Invent. math. 134, 251–299 (1998)

Inventiones mathematicae

Walks on generating sets of groups

P. Diaconis[1], **L. Saloff-Coste**[2]

[1] Cornell University, Department of Mathematics, ORIE, Ithaca, NY, 14853, USA
[2] CNRS, Université Paul Sabatier, Statistique et Probabilités, F-31062 Toulouse cedex, France

Oblatum 6-VIII-1996 & 6-XI-97

Abstract. We study a Markov chain on generating n-tuples of a fixed group which arises in algorithms for manipulating finite groups. The main tools are comparison of two Markov chains on different but related state spaces and combinatorics of random paths. The results involve group theoretical parameters such as the size of minimal generating sets, the number of distinct generating k-tuples for different k's and the maximal diameter of the group.

1 Introduction

This paper studies a new technique for generating random elements of a finite group G. Let S be a set of generators of G. The classical method for using S is to run a random walk: Starting at the identity, repeatedly pick an element of S and multiply, say on the right. The new method, suggested by work of Celler, Leedham-Greene, Murray, Niemeyer and O'Brien involves a Markov chain on n-tuples of group elements with $n > |S|$. To start, label the first $|S|$ coordinates by the generators and the remaining $n - |S|$ coordinates by the identity. At each stage, a pair of coordinates (u, v) is chosen at random and the element at u is multiplied (on the right) by the element at v or its inverse. This product is output and also replaces the uth element in the n-tuple. It is believed that the sequence of output elements "gets random" substantially faster than the classical random walk based on

Research partially supported by NATO grant CRG 950686

S. We give the first quantitative bounds for the Celler et al. algorithm. Recently, Babai [4] proved that the diameter of the graph naturally associated with the Celler et al. algorithm is at most $O(n^2)$ when $n = 2\lceil \log |G| \rceil$. This is a good indication that the chain might converge rapidly but, by itself, it is not enough to obtain any reasonable quantitative result for the convergence of the chain.

As an example, consider the symmetric group S_d with two generators, a transposition $(1, 2)$ and an d-cycle $(1, 2, \ldots, d)$. Theory and experiments developed in [15, 17] show that the classical walk based on these generators takes about $\frac{1}{8} d^3 \log d$ steps to get random. When $d = 52$ this is about 60,000 steps. Experiments reported in [9, 30] suggest that, using 10-tuples as described above, the output is random after about 190 steps.

The new algorithm is motivated by applications in computational group theory. Efficient computation with large groups often calls for a source of pseudo-random elements of the group. These are used to help find the order of the group, decompose representations and for a dozen other tasks. A good overview of the literature is in Finkelstein and Kantor [24]. See also the recent survey of L. Babai [4]. For example, the Neumann-Praeger algorithm [32] takes as input a set of $d \times d$ invertible matrices with entries in a finite field and tests if they generate a subgroup between SL_d and GL_d. In problems of interest, $d \in [30, 100]$ and the field is small (e.g., $\mathbb{Z}/2\mathbb{Z}$). The first trials of the Neuman-Praeger algorithm used the given generating set to run a classical random walk. This was run "for a while" to generate "random elements". Then, known properties of most elements of SL_d form a basis for testing. In a practical implementation, Holt and Rees [30] found that the classical random walk required a huge number of steps to get rid of obviously non-random features. They report that the new algorithm worked well.

The present paper and a companion paper [20] provide the first quantitative bounds for the convergence of the Markov chain on generating n-tuples used in the Celler et al. algorithm. Section 2 presents some background material including paths on groups and tools from Markov chain theory such as Dirichlet forms, eigenvalues and logarithmic Sobolev inequalities. Section 3 gives a careful description of the Markov chain introduced informally above. The main result of this paper, Theorem 3.2, gives a quantitative bound for the convergence of this chain to equilibrium for an arbitrary group. Theorem 3.2 is proved in Section 5. It is applied to a collection of examples in Section 7. We cite two examples here: First, for $|G| = N$ fixed and n large we show that order $C_1(N) n^2 \log n$ steps suffice for

randomness. As a second class of examples, for p-groups G of order at most p^b and with the exponent of $G/[G,G]$ equal to p^ω, we show that order $C_2(b)(np^\omega)^2(\log p)[(\log n)+(\log\log p)]$ steps suffice for randomness. The constants C_1 and C_2 are explicitly computable. These results are effective when N or b are fixed. For instance, for the Heisenberg group mod (p), $|G|=p^3$, $\omega=1$ and, for any $c>0$, $5\cdot 2^8\cdot 7^5\cdot(np)^2[(\log p)(\log\log p^n)+4c]$ steps suffice to make the walk e^{1-c} close to the uniform distribution. We also give results for the symmetric group S_d with both d and n large.

The bound in Theorem 3.2 depends on four features of the underlying group G:

(1) the minimum size $m(G)$ of a generating set,
(2) the maximum size $\overline{m}(G)$ of a minimal generating set,
(3) the number $f(k,G)$ of k-tuples that generates G,
(4) the maximum diameter $D(G)$ over all generating sets.

These features are discussed in Section 6. Bounds or exact expressions for $m,\overline{m}$ are often available. A reasonable amount is known about $f(k,G)$ thanks to P. Hall's work on abstract Möbius inversion. Less is known about $D(G)$. We show that, for p-groups of bounded class and number of generators, the diameter is essentially the exponent of $G/[G,G]$ (Theorem 6.5), independently of which minimal generating set is chosen.

Section 4 contains bounds on the least eigenvalue of the chain P. These are important for results in discrete time.

The companion paper [20] studies the same chain for some Abelian groups. See also [11, 12]. This was in fact offered as a challenge problem by David Aldous. The technique of [20] are similar, but the technical details are much easier in the Abelian case. We suggest looking at [20] before plunging into the present arguments. We also show there that for $G=\mathbb{Z}/p\mathbb{Z}$, p prime, order $n^4(\log p)^3$ steps suffice using quite different arguments. The main novel feature in the present paper is the comparison of the chain of interest with a simpler chain defined on a different state space (Proposition 5.2). This is inspired by what geometers call "rough-isometries" or "quasi-isometries" between metric spaces. Typically, a rough-isometry forgets about the local topology and preserves the large scale features of the space. For example $\mathbb{R}^d$ and $\mathbb{Z}^d$ are roughly isometric. More generally, the universal cover M of a compact manifold N is roughly isometric to the fundamental group $\pi_1(N)$. What we do here is to introduce quantitative versions of these ideas and apply them to finite state spaces.

There are other approaches to generating pseudo-random elements. Babai [2] gives a general procedure which provably works in polynomial time for general groups (see also [4]). For permutation groups, there are efficient algorithms for finding nested chains of subgroups so the subgroup algorithm [22] can be used. The Celler et al. algorithm is by far the most widely used, being implemented in both "Magma" and "Gap", two of the main packages for computer assisted manipulations of finite groups. The bounds we give for the Celler et al. algorithm are fairly good for large n. However, in cases of greatest practical interest, $|G|$ is large and n is small. Even determining the size of the state space is difficult in this case.

The algorithm we analyse is a symmetric version of the original chain proposed by Celler et al. [9]. Section 8 shows how the same analysis applies to a number of non symmetric versions of the chain including the one used by Celler et al. [9]. In the symmetric version of the algorithm, the stream of output elements has tied values. If n-tuples are used, there is chance $1/(2n)$ that outputs two apart are tied. The original non symmetric algorithm will similarly have quite correlated output. One easy fix for this problem is to multiply together larger subsets. Thus, working with n-tuples, fix $k < n$. Choose a subset of k places out of n uniformly at random, a permutation in S_k uniformly at random, a sequence of length k of ± 1 uniformly at random. Multiply the entries of the k first chosen places in the chosen order using the $\pm$ signs to indicate inverses. The result is output and also used to replace a randomly chosen one of the k entries. The techniques of the present paper can be used to analyse this algorithm for fixed k. Another easy fix for the problem of tied values is to use large n.

