# TWO RANDOM WALKS ON UPPER TRIANGULAR MATRICES

Igor Pak

Department of Mathematics
Yale University
New Haven, CT 06520
paki@math.yale.edu

December 13, 1999

Abstract. We study two random walks on a group of upper triangular matrices. In each case, we give upper bound on the mixing time by using a stopping time technique.

## Introduction

In the past decades there has been a tremendous progress in applications of Monte Carlo methods to various problems in Computer Science and Statistical Physics. In the heart of these applications there is usually a Markov chain, and analysis of its rate of convergence. Several methods for such analysis has been developed (see [AF,D3,K,SJ]). In this paper we give some nontrivial applications of the recently introduced *stopping time approach* (see [AD2,DF,P1]) which appears to be powerful when other methods fail.

Consider the following problem. Let $G = U(n, \mathbb{F}_q)$ be the group of upper triangular matrices over the finite field with $q$ elements (with matrix multiplication as a group operation). Define a random walk on $G$ as follows. Start at identity. At each step choose uniformly a pair $(i, j)$, $1 \le i < j \le n$. Then take row $j$, multiply it by a uniform element $a \in \mathbb{F}_q$, and add it to a row $i$. We prove that this random walk mixes in under $O(n^2 \log n)$ steps. More precisely, we show that after $O(n^2 \log n)$ steps the probability that the walk is at $g \in G$ is at least $1/2|G|$. We also give an upper bound on the second eigenvalue of the corresponding Cayley graph.

The second random walk we study is somewhat similar in flavor. Consider a random walk on $G = U(n, \mathbb{F}_q)$ when only pairs $(i, i+1)$ are allowed. Formally, start at identity. At each step choose uniform integer $i$, $1 \le i < n$. Then take row $i+1$, multiply it by a uniform element $a \in \mathbb{F}_q$, and add it to a row $i$. We show that this random walk mixes in under $O(n^{2.5})$ steps given $q > 2\,n^3$. This random walk has been studied before on several occasions (see [DSC2,S1,S2]). The previous

*Key words and phrases.* Random walks on groups, strong stationary time, separation distance, nilpotent group.

best upper bound is due to Stong who showed in [S2] that the walk mixes in under $O(n^3)$ steps.

While these two random walks may seem as very special examples, their importance comes from the nature of the group $G = U(n, \mathbb{F}_q)$. Random walks on abelian groups are well understood (see [D1]), but little progress has been made in a general non–commutative case. Nilpotent groups, being the closest class of group to abelian groups, probably have the best chance to be fully analyzed. Thus our work shows how far the probabilistic analysis can be pushed in two natural special cases. We refer to [AP,DSC2,S2] for general results on random walks on nilpotent groups.

The technique we employ here originated in works of Aldous and Diaconis (see [AD1,AD2]) and is based on explicit constructions of the randomized stopping rules, also called strong stationary times in this case. The requirement is that the stopping state must have a stationary distribution and must be independent of the stopping time (see definitions below). This technique was further developed in the papers [DF,LW,P1,P2,PV].

Not unlike coupling technique, construction of strong stationary times is more an art than a science. They differ for even very similar Markov chains, and their analysis can range from an almost trivial to an extremely hard. In the recent past the use of the stopping rules for bounding convergence of difficult Markov chains has been rare if not altogether nonexistent. We hope that this paper will bring this important instrument back to life.

The rest of the paper is written as follows. In sections $1, 2$ we present basic definitions and state the main results. In section 3 we describe the technique we are using along with some key examples. Both sections $4, 5$ contain constructions of the strong stationary times which give bounds on convergence.

## 1. Random walks on groups

Let $G$ be a finite group, and let $S$ be a set of generators of $G$. For a given *holding probability* $p$, $1 > p > 0$ define a probability distribution $Q^m$ on $G$ as the distribution of the product

$$g_1^{\epsilon_1} \cdot \ldots \cdot g_m^{\epsilon_m} ,$$

where $g_i$ are uniform and independent in $S$, and $\epsilon_i \in \{0, 1\}$ are obtained by independent flips of an unfair coin with $\mathbf{Pr}(0) = p$, $\mathbf{Pr}(1) = 1-p$. One can think of $Q^m$ as of a distribution after $m$ steps of a random walk $\mathcal{W}(G, S, p)$ on $G$ generated by set $S$ and with holding probability $p$. We will use $p > 0$ here to avoid periodicity problems.

It is not hard to see that $Q^m \to U$ as $m \to \infty$, where $U(g) = 1/|G|$ is the uniform distribution on $G$. There are several measures of how far $Q^m$ is from $U$, of which we will use the following:

$$\mathbf{s}(m) = |G| \max_{g \in G} \left( \frac{1}{|G|} - Q^m(g) \right)$$

Usually $\mathbf{s}(m)$ is called the *separation distance*, and can be thought of as one-sided $l_\infty$ distance (see [AD2,D1]). It is known that $0 \le \mathbf{d}(m) \le \mathbf{s}(m) \le 1$ for all $m \ge 1$

(see e.g. [AD2,AF]), where $\mathbf{d}(m)$ is the *total variation distance* :

$$\mathbf{d}(m) = \max_{A \subset G} |Q^m(A) - U(A)| = \frac{1}{2} \sum_{g \in G} \left| Q^m(g) - \frac{1}{|G|} \right|$$

Also, it is known that $s(m)$ is decreasing, submultiplicative, and

$$\mathbf{s}(2m) \leq 4\sqrt{2\mathbf{d}(m)}\,,$$

which roughly means that both distances have a similar asymptotic behavior (see [AD2,AF,D1]).

