RANDOM WALKS ON NILPOTENT GROUPS

Alexander Astashkevich

Renaissance Technologies
East Setauket, New York 11733
E-mail: ast@rentec.com

Igor Pak

Department of Mathematics
MIT, Cambridge, MA 02139
E-mail: pak@math.mit.edu

March 25, 2001

Abstract. We obtain sharp bounds on mixing time of random walks on nilpotent groups, with Hall bases as generating sets.

Introduction

In the past two decades, the study of random walks on finite groups developed into a large field with a number of powerful techniques, advanced results and interesting applications. The state of art, however, vary across different classes of groups. It is safe to say that by now we understand the behavior of random walks on abelian groups (see [D1,DF]). At the same time, despite several attempts, random walks on finite simple groups remain 'terra incognita' [Ba,D2]. While there is little hope of proving of making a breakthrough in full generality, the class of *nilpotent groups* seem the most promising to be completely understood in the foreseeable future (cf. [CP,DS2,S2]) In this paper we generalize results in [P3] to obtain sharp bounds for a class of random walks on general nilpotent groups.

From the algebraic point of view, nilpotent groups represent the next logical step in the group hierarchy, after abelian groups. In many respects, behavior of random walks on large finite nilpotent groups with a bounded number of generators, resemble that on abelian group, with mixing time being roughly the square of diameter (see [DS2]). The similarity is particularly striking for infinite nilpotent group (see e.g. [W]), although we won't explore this connection in the paper.

A completely different phenomena occur for all finite groups, when the number of generators is allowed to grow with the size of the group (say, as $\log^{\alpha} |G|$). In this case it is often possible to find rapidly mixing random walks (with mixing time, say,

Key words and phrases. Random walks, mixing time, nilpotent groups, compact groups.

$O(\log^{\beta} |G|)$). In fact, under certain assumptions, most of the sets of generators of size $O(\log |G|)$ mix in time $O(\log |G|)$ (see references and in [P2]). For nilpotent group, this represents a remarkable transition from slow to fast mixing, which begs for an explanation.

The difficulty in study of the rapidly mixing random walks is a relative lack of techniques which give sharp upper bounds. At the same time, a variety of general Markov chain techniques give upper bounds, which are away by a factor of $\log^{\alpha} |G|$, for some $\alpha \geq 1$ [AF]. Let us mention here a paper [S2], which gives sharp general bounds on the eigenvalue gap, often leading to the best known upper bounds of the mixing time. On the other hand, as was shown in several special cases, random walks on nilpotent groups can mix faster than predicted by these general bounds, and there seem to be no general technique for study these random walks [P1].

In this paper we propose a class of random walks on nilpotent group and prove rapid mixing for these. Our goal is a good understanding of random walks on particular generating sets, so that one can use a comparison technique to obtain bounds for other generating sets of nilpotent groups. We are trying to model the use of 'random transpositions' for study of random walks on S_n, championed in [DS1]. The generating sets we consider are the 'Hall bases', which extend the notion of the power commutator generating sets. We show that the mixing of the random walk on these can be reduced to the coupon collector's problem on the generators.

Our proof uses a stopping time technique, pioneered by Aldous and Diaconis in [AD1,AD2], and then employed by Diaconis and Fill [DF], Matthews [M], and others. This is, perhaps, the most rarely used approach, due to the technical difficulty of finding a desired strong uniform time (see [P1]). We are fully convinced, however, that no other approach (such as Fourier analysis, coupling, Poincare or isoperimetric inequalities, etc.) can give bounds as good as ours in this case. The construction we use extends the result of the second author for the group of upper triangular matrices [P3].

The paper is organized as follows. In section 1 we present definitions and general results for random walks on finite groups. Section 2 contains main results, Theorems 2.1 and 2.2, which prove rapid mixing of random walks on nilpotent groups, with two versions of the Hall bases as generating sets. In section 3 we illustrate the power of the theorems in several special cases. The key lemmas are given in section 4, while the proof of theorems is completed in section 5. In section 6 we present analogous results for compact nilpotent groups.

Few words about notation. Throughout the paper the random walk on G with a generating set S denotes $\mathcal{W}(G, S)$, the mixing time denotes by $\operatorname{mix} \mathcal{W}$ (see section 1. for definitions). Also, we denote $[n] = \{1, 2, \ldots, n\}$.

1. Random walks on groups

Let G be a finite group and let $S = S^{-1}$ be a symmetric generating set. A *random walk* $\mathcal{W} = \mathcal{W}(G, S)$ is defined as follows:

$$X_0 = \mathrm{id}, \quad X_{t+1} = X_t \cdot s,$$

where $s \in S$ is chosen uniformly and independently, for every $t > 0$. One can think of $\mathcal{W}(G,S)$ as of a nearest neighbor random walk on the Cayley graph $\Gamma = \Gamma(G,S)$. Let $\mathrm{Q}^t(g) = \mathbf{P}(X_t = g)$ denotes the probability distribution of the walk after t steps. Throughout the paper we assume that $\mathrm{id} \in S$, so Γ is not bipartite. Then, for all $g \in G$, we have:

$$\mathrm{Q}^t(g) \to \frac{1}{|G|}, \text{ as } t \to \infty.$$

Define a *separation distance* :

$$\mathrm{sep}(t) \,=\, |G| \max_{g \in G} \left(\frac{1}{|G|} - \mathrm{Q}^t(g) \right).$$

It is well known [AD2,D1] that $\mathrm{sep}(t)$ is monotone and submultiplicative:

$$\mathrm{sep}(t+1) \le \mathrm{sep}(t), \quad \mathrm{sep}(t+r) \le \mathrm{sep}(t) \cdot \mathrm{sep}(r), \quad \text{for all } t, r \ge 0.$$

Also, the separation distance is always bounded from below by the *total variation distance* :

$$\mathrm{sep}(t) \ge \|Q^t - \mathrm{U}\| \,=\, \frac{1}{2} \sum_{g \in G} \left| Q^t - \frac{1}{|G|} \right|.$$