Acknowledgement. We thank David Aldous, Rosemary Bailey, Jordan Ellenberg, Fan Chung, David Gluck, Susan Holmes, Ron Graham, Charles Leedham-Green, Barry Mazur, Dan Rockmore, Chris Rowley, Bálint Virág and Thomas Yan for their help with this paper. In an early draft of this paper we proved that, for any finite group G and any n large enough, order $(n|G|)^{O(\overline{m})}$ steps suffice for the chain P to reach uniformity ($\overline{m} = \overline{m}(G)$ as above). Bálint Virág's remarks on the manuscript led us to an improved bound (based on the same argument) of order $(\log|G|)|G|^{2m} M^{-2}D^2n^3 \log n$ where $m = m(G), M = f(m, G)$ and D are as above. This is a serious improvement because M is often of order $|G|^m$ so that this bound is often of order $(\log|G|)D^2n^3 \log n$. In the mean time, F. Chung and R. Graham [11] proposed a simplified and more naive version of our earlier proof and showed that order $|G|^{O(\overline{m})}n^2 \log n$ steps are enough. This motivated us to tighten our argument again and led us to the present version which gives a bound of order $(\log|G|)|G|^{2m} M^{-2}D^2n^2 \log n$.

2 Background and notation

2.A. Paths on groups

We will use paths on a group G defined in a classical way as follows. For each generating set $S \subset G$ and each $g \in G$ fix a sequence $s_i = s_i(g)$, $1 \le i \le k$, of minimal length $k = |g|_S$ such that $s_i \in S \cup S^{-1}$ and

$$g = s_1 \cdots s_k \ .$$

Given a pair $(g,h) \in G \times G$, define a path $\gamma = \gamma(S,g,h)$ of minimal length $|\gamma|_S = |g^{-1}h|_S$ by writing

$$g^{-1}h = s_1 \cdots s_k$$

where $k = |g^{-1}h|_S$, $s_i = s_i(g^{-1}h) \in S \cup S^{-1}$ and translating this path by g on the left to obtain

$$\gamma : g = x_0,\ x_1 = gs_1,\ \ldots,\ x_k = gs_1 \cdots s_k = h \ .$$

For each $s \in S \cup S^{-1}$ and $g \in G$, define

$$N_S(s,g) = \#\{i \in \{1,\ldots,|g|_S\} : s_i(g) = s\} \ .$$

Let $d_S = \max_g |g|_S$ be the diameter of G with respect to S. Let

$$D = D(G) = \max\{d_S : S \text{ generates } G\} \tag{2.1}$$

be the maximum diameter of G. We will need the following elementary lemma.

Lemma 2.1 *For any fixed generating set S, $s \in S \cup S^{-1}$ and $z, w = zs$, we have*

$$\#\{(g,h) \in G \times G : \gamma(S,g,h) \ni (z,w)\} = \sum_{u \in G} N_S(s,u) \le |G| d_S \le |G| D \ .$$

In particular,

$$\sum_{\substack{g,h: \\ \gamma(S,g,h) \ni (z,w)}} |\gamma(S,g,h)| \le |G| D^2 \ .$$

Proof. (cf. [16], p. 702) Observe that

$$\begin{aligned}\#\{(g,h) \in G \times G : \gamma(S,g,h) \ni (z,w)\} \\ = \#\{(g,u) \in G \times G : \exists\, i \text{ such that } s_i(u) = s, \\ z = g s_1(u) \cdots s_{i-1}(u)\} \ .\end{aligned}$$

The natural bijection between the two sets above is given by $(g,h) \to (g, g^{-1}h)$. For each fixed u, there are exactly $N(s,u)$ elements $g \in G$ such that (g,u) belongs to

$$\{(h,u) \in G \times G : \exists\, i \text{ such that } s_i(u) = s, z = h s_1(u) \cdots s_{i-1}(u)\} \ .$$

Hence

$$\#\{(g,h) \in G \times G : \gamma(S,g,h) \ni (z,w)\} = \sum_{u \in G} N_S(s,u) \ .$$

This proves the lemma since, clearly, $N_S(s,u) \le |u|_S \le d_S \le D$.

2.B. Markov chains

Let P be a reversible Markov chain on a finite state space $\mathscr{X}$ with reversible measure $\pi > 0$ so that $P(x,y)\pi(x) = P(y,x)\pi(y)$. Set

$$\mathrm{Var}_\pi(f) = \frac{1}{2} \sum_{x,y \in \mathscr{X}} |f(x) - f(y)|^2 \pi(x)\pi(y) \ , \tag{2.2}$$

$$\mathscr{E}_P(f,f) = \frac{1}{2} \sum_{x,y \in \mathscr{X}} |f(x) - f(y)|^2 P(x,y)\pi(x) \tag{2.3}$$

and

$$\mathscr{L}_\pi(f) = \sum_x |f(x)|^2 \log\left(\frac{|f(x)|}{\|f\|_2}\right)^2 \pi(x) \tag{2.4}$$

where $\|f\|_2 = \left(\sum_x |f(x)|^2 \pi(x)\right)^{1/2}$. The subscripts will be dropped whenever no confusion could possibly arise. For the iterated kernel of P, we will use the notation

$$P_x^\ell(y) = P^\ell(x,y) = \sum_z P^{\ell-1}(x,z) P(z,y) \ .$$

To measure distances between probability distributions, we will use the total variation distance

$$\|\pi - \mu\|_{\mathrm{TV}} = \max_{A \subset \mathscr{X}} |\pi(A) - \mu(A)| = \frac{1}{2} \sum_{x \in \mathscr{X}} |\pi(x) - \mu(x)| \ .$$

Denote by

$$\beta_0(P) = 1 \geq \beta_1(P) \geq \cdots \geq \beta_{|\mathscr{X}|-1}(P) = \beta_{\min}(P) \geq -1$$

the eigenvalues of the chain P and let

$$\beta(P) = \max\{\beta_1(P), -\beta_{\min}(P)\} \ .$$

Using (2.2) (2.3), the second largest eigenvalue $\beta_1(P)$ can be expressed as

$$1 - \beta_1(P) = \min\left\{\frac{\mathscr{E}_P(f,f)}{\mathrm{Var}_\pi(f)} : f \neq 0\right\} \ .$$

A classical and easy bound (e.g., [23, 34]) on variation distance is given by

$$\|P_x^\ell - \pi\|_{\mathrm{TV}} \leq \frac{1}{2\sqrt{\pi(x)}} \ \beta(P)^\ell \ . \tag{2.5}$$

The log-Sobolev constant $\alpha(P)$ of a reversible Markov chain (P, π) is defined as the largest non-negative number α such that

$$\alpha \mathscr{L}_\pi(f) \leq \mathscr{E}_P(f,f) \tag{2.6}$$

for any function f. We will use $\alpha(P)$ to prove mixing rates that improve upon those obtained through (2.5). More precisely, we will use the following

Theorem 2.2 *Let (P, π) be a (finite) reversible Markov chain. Then, for all $c > 0$,*

$$\|P_x^\ell - \pi\|_{\mathrm{TV}} \leq 2e^{-c} \quad \text{for } \ell \geq 1 + \frac{c}{1-\beta} + \frac{1}{4\alpha} \log\log \frac{1}{\pi(x)} \ .$$

We also consider the continuous time semigroup

$$H_t = \mathrm{e}^{-t(I-P)} = \mathrm{e}^{-t} \sum_0^\infty \frac{P^n}{n!} \ . \tag{2.7}$$

The semigroup H_t has the advantage of avoiding parity problems. In what follows the results stated for H_t could be replaced by similar

bounds on the discrete time chain $\widetilde{P}^\ell$ where $\widetilde{P} = \frac{1}{2}(I + P)$. For H_t, we have

Theorem 2.3 *Let (P, π) be a (finite) reversible Markov chain. Then, for all $c > 0$,*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \leq \mathrm{e}^{1-c} \quad \text{for } t = \frac{c}{1-\beta_1} + \frac{1}{4\alpha}\log\log\frac{1}{\pi(x)} \quad .$$

The difference between Theorem 2.3 and Theorem 2.2 is that no bound on the least eigenvalue is required in Theorem 2.3. We refer the reader to [18] for the proofs of Theorem 2.2, Theorem 2.3, and for a discussion of the use of log-Sobolev inequalities for finite Markov chains.

Remark. Although Theorem 2.2 and 2.3 are stated for total variation, their conclusions hold in fact in $\ell^2(\pi)$ (or chi-square) distance. That is, $\|P_x^\ell - \pi\|_{\mathrm{TV}}$ and $\|H_t^x - \pi\|_{\mathrm{TV}}$ can be replaced by $\|[P_x^\ell/\pi] - 1\|_2$ and $\|[H_t^x/\pi] - 1\|_2$ where $\|\cdot\|_2$ refers to the norm in $\ell^2(\pi)$. It follows that similar bounds also holds for the maximal relative errors $\sup_{x,y}\left|\frac{P^\ell(x,y)}{\pi(y)} - 1\right|$ and $\sup_{x,y}\left|\frac{H_t(x,y)}{\pi(y)} - 1\right|$ which are easily bounded in terms of the ℓ^2 distance. See [18], Section 2.D and Theorem 3.7, Corollary 3.8 of that paper. This remark applies to all the convergence results stated in the present paper including the results of Section 8 which deal with nonsymmetric chains.

