Super–character theory and comparison arguments for a random walk on the upper triangular matrices

Evita Nestoridi

1 Introduction

Recently, there has been a lot of interest in the mixing time of a specific random walk on upper triangular matrices ([3], [4], [5], [6], [9], [13], [14], [16], [17]). Let $p$ be an odd prime and let $G$ be the group of $n \times n$ upper triangular matrices with 1’s on the diagonal and elements of $\mathbb{Z}/p\mathbb{Z}$ above the diagonal. Let $E(i,i + 1)$ be the $n \times n$ matrix having a one in the $(i,i + 1)$ entry and zeros elsewhere. The set $S = \{I_n \pm E(i,i + 1), 1 \leq i \leq n - 1\}$ is a symmetric generating set for $G$. We consider the random walk on $G$ using these generators, namely we let

$$P_x(xg) = \begin{cases} \frac{1}{2(n-1)}, & \text{if } g = I_n \pm E(i,i + 1), \\ 0, & \text{otherwise}, \end{cases}$$

be the probability of moving from $x$ to $xg$ in one step. The $t^{th}$ convolution of $P$ is defined inductively as

$$P^*_x(y) = \sum_{w \in G} P_w^{t-1}(y)P_x(w)$$

and it gives the probability of moving from $x$ to $y$ in $t$ steps. According to Proposition 2.13 and Theorem 4.9 of [12], the fact that $S$ is a symmetric set of generators guarantees that $P^*_x$ converges to the uniform measure $\mu$ on $G$ with respect to the total variation distance, which is defined as

$$||P^*_x - \mu||_{T.V.} := \frac{1}{2} \sum_{g \in G} |P^*_x(g) - \mu(g)|.$$

The main result of this paper concerns the mixing time of the above walk with respect to the total variation distance, i.e.

$$t_{\text{mix}}(\varepsilon) = \inf \{ t \in \mathbb{Z} : \max_{x \in G} \{ ||P^*_x - \mu||_{T.V.} \} < \varepsilon \},$$

where $\mu$ is the uniform measure on $G$.

**Theorem 1.** There exist universal constants $0 < b, d < \infty$ such that for $c > 0$ and $t \geq cbp^2n^4$, we have that

$$4||P^*_x - \mu||_{T.V.}^2 \leq de^{-c},$$

for $p$ sufficiently large.

The dependence on $p$ is the best possible, since the entry in position $(n - 1, n)$ performs a simple random walk on $\mathbb{Z}/p\mathbb{Z}$ and the mixing time of this random walk is of order $p^2$ as explained in [8].

1
The proof of Theorem 1 relies on bounding the eigenvalues $1 = \lambda_0 > \lambda_1 \geq \ldots \geq \lambda_{|G|-1} > -1$ of the transition matrix := $(P_x(y))_{x,y \in G}$ and using the inequality

$$4\|P^t - \mu\|_{T.V}^2 \leq \sum_{i=1}^{|G|-1} \lambda_i^{2t}$$

(see Lemma 12.16 (ii) of [12]). To bound the eigenvalues of $P$, we introduce an auxiliary random walk $Q$ on $G$, which we study using the super–character theory of $G$. Then, we use comparison theory as introduced by Diaconis and Saloff-Coste [10] to bound the eigenvalues of $P$ in terms of the eigenvalues of $Q$.

This new walk $Q$ is defined as follows. Let

$$a = \begin{cases} \lfloor \sqrt{p} \rfloor, & \text{if } \lfloor \sqrt{p} \rfloor \text{ is odd}, \\ \lfloor \sqrt{p} \rfloor + 1, & \text{otherwise}, \end{cases}$$

be the closest odd integer to $\lfloor \sqrt{p} \rfloor$. Define the following probability measure on $G$:

$$Q_x(xg) = \begin{cases} \frac{1}{4(\alpha - 1)p^{\alpha - 2}}, & \text{if } g \in C_i(\pm 1) \cup C_i(\pm a) \\ 0, & \text{otherwise}, \end{cases} \quad (2)$$

where $C_i(\pm 1)$ denotes the conjugacy class of the matrix $I_n \pm E(i, i+1)$ and $C_i(\pm a)$ denotes similarly the conjugacy class of the matrix $I_n \pm aE(i, i+1)$.

**Theorem 2.** There exist uniform constants $0 < \alpha, \beta < \infty$ such that for $c > 0$ and $t = c\beta pn \log n$, then

$$4\|Q^t - \mu\|_{T.V}^2 \leq \alpha e^{-c}.$$

Section 2 gives details on the rich literature of this problem. Sections 4 and 5 provide a quick overview of the super–character theory needed. In Section 6, we present a Fourier analysis argument, which leads to the proof of Theorem 2 contained in Section 7. Section 8 provides a brief review of the comparison techniques introduced by Diaconis and Saloff-Coste [10] and then uses them to prove Theorem 1.

**Remark 3.** The case of $p = 2$ of the walk we consider has been thoroughly studied by Peres and Sly [14], who proved an upper bound for the mixing time of order $n^2$ and Stong [17] who proved a lower bound also of order $n^2$.