Now let $G$ be a compact connected Lie group with a set of generators $S$. Denote by $\mu$ the invariant measure on $G$, also known as Haar measure. Recall that $\mu$ is unique given $\mu(G) = 1$. Since in our examples will have $\mu(S) = 0$, consider a probability measure $\mu'$ on $S$. We need this measure to define a random walk (roughly: we sample element $s \in S$ according to $\mu$ and multiply the state of the walk by $s$.) Consider a natural product measure on $S^m \subset G^m$, and the probability measure $Q^m$ is defined by projection : $(g_1, \ldots, g_m) \to g_1 \cdot \ldots \cdot g_m$, where $g_i \in S$ are chosen uniformly and independently. By abuse of speech, we will say that $Q^m$ is a probability distribution of the random walk $\mathcal{W}(G, S)$ after $m$ steps. In all our examples $\mu'$ will be obvious and we will not specify it.

By analogy with a finite group case, we can define separation distance :

$$\mathbf{s}(m) = \sup_{A \subset G,\ \mu(A) > 0} \left(1 - \frac{Q^m(A)}{\mu(A)}\right)$$

Roughly, if the separation distance $\mathbf{s}(m)$ is equal to $\lambda$, then $Q^m = (1 - \lambda) \cdot \mu + \lambda \cdot \eta$ where $\eta$ corresponds to some "noise" positive measure.

Without giving all the details, let us just note here that all the properties of separation distance, its relation to the total variation distance, etc., can be translated from finite groups to compact Lie groups employing almost identical proofs.

## 2. Main results

Let $\Bbbk$ be any (finite or infinite) compact commutative ring, and let $\eta$ be an invariant measure on $\Bbbk$ (a Haar measure on an additive group of $\Bbbk$.) For example, $\Bbbk$ can be the ring of $p$-adic integers $\mathbb{Z}_p$, or a finite field $\mathbb{F}_q$. Denote by $\beta = \beta(\Bbbk)$ the measure of noninvertible elements:

$$\beta(\Bbbk) = \eta\big(\{a \in \Bbbk \,|\, \nexists\, b, \text{s.t. } a \cdot b = 1\}\big)$$

For example, $\beta(\mathbb{F}_q) = 1/q$, and $\beta(\mathbb{Z}_p) = 1/p$.

Let $G = U(n, \Bbbk)$ be the group of upper triangular $(n \times n)$-matrices over $\Bbbk$ with ones on diagonal. Denote by $\mu$ the invariant measure on $G$, also known as Haar measure. It is known that $\mu$ is given as a product measure

$$\mu = \eta \times \eta \times \cdots \times \eta$$

where the product is taken over all $\binom{n}{2}$ entries above diagonal (see e.g. [H]).

Let $S_1 \subset G$ be a set of matrices with ones on diagonal and zeroes elsewhere but one element above diagonal. Clearly, $S_1$ is a generating set of $G$. Analogously, let $S_2 \subset G$ be a set of matrices with ones on diagonal and zeroes elsewhere but one element right above diagonal. Clearly, $S_2$ is a generating set of $G$. It is easy to see that $S_2 \subset S_1$ is also a generating set of $G$.

As before, let $Q_1^k$ ($Q_2^k$) be a probability distribution of the product $M_1 \cdot \ldots \cdot M_k$ where $M_i$ are independent and uniform in $S_1$ ($S_2$)[1]. We think of $Q_1^k$, $Q_2^k$ as of probability distribution of the $k$-th step of random walks $\mathcal{W}_1 = \mathcal{W}(G, S_1)$, $\mathcal{W}_2 = \mathcal{W}(G, S_2)$ on $G$.

**Theorem 2.1** *Let $G = U(n, \Bbbk)$. Then for the separation distance of the corresponding random walk $\mathcal{W}_1$ we have*

$$\mathbf{s}(m) \le e^{-2c} \ \text{ for } m = n^2 \log n + c\, n^2.$$

*On the other hand,*

$$\mathbf{s}(m) \ge e^{-e^{-2c}} \ \text{ for } m = \frac{1}{2} n^2 \log n - c\, n^2.$$

Note that the theorem does not prove existence of the cutoff (see [D2,P1]). We conjecture that a cutoff exists in this case, though perhaps not very sharp. Note also that rate of decay in the theorem suggests that the second largest eigenvalue is of the form $\lambda_2 \sim 1 - c/n^2$ for some universal $c > 0$. It would be of interest to find a rigorous argument to prove that.

**Theorem 2.2** *Let $G = U(n, \Bbbk)$, and $\beta(\Bbbk) = 0$). Then for the separation distance of the corresponding random walk $\mathcal{W}_2$ we have*

$$\mathbf{s}(m) < \frac{B}{c^2} \ \text{ for } m = A\, n^{2.5} + c\, n^2,$$

*where $A, B$ are universal constants. Further, if $\beta(\Bbbk) < 1/(2n^2)$, then*

$$\mathbf{s}(m) < \frac{1}{2^c} \ \text{ for } m = c \cdot D\, n^{2.5},$$

*where $D$ is a universal constant. On the other hand, for any $\Bbbk$, $|\Bbbk| \ge 2$ we have*

$$\mathbf{s}(m) > 1 - 1/c^2 \ \text{ for } m = n^2 - c\, n.$$

Observe that the the first random walk $\mathcal{W}_1$ can be seen as an analog of a random walk on a permutation group $\Sigma_n$ generated by all transpositions $(i, j) \in \Sigma_n$. This random walk has been extensively studied (see [D1,DSh,M,P1,P2]) and has a mixing

[1]Note that when $\Bbbk$ is finite the identity in our set of generators has a positive measure. This corresponds to the holding probability $p = 1/|\Bbbk|$. When $\Bbbk$ is infinite, the holding probability $p$ is unnecessary (cf. [P1].)

time of the order $O(n \log n)$. Similarly, the second random walk $\mathcal{W}_1$ can be seen as an analog of a random walk on $\Sigma_n$ generated by adjacent transpositions $(i, i+1)$. Note, however, that this random walk mixes in under $O(n^3 \log n)$ (see [A,DSC1,W]), i.e. much slower than the previous walk. In contrast with the walks on $S_n$, we conjecture that when $\beta(\Bbbk) < 1/2n^2$ the random walk $\mathcal{W}_2$ mixes slightly faster than $\mathcal{W}_1$[2]

## 3. Strong uniform times

Let $X_t$ denote the trajectory of a random walk $\mathcal{W} = \mathcal{W}(G, S, p)$ on a finite group $G$. Recall that $\mathcal{W}$ has a uniform stationary distribution. A *randomized stopping rule* is an algorithm which observes the walk and stops it depending on the state passed and, perhaps, additional randomness. Denote by $\tau$ the *stopping time* of this rule. By $\varrho = X_\tau$ denote the *stopping state*. We think of $\tau$ and $\varrho$ as of random variables.