3 The two Markov chains

Let us introduce some notation. Fix a finite group G and set $\mathscr{X} = G^n$. For x, y in $\mathscr{X} \times \mathscr{X}$, write

$$x \sim y \quad \text{if } x \text{ and } y \text{ differ exactly in one coordinate} \quad ,$$

and write

$$x \approx y \text{ if } \begin{cases} x \text{ and } y \text{ differ exactly in one coordinate, say } x_i \neq y_i, \\ \text{and there exists } j \neq i \text{ such that } y_i = x_i x_j^{\pm 1} \quad . \end{cases}$$

If $x \approx y$ with $x_i \neq y_i$, let

$$N(x,y) = \begin{cases} \text{the number of } j \text{ such that } x_i^{-1}y_i = x_j^{\pm 1} & \text{if } x_i^{-1}y_i \neq (x_i^{-1}y_j)^{-1} \\ \text{twice this number if } x_i^{-1}y_i = (x_i^{-1}y_i)^{-1} \quad . \end{cases}$$

Finally, let $N(x)$ be the number of coordinates equal to the identity in x.

With this notation the chain defined informally in the introduction is given by the kernel

$$P(x,y) = \begin{cases} 0 & \text{if } x \not\approx y \text{ and } x \neq y \\ \frac{N(x,y)}{2n(n-1)} & \text{if } x \approx y \\ \frac{N(x)}{n} & \text{if } x = y \end{cases} . \tag{3.1}$$

The chain P is not irreducible on G^n. Let $S \subset G$ a set of generators. Say S is minimal if no smaller subset of S generates G. Define $\overline{m}(G)$ to be the maximum size of a minimal generating set. Define $m(G)$ to be the minimum size of a generating set. Note that often $m(G) < \overline{m}(G)$. For example, for $\mathbb{Z}/pq\mathbb{Z}$, with p, q primes, $m(G) = 1$, $\overline{m}(G) = 2$. We will constantly use the notation

$$m = m(G); \quad \overline{m} = \overline{m}(G) \ .$$

The numbers $\overline{m}(G), m(G)$, appear in the following slight generalization of a result of Celler et al. [9] which gives a useful condition for the walk at (3.1) to be irreducible on the set of generating sequences.

Lemma 3.1 *Let G be a finite group and $n \geq m(G) + \overline{m}(G)$. Then, the chain P at* (3.1) *gives an irreducible symmetric Markov chain on the set of n-tuples $(x_1, \ldots, x_n)$ which generate G.*

Proof. Fix a generating sequence $(y_1, \ldots, y_m)$ with $m = m(G)$. Any n-tuple $(x_1, \ldots, x_n)$ which generates G can be brought to $(y_1, y_2, \ldots, y_m, \mathrm{id}, \ldots, \mathrm{id})$. Indeed, a subsequence of length at most $\overline{m}(G)$ in $(x_1, \ldots, x_n)$ generates and so one can produce $y_1, \ldots, y_m$ in the complementary positions to this generating sequence. Using $y_1, \ldots, y_m$, we can set all the remaining positions to the identity. Then, it is easy to order the y_i as we wish. This shows that the Markov chain (3.1) is irreducible. Since it is symmetric and has some holding, it is ergodic.

Remark. For some classes of groups, the conclusion of Lemma 3.1 holds for all $n \geq m(G) + 1$. Diaconis and Graham [14] show this for Abelian groups. They also show that the chain need not be connected if $n = m(G)$.

We denote by $\mathscr{X} \subset \mathscr{Z}$ the set of all generating n-tuples. We assume throughout that $n \geq m(G) + \overline{m}(G)$ and consider $\mathscr{X}$ as the state space of the chain P. Thus, P is irreducible, symmetric, aperiodic on $\mathscr{X}$ and its stationary measure is $\pi(x) = |\mathscr{X}|^{-1}$. The following theorem de-

scribes our quantitative bound on the convergence of the continuous time process $H_t = e^{t(I-P)}$ associated with the chain P at (3.1) by formula (2.7). The proof is given in Section 5. It combines Theorem 2.3 and the eigenvalue and log-Sobolev estimates of Corollary 5.3.

Theorem 3.2 *For any group G and all $n \geq 2m(G) + \overline{m}(G)$ and all $c > 0$, the semigroup $H_t = e^{-t(I-P)}$ associated to the chain P at* (3.1) *on generating n-tuples satisfies*

$$\|H_t^x - \pi\|_{TV} \leq e^{1-c}$$

for

$$t = \frac{20(1+2m)^2 \binom{n}{m}\binom{n-m}{m}|G|^{2m+1}D^2n^2}{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}M^2(|G|-2)} \times [(\log(|G|-1))(\log\log|G|^n) + 4c] \ .$$

Here $M = M(G)$ is the number of distinct generating m-tuples and $D = D(G)$ is the maximum diameter of G.

For fixed G and large n, our main result simplifies to:

Theorem 3.3 *For any fixed group G and all $n \geq 2m(G) + \overline{m}(G)$, the chain P at* (3.1) *on generating n-tuples satisfies*

$$\|P_x^\ell - \pi\|_{TV} \leq 2e^{-c} \quad \text{for} \quad \ell \geq An^2[\log n + c], \quad c > 0 \ .$$

Here A depends only on G.

In order to apply Theorem 3.2 to classes of groups where the size of G is allowed to grow, it is crucial to have estimates on the four group theoretical quantities $m(G)$, $\overline{m}(G)$, $D(G)$ and $M(G)$. These quantities are studied in Section 6 which also gives pointers to the literature. Specific examples are treated in Section 7. We want to emphasize that the expression for t in Theorem 3.2 does not depend too badly on the size of G. To illustrate this point, we state two special cases that follow from Theorem 3.2 using results from Section 6:

Corollary 3.4 *Let G be a p-group of order p^b. Then, $m = \overline{m} \leq b$. For all $n \geq 3b$ and $c > 0$, the semigroup $H_t = e^{-t(I-P)}$ associated to the chain P at* (3.1) *on generating n-tuples satisfies*

$$\|H_t^x - \pi\|_{TV} \leq e^{1-c}$$

for

$$t = 320(1+2b)^{2+b} n^2 D^2 \left[(\log p^b)\left(\log\log p^{bn}\right) + 4c \right] \ .$$

Further, for such groups, $\frac{1}{4}p^\omega \le D \le (2b)^{b+1}p^\omega$ *where* p^ω *is the exponent of* $G/[G,G]$.

Corollary 3.5 *Let* G *be the symmetric group* S_d *on* d *letters. Then,* $m = 2$, $\overline{m} \le 2d$. *For* $n = 3d$ *and* $c > 0$, *the semigroup* $H_t = \mathrm{e}^{-t(I-P)}$ *associated to the chain* P *at* (3.1) *on generating* $3d$*-tuples satisfies*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \le \mathrm{e}^{1-c}$$

for $t = Ad^2D^2[d(\log d)^2 + c]$. *Here* D *is the maximum diameter of* S_d *and* A *does not depend on* d.

The proofs of these three corollaries of Theorem 3.2 are in Section 7.

The quantitative study of P proceeds by comparison with the chain Q on $\mathscr{X}$ defined by

$$Q(x,y) = \begin{cases} 0 & \text{if } x \not\sim y \text{ and } x \neq y \\ \frac{1}{n|G|} & \text{if } x \sim y \\ \frac{1}{|G|} & \text{if } x = y \end{cases} \ . \tag{3.2}$$

This chain picks a coordinate uniformly at random and multiplies this coordinate by a uniformly chosen element of G. Its stationary measure is $\mu(x) = |\mathscr{X}|^{-1} = |G|^{-n}$. It is a product chain with second largest eigenvalue

$$\beta_1(Q) = 1 - \tfrac{1}{n} \ .$$

Its log-Sobolev constant can be computed exactly using Lemma 3.2 and Corollary A.5 in [18]. It is given by

$$\alpha(Q) = \frac{(|G|-2)}{n|G|\log(|G|-1)} \ .$$

The comparison of the chains P and Q is treated in Section 5, Proposition 5.2 and Corollary 5.3.