## 2 Literature

Many people have studied similar problems, starting with Zack [20], who was interested in the Heisenberg group (which is $G$ for the case $n = 3$). Diaconis and Saloff-Coste [6] used Nash inequalities to prove that for the walk on $F$ for the case where $n$ is fixed and $p$ large, the mixing time is bounded above and below by a constant times $p^2$. See Diaconis and Hough [9] for a broader review of the $n = 3$ case and the extensions to nilpotent groups.

Stong [17] found sharp bounds for the second and last eigenvalues of the $P$–walk which allowed him to prove an upper bound of order $p^2n^3 \log p$. He also shows that at least $n^2$ steps are needed for the $P$–walk. Arias-Castro, Diaconis and Stanley [4] then used super–character theory and comparison theory to give an upper bound of order $n^4p^2 \log n + n^3p^4 \log n$, taking into consideration Stong’s
earlier bounds on the eigenvalues \cite{17}. They prove it by doing super–character theory analysis for the walk generated by \( C_i(\pm 1) \) and, then, doing comparison theory \cite{10}. They also prove a lower bound of the form \( p^2n \log n \).

Coppersmith and Pak (\cite{13}, \cite{5}) looked at the walk generated by \( \{ g = I_n + aE(i, i + 1), a \in \mathbb{Z}/p\mathbb{Z} \} \) and managed to improve the \( n \) term of the upper bound in the case where \( n \gg p^2 \). They proved that \( n^2 \log p \) steps are sufficient to reach stationarity. Peres and Sly \cite{14} proved that for \( p = 2 \) the sharp bound is of order \( n^2 \) using the east model. We refer to Peres and Sly for a more complete survey of the existing literature.

The present paper depends a lot on works by André (\cite{1}, \cite{2}, \cite{3}), Carter and Yan \cite{18}. They have developed a theory using certain unions of conjugacy classes, that we will refer to as super–classes, and sums of irreducible characters, that are called super–characters. Our work sharpens the super–character theory techniques introduced by Arias-Castro, Diaconis and Stanley and fixes the dependency on \( p \).

3 Preliminaries

We first consider the following random walk on \( \mathbb{Z}/p\mathbb{Z} \):

\[
K_x(y) := \begin{cases} 
\frac{1}{4}, & \text{if } y = x \pm a, \\
\frac{1}{4}, & \text{if } y = x \pm 1, \\
0, & \text{otherwise.}
\end{cases}
\]

This random walk, is a special case of \( Q \) for the case \( n = 2 \). Theorem 6 of \cite{7} says that the eigenvalues of the matrix \( K \) are given by the Fourier transform of the irreducible representations of \( \mathbb{Z}/p\mathbb{Z} \) with respect to \( K(g) = \frac{1}{4}, \text{ if } g = \pm a, \pm 1 \). In particular, if \( \rho_x(j) = e^{\frac{2\pi ij}{p}} \) for \( x, j \in \mathbb{Z}/p\mathbb{Z} \), then \( \hat{K}(\rho_x) := \sum_{y=\pm 1, \pm a} K(y) \rho(y) \frac{1}{2} \cos \frac{2\pi xa}{p} + \frac{1}{2} \cos \frac{2\pi x}{p} \) are the eigenvalues of the transition matrix \( (K_x(y))_{x,y \in \mathbb{Z}/p\mathbb{Z}} \).

Lemma 4. \cite{8}, Example 2.3

(a) We have that

\[
\|K^t_y - U\|_2^2 = \sum_{x=1}^{p-1} \left( \frac{1}{2} \cos \frac{2\pi xa}{p} + \frac{1}{2} \cos \frac{2\pi x}{p} \right)^2 t,
\]

where \( y \in \mathbb{Z}/p\mathbb{Z} \) and \( U \) is the uniform measure of \( \mathbb{Z}/p\mathbb{Z} \).

(b) If \( t = cp \), then

\[
\|K^t_y - U\|_2^2 \leq Ae^{-c},
\]

where \( y \in \mathbb{Z}/p\mathbb{Z} \) and \( A \) is uniform constant.

Proof. Part (a) follows from Lemma 12.18 of \cite{12} and the fact that \( \frac{1}{2} \cos \frac{2\pi xa}{p} + \frac{1}{2} \cos \frac{2\pi x}{p} \) are the eigenvalues of \( K \). Part (b) follows by the analysis presented in example 2.3 of \cite{8}.

The following lemma is a key computation for the proof of Theorem 2.

3
Lemma 5. We have that there are \( \alpha, \beta, \) uniform constants in \( p, n, \) such that if \( t = \beta p N \log N \)

\[
\sum_{(x_1, \ldots, x_n) \in \left( \mathbb{Z}/p\mathbb{Z} \right)^N \setminus \{0\}} \left( \frac{1}{2N} \sum_{j=1}^{N} \left( \cos \left( \frac{2\pi x_j a}{p} \right) + \cos \left( \frac{2\pi x_j}{p} \right) \right) \right)^{2t} \leq e^{-\beta}. \tag{3}
\]

Proof. Consider the following random walk on \( (\mathbb{Z}/p\mathbb{Z})^N \):

\[ q(x, y) = \begin{cases} \frac{1}{4N}, & \text{if } y = x \pm ae_i, x \pm e_i \text{ for } i = 1, \ldots, N \\ 0, & \text{otherwise}, \end{cases} \]

where \( e_i \) is the vector in \( (\mathbb{Z}/p\mathbb{Z})^N \) that has a one in the \( i \)th position and everywhere else zero.

