have not seen previous study of the statistics $d(D)$ or $i(D)$. From the decomposition of the regular representation (2.4) we have the generating function

$$q^{n(n-1)/2} = \sum_D q^{2d(D) - i(D)} (q-1)^{|D|}.$$

Andre [1996] had earlier proved a dual formula corresponding to the decomposition into super-classes.

The number $B(n,q)$ of super-classes equals the dimension of the algebra $\mathcal{A}$. Yan [2001, 4.1] gives the following recurrence

$$B(n+1,q)=\sum_{k=0}^{n}\binom{n}{k}(q-1)^{n-k}B(k,q),\ \ B(0,q)=1.$$

This is easy to see: a configuration counted by $B(n+1,q)$ contains some number of boxes on the super-diagonal. Call this $n-k, 0\le k\le n$. Any choice rules out $n-k$ rows and columns and leaves at most k boxes to be further placed. This can be done in $B(k,q)$ ways; of course the $(q-1)$ factor accounts for the labeling by $\mathbb{F}_q^*$. Note that when $q=2$, this becomes the usual recurrence for Bell numbers.

Lehrer [1974] has shown that the irreducible characters of maximal degree are also super-characters corresponding to boxes $(1,n),(2,n-1),(3,n-2),\ldots$ along the main anti-diagonal. He shows that 'most' representations (according to Plancherel measure) have maximal degree.

Finally, Borodin [1995] has derived elegant probabilistic limit theorems for the number of Jordan Blocks in a random element of G. These and other results are described in Fulman's Survey [2002, Sec. 4].

E. Some Fourier Analysis. Throughout, $G=G_n(q)$ and $\mathcal{A}$ is the algebra of super-class functions of G. The *Fourier Transform* of $f\in\mathcal{A}$ at the class indexed by D,ϕ is

$$\widehat{f}(D,\phi)=\sum_{g}f(g)\bar{\chi}_{D,\phi}(g)=|G|\langle f|\chi_{D,\phi}\rangle.$$

From the convolution formula (2.4) and linearity we have, for $f,\ h\in\mathcal{A}$,

$$\widehat{f*h}(D,\phi)=q^{-d(D)}\widehat{f}(D,\phi)\widehat{h}(D,\phi). \tag{2.11}$$

As usual, the Fourier transform of the uniform distribution $\pi(g)=1/|G|$ is

$$\widehat{\pi}(D,\phi)=\begin{cases}1 & \text{if } D \text{ is empty}\\ 0 & \text{otherwise.}\end{cases}$$

Also, for any probability distribution $Q\in\mathcal{A}$, $\widehat{Q}(\emptyset)=1$. The following version of the Plancherel Theorem is basic to what follows.

Proposition 3. Let $Q\in\mathcal{A}$ be a probability distribution. Then

$$\|Q^{*k}-\pi\|_2^2=\frac{1}{|G|^2}\sum_{\substack{D,\phi\\ \text{Non-empty}}}q^{-i(D)}\left|\frac{\widehat{Q}(D,\phi)}{q^{d(D)}}\right|^{2k}.$$

Proof. For any $h\in\mathcal{A}$,

$$h=\sum_{D,\phi}\frac{\langle h|\chi_{D,\phi}\rangle}{\langle\chi_{D,\phi}|\chi_{D,\phi}\rangle}\chi_{D,\phi}.$$

Thus

$$\|h\|_2^2 = \sum_{D,\phi} |\langle h|\chi_{D,\phi}\rangle|^2 q^{-i(D)}.$$

This implies

$$\|Q^{*k} - \pi\|_2^2 = \frac{1}{|G|}\sum_g |Q^{*k}(g) - \pi(g)|^2 = \sum_{\substack{D,\phi \\ \text{Non-empty}}} |\langle Q^{*k}|\chi_{D,\phi}\rangle|^2 q^{-i(D)}.$$

Now use (2.11). $\square$

Corollary. (Upper Bound Lemma) Let $Q \in \mathcal{A}$ be a probability distribution, then

$$4\|Q^{*k} - \pi\|_{TV}^2 \leq \sum_{\substack{D,\phi \\ \text{Non-empty}}} q^{-i(D)} \left|\frac{\widehat{Q}(D,\phi)}{q^{d(D)}}\right|^{2k}.$$

Proof. $4\|Q^{*k} - \pi\|_{TV}^2 = \left(\sum_g |Q^{*k}(g) - \pi(g)|\right)^2 \leq |G| \sum_g |Q^{*k}(g) - \pi(g)|^2$

$= |G|^2 \|Q^{*k} - \pi\|_2^2$ $\square$

Remark. Let us relate the analysis of this section to the class-function analysis of Diaconis [2003]. If G is any finite group and h is a class function of G,

$$h = \sum_\rho \langle h|\chi_\rho\rangle \chi_\rho$$

where the sum is over all irreducibles representations and $\chi_\rho(g) = \mathrm{Trace}(\rho(g))$. Orthonormality of characters implies $\|h\|_2^2 = \sum_\rho |\langle h|\chi_\rho\rangle|^2$.