Define the *mixing time* $\mathrm{mix} = \mathrm{mix}(G,S)$ as follows:

$$\mathrm{mix} = \min \left\{ t : \mathrm{sep}(t) \le \frac{1}{2} \right\} = \min \left\{ t : Q^t(g) \ge \frac{1}{2|G|}, \quad \text{for all } g \in G \right\}.$$

Let $A = (a_{g,h})$ be the adjacency matrix of the Cayley graph Γ, so that $P = A/|S|$ is a transition matrix of the random walk $\mathcal{W}$. Denote by $1 = \lambda_0 > \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{|G|-1} > -1$ the eigenvalues of P. Let $\beta = 1 - \lambda_1$ and $\beta' = \lambda_{|G|-1} + 1$ be the first and last eigenvalue gaps of P. It is well known and easy to see that :

$$\frac{1}{2 \min\{\beta, \beta'\}} \,\le\, \mathrm{mix} \,\le\, \frac{2 \log_2 |G|}{\min\{\beta, \beta'\}}$$

(see e.g. [AF]). Also, $\mathrm{sep}(t) \sim C\, \varrho^t$ as $t \to \infty$, where $\varrho = \min\{\beta, \beta'\}$.

Let $\tau : \{X_t\} \to \mathbb{Z}_+$ be a stopping time defined by some *stopping rule*: if $\tau(\{X_t\}) = k$, then $\tau(\{X'_t\}) = k$, given $X_i = X'_i$ for all $1 \le i \le k$. We say that τ is *strong uniform* if

$$\mathbf{P}(X_\tau = g \,|\, \tau = k) = \frac{1}{|G|}, \quad \text{for all } g \in G,\ k > 0, \text{ such that } \mathbf{P}(\tau = k) > 0.$$

If τ is strong uniform, we have:

$$(*) \quad \mathrm{sep}(t) \,\le\, \mathbf{P}(\tau > t), \text{ and } \mathrm{mix} \le 2\, E(\tau)$$

(see [AD2,AF,P1]).

Example 1.1 Let $G = \mathbb{Z}_2^m$ be an abelian group with a natural set of generators

$$S = \big\{(0, \ldots, 1_i, \ldots, 0), i \in [m]\big\}.$$

The Cayley graph $\Gamma(G, S)$ is bipartite so we shall consider $\widehat{S} = S \cup \{\mathrm{id}^m\}$. One can think of the random walk $\mathcal{W}(\mathbb{Z}_2^m, \widehat{S})$ as follows: at every $t \in \mathbb{Z}_+$ pick a random direction $i \in [m]$ and flip a fair coin. If heads, move in i-th direction. If tails, stay put.

Let us define a stopping time τ to be the first time that all directions are chosen at least once (the walker may have or have not moved in that direction). Define $\Phi_m(t) = \mathbf{P}(\phi > t)$ be given by *the coupon collectors problem* with m coupons: if at every $t = 1, 2, \ldots$ one coupon is chosen at random, what is the time ϕ when all coupons are collected? In this language, $(*)$ implies that

$$\mathrm{sep}(t) \le \mathbf{P}(\tau > t) = \Phi_m(t), \quad \text{and} \quad \mathrm{mix} \le E(\tau) = E(\phi).$$

One can show that the first inequality above becomes an equality, due to $\widehat{g} = (1, \ldots, 1) \in \mathbb{Z}_2^m$ being a halting element of τ (see [DF,P1], and subsection 5.2.)

From above, $\mathrm{mix} \le 2\,n\,\log n + O(n)$. In fact, $\mathrm{mix} \le 2\,m\,\mathcal{H}_m$, where $\mathcal{H}_m$ is a *harmonic number* $\mathcal{H}_m = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{m}$ [F]. Further, for the separation distance one obtain:

$$\mathrm{sep}\big(n\,\log n + c\,n\big) \,=\, \Phi_m(n\,\log n + c\,n) \,\to\, 1 - e^{-e^{-c}} \quad \text{as} \quad n \to \infty,$$

for all $c \in \mathbb{R}$ [ER,P1].

Example 1.2 Let $G = \mathbb{Z}_n$, $S = \{0, \pm 1\}$. It is well known and easy to see that in this case $\mathrm{mix}(G, S) \sim C\,n^2$, for some $C < 1$. Moreover, $\mathrm{sep}(c\,n^2) \le 2^{-c}$, for any $c > 1$ (see e.g. [D1]). Among many ways to prove this, one can use the stopping time technique, as shown in [DF].

Suppose, for simplicity, that $n = 2^m$. Define τ as follows. Walk until group element $a_1 = \pm n/4$ is hit. Then walk until $a_2 = a_1 \pm n/8$ is hit. Then walk until $a_2 \pm n/16$ is hit, etc. Finally walk until $a_{m-1} = a_{m-2} \pm 1$ is hit. Then walk until either 0 or 1 steps are used. Let this be a_m. Stop.

Observe by induction and symmetry, that a_{k-1} is uniformly distributed among $n/2 \pm n/4 \pm \cdots \pm n/2^k$. Thus a_{m-1} is uniform among odd integers $\mod n$, and a_m is uniform in $\mathbb{Z}_n$. A similar construction exists for general n (see [DF,P1] for details). Thus $\mathrm{mix} = E(\tau) = O(n^2)$ is bounded by the sum of the hitting times of the random walk on a line [DF].

2. Main results

Recall that a group G is *nilpotent* if there exist a subgroup chain

$$(\diamond) \quad G = G_1 \supset G_2 \supset \ldots \supset G_\ell \supset G_{\ell+1} = 1,$$

such that $[G, G_i] \subset G_{i+1}$, G_{i+1} is normal in G_i, and $H_i = G_i/G_{i+1}$ are abelian, for all $i \in [\ell]$ (see e.g. [L]). Denote by $\gamma_i : G_i \to H_i$ the natural projections of G_i onto H_i.

