AofA'12 DMTCS proc. **AQ**, 2012, 155–166

Generic properties of random subgroups of a free group for general distributions

Frédérique Bassino[1]†, Cyril Nicaud[2]‡ and Pascal Weil[3]§

[1] *LIPN, Université Paris 13, CNRS UMR 7030, France*
[2] *Université Paris-Est, LIGM, UMR 8049, France*
[3] *CNRS, LaBRI, UMR 5800, F-33400 Talence, France.*

We consider a generalization of the uniform word-based distribution for finitely generated subgroups of a free group. In our setting, the number of generators is not fixed, the length of each generator is determined by a random variable with some simple constraints and the distribution of words of a fixed length is specified by a Markov process. We show by probabilistic arguments that under rather relaxed assumptions, the good properties of the uniform word-based distribution are preserved: generically (but maybe not exponentially generically), the tuple we pick is a basis of the subgroup it generates, this subgroup is malnormal and the group presentation defined by this tuple satisfies a small cancellation condition.

Keywords: Free group, Random groups, Small cancellation

Our starting point is a classical distribution on the set of finitely generated subgroups of a free group, for which the properties of "typical" subgroups were abundantly studied in the literature. Using a probabilistic approach (rather than a more combinatorial one), we describe a vast class of generalizations of this classical distribution, for which most "typical" properties of subgroups remain valid.

Interest for the typical properties of groups can be traced to Gromov [6], who introduced several models for random groups: he considers the statistical properties of finitely presented groups given by random sets of relators of size at most n, see [13] for a survey.

A growing body of literature considers another model of random groups, by focusing on the statistical properties of random finitely generated subgroups of free groups (and not necessarily the corresponding finitely presented groups). To randomly generate a finitely generated subgroup, one may either generate a tuple of reduced words and consider the subgroup they generate (the classical case, see Arzhantseva and Ol'shanskiĭ [2], Jitsukawa [7], Kapovich, Miasnikov, Schupp and Shpilrain [8] and Section 1.1 below), or directly generate a Stallings graph [4, 3].

In these schemes, instances (say, tuples of generators) of size at most n (say, the maximum length of a generator) are considered equally likely, and one considers the probability p_n that a property $\mathcal{P}$ holds for

†Email: `bassino@lipn.univ-paris13.fr`. Work supported by ANR 2010 BLAN 0204 MAGNUM.
‡Email: `cyril.nicaud@univ-mlv.fr`. Work supported by ANR 2010 BLAN 0204 MAGNUM.
§Email: `pascal.weil@labri.fr`. Work supported by ANR 2010 BLAN 0202 01 FREC.

a randomly chosen instance of size n. The property $\mathcal{P}$ is called (exponentially) negligible if p_n tends to 0 (exponentially fast), and (exponentially) generic if its complement is (exponentially) negligible.

In [8, 7], an integer $k \geq 1$ is fixed and one draws uniformly at random k-tuples of reduced words of length at most n. This is what we call the uniform word-based distribution. It is known that, exponentially generically, the k-tuple is a basis of the subgroup it generates and this subgroup is malnormal, see [2, 7] and Section 2.1 . Proofs in this context are mostly combinatorial: the number of reduced words of length n is $2r(2r-1)^{n-1}$ (where r is the rank of the free group) and many properties can be computed directly. A bit of care and an extra injection of probability theory are however necessary to establish exponential genericity in some cases.

We propose to generalize this uniform word-based distribution as follows: the number of generators is not fixed, but determined by a random variable; the length of each generator is also determined by a random variable (excluding the possibility of a significant number of short words, and normalized so that its expected value is n); and the distribution of words of a fixed length is specified by a Markov process. Precise definitions are given in Section 2.2. We show that under rather relaxed assumptions, the good properties of the uniform word-based distribution are preserved: generically (but maybe not exponentially generically), the tuple we pick is a basis of the subgroup it generates and this subgroup is malnormal.

However, all the proofs must be revisited in a probability-theoretic spirit, as we cannot rely any more on enumeration formulas and on probabilities computed as a quotient of set cardinalities. In particular, if $\vec{h} = (h_1, \ldots, h_k)$ is a tuple of reduced words and $\vec{h}^{\pm} = (h_1, h_1^{-1}, \ldots, h_k, h_k^{-1})$, we estimate the probability that the height of the trie of $\vec{h}$ is larger than a function $\tau(n)$, according to the growth rate of τ (which may be linear, or grow much slower). This probability tends to 0 more or less fast (exponentially so, or super-polynomially so) if the random variables governing the number and the length of the generators satisfy certain technical properties. We also estimate in the same fashion the probability that long words have several occurrences in the words of $\vec{h}^{\pm}$.

Besides the consequences on the statistical properties of subgroups already mentioned (being freely generated by $\vec{h}$, malnormality), we also draw consequences on the groups presented by these tuples. We give conditions on the distribution which ensure that these groups generically satisfy the small cancellation property $C'(\frac{1}{6})$ (see [11]), implying that they are generically torsion-free, word-hyperbolic, with solvable word and conjugacy problems.

We note finally that Champetier also generalized the uniform word-based distribution, in a different fashion [5]: he also considers a fixed number of generators k, words of equal length are still equally likely, but he requires that their minimum length should tend to infinity. Proofs in that context are quite different.

1 Definitions

1.1 Free groups and reduced words

Let A be a non-empty set, which will remain fixed throughout the paper, and let $\tilde{A}$ be the symmetrized alphabet, namely the disjoint union of A and a set of formal inverses $A^{-1} = \{a^{-1} \in A \mid a \in A\}$. By convention, the formal inverse operation is extended to $\tilde{A}$ by letting $(a^{-1})^{-1} = a$ for each $a \in A$. A word in $\tilde{A}^*$ (that is: a word written on the alphabet $\tilde{A}$) is *reduced* if it does not contain length 2 factors of the form aa^{-1} ($a \in \tilde{A}$). If a word is not reduced, one can *reduce* it by iteratively removing every pattern of the form aa^{-1}. The resulting reduced word is uniquely determined: it does not depend on the order of the cancellations. For instance, $u = aabb^{-1}a^{-1}$ reduces to aaa^{-1}, and thence to a.