If $G = G_n(q)$ and h is a super-class function, Proposition 3 gives h as a sum of super-characters.

(2.12)
$$h = \sum_\psi \frac{\langle h|\chi_\psi\rangle}{\langle \chi_\psi|\chi_\psi\rangle}\chi_\psi.$$

Thus $\|h\|_2^2 = \sum_\psi |\langle h|\chi_\psi\rangle|^2 q^{-i(\psi)}$ where ψ runs over (D,ϕ) pairs. Decompose the super-character χ_ψ into irreducibles as in (2.7)

(2.13)
$$\chi_\psi = \sum_{\rho \in S(\psi)} m(\rho,\psi)\chi_\rho.$$

using (2.12), (2.13)

$$(2.14) \qquad \langle h|\chi_\rho\rangle = \frac{\langle h|\chi_\psi\rangle}{\langle \chi_\psi|\chi_\psi\rangle} m(\rho,\psi)$$

thus

$$\begin{aligned}\sum_\rho |\langle h|\chi_\rho\rangle|^2 &= \sum_\psi \sum_{\rho\in S(\psi)} \left|\frac{\langle h|\chi_\psi\rangle}{\langle \chi_\psi|\chi_\psi\rangle} m(\rho,\psi)\right|^2 = \sum_\psi \frac{|\langle h|\chi_\psi\rangle|^2}{|\langle \chi_\psi|\chi_\psi\rangle|^2} \sum_{\rho\in S(\psi)} m^2(\rho,\psi) \\ &= \sum_\psi \frac{|\langle h|\chi_\psi\rangle|^2}{\langle \chi_\psi|\chi_\psi\rangle}.\end{aligned}$$

Thus, as must be, the two formulae for $\|h\|_2^2$ agree.

From (2.14) we see that if $h \in \mathcal{A}$ and $\widehat{h}(\psi) = 0$ then $\widehat{h}(\rho) = 0$ for each ρ in $S(\psi)$.

3. Proof of Theorem 1 and Extensions

In this section we use the Fourier transform of the probability measure Q of (1.4) together with the upper bound lemma of Section 2E to prove Theorem 1. Throughout, the L^2 norms are bounded. We first treat the case when $q = 2$ with holding at the identity, both to have a theorem for this case and because the analysis is easiest here. We then treat the case of a general finite field $\mathbb{F}_q$; Theorem 1 is the special case where $q = p$. Finally we give the lower bounds which show our upper bounds are essentially sharp.

A. $\mathbf{q = 2}$. On $GL_n(2)$ consider the probability measure Q, defined in (1.4).
The Fourier transform at the super-character indexed by (D,ϕ) is

$$(3.2) \qquad \frac{\widehat{Q}(D)}{2^{d(D)}} = \frac{1}{n} + \frac{1}{n}\sum_{i=1}^{n-1} 2^{-|D_i|}(-1)^{\delta(D,i)}\delta(R_i,D).$$

When $q = 2$, ϕ doesn't enter. We write D_i for the number of positions in D strictly inside the rectangle with lower left corner at $(i,i+1)$. The indicator $\mathbf{1}(D,i)$ is one or zero as $(i,i+1)$ is in D or not and $\mathbf{1}(R_i,D)$ is one or zero as D is disjoint from positions in the row (and column) strictly to the right (above) $(i,i+1)$.

From proposition three, the L^2 or chi-square distance is given by

$$(3.3) \qquad |G|^2\|Q^{*k} - \pi\|_2^2 = \sum_{D\neq\emptyset} 2^{-i(d)}\left|\frac{\widehat{Q}(D)}{2^{d(D)}}\right|^{2k} \leq \sum_{D\neq\emptyset}\left|\frac{\widehat{Q}(D)}{2^{d(D)}}\right|^{2k}.$$

This is an upper bound for the total variation distance (1.3). Thus the following theorem proves the upper bound for Theorem 1 when $q = 2$.

Theorem 3. On $G_n(2)$, with Q defined by (3.1), let $k = n(5/2\ \log n + c/2)$, for $c > 0$. Then

$$|G|^2\|Q^{*k} - \pi\|_2^2 \leq 4e^{-c}.$$

Proof. Fix a non-empty set of positions D and consider the transform $\widehat{Q}(D)/2^{d(D)}$ at (3.2). Let m be the number of positions in D strictly above the super-diagonal and let ℓ be the number of positions in D on the super-diagonal. We always have $m + \ell \leq n - 1$. Also, since D is non-empty, $m + \ell \geq 1$.