We say that $\{h_1, \ldots, h_r\}$ is a *basis* in an abelian group H, if $\langle h_1, \ldots, h_r \rangle = H$. Fix an integer sequence $\mathbf{r} = (r_1, \ldots, r_\ell)$. Denote by $\Lambda_{\mathbf{r}}$ the set of pairs $\{(i,j) : i \in [\ell], j \in [r_i]\}$, and let $\mathrm{N}_{\mathbf{r}} = |\Lambda_{\mathbf{r}}| = r_1 + \ldots + r_\ell$. We say that a set $S = \{s_{i,j}, (i,j) \in \Lambda_{\mathbf{r}}\}$ is a *Hall basis* in a nilpotent group G, if $s_{i,j} \in G_i$ for all $(i,j) \in \Lambda_{\mathbf{r}}$, and $\{\gamma_i(s_{i,1}), \ldots, \gamma_i(s_{i,r_i})\}$ is a basis in H_i. We shall use notation $h_{i,j}$ to denote projections $\gamma_i(s_{i,j}) \in H_i$. By $d_{i,j}$ denote the order of the element $h_{i,j}$. Let us define two different bases

$$S_\circ = \big\{s_{i,j}^\alpha,\ \alpha \in \{-1,0,1\},\ (i,j) \in \Lambda_{\mathbf{r}}\big\}, \quad \text{and} \quad \widehat{S} = \big\{s_{i,j}^\alpha,\ \alpha \in [d_{i,j}],\ (i,j) \in \Lambda_{\mathbf{r}}\big\}.$$

We call these the *symmetric* and *extended* Hall bases, respectively. The main result of this paper is analysis of the random walk $\mathcal{W}(G,S)$, where S is a symmetric or and extended Hall basis.

Theorem 2.1 *Let G be a nilpotent group with a subgroup chain $(\diamond)$, let $\mathbf{r} = (r_1, \ldots, r_\ell)$ be an integer sequence, $\mathrm{N} = \mathrm{N}_{\mathbf{r}} = r_1 + \cdots + r_\ell$. Let $\widehat{S}$ be an extended Hall basis of G, defined as above. Consider a random walk $\mathcal{W} = \mathcal{W}(G, \widehat{S})$. Then, for the separation distance $\mathrm{sep}(t)$ and the mixing time $\mathrm{mix} = \mathrm{mix}(G, \widehat{S})$ of the walk $\mathcal{W}$, we have:*

$$\mathrm{sep}(t) \le \Phi_{\mathrm{N}}(t), \quad \textit{and} \quad \mathrm{mix} \le 2\,\mathrm{N}\,\mathcal{H}_{\mathrm{N}},$$

where Φ_{N} is a probability tail in the coupon collector's problem with N coupons.

Note here, that the random walk in Theorem 2.1, in a special case $G = \mathbb{Z}_2^m$, $\ell = 1$ and $r_1 = m$, is the same random walk in the Example 1.1. The proof of Theorem 2.1 will be given in section 5, and is based on an advance generalization of the strong uniform time in that example.

With every *multiplicity sequence* $\mathbf{k} = (k_1, \ldots, k_m)$ we associate a *generalized coupon collector's problem*, defined as follows. Suppose there are m types of different coupons, and at every time $t \in \{1, 2, \ldots\}$, a type $i \in [m]$ of a coupon is chosen at random, uniformly and independently. If the number of coupons of this type is k_i, no change is made. Otherwise, a fair coin is flipped. If heads, a coupon of type i is added to a collection. If tails, a coupon of type i is removed from a collection (if there are any). The process is stopped when there are exactly k_i copies of coupons of i-th type, for all $i \in [m]$.

Denote by ϕ the stopping time of the process defined above. Consider also $\Phi_{\mathbf{k}}(t) = \mathbf{P}(\phi > t)$, the probability tail of ϕ.

Theorem 2.2 *Let G be a nilpotent group with a subgroup chain $(\diamond)$, and let $\mathbf{r} = (r_1, \ldots, r_\ell)$ be an integer sequence. Let $S = \{s_{i,j},\ (i,j) \in \Lambda_{\mathbf{r}}\}$ be a Hall basis defined as above, and let $d_{i,j}$ be the orders of projections. Finally, let $S_\circ = \big\{s_{i,j}^\alpha, \alpha \in \{0, \pm 1\}\big\}$ be the symmetric Hall basis of G. Consider a random walk $\mathcal{W} = \mathcal{W}(G, \widehat{S})$. Then, for the separation distance $\mathrm{sep}(t)$ and the mixing time $\mathrm{mix} = \mathrm{mix}(G, S_\circ)$ of the walk $\mathcal{W}$, we have:*

$$\mathrm{sep}(t) \le \Phi_{\mathbf{k}}(t), \quad \textit{and} \quad \mathrm{mix} \le 2\,E(\phi),$$

where $\mathbf{k} = (\dots, \lfloor d_{i,j}/2 \rfloor, \dots)$*, is a multiplicity sequence in the generalized coupon collector's problem with* N *coupons corresponding to pairs* $(i,j) \in \Lambda_{\mathbf{r}}$*.*

Corollary 2.3 *In conditions of Theorem 2.2, for the mixing time* $\mathrm{mix} = \mathrm{mix}(G, S_\circ)$*, we have* $\mathrm{mix} = O\big(D^2 \,\mathrm{N}\,(\log \mathrm{N})^2\big)$*, where* $D = \max\{d_{i,j},\, (i,j) \in \Lambda_{\mathbf{r}}\}$*.*

Much is known about $\Phi_{\mathbf{k}}$, although, perhaps, most general bounds as in Corollary 2.3 are much weaker than that in special cases. We will elaborate on the these in the next section, where we consider several examples of interest.

The proof of Theorem 2.2 is based on a modification of the strong uniform time construction we present in the proof of Theorem 2.1. A weaker version of Theorem 2.2 can be obtained directly from Theorem 2.1 by using comparison [DS1]. In the examples, we will show that Theorem 2.2 can give much sharper bounds than comparison bound in such cases. It is interesting to see that one can obtain sharp lower bounds on the eigenvalue gap as well:

Corollary 2.4 *In conditions of Theorem 2.1, for the eigenvalue gap* $\beta = \beta(G, \widehat{S})$*, we have* $\beta = \Omega\left(\frac{1}{\mathrm{N}}\right)$*.*

We prove corollaries 2.3 and 2.4 in section 7.