The set F of reduced words is naturally equipped with a structure of group, where the product $u \cdot v$ is the (reduced) word obtained by reducing the concatenation uv. This group is called the *free group* on A. More generally, every group isomorphic to F, say, $G = \varphi(F)$ where φ is an isomorphism, is said to be a free group, freely generated by $\varphi(A)$. The set $\varphi(A)$ is called a *basis* of G. It is important to note that F has infinitely many bases: A is always a basis, but each set $\{a^n ba^m, a\}$ is one as well (if $A = \{a, b\}$). The *rank* of F (or of any isomorphic free group) is the cardinality $|A|$ of A, and one shows that this notion is well-defined in the following sense: the free groups on the sets A and B are isomorphic if and only if $|A| = |B|$.

A group G is *generated* by a subset X if every element of G can be written as a product of elements of X and their inverses. It is *finitely generated* if it admits a finite set X of generators. In this paper, we are interested especially in the finitely generated subgroups of finite rank (*i.e.*, finitely generated) free groups. Recall that every subgroup of a free group is free (Nielsen-Schreier theorem), and hence it has a rank as well, but that the rank of a subgroup may well be greater than that of the group: the free group of rank 2 has subgroups of every finite rank.

1.2 Graphical representation of subgroups of free groups

A privileged tool for the study of subgroups of free groups is the *Stallings graph* of a subgroup of H, a finite directed graph of a particular type uniquely representing H, whose computation was first made explicit by Stallings [18]. The mathematical object itself is already described by Serre [16]. The description we give below differs slightly from Serre's and Stallings', it follows [20, 9, 19, 12, 17] and it emphasizes the combinatorial, graph-theoretic aspect, which is more conducive to the discussion of algorithmic properties.

A *finite A-graph* is a pair $\Gamma = (V, E)$ with V finite and $E \subseteq V \times A \times V$, such that if both (u, a, v) and (u, a, v') are in E then $v = v'$, and if both (u, a, v) and (u', a, v) are in E then $u = u'$. Let $v \in V$. The pair (Γ, v) is said to be *admissible* if the underlying graph of Γ is connected (that is: the undirected graph obtained from Γ by forgetting the letter labels and the orientation of edges), and if every vertex $w \in V$, except possibly v, occurs in at least two edges in E.

Every admissible pair $(\Gamma, 1)$ represents a unique finitely generated subgroup H of $F(A)$ in the following sense: if u is a reduced word, then $u \in H$ if and only if u labels a loop at 1 in Γ (by convention, an edge (u, a, v) can be read from u to v with label a, or from v to u with label a^{-1}). Moreover, each finitely generated subgroup H of $F(A)$ is represented in that sense by a unique admissible pair, which we call the *Stallings graph* of H and write $(\Gamma(H), 1)$.

Some algebraic properties of H can be directly seen on its Stallings graph $(\Gamma, 1)$. For instance, the rank of H is exactly $|E| - |V| + 1$. See [18, 20, 9, 12] for more information about Stallings graphs.

The Stallings graph of a given finitely generated subgroup H can be computed effectively, and efficiently. A quick description of the algorithm is as follows. Let $\vec{h} = (h_1, \ldots, h_k)$ be a tuple of reduced words generating H. We first build a graph with edges labeled by letters in $\tilde{A}$, and then reduce it to an A-graph using *foldings*. First build a vertex 1. Then, for every $1 \leq i \leq k$, build a loop with label h_i from 1 to 1, adding $|h_i| - 1$ new vertices. Change every edge (u, a^{-1}, v) labeled by a letter of A^{-1} into an edge (v, a, u). At this point, we have constructed the so-called *bouquet* of loops labeled by the h_i.

Then iteratively identify the vertices v and w whenever there exists a vertex u and a letter $a \in A$ such that either both (u, a, v) and (u, a, w) or both (v, a, u) and (w, a, u) are edges in the graph (the

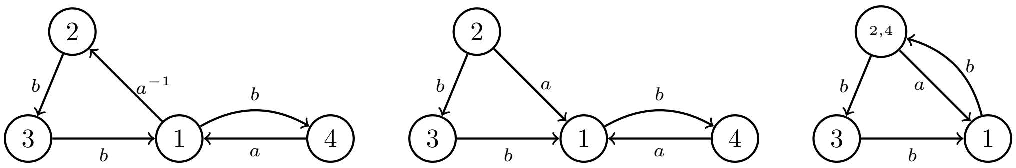

Fig. 1: Starting with $\{a^{-1}bb, ba\}$, a loop labeled by each word is built around 1; the edge $1 \xrightarrow{a^{-1}} 2$ is then changed into $2 \xrightarrow{a} 1$ to have positive labels only; since 2 and 4 both have an edge labeled with a ending in the same vertex 1, they are merged. The foldings halts here on this small example, since there are no vertex with two incoming or outgoing edges with the same label.

corresponding two edges are *folded*, in Stallings' terminology). An example is depicted on Fig. 1.

The resulting graph Γ is such that $(\Gamma, 1)$ is admissible, the reduced words labeling a loop at 1 are exactly the elements of H and, very much like in the (1-dimensional) reduction of words, that graph does not depend on the order used to perform the foldings.

1.3 Genericity

Let us say that a function f, defined on $\mathbb{N}$ and such that $\lim f(n) = 0$, is *super-polynomially small* (resp. *exponentially small*) if $f(n) = o(n^{-d})$ for every $d > 1$ (resp. if $f(n) < e^{dn}$ for some $d > 0$ and for n large enough).