Theorem 6 of chapter 3 of \cite{7} says that the eigenvalues of \( q \) are indexed by \( (x_1, \ldots, x_N) \in (\mathbb{Z}/p\mathbb{Z})^N \) and they are equal to

\[
\frac{1}{2N} \sum_{j=1}^{N} \left( \cos \left( \frac{2\pi x_j a}{p} \right) + \cos \left( \frac{2\pi x_j}{p} \right) \right) \tag{4}
\]

Lemma 4 and Theorem 1 of Section 5 of Diaconis and Saloff-Coste \cite{10} says that \( t = \beta p N \log N \)

\[
\sum_{(x_1, \ldots, x_N) \in (\mathbb{Z}/p\mathbb{Z})^N \setminus \{0\}} \left( \frac{1}{2N} \sum_{j=1}^{N} \left( \cos \left( \frac{2\pi x_j a}{p} \right) + \cos \left( \frac{2\pi x_j}{p} \right) \right) \right)^{2t} = \|q_y - \pi\|^2 \leq \alpha e^{-\beta},
\]

where \( \pi \) is the uniform measure on \( (\mathbb{Z}/p\mathbb{Z})^N \).

\[ \square \]

4 The conjugacy classes and the super–classes

While a description of general conjugacy classes (and characters in \( G \)) remains unknown \cite{11}, as explained in \cite{4}, there is a full description of the conjugacy class of \( I_n + xE(i, i + 1) \), where \( x \neq 0 \).

It consists of all matrices in \( G \) whose \( (i, i + 1) \) entry is \( x \), the entries of the \( i + 1 \) column exactly above \( (i, i + 1) \) are arbitrary elements \( a_1, a_2, \ldots, a_{i-1} \), the entries of the \( i \)th row exactly to the right of \( (i, i + 1) \) are arbitrary elements \( b_1, \ldots, b_{n-i-1} \) and in the block surrounded by these \( a_j, b_k \) the \( (j, k) \) entry is \( x^{-1}a_jb_k \). Here is an example for \( n = 6 \):

\[
\begin{bmatrix}
1 & 0 & 0 & a_1 & a_1 b_1 x^{-1} & a_1 b_2 x^{-1} \\
0 & 1 & 0 & a_2 & a_2 b_1 x^{-1} & a_2 b_2 x^{-1} \\
0 & 0 & 1 & x & b_1 & b_2 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Despite the fact that the character theory of \( G \) is unknown, André, Carter and Yan have provided a formula for the super–characters of \( G \) (certain specific sums of irreducible characters). Arias-Castro, Diaconis and Stanley \cite{4} showed how to use super–characters to prove upper bounds for mixing times. This will be explained in Section 6 after establishing the necessary notation and terminology.
Here is the general description of the super-classes of $G$ (certain unions of conjugacy classes). Let $U_n(p)$ be the set of all $n \times n$ upper triangular matrices with zeros on the diagonal and let $G \times G$ act on $U_n(p)$ by right and left multiplication. Let $\Psi$ denote the set of orbits of this action, which we refer to as transition orbits. According to Yan, each transition orbit contains a unique representative which has at most one non-zero entry per column and per row (see Theorem 3.1 of [19]). Thus $\Psi$ consists of pairs $(D, \phi)$, where $D$ is a collection of positions $(i, j)$ with $i < j$ at most one per column and per row and $\phi : D \to F_p^*$ any map.

A super-class in $G$ corresponds to a transition orbit and consists of those elements of $G$ of the form $I_n$ plus an element of the transition orbit. Yan explains at the end of Section 2 of [19] that the super-class of an element of the form $I_n + xE(i, i + 1)$ in fact coincides with its conjugacy class. For $(D, \phi) \in \Psi$, denote the corresponding super-class by $C(D, \phi)$.

5 The super-characters

The super-characters are certain sums of characters that can be used to bound the mixing time of the walk generated by $Q$, as it is later described in Lemma 6. Here is the description of the super-characters as provided by Yan [18].

Let $U_n^*(F_p)$ be the space of linear maps from $U_n(p)$ to $F_p$. Then $G$ acts on the left and right of $U_n^*(p)$ as follows:

$$(g \ast \lambda)(m) = \lambda(mg), \quad (\lambda \ast g)(m) = \lambda(gm),$$

where $g \in G, \lambda \in U_n^*(F_p)$ and $m \in U_n(p)$. The orbits of the action of $G \times G$ on $U_n^*(F_p)$ are called cotransition orbits.

The left action gives the regular representation of $G$ on $\mathbb{C}[G]$. To get an element of the group algebra $\mathbb{C}[G]$, we consider a non-trivial homomorphism $\theta : F_p \to \mathbb{C}^*$ from the additive group $F_p$ to the non-zero complex numbers. Then, for $\lambda \in U_n^*(F_p)$ we get the element of the group algebra $u_\lambda : G \to \mathbb{C}$ defined as:

$$u_\lambda(g) = \theta(\lambda(g - I))$$

The goal is to decompose regular representation of $G$ on $\mathbb{C}[G]$ into a sum of orthogonal submodules of $\mathbb{C}[G]$ (not necessarily irreducible). Proposition 2.1 of [18] says that

$$g \cdot u_\lambda = u_\lambda(g)u_{g\lambda}$$

Therefore, if $L$ is an orbit of the left action of $G$ on $U_n^*(F_p)$, then $\text{span}\{u_\lambda\}_{\lambda \in L}$ is a submodule of $\mathbb{C}[G]$. Corollary 2.3 of [18] says that the character $\chi_\lambda$ only depends on the cotransition orbit to which $\lambda$ belongs to. Theorem 3.2 of [18] adds that the cotransition orbits are indexed by pairs $(D, \phi)$ where $D$ denote the positions of the non-zero entries and $\phi : D \to F_p^*$ is the map that assigns a non-zero entry to each $(i, j)$ of $D$.