The stopping time $\tau$ is called *strong uniform* if $\varrho$ is uniform and independent of $\tau$. In other words,

$$\mathbf{Pr}(\varrho = g \,|\, \tau = k) = \frac{1}{|G|} \quad \text{for all } \; g \in G, \, k > 0$$

The main application of the strong uniform time is the following result of Aldous and Diaconis (see [AD2,D1]).

**Theorem 3.1** *Let $\tau$ be a strong uniform time for a random walk $\mathcal{W}$. Then:*

$$\mathbf{s}(k) \leq \mathbf{Pr}(\tau > k), \; k > 0$$

For other versions of this result see [F,P1].

**Example 3.2** Here is an example of a strong uniform time due to Broder (see [D1]). Let $G = \mathbb{Z}_2^n$, $S = \{(0, \ldots, 0, 1_i, 0, \ldots, 0), 1 \leq i \leq n\}$. One can think of the random walk $\mathcal{W}(G, S, 1/2)$ as of a nearest neighbor random walk on a cube. Start at $(0, \ldots, 0)$. At each step we flip a fair coin and choose a direction uniformly. If heads, move in that direction. If tails, stay.

Let us mark the directions when we choose them. Define a stopping time $\tau$ to be the time when all the directions are marked. This defines a strong uniform time. Indeed, whenever we mark a direction $i$, the walk has the $i$-the coordinate either 0 or 1 with probability $1/2$. Further walk moves do not change this condition. Thus when we mark all the coordinates we are equally likely to be at any element of $G$. This proves the claim.

**Example 3.3** Here is another example of a strong uniform time (see [DF,P1]). Let $G = \mathbb{Z}_n$, $n = 2^m$, and let $S = \{\pm 1\}$. The random walk $\mathcal{W}(G, S, 1/2)$ can be defined as follows. Start at 0. At each step flip two fair coins. The first will

[2]This conjecture has been recently resolved by D. Coppersmith and the author by a careful analysis of the board game defined in section 5. The paper is forthcoming.

determine the direction, while the second will determine whether we move in that direction or stay put.

Define a stopping time $\tau$ as follows. Walk till we hit any of the 2 points $\pm n/4$ (first stage). Then walk some more till we hit any of the 4 points $\pm n/4 \pm n/8$ (second stage). Then walk till we hit any of the 8 points $\pm n/4 \pm n/8 \pm n/16$, etc. Proceed further for $(m-1)$ such hitting times. At the final stage we walk till we hit any of the odd numbers. Now do one more step and stop.

Let us prove that $\tau$ is strong uniform. Indeed, after the first stage, once we hit either of the 2 points $\pm n/4$, by the symmetry we can be at either of these two elements with equal probability $1/2$. Analogously, after the second stage, by the symmetry we hit either of the 4 points $\pm n/4 \pm n/8$ with equal probability $1/4$, and so on. Finally, after the stage $(m-1)$ we are at a uniform odd numbered element. At the next step we either move into either direction with probability $1/2$ or stay with probability $1/2$. Thus the stopping state $\varrho$ is uniform. Use induction to show that it is independent of $\tau$. This proves the claim.

A strong uniform time $\tau$ is called *perfect* if $s(k) = \mathbf{Pr}(\tau > k)$ for all $k > 0$. It is known that a perfect time always exists (see [AD2,D1,P1]). An element $\widetilde{g} \in G$ is called *halting* for a stopping time $\tau$ if the random walk always stops whenever it gets there. If a strong uniform time has a halting element, then it is perfect (see [P1,DF]).

In example 3.2, the stopping time $\tau$ defined above is perfect. Indeed, it has a halting element $\widetilde{g} = (1, \ldots, 1)$ since the only way to get to $\widetilde{g}$ is by marking *all* the coordinates. Thus the total separation distance for the random walk $\mathcal{W}$ is given *exactly* by coupon collector's problem (see [F]). Therefore for all $c \in \mathbb{R}$ we have :

$$\mathbf{s}(m) = \mathbf{Pr}(\tau > m) \to e^{-e^{-c}} \text{ as } n \to \infty$$

where $m = n \log n + c\, n$ (see [ER]).

In example 3.3, the stopping time $\tau$ defined above is also perfect. Indeed, it has a halting element $n/2$ since the only way to get to $n/2$ is to pass $\pm n/4$, then $\pm(n/4 + n/8), \ldots, n/2 \pm 1$. That means that we must have finished the first stage, the second stage, $\ldots$, the $(m-1)$-th stage. But since $n/2$ is even, we must have also made the last additional step as well. By construction of $\tau$ we must always stop then. This implies that $n/2$ is halting indeed. Therefore we have $\mathbf{s}(k) = \mathbf{Pr}(\tau > k)$ and the latter probability can be computed by using classical results on hitting times on a line (see [DF,P1] for details).

Now let us observe that much of what was said in this section about finite groups can be translated into results about strong uniform time on compact groups. One needs to worry somewhat about sets of measure zero, but these details are easy to attend.