3. Examples and special cases

3.1 Cyclic groups.

Consider the case $G = \mathbb{Z}_n$, $\mathrm{N} = \ell = r_1 = 1$, and $S = \{0, \pm 1\}$. The generalized coupon collectors problem in this case is a problem of collecting $k_1 = n$ coupons. The expected time to collect k_1 coupons is a hitting time of k_1 of the random with a reflected boundary at 0. Thus $E(\phi) = O(n^2)$ [F]. This agrees with the mixing time of the walk $\mathrm{mix} = O(n^2)$ [D1]. Accordingly, our construction in the proof of Theorem 2.2 uses that in Example 1.2 as a building block.

3.2 Abelian groups.

Let $G = H_1$ be an abelian group, $\ell = 1$, let $S = \{s_1, \dots, s_m\}$ be a generating set, and let $d_i = \mathrm{order}(s_i)$. We claim that the product

$$a = (s_1)^{\alpha_1} \cdot \ldots \cdot (s_m)^{\alpha_m}$$

is uniform in G, given α_i is uniform in $[d_i]$. Indeed, this is trivial for a cyclic group, and for a direct product of cyclic groups can be obtained by induction. Note that we do not require the generating set to be minimal or nonredundant.

Now, consider a random walk $\mathcal{W} = \mathcal{W}(G, \widehat{S})$ in this case. A direct generalization of the stopping time τ in Example 1.1, gives a construction of the stopping time, which is strong uniform from the argument above. Thus, for the separation distance of $\mathcal{W}$, we have $\mathrm{sep}(t) \le \mathbf{P}(\tau > t) = \Phi_m(t)$. This agrees with the bound given by Theorem 1.2 in this case.

3.3 p-groups.

Suppose G is a p-group and $H_i \simeq \mathbb{Z}_p^{r_i}$. Let $\mathbf{r} = (r_1, \ldots, r_\ell)$, $\mathrm{N} = r_1 + \ldots + r_\ell$, and let $\Lambda_{\mathbf{r}}$ be as in section 2. Consider a symmetric Hall basis

$$S_\circ = \big\{\mathrm{id}^{\mathrm{N}},\, s_{i,j},\, s_{i,j}^{-1} \,:\, (i,j) \in \Lambda_{\mathbf{r}}\big\}.$$

In this case $d_{i,j} = p$ for all $(i,j) \in \Lambda_{\mathbf{r}}$. Suppose now p is fixed and $\mathrm{N} \to \infty$. It was shown in [NS,ER] that the time ψ when at least L coupons of each type is chosen satisfies

$$\mathbf{P}\big(\psi \,\le\, \mathrm{N}\log\mathrm{N} \,+\, (L-1)\,\mathrm{N}\log\log\mathrm{N} \,+\, c\,\mathrm{N}\big) \,\to\, e^{-e^{-c}},$$

for every $c \in \mathbb{R}$. Taking $L = O(p^2)$ and using [ER] one can obtain $\mathrm{mix} < C\,p^2\,\mathrm{N}\log\mathrm{N}\,\log\log\mathrm{N}$, where C is a universal constant. This is slightly better than $\mathrm{mix} = O(p^2\mathrm{N}\log^2\mathrm{N})$ bound that follows immediately when one combines bounds in Example 1.1 with those in coupon collector's problem.

3.4 Upper triangular matrices.

Let $G = \mathrm{U}(n, \mathbb{F}_p)$ be a group of upper triangular matrices over the field with p elements, with ones on diagonal. Let q be a prime. Denote by $E_{i,j}(a) \in G$ a matrix with ones on diagonal, $a \in \mathbb{F}_p$ at (i,j), and zeroes elsewhere. Consider a (redundant) generating set $S = \{E_{i,j}(a),\, 1 \le i < j \le n,\, a \in \mathbb{F}_p\}$. A random walk $\mathcal{W} = \mathcal{W}(G,S)$ was considered in [P1,P3], where

$$\frac{1}{2}\,n^2\,\log n + O(n^2) \,\le\, \mathrm{mix}\big(\mathrm{U}(n,\mathbb{F}_q), S\big) \,\le\, 2\,n^2\,\log n + O(n^2)$$

bound was shown. Note that the upper bound easily follows from Theorem 2.1. Indeed, since $E_{i,j}(a) = \big(E_{i,j}(1)\big)^a$, we can take $\mathrm{N} = \binom{n}{2}$ in the Theorem. We conclude:

$$\mathrm{mix} \le 2\binom{n}{2}\mathcal{H}_{\binom{n}{2}} = 2\,n^2\log n + O(n).$$

When one takes a smaller generating set $S_\circ = \big\{E_{i,j}(\pm 1), \mathrm{id}^{\binom{n}{2}},\, 1 \le i < j \le n\big\}$, one gets a random walk on a p-group, as above, with $\mathrm{N} = \binom{n}{2}$.

For another set of generators $S_\star = \big\{E_{i,i+1}(a),\, 1 \le i < n,\, a \in \mathbb{F}_p\big\}$, when $p = \Omega(n^2)$, it was shown in [CP] that

$$\mathrm{mix}\big(\mathrm{U}(n,\mathbb{F}_p), S_\star\big) = O(n^2).$$

Note the obvious lower bound $\mathrm{mix} \ge \binom{n}{2}$, so the upper bound above is sharp. Such a rapid mixing is rather surprising since $S_\star$ does not contain a Hall basis. Interestingly, for bounded p, the best known bound is $\mathrm{mix} = O(n^3)$, given in [S1].

4. Random elements in nilpotent groups

Let G be a nilpotent group with a subgroup chain $(\diamond)$ of length ℓ; let $\mathbf{r} = (r_1, \ldots, r_\ell)$ be an integer sequence; let $\Lambda_{\mathbf{r}}$ and $\mathrm{N} = \mathrm{N}_{\mathbf{r}}$ be as in section 2. Consider a Hall basis $S = \{s_{i,j}\}$ with projections $h_{i,j}$ and their orders $d_{i,j} = \mathrm{order}(h_{i,j})$.