4 The lowest eigenvalue

To bound the discrete time chain P^ℓ instead of the continuous time chain $H_t = \mathrm{e}^{-t(I-P)}$, we need a bound on the least eigenvalue of P. We will use a slight variation on Proposition 2 of [23], page 40.

Suppose (K, π) is a reversible Markov chain on the finite state space $\mathscr{X}$. For each $x \in \mathscr{X}$, let Σ_x be some fixed set of cycles of odd length beginning and ending at x. Let $\Sigma = \cup_x \Sigma_x$. For each cycle $\sigma \in \Sigma$, let $|\sigma|$ be its length. Finally, let θ be a non-negative function defined on Σ and such that, for each $x \in \mathscr{X}$,

$$\sum_{\sigma \in \Sigma_x} \theta(\sigma) = \pi(x) \ .$$

Such a function θ is called a flow on odd cycles (we will later encounter other kinds of flows for comparison between two chains). Then, the argument in [23], page 40, easily gives

Lemma 4.1 *With the above notation, for any finite reversible Markov chain and any flow θ on odd cycles,*

$$\beta_{\min}(K) \geq -1 + \frac{2}{I(\theta)}$$

where

$$I(\theta) = \max_{\substack{(x,y) \\ K(x,y)>0}} \left(\frac{1}{K(x,y)\pi(x)} \sum_{\substack{\sigma \in \Sigma \\ \sigma \ni (x,y)}} r(\sigma, (x,y)) |\sigma| \theta(\sigma) \right) \ .$$

Here, $r(\sigma, (x,y))$ is the number of times the edge (x,y) is used in σ (one can always assume that $r(\sigma, (x,y)) \leq 2$ and, in our applications, it will always be at most 1).

Lemma 4.1 will be used to give three lower bounds on the smallest eigenvalue.

Proposition 4.2 *The chain (P, π) at* (3.1) *has its least eigenvalue bounded by*

$$\beta_{\min}(P) \geq -1 + \frac{n - \overline{m}}{n(n-1)|G|(2D(G)+1)} \ .$$

Here $\overline{m} = \overline{m}(G)$ and $n \geq m(G) + \overline{m}(G)$.

Proof. This is a slight improvement on Proposition 3.3 of [20]. We give the proof for completeness. We use Lemma 4.1 and the following flow θ on odd cycles: If one of the coordinates of x is the identity, set

$$\Sigma_x = \{\sigma_x\} \quad \text{with} \quad \sigma_x = (x, x) \ .$$

If none of the coordinates of x is the identity, fix a generating subset S occupying $\overline{m}$ coordinates $\{i_1, \ldots, i_{\overline{m}}\}$ of x and pick a coordinate, say x_i, not in this subset. Write x_i as a word using elements in S. This describes a path $\gamma_{x,i}$ from x to x^i where x^i is the n-tuple with ith coordinate the identity and all other coordinates equal to those of x. Set

$$\Sigma_x = \{\sigma_{x,i} : i \notin \{i_1, \ldots, i_{\overline{m}}\}\}$$

where $\sigma_{x,i}$ is the cycle that goes from x to x^i along $\gamma_{x,i}$, holds at x^i for one step and goes back to x. Now, for any cycle σ, set

$$\theta(\sigma) = \begin{cases} \frac{1}{|\Sigma_x|}\pi(x) & \text{if } \sigma \in \Sigma_x \\ 0 & \text{otherwise} \end{cases} .$$

Observe that $|\Sigma_x| = n - \overline{m}$ when none of the coordinates of x is the identity. Then, we have to bound

$$I(\theta) = \max_{\substack{(x,y) \\ P(x,y)>0}} \frac{1}{P(x,y)\pi(x)} \sum_{\sigma \ni (x,y)} |\sigma|\theta(\sigma) .$$

First, examine the case where $x = y$ contains more than one coordinate equal to the identity. Then, the quantity we have to bound becomes $n/N(x) \leq n/2$.

Second, if $x = y$ contains exactly one coordinate, say x_i, equal to the identity. Then, we have to bound

$$n\left(1 + \sum_{g \in G} \frac{2|g|_x + 1}{n - \overline{m}}\right) \leq \frac{n(n - \overline{m} + |G|(2D(G) + 1))}{n - \overline{m}} .$$

Here $|g|_x$ denotes the length of g in some generating set which depends on x.

Finally, if x, y differ at exactly one coordinate, say $x_i \neq y_i$, then we have to bound

$$\frac{2n(n-1)}{N(x,y)} \sum_{g \in G} \frac{2|g|_x + 1}{n - \overline{m}} \leq \frac{2n(n-1)|G|(2D(G) + 1)}{n - \overline{m}} .$$

Hence

$$I(\theta) \leq \frac{2n(n-1)|G|(2D(G) + 1)}{n - \overline{m}} .$$

This proves Proposition 4.2.

Proposition 4.3 *Assume that any generating set of G contains an element of odd order at most T. Then the chain P at* (3.1) *with $n \geq m(G) + \overline{m}(G)$ has its least eigenvalue bounded by*

$$\beta_{\min}(P) \geq -1 + \frac{2}{nT^2} \quad .$$

Proof. Use the same technique as in the preceding proof. For any $x \in \mathscr{X}$, there exists j such that x_j has order $t \leq T$, t odd. For any $k \neq j$ define $\sigma_{x,k}$ to be the cycle of odd length from x to x obtained by changing the kth entry to $x_k x_j,\ x_k x_j^2, \ldots, x_k x_j^{t-1},\ x_k$. Set $\Sigma_x = \{\sigma_{x,k} : k \neq j\}$ and define a flow on odd cycles by setting $\theta(\sigma) = 1/[|\mathscr{X}|(n-1)]$ if $\sigma \in \cup_x \Sigma_x$ and $\theta(\sigma) = 0$ otherwise. Then,

$$I(\theta) = \max_{\substack{(x,y) \\ P(x,y)>0}} \frac{1}{P(x,y)\pi(x)} \sum_{\substack{\sigma \\ \sigma \ni (x,y)}} |\sigma| \theta(\sigma) \leq nT^2 \quad .$$

This proves the desired inequality. Observe that Proposition 4.3 applies to any group of odd order. Actually, we do not have other examples.

Given a subset S of G, let $\ell(S)$ be the shortest length of a cycle $s_1 \cdots s_{\ell(S)} = \mathrm{id}$ of odd length with $s_i \in S \cup S^{-1}$ and $\ell(S) = +\infty$ if there is no such cycle.

Proposition 4.4 *Let $L = \max_S \ell(S)$ where the maximum is taken over all generating sets of size $\overline{m}(G) = \overline{m}$. The chain P at* (3.1) *with $n \geq m(G) + \overline{m}(G)$ has its least eigenvalue bounded by*

$$\beta_{\min}(P) \geq -1 + \frac{2(n-\overline{m})}{n(n-1)L^2} \quad .$$

Proof. We can assume $L < \infty$. For any $x \in \mathscr{X}$, there exists J of size M such that $S = \{x_j : j \in J\}$ generates G. Let $\ell = \ell(S)$ and $s_1 \cdots s_\ell = \mathrm{id}$ be a cycle of odd length ℓ with $s_i \in S \cup S^{-1}$. For any $k \notin J$ define $\sigma_{x,k}$ to be the cycle of odd length from x to x obtained by changing the kth entry to $x_k s_1,\ x_k s_1 s_2, \ldots, x_k s_1 \cdots s_{\ell-1},\ x_k$. Set $\Sigma_x = \{\sigma_{x,k} : k \neq j\}$ and define a flow on odd cycles by setting $\theta(\sigma) = 1/[|\mathscr{X}|(n-\overline{m})]$ if $\sigma \in \cup_x \Sigma_x$ and $\theta(\sigma) = 0$ otherwise. Then,

$$I(\theta) = \max_{\substack{(x,y) \\ P(x,y)>0}} \frac{1}{P(x,y)\pi(x)} \sum_{\substack{\sigma \\ \sigma \ni (x,y)}} |\sigma| \theta(\sigma) \leq \frac{n(n-1)L^2}{n-\overline{m}} \quad .$$

With Lemma 4.1, this gives the desired result. This result seems difficult to apply in practice.