Given a sequence of probability laws $(\mathbb{P}_n)_n$ on a set S, we say that a subset $X \subseteq S$ is *negligible* if $\lim_n \mathbb{P}_n(X) = 0$ and *generic* if its complement is negligible.

We also say that X is *super-polynomially negligible* (resp. *exponentially negligible*) if $\mathbb{P}_n(X)$ is *super-polynomially small* (resp. *exponentially small*). And it is *super-polynomially generic* (resp. *exponentially generic*) if its complement is *super-polynomially negligible* (resp. *exponentially negligible*).

2 Generic properties of finite rank subgroups of free groups

We concentrate on the (classical) approach (following [2, 1, 7]), where a subgroup is given by a tuple of words generating it, that is, we consider a sequence of probability laws $(\mathbb{P}_n)$ on the set of tuples of elements of F.

2.1 The uniform distribution on k-tuples of words

In a situation often considered, an integer $k > 0$ is fixed, and we consider the uniform probability $\mathbb{P}_n$ on the set of k-tuples of reduced words of length at most n.

Let $\vec{h} = (h_1, \ldots, h_k)$ be a k-tuple of reduced words, let $\vec{h}^{\pm} = (h_1, h_1^{-1}, \ldots, h_k, h_k^{-1})$ and let $\mu(\vec{h}) = \min_i |h_i|$. It was observed in [2, 7] that, for each $0 < \alpha < 1$, $\mu(\vec{h}) \geq \alpha n$ exponentially generically. Moreover, let $\tau(\vec{h})$ be the length of the longest prefix common to two words in $\vec{h}^{\pm}$. Then, for every $0 < \beta < \frac{1}{2}\alpha$, $\tau(\vec{h}) \leq \beta n$ exponentially generically [7].

It follows that, exponentially generically, the Stallings graph $\Gamma(H)$ consists of a "small" *central tree* (namely the trie of $\vec{h}^{\pm}$, rooted at vertex 1) and "long" *outer loops*, one for each h_i. This geometry of the Stallings graph of H implies that H is freely generated by $\vec{h}$, and hence has rank k. Moreover, in that same geometry, the tuple $\vec{h}$ is determined by $\Gamma(H)$, up to the order of its elements and the direction in which they are read. In particular, exponentially generically, a given subgroup will be produced by a fixed number of tuples $\vec{h}$ (among the k-tuples of reduced words of length at most n) and hence, the random generation of such k-tuples is an acceptable way of randomly generating rank k subgroups of F. We refer the reader to [7, 3] for further details on these results.

We also observe that, in that situation, $\Gamma(H)$ can be computed in linear time, simply by computing the initial cancellation in the tuple $\vec{h}^{\pm}$.

Under the same sequence of uniform distributions $(\mathbb{P}_n)$, H is exponentially generically *malnormal* [7]. By definition, H is malnormal if, for every $x \notin H$, $H \cap x^{-1}Hx = \{1\}$. This algebraic property of H translates exactly to the following combinatorial property of the Stallings graph of H: no non-trivial word u labels a loop at two distinct vertices of $\Gamma(H)$ (see [9]).

In the exponentially generic situation described above, any loop in $\Gamma(H)$ must run along at least one outer loop (so it must have length at least $\mu(\vec{h}) - 2\tau(\vec{h})$), and the portions of its travel inside the central tree are each of length at most $2\tau(\vec{h})$. A more detailed analysis shows the following result.

Lemma 2.1 *Let $\vec{h}$ be a tuple of reduced words and let $H = \langle\vec{h}\rangle$. If $\tau(\vec{h}) < \frac{1}{8}\mu(\vec{h})$ and H is not malnormal, then there exists a word of length $\frac{1}{8}\mu(\vec{h})$, with two distinct occurrences in the words of $\vec{h}^{\pm}$ sitting at distance at least $\frac{1}{8}\mu(\vec{h})$ from the extremities of these words.*

We already know that, under the sequence of uniform distributions $(\mathbb{P}_n)$, $\tau(\vec{h}) < \frac{1}{8}\mu(\vec{h})$ exponentially generically. It also holds that, exponentially generically, the words of $\vec{h}^{\pm}$ do not have common factors of length $\frac{1}{8}\mu(\vec{h})$ [7, 3]. Therefore, exponentially generically, H is malnormal.

2.2 Relaxing the parameters of the distribution

We now relax all the parameters of the distributions of tuples of reduced words, with the objective of preserving the genericity of the properties discussed in Section 2.1.

In our scheme, the probability $\mathbb{P}_n$ on the set of tuples of reduced words is determined as follows. The size of the tuple and the lengths of the words are determined by random variables K_n and L_n, on which we impose the following restrictions: the number of words in our tuple, given by K_n, cannot be too large and generically a tuple does not contain short words; in addition, the average length of a word, $\mathbb{E}(L_n)$, is equal to n (a sort of normalization). More precisely:

- $\mathbb{E}(L_n) = n$ and there exists a function $\mu(n)$ such that $\lim_n \mu(n) = \infty$ and $L_n \geq \mu(n)$ generically: if we let $\mathsf{upper}_\mu = \mathbb{P}[L_n < \mu(n)]$, then $\lim_n \mathsf{upper}_\mu = 0$. One may think of $\mu(n)$ as a function of the form n^λ for some $0 < \lambda < 1$, or some logarithmic function.
- There exists a function $\nu(n)$ such that $K_n \leq \nu(n)$ generically: if we let $\mathsf{lower}_\nu = \mathbb{P}[K_n > \nu(n)]$, then $\lim_n \mathsf{lower}_\nu = 0$. We will consider situations where $\nu(n)$ is a constant function, or $\mathcal{O}(\log^d n)$ or $\mathcal{O}(n^d)$ for some $d > 0$.

Moreover, we do not consider the uniform distribution on the set of reduced words of a given length, but the distribution given by a Markovian scheme which we now proceed to describe.