We may upper bound the transform by replacing negative terms in the sum by 0 and positive terms in the sum by 1. Each of the ℓ super-diagonal positions in D contributes a zero and each of the m non-super-diagonal positions contributes a zero. This shows that

$$\frac{\widehat{Q}(D)}{2^{d(D)}} \leq 1 - \frac{m+\ell}{n}.$$

We may lower bound the transform by replacing positive terms in the sum by 0 and negative terms in the sum by -1. Each of the ℓ super-diagonal positions in D contributes a -1 and each of the m non-super-diagonal positions contributes a 0. This shows that

$$\frac{\widehat{Q}(D)}{2^{d(D)}} \geq \frac{\ell}{n}.$$

Hence,

$$\left|\frac{\widehat{Q}(D)}{2^{d(D)}}\right| \leq \max\left\{1 - \frac{m+\ell}{n}, \frac{\ell}{n}\right\}.$$

To bound the sum in (3.3) note that there are at most

$$\binom{n^2}{m}\binom{n}{\ell} \leq \min\{n^{2(m+\ell)}, n^{2(m+n-\ell)}\}.$$

such sets D.

We know can bound the rightmost sum in (3.3) by

$$\sum_{1 \leq m+\ell \leq n-1} n^{2(m+\ell)}\left(1 - \frac{m+\ell}{n}\right)^{2k} \quad + \quad \sum_{1 \leq m+\ell \leq n-1} n^{2(m+n-\ell)}\left(\frac{\ell}{n}\right)^{2k}.$$

The first sum is bounded above by

$$n\sum_{s=1}^{n-1} n^{2s}\left(1 - \frac{s}{n}\right)^{2k},$$

while the second sum is bounded above by

$$n\sum_{s=1}^{n-1} n^{4s}\left(1 - \frac{s}{n}\right)^{2k}.$$

In both cases, use $1 - x \leq e^{-x}$ to bound by e^{-c} for $k = n(5/2\ \log n + c/2)$. $\square$

Remark. The constants can be slightly improved (our estimates were made simple for didactic purposes). The lower bound in Section 3C shows they cannot be improved by much.

3B. Proof of Theorem 1 (Upper Bound). Let p be an odd prime. We want to upper bound

$$S = \sum_{D,\phi} \left| \frac{Q(D,\phi)}{p^{d(D)}} \right|^{2k}.$$

We implicitly extend ϕ to all (i,j) by zero outside D.

Let D be a set of "positions". Decompose it into $D = \mathrm{on}(D) \cup \mathrm{off}(D)$, where $\mathrm{on}(D)$ (resp. $\mathrm{off}(D)$) are the positions in D that are *on* (resp. *off*, i.e. above) the super-diagonal.

We know from Theorem 2 that

$$\frac{Q(D,\phi)}{p^{d(D)}} = \frac{1}{n-1} \sum_{i=1}^{n-1} w_i(D) \cos(2\pi\phi(i,\ i+1)/p), \tag{3.4}$$

where the "weights" $w_i(D)$ satisfy $0 \le w_i(D) \le 1$ and $w_i(D) = 0$ whenever there is s such that $(i,s) \in D$ or $(s,\ i+1) \in D$. Let $Z(D)$ be the set of $i = 1, \ldots, n-1$ such that $w_i(D) = 0$. Also, nowhere in (3.4) do the values $\phi(i,j), j > i+1$, appear.

Let $I^+(\phi)$ (resp. $I^-(\phi)$) be the set of $i = 1, \ldots, n-1$ such that $\cos(2\pi\phi(i,\ i+1)/p) > 0$ (resp. < 0). Then,

$$\begin{aligned}
\frac{Q(D,\phi)}{p^{d(D)}} &\ge \frac{1}{n-1} \sum_{i \in I^-(\phi) \cap Z(D)^c} \cos(2\pi\phi(i,\ i+1)/p), \quad \text{and,} \\
\frac{Q(D,\phi)}{p^{d(D)}} &\le \frac{1}{n-1} \sum_{i \in I^+(\phi) \cap Z(D)^c} \cos(2\pi\phi(i,\ i+1)/p).
\end{aligned}$$

Hence, $S \le S^+ + S^-$, where

$$S^{\pm} = \sum_{D,\phi} \left(\frac{1}{n-1} \sum_{i \in I^{\pm}(\phi) \cap Z(FD)^c} \cos(2\pi\phi(i,\ i+1)/p) \right)^{2k}.$$

Let us focus on S^+ – the computations for S^- are similar. What we are summing does not depend on the values that ϕ takes on $\mathrm{off}(D) \cup (\mathrm{on}(D) \cap I^-(\phi))$. Let $a(D)$ be the cardinality of $Z(D)$ and $b(D)$ be the cardinality of $\mathrm{off}(D)$. Notice that $a(D) > b(D)$. Also, let $c^{\pm}(D)$ be the cardinality on $(D) \cap I^{\pm}(\phi)$.