5 Comparison

To complete the proof of Theorem 3.2 we need some notation. For any sequence $S = \{s_1, \ldots, s_k\}$ of size k with $s_i \in G$, any n-tuple $x = (x_1, \ldots, x_n) \in \mathscr{Z}$ and any ordered k-tuple $I = (i_1, \ldots, i_k)$ with $1 \le i_1 < \cdots < i_k \le n$, let x_S^I be the n-tuple with ith coordinate $[x_S^I]_i$ given by

$$[x_S^I]_i = \begin{cases} x_i & \text{if } i \notin I \\ s_j & \text{if } i \text{ is the } j\text{th element of } I \end{cases} . \tag{5.1}$$

Given a function f on $\mathscr{X}$ we set

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in \mathscr{X} \\ \frac{1}{M\binom{n}{m}} \sum_{S,I} f(x_S^I) & \text{if } x \in \mathscr{Z} \setminus \mathscr{X} \end{cases}$$

where the sum runs over all **generating** sequences S of length $m = m(G)$ and all m-subsets $I \subset \{1, \ldots, n\}$, and $M = M(G)$ is the number of distinct generating m-tuples. The following lemma is easy.

Lemma 5.1 *The chains* (P, π) *and* (Q, μ) *defined by* (3.1), (3.2) *satisfy*

$$\mathrm{Var}_\pi(f) \le \frac{|\mathscr{Z}|}{|\mathscr{X}|} \mathrm{Var}_\mu(\tilde{f}) , \tag{5.2}$$

$$\mathscr{L}_\pi(f) \le \frac{|\mathscr{Z}|}{|\mathscr{X}|} \mathscr{L}_\mu(\tilde{f}) . \tag{5.3}$$

Proof. See Proposition 2.3 of [20]. Actually, any extension of f would do the job here.

We now reach the crucial part of the comparison argument. The Dirichlet forms $\mathscr{E}_Q(\tilde{f}, \tilde{f})$ and $\mathscr{E}_P(f, f)$ must be compared.

Proposition 5.2 *For any group* G *with* $m(G) = m$, $\overline{m}(G) = \overline{m}$ *and* $n \ge 2m + \overline{m}$, *the chains* P *and* Q *defined at* (3.1), (3.2) *satisfy*

$$\mathscr{E}_Q(\tilde{f}, \tilde{f}) \le \frac{20(1+2m)^2 \binom{n}{m}\binom{n-m}{m}|G|^{2m}}{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m} M^2 D^2 n} \frac{|\mathscr{X}|}{|\mathscr{Z}|} \mathscr{E}_P(f, f)$$

for all $f : \mathscr{X} \to \mathbb{R}$. *Here,* M *is the number of distinct generating* m*-tuples.*

Corollary 5.3 *For any group* G *with* $m(G) = m$, $\overline{m}(G) = \overline{m}$ *and* $n \geq 2m + \overline{m}$, *the chain* P *defined at* (3.1) *satisfies*

$$\beta_1(P) \leq 1 - \frac{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m} M^2}{20(1+2m)^2 \binom{n}{m}\binom{n-m}{m} |G|^{2m} D^2 n^2}$$

and

$$\alpha(P) \geq \frac{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m} M^2 (|G|-2)}{20(1+2m)^2 \binom{n}{m}\binom{n-m}{m} |G|^{2m+1} \log(|G|-1) D^2 n^2} \quad .$$

Proof. We start by writing

$$\begin{aligned} \mathscr{E}_Q(\tilde{f}, \tilde{f}) &= \frac{1}{2n|G|^{n+1}} \left(\sum_{\substack{x,y \in \mathscr{X} \\ x \sim y}} |f(x) - f(y)|^2 + 2 \sum_{\substack{x \in \mathscr{Z} \setminus \mathscr{X}, y \in \mathscr{X} \\ x \sim y}} |\tilde{f}(x) - f(y)|^2 . \right. \\ &\qquad \left. + \sum_{\substack{x,y \in \mathscr{Z} \setminus \mathscr{X} \\ x \sim y}} |\tilde{f}(x) - \tilde{f}(y)|^2 \right) \\ &= \frac{1}{2n|G|^{n+1}} (R_1 + 2R_2 + R_3) \quad . \end{aligned} \tag{5.4}$$

We are going to bound R_1, R_2 and R_3 in terms of

$$R_{\approx} = \sum_{\substack{z,w \in \mathscr{X} \\ z \approx w}} |f(z) - f(w)|^2 \quad .$$

To this end, for each $x \in \mathscr{X}$, we need to pick an ordered $\overline{m}$-tuple

$$I(x) = (i_1(x), \ldots, i_{\overline{m}}(x)), \quad 1 \leq i_\alpha(x) \leq n$$

such that

$$S(x) = (x_{i_1(x)}, \ldots, x_{i_{\overline{m}}(x)}) \in G^{\overline{m}}$$

generates G. Furthermore, we do this in a "global" way. Namely, fix any total order on the set of pairs (I, S) where I runs over all ordered $\overline{m}$-tuples with entries in $\{1, \ldots, n\}$ (i.e., $\overline{m}$-subsets of $\{1, \ldots, n\}$) and S runs over all generating $\overline{m}$-tuples in $G^{\overline{m}}$. Given $x \in \mathscr{X}$, define

$$I(x) = (i_1(x), \ldots, i_{\overline{m}}(x)), \quad S(x) = (x_{i_1(x)}, \ldots, x_{i_{\overline{m}}(x)}) \tag{5.5}$$

to be the smallest such pair built on x. The "global" property referred to above and which is an easy consequence of this construction is the following:

$$\begin{cases} \text{If } x, y \in \mathscr{X}, \text{ and } K = (\ell_1, \ldots, \ell_k) \text{ are such that} \\ x_i = y_i \text{ for } i \notin K \text{ and also } K \cap I(x) = K \cap I(y) = \emptyset, \\ \text{then } \ I(x) = I(y) \ \text{ and } \ S(x) = S(y) \ . \end{cases} \tag{5.6}$$

The heart of our argument is contained in the following technical lemma which bounds expressions such as

$$\sum_{x,I,S,K,W} |f(x_S^I) - f(x_W^K)|^2$$

in terms of $R_{\approx}$. Here, typically, x is in $\mathscr{X}$ (or $\mathscr{Z}$), I, K are ordered tuples with distinct entries in $\{1, \ldots, n\}$ and S, W are generating tuples with entries in G.

Lemma 5.4 *We have the following bounds where I, K run over all ordered m-tuples with distinct entries in $\{1, \ldots, n\}$, S, W run over all generating m-tuples in G^m, $j \in \{1, \ldots, n\}$ and $g \in G$.*

$$\sum_{x \in \mathscr{Z} \setminus \mathscr{X}} \sum_{\substack{I,S,j,g \\ j \notin I}} |f(x_S^I) - f([x_g^j]_S^I)|^2 \leq \binom{n-1}{m} |G|^{m+1} D^2 \, R_{\approx} \ . \tag{5.7}$$

$$\sum_{x \in \mathscr{Z}} \sum_{\substack{I,S,j,g : x_j^g \in \mathscr{X} \\ j \notin I : I \cap I(x_g^j) = \emptyset}} |f(x_S^I) - f(x_g^j)|^2 \leq (1+m)^2 \binom{n-1}{m} |G|^{m+1} D^2 \, R_{\approx} \ . \tag{5.8}$$

$$\sum_{x \in \mathscr{Z}} \sum_{\substack{I,S,K,W \\ I \cap K = \emptyset}} |f(x_S^I) - f(x_W^K)|^2 \leq 4m \binom{n-1}{m-1} \binom{n-m}{m} |G|^{2m} D^2 \, R_{\approx} \ . \tag{5.9}$$

$$\sum_{x \in \mathscr{X}} \sum_{\substack{I,S \\ I \cap I(x) = \emptyset}} |f(x_S^I) - f(x)|^2 \leq m \binom{n-1}{m-1} |G|^m D^2 \, R_{\approx} \ . \tag{5.10}$$

Proof. The proof starts with the basic idea of the comparison machinery, namely, constructing paths. Fix x, three ordered tuples I, K, K' (possibly empty) with distinct entries in $\{1, \ldots, n\}$ (i.e., we assume

that I, K, K' are disjoint) and cardinalities $|I|, |K|, |K'|$. Fix also three tuples S, W, W' (possibly empty) with entries in G and cardinalities $|I|$, $|K|, |K'|$. We will be only interested in cases where the following hypotheses are satisfied:

$$\begin{cases} (1) & |I| = m \text{ and } S \text{ is a generating } m\text{-tuple.} \\ (2) & \text{either } |K| = m \text{ and } W \text{ is a generating } m\text{-tuple} \\ & \text{or } K = W = \emptyset \text{ and } x^{K'}_{W'} \in \mathscr{X}, I(x^{K'}_{W'}) \cap I = \emptyset \quad . \end{cases}$$

In our application, K' will always be either empty or a singleton $K' = \{j\}$.