Let $\Psi^*$ denote the set of orbits of the action of $G$ on $U_n^*(F_p)$ and $\chi_{D, \phi}$ be the character corresponding to the above representation, where $D$ and $\phi$ determine the conjugacy class we described in Section 4. Proposition 2.2 of [18] proved that if $\lambda$ and $\lambda'$ are in the same right orbit of $G$ acting on $U_n^*(F_p)$ then $\chi_\lambda = \chi_{\lambda'}$ and therefore it makes sense to talk about $\chi_{D, \phi}$. Also, Corollary 2.8 of [18] says that $\{\chi_\lambda\}_{\lambda \in \Psi^*}$ are orthogonal characters.
6 Fourier transform setup

For $Q$ a probability measure on $G$ which is also a super-class function, which means that $Q$ is constant on the super-classes, and $\chi_{D,\phi}$, let
\[
\hat{Q}(D,\phi) = \sum_{g \in G} \chi_{D,\phi}(g)Q(g),
\]
(5)
denote the Fourier Transform of $Q$ at $\chi_{D,\phi}$. Then, E. Arias-Castro, P. Diaconis and R. Stanley \cite{arias-castro2018} proved the following upper bound lemma using Fourier Transform arguments:

**Lemma 6.** \cite{arias-castro2018}, Proposition 2.4] We have that
\[
4||Q^* - \mu||^2_{TV.} \leq \sum_{D \neq \emptyset,\phi} p^{-i(D)} \left( \frac{\hat{Q}(D,\phi)}{p^{d(D)}} \right)^{2t},
\]
(6)
where $d(D)$ is the sum of the vertical distances from the boxes in $D$ to the super diagonal $\{(i, i + 1)\}_{1 \leq i \leq n-1}$ and $i(D)$ counts the number of pairs of boxes $(i, j), (k, l)$ in $D$ with $1 \leq i < k < j < l \leq n$ so that the corner $(k, j)$ is strictly above the diagonal.

**Remark 7.** The statistics $i(D)$ and $d(D)$ are discussed in full detail in Yan \cite{yan2019}, Arias-Castro, Diaconis and Stanley \cite{arias-castro2018}.

7 Proof of Theorem 2

The proof of Theorem 2 follows the proof of Theorem 1.1 of \cite{arias-castro2018} and makes use of Lemma 6.

**Proof.** To bound the right hand side of (6), we consider each summand $\chi_{D,\phi}(g)$ of (5). For any $D$ let $D_i$ be the set of positions in $D$ in the rectangle strictly above and to the right of $(i, i + 1)$ and $R(\{i, i + 1\})$ is the complement of all the positions that are directly below $(i, i + 1)$ or directly to the right of $(i, i + 1)$. Then the formula of Remark 2 of Section 2.3 of Arias-Castro, Diaconis and Stanley says that
\[
\chi_{D,\phi}(g) = \begin{cases} p^{-|D_i|}\theta(\pm\phi(i, i + 1)), & \text{if } D \subset R(\{i, i + 1\}) \\ 0, & \text{otherwise}, \end{cases}
\]
(7)
if $g \in C_i(\pm 1)$ and
\[
\chi_{D,\phi}(g) = \begin{cases} p^{-|D_i|}\theta(\pm a\phi(i, i + 1)), & \text{if } D \subset R(\{i, i + 1\}) \\ 0, & \text{otherwise}, \end{cases}
\]
(8)
if $g \in C_i(\pm a)$. The right hand side of (6) can be bounded as follows
\[
\sum_{D \neq \emptyset,\phi} p^{-i(D)} \left( \frac{\hat{Q}(D,\phi)}{p^{d(D)}} \right)^{2t} \leq \sum_{D \neq \emptyset,\phi} \left( \frac{\hat{Q}(D,\phi)}{p^{d(D)}} \right)^{2t}.
\]
Equations (5), (7) and (8) give that
\[
\frac{\hat{Q}(D,\phi)}{p^{d(D)}} = \frac{1}{2n - 2} \sum_{i=1}^{2n-2} w_i(D) \left( \cos \frac{2\pi\phi(i, i + 1)}{p} + \cos \frac{2\pi\phi(i, i + 1)a}{p} \right)
\]
6
where the weights \( w_i(D) \) satisfy \( 0 \leq w_i(D) \leq 1 \) 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 \( i = 1, 2, ..., n - 1 \) such that \( w_i(D) = 0 \).