Lemma 4.1 *For all $i \in [\ell]$, consider a random element*

$$a_i = \big(s_{i,1}\big)^{\alpha_{i,1}} \cdot \ldots \cdot \big(s_{i,r_i}\big)^{\alpha_{i,r_i}},$$

where $\alpha_{i,j}$ are independent and uniform in $[d_{i,j}]$. Then a product of these elements $a = a_1 \cdots a_\ell$ is uniform in G.

The lemma is well known, although rarely stated in these form. Below present a short proof for completeness. But first let state a generalization of the lemma.

Lemma 4.2 *Let $\sigma : \Lambda_{\mathbf{r}} \to \Lambda_{\mathbf{r}}$ be a bijection. Consider a product*

$$(\circledast) \quad a = \big(s_{\sigma(1,1)}\big)^{\alpha_{\sigma(1,1)}} \cdot \ldots \cdot \big(s_{\sigma(\ell,r_\ell)}\big)^{\alpha_{\sigma(\ell,r_\ell)}},$$

where $\alpha_{i,j}$ are independent and uniform in $[d_{i,j}]$, for all $(i,j) \in \Lambda_{\mathbf{r}}$. Then the product a in $(\circledast)$ is uniform in G.

Note that Lemma 4.2 coincides with Lemma 4.1, when σ is the identity permutation.

Proof of Lemma 4.1. Use induction on ℓ. For $\ell = 1$ the group is abelian and the result is clear (cf. Example 3.2.) Now let us prove the inductive step. By inductive assumption, $a' = a_2 \cdots a_\ell \in G_2$ is uniform in G_2. Also,

$$\gamma_1(a_1) = \big(\gamma_1(s_{1,1})\big)^{\alpha_{1,1}} \cdot \ldots \cdot \big(\gamma_1(s_{1,r_1})\big)^{\alpha_{1,r_1}} = \big(h_{1,1}\big)^{\alpha_{1,1}} \cdot \ldots \cdot \big(h_{1,r_1}\big)^{\alpha_{1,r_1}} \in H_1$$

is uniform in H_1, since $H_1 = \langle h_{1,1}, \ldots, h_{1,r_1} \rangle$ is abelian. Since $\gamma_1(a) = \gamma_1(a_1)$ and a' is uniform in G_2, we conclude that $a = a_1 a'$ is uniform in $G_1 = G$. $\square$

Proof of Lemma 4.2. Recall that for all $h \in G_i$ and $h' \in G_j$, we have:

$$(\boxtimes) \quad h \cdot h' = h' \cdot h \cdot b(h,h'), \quad \text{where } b(h,h') = [h,h'] \in G_{\max\{i,j\}+1}.$$

Now let us rewrite $(\circledast)$ in as a product of the same type, for $\sigma = 1$, with additional terms b_*, obtained as commutators of the elements. Formally, pull the term $\big(s_{1,1}\big)^{\alpha_{1,1}}$ in $(\circledast)$ to the left, by repeatedly using commutations $(\boxtimes)$. We obtain:

$$(\circledast') \quad a = \big(s_{1,1}\big)^{\alpha_{1,1}} \cdot \big(s_{\sigma(1,1)}\big)^{\alpha_{\sigma(1,1)}} b^{(1)} \cdot \ldots \cdot \big(s_{\sigma(\ell,r_\ell)}\big)^{\alpha_{\sigma(\ell,r_\ell)}} b^{(\mathrm{N})},$$

where the term $\big(s_{1,1}\big)^{\alpha_{1,1}}$ is omitted in the product and every term preceded $\big(s_{1,1}\big)^{\alpha_{1,1}}$ in $(\circledast)$, is now followed by some $b^{(*)}$. Now pull to the left the term

$\big(s_{1,2}\big)^{\alpha_{1,2}}$ to follow $\big(s_{1,1}\big)^{\alpha_{1,1}}$, etc. The last term to be pulled is $\big(s_{r,\ell}\big)^{\alpha_{\ell,r_\ell}}$. We obtain a product of the following type:

$$(\circledcirc) \quad a = a_1\, a_2\, b_3\, a_3\, b_4\, \cdot \ldots \cdot\, b_\ell\, a_\ell\, b_{\ell+1},$$

where a_i is as in Lemma 4.1, and for the products of commutators b_i of terms $\big(s_{i-1,j}\big)^{\alpha_{i-1,j}}$ and $\big(s_{i',j'}\big)^{\alpha_{i',j'}}$, $i' \le i-1$, we have $b_i \in G_i$.

To prove uniformity of a in $(\circledcirc)$, go backwards. Observe that $b_{\ell+1} = 1$ so $a_\ell\, b_{\ell+1}$ is uniform in G_ℓ. Since $a_{\ell-1}$ and $b_\ell \in G_\ell$ are independent of $a_\ell\, b_{\ell+1}$, we have $b_\ell\, a_\ell\, b_{\ell+1}$ is uniform in G_ℓ, and $a_{\ell-1} b_\ell\, a_\ell\, b_{\ell+1}$ is uniform in $G_{\ell-1}$. Similarly, since $a_{\ell-2}$ and $b_{\ell-1} \in G_\ell$ are independent of $a_{\ell-1}\, b_\ell\, a_\ell\, b_{\ell+1}$, we obtain $a_{\ell-2}\, b_{\ell-1}\, a_{\ell-1}\, b_\ell\, a_\ell\, b_{\ell+1}$ is uniform in $G_{\ell-2}$, etc. Proceeding in this manner we obtain that $a = a_1\, a_2\, b_3\, a_3\, \ldots$ in $(\circledcirc)$ is uniform in $G_1 = G$. $\square$