Replacing $\phi(i,\ i+1)$ by h_i, we thus get

$$S^+ = \sum_D (p-1)^{b(D)} [p/2]^{c^-(D)} \sum_{h_1,\ldots,h_{c^+(D)}} \left(\frac{1}{n-1} \sum_{i=1}^{c^+(D)} \cos(2\pi h_i/p) \right)^{2k},$$

where the h_i runs through $\{-p/4, \ldots, p/4\}$, excluding the case where all h_i are zero. In the sum, $p^{b(D)}$ (resp. $[p/2]^{c^-(D)}$) comes from summing over all possibilities for the values of ϕ on $\mathrm{off}(D)$ (resp. $\mathrm{on}(D) \cap I^-(\phi)$).

Rewrite as

$$S^+ = \sum_D (p-1)^{b(D)} [p/2]^{c^-(D)} \left(\frac{c^+(D)}{n-1} \right)^{2k} \sum_{h_1,\ldots,h_{c^+(D)}} \left(\frac{1}{c^+(D)} \sum_{i=1}^{c^+(D)} \cos(2\pi h_i/p) \right)^{2k},$$

where D runs through sets of positions satisfying $c^+(D) \geq 1$.

First, we claim that, for all $1 \leq c \leq n-1$, and the range of the h_i restricted as above,

$$\sum_{h_1,\ldots,h_c} \left(\frac{1}{c} \sum_{i=1}^{c} \cos(2\pi h_i/p) \right)^{2k} \leq \alpha e^{-\beta\ k/(p^2 n \log n)},$$

for universal α, β, uniformly in p, n and c. Indeed, this follows from Theorem 1 in [Diaconis and Saloff-Coste, Section 5] with an explicit bound. See in particular, example two of Section 5. Saloff-Coste [2003, Th 8.10], gives another proof.

Second, we prove that,

$$\sum_D (p-1)^{b(D)} [p/2]^{c^-(D)} \left(\frac{c^+(D)}{n-1} \right)^{2k} \leq 1 + \eta_m,$$

where $\eta_m \to 0$ as $n \to \infty$, also for $m = \lambda p^2 n \log(n)$ with λ large enough, uniformly in p. (All we need here is to bound by a constant.)

Call the sum T. Since $a(D) + c^+(D) + c^-(D) \leq n-1$ and $a(D) > b(D)$, we have

$$T \leq \sum_D p^{b(D)+c^-(D)} \left(1 - \frac{b(D)+c^-(D)}{n-1} \right)^{2k}.$$

There are at most $\binom{n^2}{b} \times \binom{n-1}{c}$ sets of positions with $b(D) = b$ and $c^-(D) = c$. This number is bounded by $n^{2(b+c)}$. Hence,

$$T \leq 1 + \sum_{1 \leq b+c \leq n-1} n^{2(b+c)} p^{b+c} \left(1 - \frac{b+c}{n-1} \right)^{2k}.$$

(The 1 takes care of the case $b + c = 0$.)

Call T' the sum on the right. We have

$$\begin{aligned} T' &\leq n \sum_{\ell=1}^{n-1} (pn)^{2\ell} \left(1 - \frac{\ell}{n-1} \right)^{2k} \\ &\leq n \sum_{\ell=1}^{n-1} (pn)^{2\ell} e^{-2m\ell/(n-1)}. \end{aligned}$$

Now,

$$\begin{aligned}(pn)^{2\ell}e^{-2m\ell/(n-1)} &\leq e^{-2\ell(\lambda p^2\log(n)-\log(p)-\log(n))}\\ &\leq e^{-2\ell\log(n)(\lambda p^2-\log(p))}.\end{aligned}$$

Choose $\lambda > 0$ so that $\lambda p^2 - \log(p) \geq 1$, for all $p \geq 3$. Then,

$$T' \leq n\ \frac{e^{-2\log(n)}}{1-e^{-2\log(n)}} \leq 2/n,$$

and that tends to zeros as n increases. This completes the proof of the upper bound for Theorem 1.

Remark. It is straight-forward to give a bound for the analogous walk over $\mathbb{F}_q$. Let $q = p^u$. Let $a_1, a_2, \ldots, a_u \in \mathbb{F}_q$ be a basis for $\mathbb{F}_q$ as a vector space over $\mathbb{F}_p$. For $a \in \mathbb{F}_q$, define $Tr(a) = a + a^p + a^{p^2} + \ldots + a^{p^{u-1}}$. As in Lidl and Niederreiter [1997, 2.30], let $b_1, b_2, \ldots, b_u \in \mathbb{F}_q$ be a dual basis, thus $Tr(a_ib_j) = \mathbf{1}_{i=j}$. Choose θ in Theorem 2 as

$$\theta(a) = e^{i\frac{2\pi}{p}Tr(a)}.$$

In Theorem 2, field elements

$$\phi(i,j) = \sum_{k=1}^{u} \alpha_k a_k,$$

are written in basis a_k and transform variables

$$\phi'(i,j) = \sum_{k=1}^{u} \beta_k b_k,$$

are written in basis b_k. Then

$$\theta(\phi(i,j)\phi'(i,j)) = e^{\frac{2\pi i}{p}\sum\alpha_k\beta_k}.$$

From here, the analysis follows more or less as above with n replaced by nu, if Q_0 is defined on $G_n(q)$ by

$$Q_0(g) = \begin{cases}\frac{1}{2u(n-1)} & \text{if} \quad g = I \pm a_jE_{i,\ i+1} \quad 1 \leq j \leq a,\ 1 \leq i \leq n-1\\ \\ 0 & \quad\text{Otherwise}\end{cases}$$

Theorem 1 holds as stated provided q is odd and $m = p^2nu\log(nu)$. Further details are omitted.