For \( x \in \mathbb{Z}/p\mathbb{Z} \), let \( I_x^+ (\phi) \) (respectively \( I_x^- (\phi) \)) be the set of \( i = 1, 2, 3, ..., n - 1 \) such that \( \cos \frac{2\pi \phi(i, i+1)x}{p} > 0 \) (resp. \( \cos \frac{2\pi \phi(i, i+1)x}{p} < 0 \)). The following will be the dominating terms:

\[
A^+ (D, \phi) = \frac{1}{2n - 2} \left( \sum_{i \in I^+ (\phi) \cap Z(D)^c} \cos \frac{2\pi \phi(i, i+1)}{p} + \sum_{i \in I^+_a (\phi) \cap Z(D)^c} \cos \frac{2\pi \phi(i, i+1)a}{p} \right)
\]

and

\[
A^- (D, \phi) = \frac{1}{2n - 2} \left( \sum_{i \in I^- (\phi) \cap Z(D)^c} \cos \frac{2\pi \phi(i, i+1)}{p} + \sum_{i \in I^- (\phi) \cap Z(D)^c} \cos \frac{2\pi \phi(i, i+1)a}{p} \right)
\]

Since

\[
A^- (D, \phi) \leq \frac{\tilde{Q}(D, \phi)}{p^{\theta(D)}} \leq A^+ (D, \phi),
\]

\[
\sum_{D \neq \emptyset, \phi} \left( \frac{\tilde{Q}(D, \phi)}{p^{\theta(D)}} \right)^{2t} \leq S^+ + S^-
\]

where

\[
S^\pm = \sum_{D \neq \emptyset, \phi} \left( A^\pm (D, \phi) \right)^{2t}
\]

To bound \( S^+ \):

Let \( b(D) \) be the cardinality of the elements of \( D \) that are off the super diagonal and \( c^\pm (D) = |I^\pm_a (\phi) \cap on(D)| + |I^\pm (\phi) \cap on(D)| \) where \( on(D) \) are the elements of the super diagonal of \( D \).

Then, replacing \( \phi(i, i+1) \) by \( h_i \),

\[
S^+ \leq \sum_{D} p^{b(D)} p^{c^- (D)} \left( \frac{c^+ (D)}{2n - 2} \right)^{2t} \sum_{h_1, h_2, ..., h_c+ (D)} \left( \frac{1}{c^+ (D)} \left( \sum_{i \in I^+ (\phi) \cap Z(D)^c} \cos \frac{2\pi h_i}{p} + \sum_{i \in I^- (\phi) \cap Z(D)^c} \cos \frac{2\pi h_i a}{p} \right) \right)^{2t}
\]

Lemma 4 says that after \( t = cp \log n \) steps, there are uniform constants \( \alpha \) and \( \beta \) such that

\[
\sum_{h_1, h_2, ..., h_c+ (D)} \left( \frac{1}{c^+ (D)} \left( \sum_{i \in I^+ (\phi) \cap Z(D)^c} \cos \frac{2\pi h_i}{p} + \sum_{i \in I^- (\phi) \cap Z(D)^c} \cos \frac{2\pi h_i a}{p} \right) \right)^{2t} \leq \alpha e^{-\beta c}.
\] (9)

We, also, need to bound the term \( T = \sum_D p^{b(D)} p^{c^- (D)} \left( \frac{c^+ (D)}{2n - 2} \right)^{2t} \). Following the second half of the proof of Theorem 1.1 of [4], for \( t = 2n(p + 2) \log n + dn \)

\[
T \leq 1 + 2e^{-d}
\]
where $d > 0$. To prove this, notice that if $a(D)$ is the cardinality of $Z(D)$ then $a(D) + c^+(D) + c^-(D) \leq 2n - 2$ and $a(D) > b(D)$ so

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

since there are at most $\binom{n-1}{c} \times \binom{n^2}{c}$ sets of positions with $b = b(D)$ and $c = c^-(D)$, which is bounded by $n^{2(b+c)}$

$$T \leq 1 + \sum_{1 \leq b+c \leq n-1} (np)^{2(b+c)} \left( 1 - \frac{b + c}{2n - 2} \right)^{2t} \leq 1 + \sum_{l=1}^n (pn)^{2l}e^{-d/2(n-1)}$$

For $t > 2np \log n + dn$, we have that

$$T \leq 1 + e^{-d} \sum_{l=1}^{n-1} \left( \frac{pn}{n^p} \right)^{2l} \leq 1 + \frac{e^{-d}}{1 - \frac{2}{n^p}} \leq 1 + 2e^{-d}$$

since $n^{p+1} \geq 2p$ for $n > 2$.

Therefore, overall there are new $\alpha, \beta$, uniformly in $p, n$ such that for $c > 0$ such that if $t = c\beta pn \log n$

$$\sum_D p^{b(D)}p^{c^-(D)} \left( \frac{c^+(D)}{2n - 2} \right)^{2t} \sum_{h_1^+, h_2^-, \ldots, h_i^+(D)} \left( \frac{1}{c^+(D)} \sum_{i=1}^{c^+(D)} \cos \frac{2\pi h_i^+a}{p} \right)^{2t} \leq \alpha e^{-c}$$

Similar arguments can be used to bound $S^-$.

\[ \square \]

### 8 The Comparison Argument

A comparison argument allows to use theorem 2 in order to prove theorem 1. The $L^2$ distance of $P^{st}$ from $\mu$ is defined as

$$||P^{st} - \mu||_2 := \left( \sum_{g \in G} |P^{st}(g) - \mu(g)|^2 \right)^{1/2}.$$ 

The Cauchy-Schwartz inequality gives that

$$4||Q^{st} - \mu||^2_{T.V.} \leq |G||P^{st} - \mu||_2^2.$$ 

A direct application of Lemma 8 of [10] gives the following tool.