Let $G$ be a compact group, and let $\mu$ be an invariant measure. Consider a stopping time $\tau$ defined by a randomized stopping rule. As before, denote by $\varrho$ the stopping state of the walk. Define a stopping time $\tau$ to be *strong uniform* if for all $A \subset G$ and $k > 0$ we have

$$\mathbf{Pr}(\varrho \in A \,|\tau = k) = \mu(A)$$

In a finite case, when $\mu(A) = |A|/|G|$, this definition reduces to the old definition. To simplify the notation, we say $g$ is uniform in $G$ if $g$ is the value of a random variable $X$ such that $\mathbf{Pr}(X \in A) = \mu(A)$ for all $A \subset G$.

**Theorem 3.4** *Let $\tau$ be a strong uniform time for a random walk $\mathcal{W}$ on a compact group $G$. Then:*

$$\mathbf{s}(k) \leq \mathbf{Pr}(\tau > k), \ k > 0$$

*Proof.* Follows verbatim from the proof of Theorem 3.1 in [AD] (see also [D1, AF]). □

**Example 3.5** Let $\Bbbk$ be as above, $\Bbbk_+$ be its additive group. Consider a compact group $G = (\Bbbk_+)^n$. Let $S$ contain all the elements $g$ as follows : $g = (0, 0, \ldots, a, \ldots, 0)$ where $a \in \Bbbk$, and $a$ in $i$-th position, $1 \leq i \leq n$. The random walk $\mathcal{W}(G, S)$ can be thought as of randomly choosing a coordinate and changing it to a random number.

It is easy to see that this example is similar to the example 3.2. Indeed, consider a stopping time $\tau$ : *wait till all coordinates are marked, and then stop.* This is again a strong uniform time. Thus the separations distance is again given by coupon collector's problem. Observe that $\tau$ is also a perfect time as an element $(1, 1, \ldots, 1)$ is halting (cf. Example 3.2.) Therefore by coupon collector's problem we have:

$$\mathbf{s}(k) = \mathbf{Pr}(\tau > k) > \exp\left(-e^{-c}\right),$$

where $k = n \ln n - c \cdot n$ (see e.g. [D1, F].)

Note also that if $\beta(\Bbbk) = 0$, we can define a formally different stopping time $\tau'$ : *wait till all coordinates are not divisors of zero.* Since the set $A = (\mathbb{Z}_p \setminus 0)^n$ has a measure $\mu(A) = 1$, we can disregard $(G \setminus A)$ and check that $\tau'$ is strong uniform. Simply note that the probability that $\tau \neq \tau'$ is zero.

## 4. Proof of Theorem 2.1

Let $\Bbbk$ be a compact commutative ring. Denote by $U(n, \Bbbk)$ the group of upper triangular matrices over $\Bbbk$. Define *elementary transvections* $E_{i,j}(a)$, $1 \leq i < j \leq n$ to be matrices with 1-s on diagonal, $a \in \Bbbk$ in the entry $(i, j)$, and 0 elsewhere. We will need the following lemma.

**Lemma 4.1** *Let $E \in U(n, \Bbbk)$ be given by a product of matrices*

$$(*) \quad E = E_{i_1,j_1}(a_1) \cdot \ldots \cdot E_{i_k,j_k}(a_k)$$

*where $1 \leq i_l < j_l \leq n$ for all $l = 1, \ldots, n$. Suppose also all possible pairs $(i, j)$, $1 \leq i < j, \leq n$ occur in $(*)$. Then $E$ is uniform in $U(n, \Bbbk)$ given $a_l$ are uniform and independent in $\Bbbk$, $1 \leq l \leq k$.*

The lemma in a slightly different form was introduced in [P1]. It was recently generalized in [AP] to all nilpotent groups. Before we prove the lemma let us deduce the theorem from it.

*Proof of Theorem* 2.1. Let $n$ be fixed. Denote by $\Xi = \Xi_n$ the set of all pairs $(i,j)$, $1 \le i < j \le n$. Think of our random walk $\mathcal{W}_1 = \mathcal{W}(G, S^{(1)})$ as of successive multiplication by a matrix $E_{i,j}(a)$ where each pair $(i,j)$ is chosen uniformly and independently in $\Xi$ and $a$ is uniform in $\Bbbk$ and independent of $(i,j)$. Denote by $\gamma$ the sequence of elements of $\Xi$ as in decomposition $(*)$. We say that $\gamma$ *contains* $(i,j)$ if $(i_l, j_l) = (i,j)$ for some $l$ with $1 \le l \le k$. We say that $(*)$ *contains* $\Xi$ if it contains all $(i,j) \in \Xi$. By definition, all decompositions in Lemma 3.1 contain $\Xi$.

Consider the following stopping time $\tau$. Mark an element $(i,j) \in \Xi$ the first time we choose a generator $E_{i,j}(a)$. Once the elements are marked, they remain marked forever. Stop when all elements in $\Xi$ are marked. We claim that the stopping time $\tau$ is strong uniform. Indeed, given we stopped at time $k$, we know that all the pairs $(i,j)$ are marked. But by Lemma 3.1 for *any* such decomposition, the product $E$ is uniform in $G = U(n, \Bbbk)$. Therefore $\tau$ is strong uniform. Formally, let $\gamma = \big((i_1,j_1), \ldots, (i_k,j_k)\big)$ be a sequence of elements of $\Xi$. By $E_\gamma$ denote a random variable defined by a product $(*)$, where $a_t$ are uniform and independent. For every subset $A \subset G$ we have

$$\mathbf{Pr}(\varrho \in A \,|\, \tau \le k) = \frac{\sum_{\text{sequence } \gamma \text{ of length } k \text{ which contains } \Xi} \mathbf{Pr}(E_\gamma \in A)}{\#\text{ of sequences } \gamma \text{ of length } k \text{ which contain } \Xi} = \mu(A)$$

The first equality follows by definition of $\tau$. Indeed, if $\tau = k$ the sequence $\gamma$ must contain $\Xi$, and each such a sequence is equally likely. The second equality follows from Lemma 4.1. This immediately proves that $\tau$ is strong uniform. Now to get an upper bound use coupon collector's problem with $|\Xi| = \binom{n}{2}$ coupons (see [F]).