2.3 Markovian automata

A *Markovian automaton*[(i)] $\mathcal{A}$ consists of

[(i)] This notion is different from the two notions of probabilistic automata, introduced by Rabin [14] and Segala and Lynch [15], respectively.

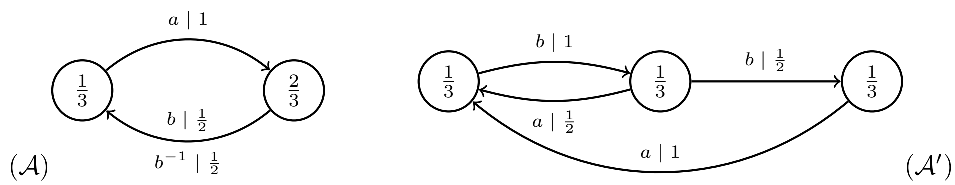

Fig. 2: Markovian automata $\mathcal{A}$ and $\mathcal{A}'$.

- a deterministic transition system $(Q, \cdot)$ on alphabet X, where Q is a finite non-empty set called the *state set*, and for each $q \in Q$, $x \in X$, $q \cdot x \in Q$ or $q \cdot x$ is undefined;
- an initial probability vector $\gamma_0 \in [0,1]^Q$, i.e. a vector such that $\sum_{q\in Q} \gamma_0(q) = 1$;
- for each $p \in Q$, a probability vector $(\gamma(p,x))_{x\in X} \in [0,1]^X$, such that $\gamma(p,x) = 0$ if and only if $p \cdot x$ is undefined.

If $u = x_1 \cdots x_n \in X^*$, we write $\gamma(q,u) = \gamma(q,x_1)\gamma(q\cdot x_1, x_2)\cdots\gamma(q\cdot(x_1\cdots x_{n-1}), x_n)$ for $n \geq 1$ and $\gamma(q,u) = 1$ if $n = 0$. We also write $\gamma_0(u) = \sum_{q\in Q}\gamma_0(q)\gamma(q,u)$.

For each $n \geq 0$, γ_0 determines a distribution on the set of elements of X^* of length n, and hence a probability law on that set. We denote this probability law by $\mathbb{P}_n$. In particular, we are only defining the probability of a subset of X^* that consists of equal length elements.

Markovian automata are very similar to hidden Markov chain models, except that symbols are output on transitions instead of on states. In our context, Markovian automata are more convenient since sets of words (languages) are naturally described by automata. See the examples below, all set in the context of group theoretic applications, and Section 2.4.

Uniform distribution on length n elements of F: We exploit the fact that the set of reduced words is rational, and hence is accepted by an automaton, and we set uniform probabilities on the transitions of that automaton. The state set is $Q = \tilde{A}$. For each $a \in \tilde{A}$, there is an a-labeled transition from every state except a^{-1}, ending in state a. If all of these transitions have the same probability, namely $\frac{1}{2r-1}$, r being the rank of F, and if the initial probability vector is uniform as well, with each component equal to $\frac{1}{2r}$, then the Markovian automaton yields the uniform distribution on length n elements of F. We can also tweak these probabilities, to favor certain letters over others, or to favor positive letters (the letters in A) over negative letters.

Distributions on rational subsets of F: The support of the distribution (the words with non-zero probability) does not have to be equal to the set of all reduced words. We can consider a rational subset L of F, or rather a deterministic automaton accepting only reduced words, and impose probabilistic weights on its transitions to form a Markovian automaton. The resulting distribution gives non-zero weights only to prefixes of elements of L. This can be applied to the case where $L = A^*$, the set of positive words, or when L is a finitely generated subgroup of F and the automaton is constructed from the Stallings graph of that subgroup.

Distributions related to the group $PSL(2,\mathbb{Z}) = \langle a, b \mid a^2, b^3\rangle$: Figure 2 represents two Markovian automata: the transitions are labeled by a letter and a probability, and each state is decorated with the corresponding initial probability. The support of the distribution defined by automaton $\mathcal{A}$ is the set of words over alphabet $\{a, b, b^{-1}\}$ without occurrences of the factors a^2, b^2, $(b^{-1})^2$, bb^{-1} and $b^{-1}b$, and the support of the distribution defined by $\mathcal{A}'$ consists of the words on alphabet $\{a, b\}$, without occurrences of a^2 or b^3. Both are regular sets of unique representatives of the elements of $PSL(2,\mathbb{Z})$ (the first is the set of geodesics of $PSL(2,\mathbb{Z})$, and also the set of Dehn-reduced words with respect to the given presentation of that group; the second is a set of quasi-geodesics of $PSL(2,\mathbb{Z})$). Note that the underlying graphs of the markovian automata of Fig.2 are parts of de Bruijn graphs since they are defined by forbidden finite factors. Notice that the distribution produced by $\mathcal{A}'$ is not uniform on words of length n of its support.

2.4 Markov chains and Markovian automata

A Markovian automaton hides (rather poorly) a Markov chain. If $\mathcal{A}$ is a Markovian automaton on alphabet X, with state set Q, we define a Markov chain $M(\mathcal{A})$ on Q as follows: its transition matrix is given by $M(p,q) = \sum_{x\in X \text{ s.t. } p\cdot x=q} \gamma(p,x)$ for all $p,q \in Q$, and its initial vector is γ_0.

Recall that the Markov chain $M(\mathcal{A})$ is *irreducible* if, for all $p,q \in Q$, $M(\mathcal{A})^n(p,q) > 0$ for some $n > 0$: this is equivalent to the strong connectedness of $\mathcal{A}$. Recall also that $M(\mathcal{A})$ is *aperiodic* if, if for each $q \in Q$, $M(\mathcal{A})^n(q,q) > 0$ for all large enough n: this is equivalent to stating that $\mathcal{A}$ has loops of every large enough length at every state, equivalent again to stating that $\mathcal{A}$ has a collection of loops of relatively prime lengths. If both these properties hold, we say that $\mathcal{A}$ itself is irreducible and aperiodic and we can apply the classical theorem on Markov chains: there exists a *stationary vector* $\tilde\gamma$ and the distribution defined by $\mathcal{A}$ converges to that stationary vector exponentially fast (see [10, Thm 4.9]). In the vocabulary of Markovian automata, this yields the following theorem.