**Lemma 8.** Let $P, Q$ be the probability measures on $G$, that were defined in the introduction. Let $g \in C_i(\pm 1) \cup C_i(\pm a)$. Fix a way of writing $g$ as a product of the generators $\{I_n \pm E(i, i+1)\}$ of odd length. Let $|g|$ be the length of this word and if $z \in \{I_n \pm E(i, i+1)\}$ and let $N(g, z)$ be the number of times $z$ appears in this word. Then

$$|G||P^{st} - \mu||_2^2 \leq |G|(e^{-t/A} + ||Q^{st/2A} - \mu||_2^2)$$

8
where

\[ A = \max_{\{I_n \pm E(i,i+1)\}} \frac{1}{P(z)} \sum_{g \in C_i(\pm 1) \cup C_i(\pm a)} |g|N(g,z)Q(g). \]

Notice that each element of \( C_i(\pm 1) \cup C_i(\pm a) \) must be expressed as a product of elements of the form \( \{I_n \pm E(i,i + 1)\} \). It so happens that the paths considered in the following section are of odd length exactly because \( a \) and one are odd integers.

### 8.1 Building up \( I_n + bE(i, i + 2) \) in \( O(\sqrt{b}) \) steps

At first, the goal is to create the element \( I + E(i, i + 2) \) which has the entry 1 in position \((i, i + 2)\). But that simply occurs by considering the following commutator:

\[ I + E(i, i + 2) = [I + E(i + 1, i + 2), I - E(i, i + 1)], \]

where \([x, y] = x^{-1}y^{-1}xy\) for \( x, y \in G \). For example, if \( n = 4 \) and \( i = 2 \), we have that:

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}.
\]

Then just notice the following identity holds if the entries are over \( \mathbb{R} \):

\[ I + bE(i, i + 2) = [I + \sqrt{b}E(i + 1, i + 2), I - \sqrt{b}E(i, i + 1)]. \quad (10) \]

This idea gives rise to the following lemma.

**Lemma 9.** We can express \( I_n + bE(i, i + 2) \) as a word in the generators \( S \) whose length is even and is at most \( 12 \lfloor \sqrt{b} \rfloor + 10 \).

**Proof.** Equation (10) is over \( \mathbb{R} \). To get something similar over \( \mathbb{Z}/p\mathbb{Z} \), consider the following identity

\[ I + bE(i, i + 2) = (I + (b - \lfloor \sqrt{b} \rfloor^2)E(i, i + 2))(I + \lfloor \sqrt{b} \rfloor^2 E(i, i + 2)), \quad (11) \]

where we use (10) to get \( E(i, i + 2))(I + \lfloor \sqrt{b} \rfloor^2 E(i, i + 2)) \) and we write

\[ (I + (b - \lfloor \sqrt{b} \rfloor^2)E(i, i + 2)) = [(I + E(i, i + 1)(I + E(i, i + 2))b - \lfloor \sqrt{b} \rfloor^2]. \]

The length of the word that occurs by (10) and (11) is at most \( 4(b - \lfloor \sqrt{b} \rfloor^2) + 4\lfloor \sqrt{b} \rfloor \leq 12\lfloor \sqrt{b} \rfloor + 10 \). This means that we can achieve to express \( I_n + bE(i, i + 2) \) as a word in the generators \( S \), whose length is at most \( O(\sqrt{b}) \). Notice that we expressed \( I + bE(i, i + 2) \) as a product of one or two commutators, therefore the length of the word is even. \( \square \)
8.2 Building up $C_i(±1)$

In this section, we show how to express an element of $C_i(±1)$ as a word in the generators $S$.

**Lemma 10.** Let $B ∈ C_i(±1)$. Then, $B$ can be expressed as a word in the generators $S$, whose length is odd and is at most $O(n, √n)$. Also, each generator $I ± E(j, j + 1)$ appears at most $O(√n)$ times in such a word.

To build an element $B$ of the conjugacy class of $I ± E(i, i + 1)$, the main idea is to build two matrices $B_1$ and $B_2$ in $G$ whose product is $B$. More specifically, the $i + 1, i + 3, i + 5$... columns of $B_1$ are the same as the ones of $B$ and the rest zeros and similarly $B_2$ has the same $i + 2, i + 4$,... column as $B$ and the rest zeros. $B_1$ will have odd length when expressed as a word in the generators $S$, while $B_2$ will have even length. This way, $B$ will have indeed odd length, something that is needed to do comparison.

**Building up the odd columns.** This section describes how to build $B_1$.

**Definition 11.** Let $A_i$ be the matrix that has ones on the diagonal and on the positions $(i, i + 1), (i − 1, i + 1), . . . (1, i + 1)$.

For example, if $n = 6$ and $i = 3$, we have that

$$A_3 = \begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}.$$ 

Firstly, notice that to express $A_i$ as a word in the generators $S$, conjugate $I + E(i, i + 1)$ by $I − E(i − 1, i)$ to get a one exactly above the position $(i, i + 1)$. Continue conjugating by $I − E(j − 1, j), j = 2, 3, . . . i − 1$ to get ones everywhere above the position $(i, i + 1)$. This can be formally written as

$$A_i = \left(\prod_{j=1}^{i-1}(I + E(j, j + 1))\right)(I + E(i, i + 1))\left(\prod_{j=1}^{i-1}(I − E(i − j, i − j + 1))\right), \quad (12)$$

where $\prod_{i=1}^{n} x_i = x_1 x_2 . . . x_n$.