For the lower bound, consider the entries right above the diagonal. Observe that we touch these entries with probability $(n-1)/|\Xi| = 2/n$. Thus the random walk projected on these entries is equivalent to the random walk on $(\Bbbk_+)^{n-1}$ with holding probability $1 - 2/n$. But for this walk the lower bound is again given by coupon collector's problem with $(n-1)$ coupons (see Example 3.5). This easily implies the result. We omit the obvious details. $\square$

*Proof of Lemma* 4.1 Suppose our decomposition is fixed and contains $\Xi$ (see above). To simplify the notation, we adopt the following conventions. We use $B_l$ to denote matrices in $G$ with zero entries in $(i,j) \in \Xi$, $1 \le j - i < l$. Respectively use $R_l$ to denote matrices with zeroes everywhere but main diagonal and $l$-th diagonal: entries $(i,j)$, $i - j = l$. The main idea of the proof is to rewrite a product $(*)$ using group structure of $G = U(n, \Bbbk)$. Denote by $G_l = U(n, l, \Bbbk) \subset G$ the group of upper triangular matrices with 1's on main diagonal and 0's in all diagonals below $l$-th diagonal. In our notation $R_l, B_l \in U(n, l, \Bbbk)$.

Our main tool is the following identity:

$$(\circ) \quad R_l \cdot R_m = R_m \cdot R_l \cdot B_{\max(l,m)+1}$$

This identity follows immediately from definition of matrix multiplication. More generally,

$$(\circ') \quad B_l \cdot B_m = B_m \cdot B_l \cdot B_{\max(l,m)+1}$$

When $l = m = 1$ it means that commutator of two matrices in $G = G_1$ have zeroes in the first diagonal.

Now let us rewrite $(*)$. Start with a product

$$R_{l_1} \cdot \ldots \cdot R_{l_k}$$

Take all $R_1$ one by one and move to the left using $(\circ)$. Note that $B$'s can appear only to the right of $R_l$, $l > 1$. Then, take all $R_2$ and move then to the left using $(\circ)$ leaving them to the right of $R_1$'s. Then, move $B_3$'s (which appeared after we used $(\circ)$ in the first stage) to the left (using $(\circ')$ now). Then, move $R_3$'s to the left, and so forth. At the end we obtain a product of the following type

$$E = R_1 \cdot R_1' \cdot \ldots \cdot R_2 \cdot R_2' \cdot \ldots \cdot B_3 \cdot B_3' \cdot \ldots \cdot R_3 \cdot \ldots \ldots \cdot B_{n-1} \cdot B_{n-1}' \cdot \ldots \cdot R_{n-1} \cdot R_{n-1}' \cdot \ldots$$

Now, start reading this product from right to left.

First, take the product $D_{n-1}$ of all $R_{n-1}(a)$'s given the corresponding $a$'s are uniform in $\Bbbk$. This product generates a uniform element of $G_{n-1} = U(n, n-1, \Bbbk)$. Indeed, simply use the condition that the original product contains $(1, n)$. Now, when we multiply $D_{n-1}$ on the left by $B_{n-1}$'s we are still going to have a uniform element of $G_{n-1}$. Denote by $D_{n-2}$ the product of all $R_{n-2}$'s. Observe that $D_{n-2}$ is a matrix with a uniform $(n-2)$-th diagonal. Indeed, $R_{n-2}$'s commute when applying to that diagonal and since this product contains $(1, n-1)$ and $(2, n)$ will generate it uniformly. But then, when $D_{n-2}$ is multiplied by a uniform element of $G_{n-1}$, we get a uniform element of $G_{n-2}$. Again, when $B_{n-2}$'s are multiplied from the left on uniform element of $G_{n-2}$, we still get a uniform element of $G_{n-2}$. Now proceed by induction till we get a uniform element of $G_1 = G$. $\square$

## 5. Proof of Theorem 2.2

We need to introduce several definitions. Consider the following *board game* (for just one player). Define *board* to be an interval of integers $[1, n]$. Place $n$ pieces numbered from 1 up to $n$ on the first space. At each step pick a uniform integer $i$ from $[1, n-1]$. If the space $i > 1$ is occupied *and* space $i+1$ is unoccupied, move the piece from $i$ to $i+1$. If $i = 1$ move the smallest of the pieces from space 1 to 2. In all other possibilities (space $i$ is unoccupied, or both $i$ and $i+1$ are occupied) stay put. Note that according to the rules we can never have more than one piece on space $i > 1$. The game is over when all spaces are occupied, i.e. no moves are allowed. The problem is to analyze this game.

One can ask the following question : *What is the probability that the game is not over after $m$ steps?* It turns out that this probability gives an upper bound for the separation distance of the random walk $\mathcal{W}_2$. Formally, denote by $\varkappa$ be the stopping time of the game above.

**Lemma** 5.1 *Let $\beta(\Bbbk) = 0$, and $\mathbf{s}(m)$ be the separation distance of the random walk $\mathcal{W}_2$ defined above. Then for all $m > 0$*

$$\mathbf{s}(m) \leq \mathbf{Pr}(\varkappa > m)$$

*Proof.* The idea of the proof is based on construction of a strong uniform time $\tau$ for the random walk $\mathcal{W}_2$ and then show that $\mathbf{Pr}(\tau = m) = \mathbf{Pr}(\varkappa = m)$ for all $m$. By Theorem 3.4 this implies the lemma.

For convenience, number rows of the upper triangular matrices upside down. Namely, the bottom row is number 1, the next row is 2, etc., the top row is $n$. We can define the random walk $\mathcal{W}_2$ as follows: At time $t$ choose uniformly an integer $i = i(t)$ between 1 and $n-1$, and add to the $i+1$-th row the $i$-th row multiplied by a uniform number $a = a(t) \in \Bbbk$. Now let us move pieces on the board according to the same choices of integers $i(t)$. Note that in the board game we disregard the number $a = a(t)$ we used in the random walk.