If $u \in X^*$ has length n, let $Q_n^p(u)$ be the state of $\mathcal{A}$ reached after reading the word u starting at state p.

Theorem 2.2 *Let $\mathcal{A}$ be an irreducible and aperiodic Markovian automaton on alphabet X, with state set Q ($|Q| \geq 2$) and stationary vector $\tilde\gamma$. For each $q \in Q$, $\tilde\gamma(q) = \lim_{n\to\infty} \mathbb{P}[Q_n^p = q]$. More precisely, there exist $K > 0$ and $0 < c < 1$, such that $|\mathbb{P}[Q_n^p = q] - \tilde\gamma(q)| < Kc^n$ for all n large enough.*

Remark 2.3 The constant c in Theorem 2.2 is the maximal modulus of the non-1 eigenvalues of $M(\mathcal{A})$.

The Markovian automaton discussed in Section 2.3, relative to the uniform distribution on reduced words of length n is aperiodic and irreducible, as well as the two Markovian automata related to $PSL(2,\mathbb{Z})$. The respective values for c are $\frac{1}{2r-1}$, $\frac{1}{2}$ and $\frac{1}{\sqrt{2}}$.

We also record the following statement, on the exponential decrease of the probability of a word.

Lemma 2.4 *Let $\mathcal{A}$ be an irreducible and aperiodic Markovian automaton on alphabet X. There exist constants $K_1 > 0$ and $0 < c_1 < 1$ such that, for each state q and for each word v of length n, $\gamma_0(v), \tilde\gamma(v), \gamma(q,v) \leq K_1 c_1^n$.*

Proof. Let ℓ be the maximum length of an elementary cycle (one that does not visit twice the same state) and let δ be the maximum value of $\gamma(q,\kappa)$ where κ is an elementary cycle at state q. Since $\mathcal{A}$ is irreducible and aperiodic, we always have $\gamma(q,\kappa) < 1$, so $\delta < 1$.

Every cycle κ can be represented as a composition of at least $|\kappa|/\ell$ elementary cycles (here, the composition takes the form of a sequence of insertions of a cycle in another). Consequently $\gamma(q,\kappa) \leq \delta^{\frac{|\kappa|}{\ell}}$. Finally, every path π can be represented as a product of cycles and at most $|Q|$ individual edges. So, if π starts at state q, $\gamma(q,\pi) \leq \delta^{\frac{|\pi|-|Q|}{\ell}}$. We get the announced result by letting $K_1 = \delta^{\frac{-|Q|}{\ell}}$ and $c_1 = \delta^{\frac{1}{\ell}}$. $\square$

3 Cancellation properties

The random variables K_n and L_n, and the irreducible and aperiodic Markovian automaton $\mathcal{A}$ are now fixed, and the constants K, K_1, c and c_1 are those discussed in Section 2.4.

Let $\vec{h} = (h_1, \dots, h_s)$ be a randomly chosen tuple of reduced words and let $H = \langle \vec{h} \rangle$. We first consider *initial cancellation* on $\vec{h}$, that is, the existence of common prefixes between the words in $\vec{h}^{\pm}$, measured by the parameter $\tau(\vec{h})$ (see Section 2.1). In Section 3.2, we will also estimate the probability of the existence of long common factors in the middle part of the words of $\vec{h}^{\pm}$. Applications to subgroup properties are discussed in Section 4.

3.1 Initial cancellation

Let T_n be the random variable, relative to the probability law $\mathbb{P}_n$, given by $T_n(\vec{h}) = \tau(\vec{h})$. Our main theorem on initial cancellation is the following.

Theorem 3.1 *Let $0 < \alpha < 1$ and let $\tau(n)$ be a function such that $\tau(n) \le \alpha\mu(n)$.*

- *Exponential genericity case: If $\tau(n)$ grows at least linearly, ν grows sub-exponentially ($\nu(n) = o(d^n)$ for every $d > 1$) and upper_μ and lower_ν are exponentially small, then $T_n \le \tau(n)$ exponentially generically.*
- *Super-polynomial genericity case:If $\tau(n)$ grows faster than $\log n$ ($\log n = o(\tau(n))$), ν grows at most polynomially ($\nu(n) = \mathcal{O}(n^d)$ for some $d > 1$) and upper_μ and lower_ν are super-polynomially small, then $T_n \le \tau(n)$ super-polynomially generically.*
- *Genericity case: Suppose now that $\lim \tau(n) = \infty$. Any one of the following conditions implies that $T_n \le \tau(n)$ generically:*
  - $\nu(n)$ is bounded;
  - $\nu(n) = \mathcal{O}(\log^d n)$ for some $d > 0$, $\mathsf{upper}_\mu = o(\frac{1}{\log^{2d} n})$ and $\tau(n)$ grows faster than $\log\log n$;
  - $\nu(n) = \mathcal{O}(n^d)$ for some $d > 0$, $\mathsf{upper}_\mu = o(n^{-2d})$ and $\tau(n)$ grows faster than $\log n$.

Here we only sketch the main steps of the proof, which is rather technical. The first step is to observe that the probability that $T_n(\vec{h}) > \tau(n)$ is bounded above by the sum of the following probabilities:

- the probability lower_ν that $K_n > \nu(n)$;

- the probability P_1 that $K_n \le \nu(n)$ and for some $i < j$, h_i and h_j have a common prefix of length greater than $\tau(n)$ (or h_i and h_j^{-1}, or h_i^{-1} and h_j^{-1}, or h_i^{-1} and h_j^{-1});

- the probability P_2 that $K_n \le \nu(n)$ and for some i, h_i and h_i^{-1} have a common prefix of length greater than $\tau(n)$.