**Example 12.** For $n = 6$, we can get $A_3$ by conjugating $I + E(3, 4)$ by $(I − E(2, 3))(I − E(1, 2))$:

$$A_3 = \begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix} \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}.$$
We now explain why the matrix
\[
\left( \prod_{j=1}^{i-2} \left( I - (a_j - 1)E(j,j+2) \right) \right)^{-1} A_i \prod_{j=1}^{i-2} \left( I - (a_j - 1)E(j,j+2) \right),
\]
has the same \( i + 1 \) column as \( B \). Conjugating \( A_i \) by \( I - (a_1 - 1)E(1,3) \) turns the 1 in position \((1, i + 1)\) into \( a_1 \) in \( O(\sqrt{p}) \) steps, as explained in Lemma 9 and by the following computation:

\[
\begin{bmatrix}
1 & 0 & 0 & a_1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & a_1 - 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & a_1 + 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

Conjugating by elements of the form \( I - (a_j - 1)E(j,j+2), j = 2, 3 \ldots i - 2 \) and at the end multiplying from the left by \( I - (a_{i-1} - 1)E(i-1, i+1) \) builds the first column of the box in \( O(n\sqrt{p}) \) steps. Conjugating doesn’t change the parity of the length. Lemma 9 says that multiplying by elements of the form \( I - (a_{i-1} - 1)E(i-1, i+1) \) doesn’t affect the parity of the word either. That is so far we have a word of odd length.

**Example 13.** The following calculation illustrates how we conjugate \( A_3 \) by \( \prod_{j=1}^{i-2} I - (a_j - 1)E(j,j+2) \), to create the column elements that we desire.

\[
\begin{bmatrix}
1 & 0 & 0 & 2 & 0 & 0 \\
0 & 1 & 0 & 3 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 2 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 2 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
1 & 0 & -1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & -2 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

Now conjugating by \( I + b_3E(i+1,i+3) \) will force the third column of the box to be exactly what we want. Continue conjugating by elements of the form \( I + b_{j+2}E(j,j+2), j = i + 1, i + 3, \ldots \) to create the odd columns of the box in \( O(n\sqrt{p}) \) steps.
Remark 16. The construction presented in this section does not work for the conjugacy class of $C_8$. To explain how to express an element of $C_8$ as a word in the generators $S$, we begin with building up $A_{i+1}$, (see Definition 11), following the construction of (12). Imitate the construction of the first column to create the second column of the box as wished. Then multiply with $I - E(i + 1, i + 2)$ to get rid of the 1 in position $(i + 1, i + 2)$. And then conjugate by $I + b_j E(j + 2, j + 4)$ for $j \geq 1$ as many times as needed to create the even columns. This way $B_2$ is expressed as a word in the generators $S$ of even length of order $O(n\sqrt{p})$.

### Building up the even columns

The next step creates the even columns in a separate, new matrix $B_2$. We begin with building up $A_{i+1}$, (see Definition 11), following the construction of (12). Imitate the construction of the first column to create the second column of the box as wished. Then multiply with $I - E(i + 1, i + 2)$ to get rid of the 1 in position $(i + 1, i + 2)$. And then conjugate by $I + b_j E(j + 2, j + 4)$ for $j \geq 1$ as many times as needed to create the even columns. This way $B_2$ is expressed as a word in the generators $S$ of even length of order $O(n\sqrt{p})$.

Building up $B$. At this point we notice that $B = B_1 B_2$. This gives the conjugacy class wanted in $O(n\sqrt{p})$ steps and each generator has multiplicity at most $O(\sqrt{p})$.

The following example illustrates how an element of the conjugacy class of $S$ can be obtained if the even columns and the odd columns are constructed in two separate matrices.

Example 15. In this example, we illustrate why the product of $B_1$ and $B_2$ is $B$.

$$
\begin{bmatrix}
1 & 0 & 0 & 2 & -4 & 10 \\
0 & 1 & 0 & 3 & -6 & 15 \\
0 & 0 & 1 & 1 & -2 & 5 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 & 2 & 0 & 10 \\
0 & 1 & 0 & 3 & 0 & 15 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 0 & -4 & 0 \\
0 & 1 & 0 & 0 & -6 & 0 \\
0 & 0 & 1 & 0 & -2 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

Remark 16. The construction presented in this section does not work for the conjugacy class of $C_1(\pm 1)$. But in this case, building up the first column of the box is easy, because we can begin with $A_1 = I \pm E(1, 2)$, which has the same first column as $B$. Then we continue building $B_1$ as explained directly above Example 13. $B_2$ is built on the same way as before.

### 8.3 Building up the conjugacy class of $C_4(\pm a)$

In this section, we explain how to express an element of $C_4(\pm a)$ as a word in the generators $S$.