For technical reason mark the row 1 at the start of the walk. We claim that at any time $t$, given pieces are positioned in spaces $1 \le i_1 < i_2 < \dots$, then the corresponding rows $i_1$, $i_2$, ... of the obtained upper triangular matrix are uniform and independent. Call the rows $i_1$, $i_2$, ... , *marked rows*. Use induction. It convenient to make a somewhat stronger inductive assumption : *every marked row is uniform and independent of all the rows above.*

The claim is trivial when $t = 0$ : the only marked row is 1 and it always stays $(0, 0, \dots, 1)$. Suppose the claim is true when $t = m$. Say at the next step we choose $i = i(t)$. We either move a game piece or stay put. If we stay put and do not add anything to any of the marked rows ($i+1 \neq i_l$ for any $l$), there is nothing to check (we change only unmarked rows). Suppose we add an unmarked row $i$ to a marked row $i+1$. We stay put in the board game. The row $i+1$ will remain uniform. Also, the row $i+1$ will remain independent of rows above. Now suppose that both rows $i$ and $i+1$ are marked, which means they are uniform and independent. Again, we stay put in the board game. We claim that after addition the rows are still uniform and independent of each other and all rows above. Indeed, call these rows $X_1$, $X_2$ and think of them as $n$-vectors over $\mathbb{Z}_p$. Clearly, $(X_1, X_2)$ are uniform and independent is *equivalent* to $(X_1, X_2 + a\,X_1)$ are uniform and independent for *any* $a$. Independence of other (unchanged) rows above is obvious. This implies the claim.

Now, suppose we add marked row $i$ to an unmarked row $i+1$. This is the only way to make a move in the board game. The row $i$ looks like $(1, x_1, \dots, x_{i-1})$. Then row $i+1$ will look $(1, a, a \cdot x_1, \dots, a \cdot x_{i-1})$. We claim that since row $i$ is uniform, and $a$ is uniform in $\Bbbk$, the row $i+1$ becomes also uniform. Indeed, this is clear if $a$ is invertible. On the other hand, if $\beta(\Bbbk) = 0$, $a$ is invertible with probability 1, which proves the claim. However now row $i$ is not *independent* on row $i+1$. We conclude that the (new marked) row $i+1$ becomes uniform and independent of rows above. This confirms the step of induction and proves the main claim.

Now consider what happens when the board game is over. Define a stopping time $\tau$ to be this time, or, equivalently, the time when all rows are marked. Hence

$$\mathbf{Pr}(\varkappa = m) = \mathbf{Pr}(\tau = m)$$

By the claim above, when the game is over we get all the rows uniform and independent on each other. Recall that the measure on $U(n, \mathbb{Z}_p)$ is a product measure. We have

$$\mathbf{Pr}(\varrho = (X_1, \dots, X_{n-1} \in A_1 \times \dots \times A_n \,|\, \tau = k) = \mu_1(A_1) \times \dots \times \mu_{n-1}(A_{n-1}) = \mu(A)$$

where $X_i$ are rows of the obtained matrix and $\mu_i$ are the corresponding measures. This finishes the proof. $\square$

**Lemma 5.2** *Let $\varkappa$ be the stopping time of the board game defined above. Then $E(\varkappa) = O(n^{2.5})$ and $Var(\varkappa) = O(n^4)$.*

*Proof.* Indeed, let us analyze the game. Denote by $s_i$ the time (the number of flips required) for the $i$-th piece to get into space $n+1-i$, from where it cannot move any further. Denote by $t_i$ the difference $s_i - s_{i-1}$, where $i = 2, \dots, n$, and let $t_1 = s_1$. Consider the time $t_1$ for the first piece to get to space $n$. Clearly, nothing is ever ahead of this piece and it moves with probability $1/n$. Since it has $n-1$ steps to go, we have

$$E(t_1) = n \cdot (n-1) \le n^2, \quad Var(t_1) \le n^3.$$

Indeed, $E(t_1) = E(y_1) + \cdots + E(y_{n-1})$, $Var(t_1) = Var(y_1) + \cdots + Var(y_{n-1})$, where $y_i$ is the time for the first piece to move from $i$ to $i+1$. By definition, $y_i$ are identical independent geometric distributions with probability of success $1/n$. Thus $E(y_i) = n$, $Var(y_i) = (1-1/n)\, n^2$ (see e.g. [F]), and we obtain the formulas above.

For the second piece, observe that since the start of the game and before $t_1$, whenever two pieces are apart, the difference in positions between the first and the second piece with equal probability $1/n$ would either increase by 1, decrease by 1, or with probability $1-2/n$ remains the same. On the other hand, the second piece can never reach the first piece which means that the above difference never fall below 1. Thus the difference changes according to a simple reflecting random walk on $\mathbb{Z}_+$. By reflection principle, direct computation, or otherwise we obtain that at time $t_1$ the second piece is about $O(\sqrt{n})$ spaces behind the first piece. Although this is a classical result, let us present here a short proof for completeness.

Let's slightly change rules of the game, but concentrate only on pieces 1 and 2. Assume that when two pieces are at positions $x$ and $x+1$ and we choose $i(t) = x$ we *exchange* these pieces. Clearly, this can only change the labels of pieces but not the positions they occupy. But now the distance between positions of the first and the second piece changes like a traditional random walk on a line (with a holding probability of $1-2/n$.) from the beginning and until time $t_1$. By then, the random walk above will make at most $2n$ steps. But as well known, after $O(n)$ steps the random walk is at a distance $O(\sqrt{n})$ with high probability (see e.g. [D1,F]). We conclude that the lagging piece (the second piece in our board game) is indeed at a distance about $O(\sqrt{n})$ spaces behind the first piece.

Note that between $s_1 = t_1$ and $s_2$ the second piece will move ahead freely since the first piece is already in position $n$ and cannot "block" it. Thus we have

$$E(t_2) = O(n\sqrt{n}), \quad Var(t_2) = O(n^3)$$

The calculation is identical to the calculation for the first piece and is omitted.