Under these hypotheses, we construct a path

$$\gamma(x, I, S, K, W, K', W') \text{ from } x^I_S \text{ to } [x^K_W]^{K'}_{W'}$$

as follows. Starting at x^I_S we use the generating set S and the group-paths $\gamma(S, h, h')$, $h, h' \in G$, of Section 2 to set the entries of x^I_S at K to the desired values given by W unless $I = K, S = W$ in which case there is nothing to do. We always proceed from left to right to reach $[x^I_S]^K_W$.

Then, if W is generating, we use the entries at W (i.e., the generating set W) to set the entries of $[x^I_S]^K_W$ at I to their desired value x_i, $i \in I$ (proceeding from left to right and using the group-paths $\gamma(W, h, h')$, $h, h' \in G$). Thus, we reach x^K_W and again, if $I = K, S = W$, we skip this phase. Now, we set the entries at K' to the desired values given by W' (using W and proceeding from left to right).

If W is not generating, then our hypotheses imply that $K = W = \emptyset$. In this case, we use the generating set S contained in x^I_S and the group-paths $\gamma(S, h, h')$, $h, h' \in G$, to set the entries at K' to their desired values given by W'. We thus reach $[x^I_S]^{K'}_{W'}$. By hypothesis, $K' \cap I = \emptyset$, $x^{K'}_{W'} \in \mathscr{X}$ and $I(x^{K'}_{W'}) \cap I = \emptyset$. By definition, the entries at $I(x^{K'}_{W'})$ form a generating set and we use them to set the entries at I to their final desired values x_i, $i \in I$. The paths $\gamma(x, I, S, K, W, K', W')$ all have the following properties. Any edge (z, z') along $\gamma(x, I, S, K, W, K', W')$ satisfies $z \approx z'$ and the length of these paths is bounded by

$$|\gamma(x, I, S, K, W, K', W')| \leq \begin{cases} |K'|D & \text{if } I = K, \ S = W \\ (|K| + |I| + |K'|)D & \text{otherwise} \quad . \end{cases} \tag{5.11}$$

To prove the inequalities of Lemma 5.4, we start by bounding the differences

$$|f(x^I_S) - f([x^K_W]^{K'}_{W'})|^2$$

appearing on the left hand-side of each of these inequalities by

$$|f(x_S^I) - f([x_W^K]_{W'}^{K'})|^2 \le |\gamma| \sum_{(z,z')\in\gamma} |f(z) - f(z')|^2 \ ,$$

where $\gamma = \gamma(x, I, S, K, W, K', W')$ and $|\gamma|$ is the length of this path. This simply uses a telescoping sum and Cauchy-Schwarz. We now proceed case by case.

Proof of 5.7. In this case, $I = K, S = W$, $K' = \{j\}$, $W' = \{g\}$ and the paths introduced above have length at most D by (5.11). We write $\gamma(x, I, S, j, g)$ for $\gamma(x, I, S, I, S, \{j\}, \{g\})$. Thus

$$|f(x_S^I) - f([x_S^I]_g^j)|^2 \le D \sum_{(z,z')\in\gamma(x,I,S,j,g)} |f(z) - f(z')|^2 \ ,$$

and

$$\sum_{x\in\mathscr{X}\setminus\mathscr{X}} \sum_{\substack{I,S,j,g \\ j\notin I}} |f(x_S^I) - f([x_g^j]_S^I)|^2 \le D \sum_{\substack{(z,z') \\ z\approx z'}} \sum_{\substack{x,I,S,j,g \\ j\notin I;\gamma(x,I,S,j,g)\ni(z,z')}} |f(z) - f(z')|^2 \ .$$

Given (z, z') with $z \approx z'$ we have to count how many (x, I, S, j, g) there are such that $j \notin I$ and $\gamma(x, I, S, j, g) \ni (z, z')$. Since $z \approx z'$, these two n-tuples differ exactly at one entry and, in the present case, it has to be the jth. Thus, we know j. Now, we pick I among the $\binom{n-1}{m}$ possible choices (recall that $j \notin I$). Knowing I, we can find S just by scanning z (because z differs from x_S^I only at j). By the same token we find all the entries of x outside $I \cup \{j\}$. Further, using the notation of Section 2 concerning group-paths, x_j and g must satisfy

$$(z_j, z_j') \in \gamma(S, x_j, g) \ .$$

By Lemma 2.1, there are at most $|G|d_S \le |G|D$ possible pairs (x_j, g) having this property. Finally, we obtain

$$\sum_{x\in\mathscr{X}\setminus\mathscr{X}} \sum_{\substack{I,S,j,g \\ j\notin I}} |f(x_S^I) - f([x_g^j]_S^I)|^2 \le \binom{n-1}{m} |G|^{m+1} D^2 R_{\approx}$$

which is (5.7).

Proof of 5.8. Here, $K = W = \emptyset$, $K' = \{j\}$, $W' = \{g\}$ and we set

$$\gamma(x, I, S, \emptyset, \emptyset, \{j\}, \{g\}) = \gamma(x, I, S, j, g)$$

(which has a different meaning than in the proof of (5.7)). These paths have length at most $(|I|+1)D = (m+1)D$ by (5.11). We obtain

$$\sum_{x\in\mathscr{Z}} \sum_{\substack{I,S,j,g:x_j^g\in\mathscr{X}\\ j\notin I; I\cap I(x_g^j)=\emptyset}} |f(x_S^I) - f(x_g^j)|^2 \le (1+m)D \sum_{\substack{(z,z')\\ z\approx z'}} \sum_{\substack{x,I,S,j,g:x_j^g\in\mathscr{X}\\ j\notin I; I\cap I(x_g^j)=\emptyset\\ \gamma(x,I,S,j,g)\ni(z,z')}} |f(z)-f(z')|^2 \ .$$

Fix an edge (z,z') between points that differ at one entry only, say the kth. Observe that either $k=j\notin I$ or $j\neq k\in I$ and consider each of these cases separately.

If $k=j$, we pick I among the the $\binom{n-1}{m}$ possible choices with $j\notin I$. Scanning z, we can now find S and all the entries of x outside $I\cup\{j\}$. Since we know that $I\cap I(x_g^j)=\emptyset$, we can also find $I(x_g^j)$ and $W = S(x_g^j)$ (here we are using property (5.6)). As in the proof of (5.7), the number of possible choice for (x_j, g) is bounded by $|G|D$ because x_j and g must satisfy

$$(z_j, z_j') \in \gamma(W, x_j, g) \ .$$

Thus, the case $k=j$ will contribute at most a factor of $\binom{n-1}{m}|G|^{m+1}D$ to our sum.

If $k\in I = \{i_1,\dots,i_m)$, we have to pick the remaining entries of I among the $\binom{n-1}{m-1}$ possible choices. We also have to pick j among the remaining $n-m$ entries. Let α be such that $k=i_\alpha$. By scanning z, we now easily find g, the entries $s_\beta\in S$ with $\beta>\alpha$, the entries x_{i_β} with $\beta<\alpha$, and the entries of x outside $I\cup\{j\}$. Further, the pair (s_α, x_k) must satisfy (in the notation of Section 2)

$$(z_k, z_k') \in \gamma(S, s_\alpha, x_k) \ .$$

By Lemma 2.1, there are at most $|G|D$ such pairs. Thus, the case $k\in I$ will contribute at most a factor of $(n-m)\binom{n-1}{m-1}|G|^{m+1}$.

Putting the two cases together yields

$$\sum_{x\in\mathscr{Z}\setminus\mathscr{X}} \sum_{\substack{I,S,j,g:x_g^j\in\mathscr{X}\\ j\notin I; I\cap I(x_g^j)=\emptyset}} |f(x_S^I) - f(x_g^j)|^2$$

$$\le (1+m)\left(\binom{n-1}{m} + (n-m)\binom{n-1}{m-1}\right)|G|^{m+1}D^2R_\approx$$

which yields (5.8).