3C. Lower Bounds. A lower bound on the L^2 or chi-squared distance which matches the upper bound of Theorems 2 and 3 can be obtained from the expression for $|G|^2\|Q^{*m} - \pi\|_2^2$

in terms of the Fourier transform (3.4). Keep only terms corresponding to D having a single position on the super diagonal and $\phi = 1$ on that entry. Then

$$|G|^2 \|Q^{*m} - \pi\|_2^2 \geq (n-1)\left[1 - \frac{1}{n-1}\left(1 - \cos\left(\frac{2\pi}{p}\right)\right)\right]^{2k}.$$

Elementary calculus estimates show that the right side is not small when $m \geq cp^2 n \log n$ for c fixed.

A lower bound for total variation comes from considering the quotient walk on G/Φ. As explained in the introduction, this evolves as the walk on $\mathbb{F}_p^{n-1}$ which proceeds by picking a coordinate at random and adding ± 1 to this coordinate. For this walk a $p^2 n \log n$ lower bound (for total variation) is well known. See e.g. Saloff-Coste [2003, Th 8.10]. Further details are omitted.

4. A Probabilistic Argument.

In this section we give a conceptually simple probabilistic proof of Theorem 1 for the walk based on generating conjugacy classes. The argument is a hybrid of strong stationary times as in Aldous and Diaconis [1986], Diaconis and Fill [1990] and Fourier analysis on $\mathbb{F}_p^{n-1}$. It is related to the stopping time arguments used by Pak [2000] and Uyemura-Reyes [2002].

Consider the measure Q defined at (1.4). As explained there, the random walk based on multiplying by successive choices from Q may be described as follows: If the current position of the walk is $X_t \in G_n(p)$, the next position is determined by multiplying on the left by a matrix having $\epsilon = \pm 1$ in position $(i, i+1)$, independent, uniformly chosen field elements $\alpha_1, \dots, \alpha_{i-1}$ in the column above $(i, i+1)$, independent uniformly chosen field elements $\beta_1, \dots, \beta_{n-(i+1)}$ in the row to the right of $(i, i+1)$. The entries in the (k, ℓ) position in the rectangle with corner at $(i, i+1)$ are $\epsilon\alpha_k\beta_\ell$. The first proposition shows that the elements in the row above $(i, i+1)$ and in the column to the right of $(i, i+1)$ in X_{t+1} are independent and identically distributed and remain so in successive steps of the walk.

Proposition 1. Let S be a subset of $\{(i,j), 1 \leq i < j \leq n\}$. Let M be a random matrix in $G_n(q)$ with $\{M_{ij}\}_{(i,j)\in S}$ uniformly distributed and independent of each other and other other entries in M. Let N be a second random matrix independent of M. Then, the entries in positions of S in the product MN (or NM) are uniformly distributed, and independent of each other and the other entries in the product.

Proof. Start with

$$(MN)_{ij} = \sum_k M_{ik} N_{kj} = M_{ij} + T_{ij},$$

where T_{ij} is a term involving elements of M and N distinct from M_{ij}. It follows that $(MN)_{ij}$ is uniform for all $(i,j) \in S$. To prove independence, argue column by column, working from the right. Entries in (MN) with the largest values of j occurring in S have unique entries which do not occur in other terms in S. These are thus independent of each other and the rest of the entries. Then consider entries with the second largest value of j in S, and so on. The argument for NM is similar. $\square$

The above proposition says, once an entry is random, it stays random. Returning to the random walk generated by Q, let T be the first time each position $(i, i+1)$ $1 \leq i \leq n-1$ has been chosen at least once. It follows from the proposition that at time $T = t$, all the entries at or above the second diagonal are independent and uniformly distributed, even given $T = t$. This last is a partial analog of strong stationarity.

Let $\Phi = \Phi(G_n(q))$ be the Frattini subgroup. This consists of matrices M in G with $M_{i,i+1} = 0$ $1 \leq i \leq n-1$. We thus see that for any s, t with $n-1 \leq t \leq s$, $P\{X_s \in A | T \leq t\}$ is right Φ invariant. The following proposition gives a precise sense in which the distribution of T and the rate of convergence of the the induced walk on G/Φ combine to give a bound on the rate of convergence of the walk on $G_n(p)$ to the uniform distribution π. The proposition is a variation of proposition (2.2) of Uyemura-Reyes (2002).