For the third and the other pieces, again consider the difference in positions between the them and the previous piece. Say, we have a piece $i$, and $\delta_i$ is the difference between pieces $i$ and $i-1$. Then $\delta_i$ can again change only by $\pm 1$ and must be nonnegative. Now, however, $\delta$ is more inclined to decrease than increase since the previous piece $i-1$ may not be able to move forward when chosen because it is directly behind $i-2$. Still, $\delta_i$ at time $s_{i-1}$ can be bounded from above by the reflecting random walk. Note also that piece $i-1$ had only $n-i+2$ spaces to go. Therefore

$$E(t_i) = O\big(n\sqrt{n+1-i}\big), \quad Var(t_i) = O\big(n^2(n+1-i)\big)$$

Note that each of the processes which happen in time $t_i$ (of the piece $i$ reaching the space $n+1-i$ *after* the piece $i-1$ reached $n-i+2$) are independent with length bounded from above. Thus we obtain:

$$E(\varkappa) = E(t_1) + E(t_2) + \cdots + E(t_n) = O(n^{2.5})$$

$$Var(\varkappa) = Var(t_1) + Var(t_2) + \cdots + Var(t_n) = O(n^4)$$

This completes proof of the lemma. $\square$

*Proof of the first part of Theorem 2.2.* From the Lemma 5.2 and one-sided Chebyshev inequality we have :

$$\mathbf{Pr}(\varkappa > m) = \mathbf{Pr}\big(\varkappa > E(\varkappa) + (m - E(\varkappa))\big) < \frac{Var(\varkappa)}{\big(m - E(\varkappa)\big)^2} \leq \frac{C_2 n^4}{c^2 n^4} = \frac{C_2}{c^2}\,,$$

where $m = C_1 n^{2.5} + c\,n^2$, $C_1, C_2$ are universal constants.

Now assume that $\beta(\Bbbk) = 0$. In this case we immediately obtain

$$\mathbf{s}(m) \leq \mathbf{Pr}(\varkappa > m) < \frac{C_2}{c^2}\,,$$

given $m = C_1 n^{2.5} + c\,n^2$. The first inequality here follows from Lemma 5.1. This implies the upper bound in Theorem 2.2 in case $\beta(\Bbbk) = 0$. $\square$

*Proof of the lower bound.* Denote by $\xi$ the first time we have a nonzero element in the matrix entry $(1, n)$. We think of $\xi$ as of random variable which depends on the random walk $\mathcal{W}_2$. We claim that $E(\xi) = n(n-1)$, $Var(\xi) \leq n^3$. Indeed, from the proof of Lemma 5.2 the distribution of $\xi$ is exactly the same as distribution of $t_1$ (the time for the first piece to reach position $n$). The computation there implies the claim.

From here and Chebyshev inequality we immediately obtain:

$$\mathbf{Pr}(\xi < n(n-1) - cn\sqrt{n}) < \frac{1}{c^2}$$

We conclude that after $m = n(n-1) - c\,n\sqrt{n}$ steps entry $(1, n)$ is zero with probability $> 1 - 1/c^2$. Denote by $A$ the set of elements in $G = U(n, \Bbbk)$ with nonzero entry $(1, n)$. Consider two cases. If $|\Bbbk| = \infty$, then $\mu(A) = 1$, and

$$\mathbf{s}(m) \geq \left(1 - \frac{Q^m(A)}{\mu(A)}\right) = 1 - \frac{1}{c^2}$$

If $|\Bbbk| < \infty$, then $\mu(A) = |A|/|G| = 1 - 1/|\Bbbk| > 1/2$. Therefore

$$\mathbf{s}(m) \geq \left(1 - \frac{Q^m(A)}{\mu(A)}\right) \geq 1 - \frac{2}{c^2}.$$

This completes the proof of the lower bound and finishes proof of Theorem 2.2 in case $\beta(\Bbbk) = 0$. $\square$.

**Remark 5.3** Let us sketch here also the dimension argument which gives an alternative proof of the lower bound when $|\Bbbk| = \infty$. Observe that the Hausdorf dimension of the topological space of group elements obtained after each new rotation can increase by at most 1. Thus after $i < \binom{n}{2}$ steps we have $\mathbf{s}(i) = 1$. We skip the details. Note also that this argument gives weaker bound than that in Theorem 2.2.

We have a reason to believe that our analysis of the board game was too crude. While estimates on $E(t_1)$ and $E(t_2)$ cannot be improved, for the subsequent pieces the expected times $E(t_i)$ probably decrease asymptotically faster than we estimate. Overall, we believe in the following result.

**Conjecture 5.4** *Let $\varkappa$ be the stopping time of the board game defined above. Then $E(\varkappa) \leq C\, n^2$ for some universal constant $C \geq 0$.*

Of course, the positive solution of this conjecture, or any improvement of our bound on $E(\varkappa)$ will give immediately improvement of the bounds on the separation distance for this walk. We hope to return to this problem in subsequent publications.

*Proof of the second part of Theorem* 2.2 Let $\beta = \beta(\Bbbk)$. We shall deduce Theorem 2.3 from the same stopping time $\tau$ defined in Lemma 4.1. Observe that $\tau$ can no longer be shown a strong uniform. The proof breaks when we move a piece and claim that if $i$-th row is uniform, then the next row is uniform. It is no longer true since when we add a row multiplied by a noninvertible element we obtain a uniform row. By definition, an element is noninvertible with probability $\beta$.

Still, consider the distribution of the stopping state $\varrho$ we obtain. Observe that at each addition as above we can "mess up" at most $1-\beta$ portion of the row. Thus after $\binom{n}{2}$ additions we obtain a distribution $Q^\tau$ which will be at least uniform on at least $\psi = (1-\beta)^{\binom{n}{2}}$ fraction of elements. Formally, we claim that there exist a subset $A \subset G$ such that $\mu(A) \geq \psi$, and $Q^\tau(B) \geq \mu(B)$ for all $B \subset A$. The claim is proved by induction.