Proof of 5.9. Here, I, K are disjoint ordered m-tuples, S, W are generating m-tuples and $K' = W' = \emptyset$. We write $\gamma(x, I, S, K, W)$ for $\gamma(x, I, S, K, W, \emptyset, \emptyset)$. These paths have length at most $2mD$ by (5.11). Starting as for (5.7) and (5.8), we have

$$\sum_{x\in\mathscr{X}} \sum_{\substack{I,S,K,W \\ I\cap K=\emptyset}} |f(x_S^I) - f(x_W^K)|^2 \le 2mD \sum_{\substack{(z,z') \\ z\approx z'}} \sum_{\substack{x,I,S,K,W \\ I\cap K=\emptyset;\gamma(x,I,S,j,g)\ni(z,z')}} |f(z) - f(z')|^2 \ .$$

We fix points z and $z' \approx z$ that differ at exactly one entry, say the jth. Then, by construction, either $j \in I$ or $j \in K$. We treat these two cases separately. As they are very similar, we only give the details when $j \in I$. Assuming $j \in I$, we pick the remaining entries of I, and the entries of K: there are $\binom{n-1}{m-1} \times \binom{n-m}{m}$ possible choices. Now that we know $I = \{i_1, \ldots, i_m\}$, $j = i_\alpha$ and K, we can find W, all the entries of x outside $I \cup K$, the entries x_{i_β} with $\beta < \alpha$ and the $s_\beta \in S$ with $\beta > \alpha$. Further, the pair (s_α, x_j) must satisfy (in the notation of Section 2)

$$(z_j, z'_j) \in \gamma(W, s_\alpha, x_j) \ .$$

By Lemma 2.1, there are at most $|G|D$ such pairs. Thus the case where $j \in I$ contributes at most a factor

$$\binom{n-1}{m-1}\binom{n-m}{m}|G|^{2m}D \ .$$

The same is true for the case where $j \in K$. Hence

$$\sum_{x\in\mathscr{X}} \sum_{\substack{I,S,K,W \\ I\cap K=\emptyset}} |f(x_S^I) - f(x_W^K)|^2 \le 4m\binom{n-1}{m-1}\binom{n-m}{m}|G|^{2m}D^2R_\approx$$

which is (5.9).

Proof of 5.10. Here, $K = K' = W = W' = \emptyset$. We write $\gamma(x, I, S)$ for $\gamma(x, I, S, \emptyset, \emptyset, \emptyset, \emptyset)$. These paths have length at most mD by (5.11). We have

$$\sum_{x\in\mathscr{X}} \sum_{\substack{I,S \\ I\cap I(x)=\emptyset}} |f(x_S^I) - f(x)|^2 \le mD \sum_{\substack{(z,z') \\ z\approx z'}} \sum_{\substack{x,I,S \\ I\cap I(x)=\emptyset;\gamma(x,I,S)\ni(z,z')}} |f(z) - f(z')|^2 \ .$$

We fix z and z' that differ at exactly one entry, say the jth. Then, necessarily, $j \in I$. We pick the remaining entries of $I = (i_1, \ldots, i_m)$ among the $\binom{n-1}{m-1}$ possible choices. Define α by $j = i_\alpha$. Now, scanning z

yields the entries of x outside I, the entries x_{i_β} with $\beta < \alpha$, the $s_\beta \in S$ with $\beta > \alpha$. We can also find $I(x)$ and $S(x)$ just by looking at z because of property (5.6) and the fact that $I \cap I(x) = \emptyset$. Further the pair (s_α, x_j) must satisfy (in the notation of Section 2)

$$(z_j, z'_j) \in \gamma(S(x), s_\alpha, x_j).$$

By Lemma 2.1, there are at most $|G|D$ such pairs. Observe that it is important that $S(x)$ is known and fixed in order to apply Lemma 2.1: property (5.6) is used here. Finally, we get

$$\sum_{x \in \mathscr{X}} \sum_{\substack{I,S \\ I \cap I(x) = \emptyset}} |f(x_S^I) - f(x)|^2 \le m \binom{n-1}{m-1} |G|^m D^2 \, R_\approx$$

which is (5.10). This ends the proof of Lemma 5.4.

We now return to (5.4), i.e.,

$$\mathscr{E}_Q(\tilde{f}, \tilde{f}) = \frac{1}{2n|G|^{n+1}} (R_1 + 2R_2 + R_3)$$

and estimate R_1, R_2, R_3 in terms of $R_\approx$ using Lemma 5.4.

Estimating R_3: We start with R_3 which is the easiest to deal with. Write

$$\begin{aligned} R_3 &= \sum_{\substack{x,y \in \mathscr{Z} \setminus \mathscr{X} \\ x \sim y}} |\tilde{f}(x) - \tilde{f}(y)|^2 \\ &= \frac{1}{\binom{n}{m}^2 M^2} \sum_{\substack{x,y \in \mathscr{Z} \setminus \mathscr{X} \\ x \sim y}} \left| \left(\sum_{I,S} f(x_S^I) \right) - \left(\sum_{I,S} f(y_S^I) \right) \right|^2 \\ &\le \frac{1}{\binom{n}{m} M} \sum_{\substack{x,y \in \mathscr{Z} \setminus \mathscr{X} \\ x \sim y}} \sum_{I,S} |f(x_S^I) - f(y_S^I)|^2 \ . \end{aligned}$$

The last step uses the Cauchy-Schwarz inequality. Now observe that for $x, y \in \mathscr{Z} \setminus \mathscr{X}$ with $x \sim y$, x_S^I and y_S^I are either equal or differ only at one position $j \notin I$. Let $g = y_j$ be the jth coordinate of y. Then, $y = x_g^j$ and $y_S^I = [x_S^I]_g^j$. Thus, using the first inequality (5.7) of Lemma 5.4, we get

$$R_3 \le \frac{1}{\binom{n}{m}M} \sum_{x\in\mathscr{Z}\setminus\mathscr{X}} \sum_{\substack{j,g\\ x_g^j\in\mathscr{X}}} \sum_{I,S} |f(x_S^I) - f([x_S^I]_g^j)|^2$$
$$\le \frac{1}{\binom{n}{m}M} \times \binom{n-1}{m} |G|^{m+1} D^2 R_{\approx}$$
$$\le \frac{(n-m)|G|^{m+1}D^2}{nM} R_{\approx} \ . \tag{5.12}$$

Estimating R_2: We now look at R_2. We have

$$R_2 \le \frac{1}{\binom{n}{m}M} \sum_{x\in\mathscr{Z}\setminus\mathscr{X}} \sum_{\substack{j,g\\ x_g^j\in\mathscr{X}}} \sum_{I,S} |f(x_S^I) - f(x_g^j)|^2 \ . \tag{5.13}$$

We are going to break the sum in (5.13) into two pieces. Call Ξ_1 the set of (x,j,g,I) such that $x \in \mathscr{Z}\setminus\mathscr{X}$, $x_g^j \in \mathscr{X}$ and $I \cap I(x_g^j) = \emptyset$. Call Ξ_2 the set of (x,j,g,I) such that $x \in \mathscr{Z}\setminus\mathscr{X}$, $x_g^j \in \mathscr{X}$ and $I \cap I(x_g^j) \neq \emptyset$ where $I(x_g^j)$ is defined at (5.5). Write

$$\sum_{x\in\mathscr{Z}\setminus\mathscr{X}} \sum_{\substack{j,g\\ x_g^j\in\mathscr{X}}} \sum_{I,S} |f(x_S^I) - f(x_g^j)|^2 = \sum_{\Xi_1} \sum_{S} |f(x_S^I) - f(x_g^j)|^2$$
$$+ \sum_{\Xi_2} \sum_{S} |f(x_S^I) - f(x_g^j)|^2 \ . \tag{5.14}$$

For Ξ_1, observe that $x \in \mathscr{Z}\setminus\mathscr{X}$ and $x_g^j \in \mathscr{X}$ imply $j \in I(x_g^j)$. It follows that $j \notin I$. Thus, we can use the second inequality (5.8) of Lemma 5.4 which yields

$$\sum_{\Xi_1} \sum_{S} |f(x_S^I) - f(x_g^j)|^2 \le (1+m)^2 \binom{n-1}{m} |G|^{m+1} D^2 \ R_{\approx} \ . \tag{5.15}$$