5. Proof of Theorems 2.1, 2.2

5.1 Proof of Theorem 2.1.

The proof is based on an explicit construction of a strong uniform time with tail probabilities $\Phi_{\mathrm{N}}(t)$. Namely, for a random walk $\mathcal{W} = \mathcal{W}(G, \widehat{S})$ consider the following stopping rule: *walk until for every* $(i,j) \in \Lambda_{\mathbf{r}}$*, a generator* $s_{i,j}^{\alpha}$ *is used, for some* $\alpha \in [d_{i,j}]$*. Stop.*

We claim that the stopping time τ defined by the rule above is strong uniform. This would prove the theorem. Indeed, consider the usual coupon collector's problem with $\mathrm{N} = |\Lambda_{\mathbf{r}}|$ coupons corresponding to pairs $(i,j) \in \Lambda_{\mathbf{r}}$. By construction of τ and from the property $(*)$ of strong uniform time (see section 1), we immediately have:

$$\mathrm{sep}(t) \le \mathbf{P}(\tau > t) = \Phi_N(t), \quad \text{and} \quad \mathrm{mix} \le 2\,\mathrm{N}\,\mathcal{H}_{\mathrm{N}},$$

(cf. Example 1.1), which is exactly what we needed to prove.

Now, we will show that strong uniformity follows from Lemma 3.2. Indeed, fix a sequence of pairs

$$\triangle : \quad (i_1,j_1),\, (i_2,j_2),\, \ldots\,,\, (i_k,j_k),$$

where $\tau = k$. By definition of τ, this sequence $\triangle$ contains every pair $(i,j) \in \Lambda_{\mathbf{r}}$. Therefore a set of elements $S' = \big\{s_{i_1,j_1}, \ldots, s_{i_k,j_k}\big\}$ is a Hall basis (perhaps, very redundant). Thus $\triangle$ is a permutation of the corresponding set of pairs $\Lambda_{\mathbf{r}'}$. We have:

$$\mathbf{P}\big(X_\tau = g \,|\, \tau = k,\, \triangle\big) \;=\; \mathbf{P}(a = g),$$

where

$$(\boxplus) \quad a \;=\; s_{i_1,j_1}^{\alpha_1} \cdot\, s_{i_2,j_2}^{\alpha_2} \,\cdot \ldots \cdot\, s_{i_k,j_k}^{\alpha_k}\,,$$

and $\alpha_m \in \big[d_{i_m,j_m}\big]$, $1 \le m \le k$. By Lemma 3.2, the product a in $(\boxplus)$ is uniform in G. This implies that τ is strong uniform, which completes the proof. $\square$

5.2 Proof of Theorem 2.2.

We define a construction of the strong uniform time in this case, based on an inexplicit construction of strong uniform time for a random walk on a cycle (see Example 1.2, [AD2,DF,P1]).

First, assume we have any construction of the strong uniform time for each of the cyclic groups $H_{i,j} = \langle h_{i,j} \rangle \simeq \mathbb{Z}_{d_{i,j}}$, where $(i,j) \in \Lambda_{\mathbf{r}}$. Denote this stopping time by $\tau_{i,j}$.

Now, while running a random walk on G by $S_\circ$, let us run N parallel random walks on $H_{i,j}$ by applying generators $h_{i,j}^{\alpha}$ whenever $s_{i,j}^{\alpha}$ is used by a random walk on G, $\alpha \in \{0, \pm 1\}$. Let us define $\tau = \max\{\tau_{i,j},\, (i,j) \in \Lambda_{\mathbf{r}}\}$ to be the time when all parallel walks on $H_{i,j}$ are stopped. Conditioned on the sequence $\triangle$ of the walk (but not on α again), proceed with a rewriting procedure as in the proof of Lemma 4.2. We obtain a random element a of the form $(\circledcirc)$. By construction of τ, the term a_i here is a product $\prod_j \left(s_{i,j}\right)^{\nu_{i,j}}$, where $\nu_{i,j}$ is uniform in $[d_{i,j}]$ by definition of $\tau_{i,j}$. Repeating the proof of Lemma 4.2 verbatim, we conclude that a is uniform in G. Now proceed as in the proof of Theorem 2.1 to obtain that τ is strong uniform. From here, the separation distance of the random walk $\mathcal{W}$ satisfies $\operatorname{sep}(t) \le \mathbf{P}(\tau > t)$.

Let us return now to the generalized coupon collector's problem with multiplicity sequence $\mathbf{k} = \left(d_{1,1},\, \ldots\, ,\, d_{\ell,r_\ell}\right)$. We claim that

$$\mathbf{P}(\tau > t) \le \Phi_{\mathbf{k}}(t) = \mathbf{P}(\phi > t),$$

which suffices to prove the theorem. By definition, ϕ is the maximum of the stopping times $\phi_{i,j}$ for every type of a coupon $(i,j) \in \Lambda_{\mathbf{r}}$. Thus it remains to prove that

$$\mathbf{P}(\tau_{i,j} > t) \le \mathbf{P}(\phi_{i,j} > t), \quad \text{for all } \ t \ge 0.$$

At this point we need to recall extra properties of the strong uniform times [AD2,P1]. We say that $\widehat{g} \in G$ is *halting*, if $\tau = k$ whenever $X_k = \widehat{g}$ and $k \le \tau$. We say that $\widetilde{g}$ is *minimal*, if $Q^t(\widetilde{g}) \le Q^t(g)$ for all $g \in G$.

It is known that for every $\mathcal{W}(G,S)$ there exists a strong uniform time τ such that the separation distance of $\mathcal{W}$ satisfies $\operatorname{sep}(t) = \mathbf{P}(\tau > t)$, for all $t > 0$ (see [AD2]). Moreover, if $\mathcal{W}$ has a minimal element $\widehat{g}$, then this is a halting element for the stopping time τ [P1].