Lemma 17. Let $B \in C_4(\pm a)$. Then $B$ can be expressed as a word in the generators $S$, whose length is odd and is at most $O(n\sqrt{p})$. Also, each generator $I \pm E(j, j + 1)$ appears at most $O(\sqrt{p})$ times in such a word.
Proof. To construct the conjugacy class of $I + a E(i, i + 1)$, things are similar to Section 8.2. Let $B \in C_i(\pm a)$. We start by constructing $A_i$, as indicated by Lemma 12 in at most $2n + 1$ steps. Then we create the entries in positions $(j, i + 1), j \in \{1, \ldots, i - 1\}$, just as described directly above Example 13. Then, we multiply by $I + (a - 1)E(i, i + 1)$ from the left to set the entry on position $(i, i + 1)$ equal to $a$. This is illustrated as

$$
\begin{pmatrix}
1 & 0 & 0 & a_1 & 0 & 0 \\
0 & 1 & 0 & a_2 & 0 & 0 \\
0 & 0 & 1 & a & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} = (I + (a - 1)E(i, i + 1))
$$

Now conjugate by $(I + E(i, i + 2))\frac{b_2}{a} = I + \frac{b_2}{a}E(i, i + 2)$ to get exactly the third column. Notice that the third column of the box occurs by multiplying the first column of the box by $\frac{b_2}{a}$.

$$
\begin{pmatrix}
1 & 0 & 0 & a_1 & 0 & a_1 b_2 a^{-1} \\
0 & 1 & 0 & a_2 & 0 & a_2 b_2 a^{-1} \\
0 & 0 & 1 & a & 0 & b_2 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} = (I + \frac{b_2}{a}E(i, i + 2)) = (I - \frac{b_2}{a}E(i, i + 2)).
$$

Conjugating by $(I - \frac{b_2}{a}E(i + 2, i + 4))$, we get the fifth column. Continuing this way, we construct $B_1$, the matrix of the odd columns. $B_2$ is constructed just like in Section 8.2. Therefore, we have that $B = B_1B_2$. Notice that again, the length of the word is at most $O(n \sqrt{p})$ and the multiplicity of each generator is at most $O(\sqrt{p})$.

**Remark 18.** The construction presented in this section does not work for the conjugacy class of $C_1(\pm a)$. But in this case, building up the first column of the box is easy, because we can begin with $A_1 = I \pm aE(1, 2)$, which has the same first column as $B$. Then we continue building $B_1$ as explained directly above Example 13. $B_2$ is built on the same way as before.

## 9 Proof of Theorem 1

Proof. Lemmas 8, 10 and 17 prove that $A = O(pn^2)$. We consider $A = kp^m$, where $k$ is a universal constant. Therefore, Lemma 8 gives that there exist universal constants $0 < b, d < \infty$ such that for $c > 2$ and $t \geq cbp^2n^4$, we have that

$$
|G||P^\ast - \mu||^2 \leq |G|\left(e^{-t/A} + ||Q^\ast A - \mu||^2\right) \leq p^m e^{-ckn^2p} + \alpha e^{-c} \leq de^{-c},
$$

since $p^n e^{-ckn^2p} \leq e^{-c}$.
10 Acknowledgments

I would like to thank Persi Diaconis for all the useful comments and guidance he provided.
This paper is adapted from my PhD thesis, for the completion of which I was supported by Stanford University.

References

[1] Carlos A. M. André. Basic sums of coadjoint orbits of the unitriangular group. J. Algebra, 176(3):959–1000, 1995.
[2] Carlos A. M. André. The regular character of the unitriangular group. J. Algebra, 201(1):1–52, 1998.
[3] Carlos A. M. André. Basic characters of the unitriangular group (for arbitrary primes). Proc. Amer. Math. Soc., 130(7):1943–1954 (electronic), 2002.
[4] Ery Arias-Castro, Persi Diaconis, and Richard Stanley. A super-class walk on upper-triangular matrices. J. Algebra, 278(2):739–765, 2004.
[5] P. Diaconis and L. Saloff-Coste. Nash inequalities for finite Markov chains. J. Theoret. Probab., 9(2):459–510, 1996.
[6] Persi Diaconis. Group representations in probability and statistics, volume 11 of Institute of Mathematical Statistics Lecture Notes—Monograph Series. Institute of Mathematical Statistics, Hayward, CA, 1988.
[7] Persi Diaconis. Threads through group theory. In Character theory of finite groups, volume 524 of Contemp. Math., pages 33–47. Amer. Math. Soc., Providence, RI, 2010.
[8] Persi Diaconis and Bob Hough. Random walk on unipotent matrix groups. arXiv:1512.06304, 2015.
[9] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. Markov chains and mixing times. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.
[15] Laurent Saloff-Coste. Random walks on finite groups. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 263–346. Springer, Berlin, 2004.

[16] Ganguly Shirshendu and Fabio Martinelli.

[17] Richard Stong. Random walks on the groups of upper triangular matrices. *Ann. Probab.*, 23(4):1939–1949, 1995.

[18] Ning Yan. *Representation theory of the finite unipotent linear groups*. ProQuest LLC, Ann Arbor, MI, 2001. Thesis (Ph.D.)–University of Pennsylvania.

[19] Ning Yan. Representations of finite unipotent linear groups by the method of clusters. *arXiv:1004.2674*, 2010.

[20] Maria Russell Zack. *Convergence to uniform on the finite Heisenberg group and applications to random number generators*. ProQuest LLC, Ann Arbor, MI, 1989. Thesis (Ph.D.)–University of California, San Diego.