Proposition 2. Let H be a normal subgroup of the finite group G. Let Q be a probability on G with X_t, $0 \leq t < \infty$ the associated random walk. Let $\bar{Q}$ be the induced probability on G/H with Z_t, $0 \leq t < \infty$ the associated random walk. Suppose T is a stopping time for X_t, with

$$P\{X_t \in A | T \leq t\}$$

right H invariant. Then, for $1 \leq t < \infty$,

$$\|Q^{*t} - \pi\| \leq \|\bar{Q}^{*t} - \bar{\pi}\| + 2P\{T > t\}.$$

Proof. Choose coset representatives z_i $1 \leq i \leq |G/H|$. Write the walk as $X_t = (Z_t, H_t)$. Observe

$$\begin{aligned} P\{Z_t = z, H_t = h\} - \frac{1}{|G|} &= P\{T \leq t\}[P\{Z_t = z, H_t = h | T \leq t\} - \tfrac{1}{|G|}] + \\ &\quad P\{T > t\}[P\{Z_t = z, H_t = h | T > t\} - \tfrac{1}{|G|}]. \end{aligned}$$

Thus,

$$\begin{aligned} 2\|Q^{*t} \ -\pi\| &\leq P\{T \leq t\} \textstyle\sum_{z,h} |P\{Z_t = z, H_t = h | T \leq t\} - \frac{1}{|G|}| + \\ &\quad P\{T > t\} \textstyle\sum_{z,h} |P\{Z_t = z, H_t = h | T > t\} - \frac{1}{|G|}|. \end{aligned}$$

The second term is bounded by $2P\{T > t\}$. For the first sum use

$$\begin{aligned} (P\{Z_t = z | T \leq t\} - 1/|G/H|)P(T \leq t) &= \left(P(Z_t = z) - \frac{1}{|G/H|}\right) - \\ &\quad (P(Z_t = z | T > t) - 1/|G/H|)P(T > t) \end{aligned}$$

combining bounds (and dividing by two) gives the result $\square$

Propositions one and two lead to another proof of Theorem 1. Indeed, use Proposition 2 with t as given. For the stopping time T take the first time all positions $(i, i+1)$ have been chosen at least once. The classical coupon collectors problem (Feller [1968]) gives $P\{T > t\} \leq e^{-c}$. The process Z_i on G/Φ was analyzed in Diaconis and Saloff-Coste [1993A, Sec. 6.1]. They show universal α, β with

$$\|P\{Z_t \in \cdot\} - \pi_{G|\phi}\| \leq \alpha e^{-\beta t/p^2 n \log n}.$$

Combining bounds completes the proof.

Remark. Our first proof of Theorem 1 used character theory to prove an approximation in $L^2(\pi)$. This allows the walk to stand as a base of comparison. There is no sharp comparison based on total variation bounds.

5. A Comparison Argument

This section uses comparison techniques and the bounds on the conjugacy walk Q in Theorem 1 to get rates for the original walk Q_0 supported on generators $I \pm E_{i,\ i+1}$, $1 \leq i \leq n-1$, as at (1.1). Throughout, p is an odd prime, G is $G_n(p)$, and π is the uniform distribution on G. Let $L^2(\pi)$ be the real functions of G with inner product $\langle f_1 | f_2 \rangle = \sum_g f_1(g) f_2(g) \pi(g)$. We caution the reader that we use results from Diaconis and Saloff-Coste [1993A] which uses this inner product multiplied by $|G|$.

The quadratic form $\mathcal{E}$ (resp. $\mathcal{E}_0$) associated with Q (resp. Q_0) is

$$\mathcal{E}(f|f) = \sum_{s,t} (f(s) - f(t))^2 \pi(s) Q(ts^{-1}),$$

(resp. Q_0 in place of Q).

Lemma 5 of Diaconis and Saloff-Coste [1993A] shows that if there is a constant A such that $\mathcal{E} \leq A\mathcal{E}_0$ then

$$(5.1) \qquad |G|^2 \|Q_0^{*k} - \pi\|_2^2 \leq |G|^2 (\lambda_{\min}^{2k} + e^{-k/A} + \|Q^{*\lfloor k/2A \rfloor} - \pi\|_2^2)$$

with $\lambda_{\min}$ the smallest eigenvalue of the Q_0-walk. To give a suitable A, write each element in the support of Q as a product of generators $(I \pm E_{i,\ i+1})$. Let $|g|$ be the length of $g \in G$ and $N(\pm i, g)$ the number of times $I \pm E_{i,\ i+1}$, is used in the chosen representation for g. Theorem 1 of Diaconis and Saloff-Coste [1993A] shows that

$$(5.2) \qquad \mathcal{E} \leq A\mathcal{E}_0 \quad \text{with } A = \max_s \frac{1}{Q_0(s)} \sum_g |g| N(s,g) Q(g),$$

with the maximum taken over $s = \pm i,\ 1 \leq i \leq n-1$.