The proof now consists in bounding P_1 and P_2.

First we have $P_1 \le \binom{\nu(n)}{2}(P_{1,1} + P_{1,2} + P_{1,3} + P_{1,4})$, where $P_{1,1}$ (resp. $P_{1,2}$, $P_{1,3}$, $P_{1,4}$) is the probability for a pair of words (h, h') that h and h' (resp. h and ${h'}^{-1}$, h^{-1} and h', h^{-1} and ${h'}^{-1}$) have a common prefix of length $t = 1 + \tau(n)$.

Since h and h' are drawn independently, we have $P_{1,1} = \sum_{|u|=t} \mathbb{P}_n[h \in u\tilde{A}^*]\mathbb{P}_n[h' \in u\tilde{A}^*]$. Note that $\sum_{|u|=t} \mathbb{P}_n[h \in u\tilde{A}^*] = \sum_{|u|=t} \mathbb{P}_n[h \in \tilde{A}^* u^{-1}] = 1$. So $P_{1,1} \le \max_{|u|=t} \mathbb{P}_n[h \in u\tilde{A}^*] = \max_{|u|=t} \gamma_0(u)$. In view of Lemma 2.4, it follows that $P_{1,1} \le K_1 c_1^t$.

The same reasoning yields the same bound for $P_{1,2}$ and $P_{1,3}$. It also yields the inequality $P_{1,4} \leq \max_{|u|=t} \mathbb{P}[h \in \tilde{A}^* u]$. This can be bounded using Theorem 2.2, provided we know that h is long enough. Thus we get, for all n large enough, $P_{1,4} \leq \mathsf{upper}_\mu + K_1 c_1^t + K|Q|c^{\mu(n)-t}$.

Next we bound P_2. We first get $P_2 \leq \nu(n) \sum_{|u|=t} \mathbb{P}_n[h \in u\tilde{A}^* u^{-1}]$.
For each u of length t and for all h long enough, using Theorem 2.2, we have

$$\begin{aligned}
\mathbb{P}_n[h \in u\tilde{A}^* u^{-1}] &= \sum_{p \in Q} \gamma_0(p)\gamma(p,u) \left(\sum_{q \in Q} \mathbb{P}[Q^{p \cdot u}_{|h|-2t} = q] \gamma(q, u^{-1}) \right) \\
&\leq \sum_{p \in Q} \gamma_0(p)\gamma(p,u) \left(\sum_{q \in Q} (\tilde{\gamma}(q) + Kc^{|h|-2t}) \gamma(q, u^{-1}) \right) \\
&\leq \gamma_0(u) \left(\tilde{\gamma}(u^{-1}) + Kc^{|h|-2t} \sum_{q \in Q} \gamma(q, u^{-1}) \right) \leq \gamma_0(u) \left(K_1 c_1^t + |Q| K c^{|h|-2t} K_1 c_1^t \right).
\end{aligned}$$

Therefore, summing for all u of length t:

$$P_2 \;\leq\; \nu(n) \left(\mathsf{upper}_\mu + K_1 c_1^t + |Q| K c^{\mu(n)-2t} K_1 c_1^t \right).$$

The remainder of the proof of Theorem 3.1 is a simple application of these upper bounds, to make sure that all are exponentially small, super-polynomially small, or simply tend to 0.

3.2 Probability of multiple occurrences of long factors

In view of Lemma 2.1, we are interested in bounding the probability that a word of length $\beta\mu(n)$ has several occurrences in the words of $\vec{h}^{\pm}$, at distances at least $\beta\mu(n)$ from the extremities (and then take $\beta = \frac{1}{8}$, another value for β will be chosen in Section 4). This probability itself is bounded by the sum of the following probabilities:

- the probability P_1 that such a word has two occurrences in $h_i^{\pm 1}$ and $h_j^{\pm 1}$ for some $i \neq j$;
- the probability P_2 that such a word has two occurrences in h_i (or in h_i^{-1}) for some i;
- the probability P_3 that such a word has an occurrence in h_i and an occurrence in h_i^{-1} for some i.

Proposition 3.2 *There exist constants $b_1, b_2, b_3 > 0$, $L_1, L_2, L_3 > 0$ and $0 < d_1, d_2, d_3 < 1$, depending on the Markovian automata, the constant β and the size of the alphabet, such that, for $i = 1, 2, 3$:*

- *if $\mu(n) > b_i \log n$, then P_i tends to 0;*
- *if the length of the words generated is at most $d_i^{-\frac{1}{4}\mu(n)}$, then $P_i \leq L_i d_i^{\frac{1}{2}\mu(n)}$.*

Sketch of proof. We first sketch the proof concerning P_1. Let $0 < a < \beta$: the probability that h_1 and h_2 have a common factor of length $a\mu(n)$ at distance at least $\beta\mu(n)$ from the extremities, is greater than or equal to P_1. Let v be a word of length $a\mu(n)$ and let $i > \beta\mu(n)$.

Using Theorem 2.2 and Lemma 2.4, we find that the probability that v occurs as a factor of h starting at position i is

$$\sum_{p,q \in Q} \gamma_0(p) \mathbb{P}[Q_i^p = q] \gamma(q, v) \;\leq\; \tilde{\gamma}(v) + Kc^{\beta\mu(n)} \leq K_1 c_1^{a\mu(n)} + Kc^{\beta\mu(n)}.$$

It follows that the probability $P_1(v,i,j)$ that v occurs as a factor in h_1 and h_2, in positions i and j respectively, $i,j > \beta\mu(n)$, satisfies

$$P_1(v,i,j) \leq \left(\tilde{\gamma}(v) + Kc^{\beta\mu(n)}\right)^2 \leq \tilde{\gamma}(v)\left(K_1c_1^{a\mu(n)} + 2Kc^{\beta\mu(n)}\right) + K^2c^{2\beta\mu(n)}.$$