Recall that $\lfloor m/2 \rfloor$ is a minimal element of $\mathcal{W}(\mathbb{Z}_m, \{0, \pm 1\})$ considered in Example 1.1. We conclude that for every $m > 1$, there exists a strong uniform time τ on $\mathbb{Z}_m$ such that $\pm\lfloor m/2 \rfloor$ are halting element. Therefore, for a hitting time η of the elements $\pm\lfloor m/2 \rfloor$, we have $\mathbf{P}(\tau > t) \le \mathbf{P}(\eta > t)$, for all $t \ge 0$. Finally, by the reflection principle, we have $\mathbf{P}(\eta = t) = \mathbf{P}(\phi_{i,j} = t)$, given $m = d_{i,j}$ [F]. We conclude:

$$\mathbf{P}(\tau_{i,j} > t) \,\le\, \mathbf{P}(\eta > t) \,=\, \mathbf{P}(\phi_{i,j} > t), \quad \text{for all } \ t \ge 0, \ (i,j) \in \Lambda_{\mathbf{r}}.$$

This implies the result. $\square$

6. Compact nilpotent groups

In this section we show that Theorem 2.1 has a straightforward analogue for compact nilpotent groups.

Let G be a compact connected real Lie group, with a set of generators S. Denote by μ the invariant measure on G, also known as Haar measure [Bu,H]. Recall that μ is unique, given that $\mu(G) = 1$. Since in our examples we have $\mu(S) = 0$, we consider a different probability measure ϑ on S, $\vartheta(S) = 1$. We need this measure to define a random walk (roughly: we sample element $s \in S$ according to μ 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 $\iota_m : (g_1, \ldots, g_m) \to g_1 \cdot \ldots \cdot g_m$, where $g_i \in S$ are sampled independently from ϑ. By abuse of speech, we refer to probability measure Q^m as a probability distribution of the random walk $\mathcal{W}(G, S)$ after m steps.

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

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

Roughly, if the separation distance $\operatorname{sep}(m)$ is equal to λ, then $Q^m = (1-\lambda)\cdot\mu + \lambda\cdot\eta$, where η is some positive measure, which can be interpreted as "noise". As before, we can define $\operatorname{mix} = \min\{t : \operatorname{sep}(t) \le 1/2\}$.

Let G be a compact nilpotent Lie group, with a subgroup chain $(\diamond)$ of length ℓ. Define $\mathbf{r} = (r_1, \ldots, r_\ell)$ and $\Lambda_{\mathbf{r}} = \{(i,j) : i \in [\ell],\ j \in [r_i]\}$, $\mathrm{N} = \mathrm{N}_{\mathbf{r}} = |\Lambda_{\mathbf{r}}|$ to be as before. Let $H_i = G_i/G_{i+1}$ be abelian quotients and let $S = \{s_{i,j},\ (i,j) \in \Lambda_{\mathbf{r}}\}$ be the generating set of G, such that $s_{i,j} \in G_i$. Again, let $h_{i,j} = \gamma_i(s_{i,j}) \in H_i$ to be a projection of $s_{i,j}$ onto H_i. We say that S is a Hall basis if $\langle h_{i,1}, \ldots, h_{i,r_i}\rangle = H_i$ for all $i \in [\ell]$.

Let $H = \langle h_1, \ldots, h_r\rangle$ be a compact abelian groups. Then $H_i \simeq \mathbf{S}^1 \times \cdots \times \mathbf{S}^1$ (m times), for some integer m, and where $\mathbf{S}^1 \simeq \mathbb{R}/\mathbb{Z}$. Therefore, there exists real numbers $d_1, \ldots, d_r > 0$, which we call *periods*, such that

$$(\boxdot) \quad h = h_1^{\alpha_1} \cdot \ldots \cdot h_r^{\alpha_r} \ \text{ is uniform in } \ H,$$

given α_i is uniform in $[0, d_i]$ for all $1 \le i \le r$. By abuse of speech, we refer to $d_{i,j}$ as periods of generators $s_{i,j}$.

Now, let $d_{i,j} > 0$ be periods of $s_{i,j}$. Let $\vartheta_{i,j} = \mathrm{U}\big[0, d_{i,j}\big]$ be a uniform measure of the interval, and let

$$\vartheta = \bigcup \frac{1}{\mathrm{N}_{\mathbf{r}}}\ \vartheta_{i,j}$$

be a measure on an extended Hall basis $\widehat{S} = \{s_{i,j},\ (i,j) \in \Lambda_{\mathbf{r}}\}$. One can think of the random walk $\mathcal{W} = \mathcal{W}(G, \widehat{S}, \vartheta)$ as follows: choose a random pair (i,j) uniformly and independently from $\Lambda_{\mathbf{r}}$; choose a random point $\alpha \in [0,1]$; let

$$X_{t+1} = X_t \cdot \big(s_{i,j}\big)^{\alpha\, d_{i,j}}.$$

Theorem 6.1 *Let G be a compact connected real nilpotent group Lie group with a subgroup chain $(\diamond)$ of length ℓ. Let $\mathbf{k} = (k_1, \ldots, k_\ell)$ be an integer sequence, let $\mathrm{N} = |\Lambda_{\mathbf{r}}|$, and let $S = \{s_{i,j},\, (i,j) \in \Lambda_{\mathbf{r}}\}$ be a Hall basis. Suppose $d_{i,j}$ are the periods of elements in S. Consider a random walk $\mathcal{W} = \mathcal{W}(G, \widehat{S}, \vartheta)$ defined as above. Then*

$$\mathrm{sep}(t) \le \Phi_{\mathrm{N}}(t), \quad \textit{and} \quad \mathrm{mix} \le 2\,\mathrm{N}\,\mathcal{H}_{\mathrm{N}}.$$

The theorem is a direct analogue of Theorem 2.1 for finite nilpotent groups. Much of the proof goes verbatim, although one has to be careful and develop the theory of strong uniform times for compact groups. This will be done in a sequel paper [P4]. Without giving all the details, let us just note here that all the properties of separation distance, its relation to strong uniform times, etc., can be translated from finite groups to compact Lie groups employing almost identical proofs. Here we give a sketch of the proof of Theorem 6.1.