Lemma 1. Any element $g \in \text{supp}(Q)$ can be written with $|g| \leq 2np$ and $N(\pm i, g) \leq 4p$.

Proof. The elements of the conjugacy classes $C_i(\pm 1)$ are described in Remark Three following Theorem 1. They are matrices in G with ± 1 in position $(i,\ i+1)$, arbitrary field elements $\alpha_1, \alpha_2, \ldots, \alpha_{i-1}$ in the column above $(i,\ i+1)$, arbitrary field elements β_j $i+2 \leq j \leq n$ in the row to the right of $(i,\ i+1)$ and entry $\pm\alpha_a\beta_b$ in position (a,b) $1 \leq a \leq i-1, x+2 \leq b \leq n$, with zeros elsewhere.

It is straight-forward to write such an element as a product of generators. Begin by writing down $I + E_{i,\ i+1}$. Conjugating this by $I - E_{i-1,\ i}$ puts a one in position $(i-1,\ i+1)$ leaving remaining entries unperturbed. Next conjugating by $I - E_{i-2,\ i-1}$ puts a one in position $(i-2,\ i+1)$. Continuing, gives a matrix with ones above entry $(i,\ i+1)$. With these ones, general entries $\alpha_1, \alpha_2, \ldots \alpha_{i-1}$ can now be built up, working from the top down. This results in a matrix with ± 1 in position $(i,\ i+1)$, $\alpha_1, \ldots, \alpha_{i-1}$ in the column above this entry and zeros elsewhere.

From here, conjugate by $(I + E_{i+1,\ i+2}), \ldots, (I + E_{n-1,\ n})$ to put ones in the i^{th} row. Then, working from the right, build up the required pattern of β_j. The remaining entries in the matrix are all as they need to be to give the general entry of $C_i(\pm 1)$.

Each conjugation uses two generators so the final representing word has length at most $2np$. Further, any fixed generator is used at most $4p$. $\square$

Using the bounds in Lemma 1 in (5.2) gives

$$\mathcal{E} \leq A\mathcal{E}_0 \quad \text{with } A = 8n^2p^2. \tag{5.3}$$

The final ingredient needed is a bound of Stong for the smallest eigenvalue. Using basic path arguments, Stong [1995] shows

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

Combining bounds we see that

$$|G|^2 \|Q_0^{*k} - \pi\|_2^2 \leq |G|^2 \Big\{ (1 - 2/p^2)^{2k} + \bar{e}^{k/8n^2p^2} + \|Q^{\lfloor k/16n^2p^2 \rfloor} - \pi\|_2^2 \Big\}.$$

This is small provided $k \gg (n^4 \log n)(p^2 \log p)$.

Remarks.

1. The final result is "off". Stong's results show order n^3 steps suffice for fixed p, and Pak [2000] shows that $n^{2.5}$ steps suffice when $p = 2$. It is possible to improve the dependence on p by building up α_a/β_b in Lemma 1 more cleverly. An indication of the problem can be seen in the bound (5.3). From our work on Theorem 1, we know that the second eigenvalue of the Q chain is from the super-character with $D = \{(1,2)\}$ and $\phi(1,2) = 1$; this eigenvalue is $\widehat{\lambda}_1 = 1 - \frac{1}{n-1}\left(1 - \cos\left(\frac{2\pi}{p}\right)\right) = 1 - \frac{2\pi^2(1+o(1))}{np^2}$. The minimax characterization of eigenvalues shows that (5.3) implies $\lambda_i \leq 1 - \frac{(1-\widetilde{\lambda}_i)}{A}$ this gives $\lambda_1 \leq 1 - \frac{c}{n^3p^4}$ while Stong's results show $1 - \frac{c_1}{np^2} \leq \lambda_1 \leq 1 - \frac{c_2}{np^2}$. This suggests that the paths we have chosen can be improved, perhaps by randomization.

2. There is an amazing development of Mathematics connected to minimal factorizations in Berenstein, Fomin and Zelevinsky [1996]. Las, this does not seem to help improve our bounds.

We have included this section to show what a straight-forward use of comparison yields. We hope that someone will be motivated to improve our results.

Acknowledgments

We thank Alexi Borodin, Dan Bump, Roger Carter, Marty Isaacs, Jan Saxl, Laurent Saloff-Coste and J.C. Uyemura-Reyes for their help.

REFERENCES

Aldous, D. and Diaconis, P. (1986). Shuffling Cards and Stopping Times. *Amer. Math. Monthly* **93**, 333-348.

Aldous, D. and Diaconis, P. (2002). The Asymetric One-Dimensional Constrained Ising Model: Rigorous Results. *Jour. Statist. Physics* **107**, 945-975.