Therefore, the probability that h_1 and h_2 have a common factor of length $a\mu(n)$ starting at fixed positions $i,j > \beta\mu_n$, $P_1(i,j) = \sum_{|v|=a\mu(n)} P_1(v,i,j)$, satisfies

$$P_1(i,j) \;\; \leq \;\; K_1c_1^{a\mu(n)} + 2Kc^{\beta\mu(n)} + \frac{2r}{2r-1}(2r-1)^{a\mu(n)}K^2c^{2\beta\mu(n)}.$$

Thus, if $a < \frac{-2\beta\log c}{\log(2r-1)}$ (so that $(2r-1)^ac^{2\beta} < 1$), we have $P_1(i,j) \leq L_1'd_1^{\mu(n)}$ for some constants L_1 and d_1 (with $d_1 = \max(c_1^a, c^\beta, (2r-1)^ac^{2\beta})$).

Now $P_1 \leq \sum_{i,j>\beta\mu(n)} P_1(i,j)$, and bounding this sum raises a difficulty as we have not assumed, so far, that L_n is bounded.

We can get simple convergence to 0 without assuming bounded length, if $\mu(n)$ grows sufficiently fast: if $\mu(n) > b\log n$ for some $b > -\frac{2}{\log d_1}$ and if $1 < d < -\frac{b}{2}\log d_1$, then $\mathbb{P}[L_n > n^d] < n^{-d}\mathbb{E}(L_n) = \frac{1}{n^{d-1}}$ (using Markov's inequality and the fact that $\mathbb{E}(L_n) = n$). So

$$P_1 \leq \mathbb{P}[L_n > n^d] + \binom{n^d}{2}L_1'd_1^{\mu(n)} \leq L_1'n^{2d}d_1^{\mu(n)} \leq L_1'n^{b\log d_1+2d},$$

and our assumptions guarantee that $b\log d_1 + 2d < 0$.

We also obtain a genericity result by imposing a generic bound on the length of the words under consideration. More precisely, if $|h_1|, |h_2| \leq d_1^{-\frac{1}{4}\mu(n)}$, then the probability P_1 satisfies

$$P_1 \leq \binom{d_1^{-\frac{1}{4}\mu(n)}}{2}L_1'd_1^{\mu(n)} \leq L_1'd_1^{-\frac{1}{2}\mu(n)}.$$

This is sufficient to conclude the proof concerning P_1.

To bound P_2, we observe that if a word w has two occurrences in some h, these occurrences may overlap but whatever the situation, the prefix of w of length $\frac{1}{4}|v|$ has two occurrences separated by a gap of at least $\frac{1}{4}|v|$, using classical combinatorics on words. So we want to bound the probability that h_1 contains two occurrences of a word v of length $a\mu(n)$, $a < \frac{\beta}{4}$ has occurrences at positions $i > \beta\mu(n)$ and $j > i + \frac{\beta}{2}\mu(n)$. Using again Theorem 2.2 and Lemma 2.4, we find that for v, i and j fixed, this probability satisfies

$$\begin{aligned} P_2(v,i,j) &= \sum_{p,q,q'\in Q} \gamma_0(p)\mathbb{P}[Q_i^p = q]\gamma(q,v)\mathbb{P}[Q_{j-i-a\mu(n)}^{q\cdot v} = q']\gamma(q',v) \\ &\leq \tilde{\gamma}(v)^2 + 2KK_1c_1^{\beta\mu(n)} + K^2c^{2\beta\mu(n)}. \end{aligned}$$

The proof then follows the same steps as for the bound of P_1.

As for P_3, we note that if w and w^{-1} both have occurrences in some reduced word h, then these occurrences may not overlap: if v is the prefix of w of length $\frac{2}{3}|w|$, then v and v^{-1} have occurrences in h, separated by a gap of at least $\frac{2}{3}|w|$. The proof then follows the same steps as for the bound of P_2. □

4 Applications

As observed in Section 2.1, if $T_n(\vec{h}) < \frac{1}{2}\min|h_i|$, then the Stallings graph $\Gamma(H)$ is in the shape of a central tree, of height $T_n(\vec{h})$, and of a collection of outer loops, one for h_i. In that situation, $\vec{h}$ freely generates the subgroup H. Moreover, $\Gamma(H)$ is constructed in linear time, simply by computing the initial cancellation in the tuple $\vec{h}^{\pm}$.

Theorem 4.1 (Base) *Under the hypotheses of Theorem 3.1, a randomly chosen tuple is a basis of the subgroup it generates exponentially generically (resp. super-polynomially generically, generically).*

Touikan [19] proposed an algorithm that computes the folding of the bouquet (see Section 1.2) of $\vec{h}$ in time $\mathcal{O}(m\log^* N)$, where N is the size of the bouquet, and m is the number of states that have been merged in the process. Using this result in our generic settings, we obtain the following theorem.

Theorem 4.2 (Stallings graph computation) *Under the hypotheses of Theorem 3.1 and assuming furthermore that $\mu(n) = \mathcal{O}(\frac{n}{\log^* n})$, the Stallings graph of a random subgroup is computed in linear time exponentially generically (resp. super-polynomially generically, generically).*

By Lemma 2.1, the probability that H is not malnormal is bounded by the sum that $T_n(\vec{h}) > \frac{1}{8}\mu(n)$ and the probabilities that, assuming $T_n(\vec{h}) \leq \frac{1}{8}\mu(n)$, a word of length $\frac{1}{8}\mu(n)$ has two occurrences in the words of $\vec{h}^{\pm}$, at distance at least $\frac{1}{8}\mu(n)$ from the extremities. In view of Proposition 3.2, this leads to the following statement. Let lbound_L be the probability that $L_n > \max(d_1, d_2, d_3)^{\frac{1}{4}\mu(n)}$.