Now, if $\beta < 1/2n^2$, we have

$$\psi > \left(1 - \frac{1}{2n^2}\right)^{n^2/2} > (1/e)^{1/4} > 3/4$$

Let $A \subset G$, $\mu(A) \geq 3/4$ be as above. Consider a distribution $Q^{2\tau}$. Observe that for *every* element $g \in G$ the measure of those $(a_1, a_2) \in A^2$ such that $a_1 \cdot a_2 = g$ is at least $1/2$. Thus we have

$$Q^{2\tau}(B) \geq \frac{1}{2}\mu(B)$$

for all $B \subset G$. The rest of the theorem follows from the *incomplete stopping principle* which in different for appeared several times in the literature (see [DF], section 2.5; [LW], section 2.2; [P1] section 3.5). We present below one version of the principle.

We will use here approach introduced in [P1]. First we need several definitions. Define

$$\zeta = 1 + \mathbf{s}(1) + \mathbf{s}(2) + \dots$$

Then, by Corollary 2.2.9 in [P1] we have $\mathbf{s}(2\zeta) \le 1/2$ and further

$$\mathbf{s}(2r\zeta) \le \frac{1}{2^r}\,,$$

where $r$ is an integer. Recall that $E(\tau) = O(n^{2.5})$. By Theorem 3.5.2 in [P1] we have

$$\zeta \le 2\,E(\tau) = O(n^{2.5}).$$

Therefore for $m = O(cn^{2.5})$ we have $\mathbf{s}(m) \le 2^{-c}$, which completes the proof. $\square$

**Remark 5.4** Note that when $1/\beta$ is small compared to $n^2$ the idea used in the last proof gives exponential upper bounds for the mixing time, which are weak compared to $O(n^3)$ bounds in [S1,S2].

**Acknowledgements**

I would like to thank my graduate advisor Persi Diaconis for suggesting the problem and for his continuing interest and encouragement. A weaker version of the results was a part of the author's Ph.D. thesis.

I am grateful to Alex Astashkevich, Jim Fill, László Lovász, Gregory Margulis, David Wilson, Peter Winkler and anonymous referee for helpful remarks.

Special thanks to Center for the Mathematical Sciences at University of Wisconsin-Madison for the hospitality during author's stay there in July of 1998, when the paper was written. The author was supported in part by the NSF Postdoctoral Research Fellowship.

## References

[A] D. Aldous, *Random walks on finite groups and rapidly mixing Markov chains*, Lect. Notes in Math. **986** (1983).

[AD1] D. Aldous, P. Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), 333–348.

[AD2] D. Aldous, P. Diaconis, *Strong uniform times and finite random walks*, Adv. Appl. Math. **8** (1987), 69–97.

[AF] D. Aldous, J. Fill, *Reversible Markov Chains and Random Walks on Graphs*, monograph in preparation, 1996.

[AP] A. Astashkevich, I. Pak, *Random walks on nilpotent and supersolvable groups*, preprint (1997).

[B] N. Bourbaki, *Éléments de mathématique: Intégration, Mesure de Haar; Topologie générale, Groupes topologiques*, Hermann, Paris, 1960, 1963.

[D1] P. Diaconis, *Group Representations in Probability and Statistics*, IMS, Hayward, California, 1988.

[D2] P. Diaconis, *The cutoff phenomenon in finite Markov chains*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), 1659–1664.

[D3] P. Diaconis, *An introduction to modern Markov chain theory*, Proc. ICM Berlin **1** (1998), 187–204.

[DF] P. Diaconis, J. A. Fill, *Strong stationary times via new form of duality*, Ann. Prob. **18** (1990), 1483–1522.

[DSC1] P. Diaconis, L. Saloff–Coste, *Comparison techniques for random walk on finite groups*, Ann. Prob. **21** (1993), 2131–2156.

[DSC2] P. Diaconis, L. Saloff–Coste, *Moderate growth and random walk on finite groups*, Geom. Funct. Anal. **4** (1994), 1–36.

[DSh] P. Diaconis, M. Shahshahani, *Generating a random permutation with random transpositions*, Z. Wahr. verw. Gebiete **57** (1981), 159–179.

[ER] P. Erdős, A. Rényi, *On classical problem of probability theory*, MTA Tat. Kut Int. Közl. **6A** (1961), 215–220.

[F] W. Feller, *An introduction to Probability theory and its applications* (third edition), John Wiley, New York, 1970.

[H] J. Humphreys, *Linear algebraic groups*, Springer, Berlin, 1975.

[K] R. Kannan, *Markov chains and polynomial time algorithms* (1994), 35-th FOCS, IEEE Comput. Soc. Press, Los Alamitos, CA, 656–671.

[LW] L. Lovász, P. Winkler, *Mixing Times* (1998), AMS DIMACS Series, vol. 41, 189–204.

[M] P. Matthews, *A strong uniform time for random transpositions*, J. Theor. Prob. **1** (1988), 411–423.

[P1] I. Pak, *Random walks on groups: strong uniform time approach*, Ph.D. Thesis, Harvard U., 1997.

[P2] I. Pak, *When and how n choose k* (1998), AMS DIMACS Series, vol. 43, 191–238.

[PV] I. Pak, V. H. Vu, *On finite geometric random walks*, Discrete Applied Math. (to appear).

[R] D. Revuz, *Markov chains,* in North Holland Mathematical Library, vol. 11, Elsevier, New York, 1975.

[SJ] A. Sinclair, M. Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Inform. and Comput. **82** (1989), 93–133.

[S1] R. Stong, *Random walk on the upper triangular matrices*, Ann. Prob. **23** (1995), 1939–1949.

[S2] R. Stong, *Eigenvalues of random walks on groups*, Ann. Prob. **23** (1995), 1961–1981.