Andre, C. (1995A). Basic Characters of the Unitriangular Group. *J. Algebra* **175**, 287-319.

Andre, C. (1995B). Basic Sums of Cordjoint Orbits of the Unitriangular Group. *J. Algebra* **176**, 959-1000.

Andre, C. (1996). On the Cordjoint Orbits of the Unitriangular Group. *J. Algebra* **180**, 587-630.

Andre, C. (1998). The Regular Character of the Unitriangular Group. *J. Algebra* **201**, 1-52.

Astashkevich, A. and Pak, I. (2001). Random Walk on Nilpotent Groups. Technical Report, Department of Mathematics, MIT.

Borodin, A. (1995). Limit Jordan Normal Form of Large Triangular Matrices Over a Finite Field. *Func. Anal. Appl.* **29**, 279-281.

Coppersmith, D. and Pak, I. (2000). Random Walk on Upper Triangular Matrices Mixes Rapidly. *Prob. Th. Related Fields* **117**, 467-417.

Diaconis, P. (1988). *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics, Hayward, CA.

Diaconis, P. (2003). Random Walk on Groups: Geometry and Character Theory. In *Groups St. Andrews 2001*, M. Liebeck et al., Eds., Cambridge University Press.

Diaconis, P. and Fill, J. (1990). Strong Stationary Times Via a New Form of Duality. *Ann. Probab.* **18**, 1483-1522.

Diaconis, P. and Isaacs, I.M. (2003). Super-Character Theory for Algebra Groups. Forthcoming.

Diaconis, P. and Saloff-Coste, L. (1993A). Nash Inequalities and Finite Markov Chains. *Jour. Theoret. Probab.*

Diaconis, P. and Saloff-Coste, L. (1993B). Comparison Theorems for Random Walk on Groups. *Ann. Probab.* **21**.

Diaconis, P. and Saloff-Coste, L. (1994A). An Application of Harnack Inequalities to Random Walk on Nilpotent Quotients,

Diaconis, P. and Saloff-Coste, L. (1994B). Moderate Growth and Random Walk on Finite Groups. *Geometric and Functional Analysis* **9**, 1-36.

Ellenberg, J. (1993). A Sharp Diameter Bound for Upper Triangular Matrices. Senior Honors Thesis, Department of Mathematics, Harvard University.

Feller, W. (1968). *An Introduction to Probability and its Applications*, Vol. 1, 3rd Ed., Wiley, N.Y.

Fristedt, B. (1987). The Structure of Random Partitions of Large Sets. Technical Report, Department of Mathematics, University of Minnesota.

Fulman, J. (2002). Random Matrix Theory Over Finite Fields. *Bull. Amer. Math. Soc.* **34**, 51-85.

Isaacs, M. (1976). Character Theory of Finite Groups. Dover, N.Y.

Isaacs, I.M. (1995). Characters of Groups Associated with Finite Algebras. *Jour. Algebra* **177**, 708-730.

Kirillov, H. (1995). Variations on the Triangular Theme. *Translations Amer. Math. Soc.* **164**, 43-73.

Kirillov. A. (1999). Merits and Demerits of the Orbit Method. *Bull. Amer. Math. Soc.* **36**, 433-488.

Lehrer, C. (1974). Discrete Series and the Unipotent Subgroup S. Composito. *Math.* **28**, 9-19.

Pak, I. (2000). Two Random Walks on Upper Triangular Matrices. *Journ. Theoret. Probab* **13**, 1083-1114.

Pitman, J. (2003). Combinatorial Stochastic Processes. Lecture Notes: St. Flour.

Poljak, S.S. (1966). On the Characters of Irreducibla Comlex Representations of the Group of Triangular matrices over a price Finite Field. *Dopovidt Akao*, ??nuk Ukraine. RSR 1996, 434-436.

Ritort, F. and Sollich, P. (2002). Glassy Dynamics of Kinetically Constrained Models ARXIV: Condmat 10216382.

Saloff-Coste, L. (1997). Lectures on Finite Markov Chains. *Springer Lecture Notes in Math* **1665**, 301-408.

Saloff-Coste, L. (2003). Random Walk on Finite Groups. To appear, *Encyclopedia of Math.* Springer.

Stong, R. (1995). Random Walks on the Groups of Upper Triangular Matrices. *Annals of Probab.* **23**, 1939-1949.

Thompson, J. (2003). $U_n(q)$. Web Manuscript.

Uyemura-Reyes, J.C. (2002). Random Walk, Semi-Direct Products and Card Shuffling. Ph.D. Thesis, Department of Mathematics, Stanford University.

Yan, N. (2001). Representation Theory of the Finite Unipotent Linear Group. Preprint, Department of Mathematics, University of Pennsylvania.

Zack, M. (1984). A Random Walk on the Heisenberg Group. Ph.D. Thesis, Department of Mathematics, University of California, La Jolla.