Theorem 4.3 (Malnormality) *Let $\vec{h}$ be a tuple of elements of F generated at random and let $H = \langle\vec{h}\rangle$.*

If μ grows at least linearly, ν grows sub-exponentially and upper_μ, lower_ν and lbound_L are exponentially small, then H is exponentially generically malnormal.

If μ grows faster than $\log n$, ν grows at most polynomially and upper_μ, lower_ν and lbound_L are super-polynomially small, then H is exponentially generically malnormal.

Each of the following conditions is sufficient to guarantee that H is generically malnormal:

- *ν is bounded and upper_μ, lower_ν and lbound_L tend to 0;*
- *$\nu(n) = \mathcal{O}(\log^d n)$ for some $d > 0$, $\mathsf{upper}_\mu = o(\frac{1}{\log^{2d} n})$, and lower_ν and lbound_L tend to 0,*
- *$\nu(n) = \mathcal{O}(n^d)$ for some $d > 0$, $\mathsf{upper}_\mu = o(n^{-2d})$, and lower_ν and lbound_L tend to 0;*
- *ν is bounded, $\mu(n) > \max(b_1, b_2, b_3)\log n$ and upper_μ and lower_ν tend to 0.*

We finally discuss the properties of the quotient of F by the normal closure of a random subgroup under our model. Recall that a reduced word h its *cyclically reduced* when every cyclic permutation of u is reduced. To any reduced word h we associate its cyclic reduction $c(h)$ obtained by repeatedly remove the first and the last letter while they are the inverse of each other. The normal subgroup generated by a set $\vec{h} = \{h_1, \ldots, h_k\}$ is the same as as the one generated by $\vec{c} = \{c(h_1), \ldots, c(h_k)\}$. The small cancellation theory [11] can greatly help in studying the properties of the quotient: if whenever a word u is a factor of two distinct cyclic conjugates of $\vec{c}$ then $|u| \leq \lambda\min_{i\in\{1\ldots k\}}|c(h_i)|$, then $\vec{h}$ is said to satisfy the small cancellation property $C'(\lambda)$. If $\vec{h}$ satisfies $C'(1/6)$, then the quotient has a lot of nice algebraic properties; we state some in the following theorem and, due to the lack of space, refer to [11] for the definitions.

Theorem 4.4 (Quotient) *Under the hypotheses of Theorem 3.1 and Proposition 3.2, the small cancellation condition $C'(1/6)$ holds generically. In particular, the group $G = \langle A \mid \vec{h}\rangle$ is generically torsion-free, word-hyperbolic and has solvable word problem and conjugacy problem.*

References

[1] Goulnara N. Arzhantseva. A property of subgroups of infinite index in a free group. *Proc. Amer. Math. Soc.*, 128(11):3205–3210, 2000.

[2] Goulnara N. Arzhantseva and Alexander Yu. Olʹshanskiĭ. Generality of the class of groups in which subgroups with a lesser number of generators are free. *Mat. Zametki*, 59(4):489–496, 638, 1996.

[3] Frédérique Bassino, Armando Martino, Cyril Nicaud, Enric Ventura, and Pascal Weil. Statistical properties of subgroups of free groups. *Random Structures and Algorithms*, to appear.

[4] Frédérique Bassino, Cyril Nicaud, and Pascal Weil. Random generation of finitely generated subgroups of a free group. *Internat. J. Algebra Comput.*, 18(2):375–405, 2008.

[5] Christophe Champetier. Propriétés statistiques des groupes de présentation finie. *Journal of Advances in Mathematics*, 116(2):197–262, 1995.

[6] Mikhail Gromov. *Essays in Group Theory*, chapter Hyperbolic groups, pages 75–265. Springer, 1987.

[7] Toshiaki Jitsukawa. Malnormal subgroups of free groups. In *Computational and statistical group theory (Las Vegas, NV/Hoboken, NJ, 2001)*, volume 298 of *Contemp. Math.*, pages 83–95. Amer. Math. Soc., Providence, RI, 2002.

[8] Ilya Kapovich, Alexei Miasnikov, Paul Schupp, and Vladimir Shpilrain. Generic-case complexity, decision problems in group theory, and random walks. *J. Algebra*, 264(2):665–694, 2003.

[9] Ilya Kapovich and Alexei Myasnikov. Stallings foldings and subgroups of free groups. *J. Algebra*, 248(2):608–668, 2002.

[10] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009.

[11] Roger C. Lyndon and Paul E. Schupp. *Combinatorial group theory*. Springer-Verlag, Berlin, 1977. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89.

[12] Alexei Miasnikov, Enric Ventura, and Pascal Weil. Algebraic extensions in free groups. In *Geometric group theory*, Trends Math., pages 225–253. Birkhäuser, Basel, 2007.

[13] Yann Ollivier. *A January 2005 invitation to random groups*, volume 10 of *Ensaios Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, 2005.

[14] Michael O. Rabin. Probabilistic automata. *Information and Computation*, 6(3):230–245, 1963.

[15] Roberto Segala and Nancy Lynch. Probabilistic simulations for probabilistic processes. *Nordic Journal of Computing*, 2(2):250–273, 1995.

[16] Jean-Pierre Serre. *Trees*. Springer-Verlag, Berlin, 1980. Translated from the French by John Stillwell.

[17] Pedro V. Silva and Pascal Weil. On an algorithm to decide whether a free group is a free factor of another. *Theor. Inform. Appl.*, 42(2):395–414, 2008.

[18] John R. Stallings. Topology of finite graphs. *Invent. Math.*, 71(3):551–565, 1983.

[19] Nicholas W. M. Touikan. A fast algorithm for Stallings' folding process. *Internat. J. Algebra Comput.*, 16(6):1031–1045, 2006.

[20] Pascal Weil. Computing closures of finitely generated subgroups of the free group. In *Algorithmic problems in groups and semigroups (Lincoln, NE, 1998)*, Trends Math., pages 289–307. Birkhäuser Boston, Boston, MA, 2000.