Example 6.2 Let $\Bbbk$ be an infinite compact ring over $\mathbb{R}$ with identity, and let $G = \mathrm{U}(n, \Bbbk)$ be a group of upper triangular matrices with elements in $\Bbbk$, and with id on the diagonal. Let $E_{i,j}(a) \in \mathrm{U}(n, \Bbbk)$ be a matrix with id on the diagonal, with $a \in \Bbbk$ at (i,j), and zeroes elsewhere. Consider a natural uniform measure ϑ on $\widehat{S} = \{E_{i,j}(a), 1 \le i < j \le n\} = \{(i,j), 1 \le i < j \le n\} \times \Bbbk$. Let $\mathcal{W} = \mathcal{W}(G, \widehat{S}, \vartheta)$ be be the corresponding random walk on G.

This random walk was first considered in [P3], where the author proved that

$$\mathrm{sep}(t) \,\le\, \Phi_{\binom{n}{2}}(t), \quad \text{and} \quad \mathrm{mix} \,\le\, 2\binom{n}{2}\mathcal{H}_{\binom{n}{2}} = 2\,n^2\,\log n \,+\, O(n^2).$$

This result is a direct analogue of that in section 3.4. It follows immediately from Theorem 6.1, for $\mathbf{r} = (n-1, n-2, \ldots, 2, 1)$, $\mathrm{N} = \binom{n}{2}$, and $d_{i,j} = \mathrm{const}$. We refer to [P3] for the details.

Sketch of proof of Theorem 6.1. First start with analogues of Lemmas 4.1 and 4.2. When G is abelian, the analogue of Lemma 4.1 is now a definition $(\boxdot)$ of the periods. For nonabelian G, the proof is elementary but requires understanding of Haar measure μ on G [Bu]. One can show in this case that $\mu = \mu_1 \times \ldots \mu_\ell$, where μ_i is a uniform measure on H_i.

The analogue of Lemma 4.2 is exactly the same. For the proof, we again rewrite $(\circledast)$ by using $(\boxtimes)$ a number of times until we arrive to $(\circledcirc)$. The product a in $(\circledcirc)$ is uniform by the analogue of Lemma 4.1.

Now, define a stopping time for $\mathcal{W}$ as in the proof of Theorem 2.1 (see section 5.1), by stopping when the generators of the type $(s_{i,j})^{\alpha}$ are used, for all $(i,j) \in \Lambda_{\mathbf{r}}$. Then use the analogue of Lemma 4.2 to obtain measure μ on G of the product a in $(\boxplus)$, even conditioned on the sequence $\triangle$. We omit the details. $\square$

7. Proof of Corollaries

7.1 Proof of Corollary 2.2.

The probability that after t steps a simple random walk on $\mathbb{Z}$ is within distance d is $e^{-t/c\,d^2}$, where c is a universal constant [F]. For N independent random walks, the probability P that after $Cd^2 \log \mathrm{N}$ steps at least one of the walks is within d satisfies:

$$\mathrm{P} \leq \mathrm{N} \cdot e^{-C^2 \log \mathrm{N}/c} < \varepsilon,$$

for some $C = C(\varepsilon)$.

Now, by coupon collector's problem, it takes on average at most

$$O(D^2 \log N) \cdot (\mathrm{N} \log \mathrm{N} + O(\mathrm{N}))$$

steps to collect $d_{i,j} \leq D$ copies of every coupon (i, j). This implies the result. $\square$

7.2 Proof of Corollary 2.4.

For coupon collector's problem, it is well known that $\Phi_{\mathrm{N}}(t) < e^{-c\,t/\mathrm{N}} \sim (1 - c'/\mathrm{N})^t$, as $t \to \infty$. Indeed, this follows easily from Chernoff bound for geometric distributions with exponents $(n-1)/n$, $(n-2)/n$, $\ldots$, $2/n$, $1/n$ [AS]. Now, we have $\mathrm{sep}(t) \sim (\lambda_1)^t$, as $t \to \infty$, t - odd. By Theorem 2.1, we have $\mathrm{sep}(t) \leq \Phi(t)$ which implies that $(1 - \lambda_1) = \beta = \Omega(1/\mathrm{N})$. $\square$

Acknowledgements

The results of this paper were obtained several years ago, when both authors were graduate students at MIT and Harvard University, respectively. The second author is grateful to Persi Diaconis for the introduction to the subject, help and encouragement. Discussions with Jason Fulman, Laurent Saloff-Coste and David B. Wilson were also helpful.

This paper was written when the second author was a Fellow at the Mathematical Sciences Research Institute at Berkeley, California.

References

[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*, Advances Applied Math. **8** (1987), 69–97.

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

[AS] N. Alon, J.H. Spencer, P. Erdős, *The Probabilistic Method*, Wiley, New York, 1992.

[Ba] L. Babai, *Automorphism groups, isomorphism, reconstruction*, in Handbook of Combinatorics (R. L. Graham, M. Groetschel, and L. Lovasz, eds.), Elsevier, 1996.

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

[CP] D. Coppersmith, I. Pak, *Random walk on upper triangular matrices mixes rapidly*, Probability Theory and Related Fields **117** (2000), 407–417.

[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.

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

[DS1] P. Diaconis, L. Saloff–Coste, *Comparison techniques for random walk on finite groups*, Annals of Probability **21** (1993), 2131–2156.

[DS2] P. Diaconis, L. Saloff–Coste, *Moderate growth and random walk on finite groups*, Geom. Funct. Anal. **4** (1994), 1–36.

[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.

[L] S. Lang, *Algebra* (Second edition), Addison-Wesley, Reading, MA, 1984.

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

[NS] D.J. Newman, L. Shepp, *The double dixie cup problem*, Amer. Math. Monthly **67** (1960), 58–61.

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

[P2] I Pak, *Random walks on finite groups with few random generators*, Electronic J. of Probability **4** (1999), 1–11.

[P3] I. Pak, *Two random walks on upper triangular matrices*, J. Theoretical Probability **13** (2000), 1083–1100.

[P4] I. Pak, *Random walk by rotations,* in preparation.

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

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

[W] W. Woess, *Random walks on infinite graphs and groups*, Cambridge University Press, Cambridge, 2000.