We now pass to Ξ_2. For each (x,j,g,I,S) with $|I \cap I(x_g^j)| = v \ge 1$, write

$$|f(x_S^I) - f(x_g^j)|^2 \le \frac{2}{\binom{n-m-\overline{m}+v}{m}M} \sum_{\substack{K,W\\ (I\cup I(x_g^j))\cap K=\emptyset}} \Big(|f(x_S^I) - f(x_W^K)|^2$$
$$+ |f(x_W^K) - f(x_g^j)|^2 \Big)$$

where K runs over all ordered m-tuples such that $(I \cup I(x_g^j)) \cap K = \emptyset$ and $W = (w_1, \ldots, w_m) \in G^m$ runs over all generating m-tuples (there

are M of them). Here, we use the hypothesis that $n \geq 2m(G) + \overline{m}(G)$. This gives

$$\begin{aligned}
&\sum_{\Xi_2} \sum_{S} |f(x_S^I) - f(x_g^j)|^2 \\
&\leq \frac{2}{\binom{n-m-\overline{m}+1}{m} M} \left(\sum_{\Xi_2} \sum_{\substack{S,K,W \\ j \notin K; I \cap K = \emptyset}} |f(x_S^I) - f(x_W^K)|^2 \right. \\
&\quad \left. + \sum_{\Xi_2} \sum_{\substack{S,K,W \\ I(x_g^j) \cap K = \emptyset}} |f(x_W^K) - f(x_g^j)|^2 \right) \\
&= \frac{2}{\binom{n-m-\overline{m}+1}{m} M} (\Sigma + \Sigma') \ . \qquad (5.16)
\end{aligned}$$

For Σ', we have

$$\Sigma' \leq \binom{n-m}{m} M \sum_{(x,j,g,K) \in \Xi_1} \sum_{W} |f(x_W^K) - f(x_g^j)|^2 \ .$$

Hence the analysis used for Ξ_1 applies and yields

$$\Sigma' \leq (1+m)^2 \binom{n-1}{m} \binom{n-m}{m} M |G|^{m+1} D^2 \, R_{\approx} \ . \qquad (5.17)$$

We are left with the task of bounding

$$\begin{aligned}
\Sigma &= \sum_{(x,j,g,I) \in \Xi_2} \sum_{\substack{S,K,W \\ j \notin K; I \cap K = \emptyset}} |f(x_S^I) - f(x_W^K)|^2 \\
&\leq (n-m)|G| \sum_{\substack{x,I,S,K,W \\ j \notin K; I \cap K = \emptyset}} |f(x_S^I) - f(x_W^K)|^2 \ .
\end{aligned}$$

The factor $(n-m)|G|$ counts the number of (j, g), $j \notin K$. It follows from the third inequality (5.9) of Lemma 5.4 that

$$\Sigma \leq 4m(n-m) \binom{n-1}{m-1} \binom{n-m}{m} |G|^{2m+1} D^2 R_{\approx} \ . \qquad (5.18)$$

Using (5.17) and (5.18) in (5.16), we get

$$\sum_{\Xi_2} \sum_{S} |f(x_S^I) - f(x_g^j)|^2 \leq \frac{2}{\binom{n-m-\overline{m}+1}{m} M} (\Sigma + \Sigma')$$

$$\leq \frac{2\left(1+2m+5m^2\right)\binom{n-1}{m}\binom{n-m}{m}|G|^{2m+1}D^2}{\binom{n-m-\overline{m}+1}{m}M} R_{\approx} \ .$$

Using this and (5.15) in (5.14), (5.13), we conclude that

$$R_2 \leq \left(3+6m+11m^2\right)\frac{\binom{n-m}{m}}{\binom{n-m-\overline{m}+1}{m}}\frac{|G|^{2m+1}D^2}{M^2}R_{\approx} \ . \tag{5.19}$$

Estimating R_1: We now bound

$$R_1 = \sum_{\substack{x,y\in\mathscr{X}\\ x\sim y}} |f(x)-f(y)|^2 \ .$$

To each x and y in $\mathscr{X}$ correspond by (5.5) two $\overline{m}$-tuples $I=I(x)$ and $I(y)$ such that the associated $\overline{m}$-tuples $S(x)$ and $S(y)$ generate G. Moreover, if $x\sim y$, then they differ only at one place, say j, and $y=x_g^j$ where $g=y_j\in G$. If $j\notin I(x)$, change the value at j using the generating set corresponding to I and the group-paths $\gamma(S(x),x_j,y_j)$. Similarly, if $j\notin I(y)$, change the value at j using the generating set corresponding to $I(y)$ and the group-paths $\gamma(S(y),x_j,y_j)$. Call this path $\gamma(x,y)$. Its length is at most D. Let Θ_1 be the set of the (x,y) for which this construction works and Θ_2 be the set of (x,y) such that $j\in I(x)\cap I(y)$. For Θ_1, using Cauchy-Schwarz, we have

$$\begin{aligned}\sum_{(x,y)\in\Theta_1}|f(x)-f(y)|^2 &\leq D\sum_{(x,y)\in\Theta_1}\sum_{(z,z')\in\gamma(x,y)}|f(z)-f(z')|^2\\ &\leq D\sum_{\substack{z,z'\in\mathscr{X}\\ z\approx z'}}\sum_{\substack{(x,y)\in\Theta_1\\ \gamma(x,y)\ni(z,z')}}|f(z)-f(z')|^2 \ .\end{aligned}$$

We have to count how many times each (z,z') appears. By construction, (z,z') determines j and all the coordinates of x and y except the jth. Since $(x,y)\in\Theta_1$, we know that either $j\notin I(x)$ or $j\notin I(y)$. In the first case, $S(x)=S_{z,z'}$ is independent of x_j by (5.6) and (x_j,y_j) must belong to

$$\{(u,v):\gamma(S_{z,z'},u,v)\ni(z_j,z_j')\} \ .$$

Lemma 2.1 shows that there are at most $|G|D$ such (x_j,y_j). Similarly, if $j\notin I(y)$, there are at most $|G|D$ possible choices for (x_j,y_j). It follows that

$$\sum_{(x,y)\in\Theta_1} |f(x)-f(y)|^2 \le 2|G|D^2 R_{\approx} \ . \tag{5.20}$$

We must now deal with Θ_2 where $j \in I(x)\cap I(y)$. Fix $(x,y)\in\Theta_2$. Let K, W be such that K is an ordered m-tuple of distinct integers between 1 and n satisfying $K\cap I(x)=\emptyset$, and $W=(w_1,\dots,w_m)\in G^m$ is a generating m-tuple. Similarly, let K', W' be m-tuples such that $K'\cap(I(y)\cup K)=\emptyset$ and W' generates. This is possible because $n\ge 2m+\overline{m}$. We pick uniformly at random among all the possible (K,W,K',W') and write

$$\begin{aligned}|f(x)-f(y)|^2 \le{}& \frac{3}{M^2\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}}\\ &\times \sum_{\substack{K,W\\ K\cap I(x)=\emptyset}} \sum_{\substack{K',W'\\ K'\cap(K\cup I(y))=\emptyset}} \Big(|f(x)-f(x_W^K)|^2+|f(x_W^K)-f(x_{W'}^{K'})|^2\\ &+|f(x_{W'}^{K'})-f(y)|^2\Big) \ .\end{aligned}$$

This gives

$$\sum_{\Theta_2}|f(x)-f(y)|^2 \le \frac{3}{M^2\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}}$$

$$\times\Big(\overline{m}M|G|\binom{n-m}{m}\Delta_1+\overline{m}|G|\Delta_2+M\binom{n-m}{m}\Delta_3\Big) \tag{5.21}$$

where

$$\Delta_1=\sum_{\substack{x,K,W\\ K\cap I(x)=\emptyset}}|f(x)-f(x_W^K)|^2$$

$$\Delta_2=\sum_{\substack{x,K,W\\ K\cap I(x)=\emptyset}}\sum_{\substack{K',W'\\ K'\cap(K\cup I(y))=\emptyset}}|f(x_W^K)-f(x_{W'}^{K'})|^2$$

$$\Delta_3=\sum_{\substack{(x,y)\in\Theta_2,K',W'\\ K'\cap I(y)=\emptyset}}|f(y)-f(x_W^K)|^2 \ .$$

In (5.21), the factor $\overline{m}M|G|\binom{n-m}{m}$ in front of Δ_1 accounts for the variables j, W', g, K' where $y=x_g^j$ For j and K', we have taken into account the facts that $j\in I(x)$ and $K'\cap K=\emptyset$. These variables do not appear in Δ_1. Similarly, the factor $\overline{m}|G|$ in front of Δ_2 accounts for the variables j, g where $y=x_g^j$, taking into account the fact that $j\in I(x)$.

Finally the factor $M\binom{n-m}{m}$ in front of Δ_3 accounts for the variables W and K (recall that $K \cap I(x) = \emptyset$).

Now, Δ_1 can be bounded using the fourth inequality (5.10) in Lemma 5.4. For Δ_2, we can use inequality (5.9) of Lemma 5.4 whereas, for Δ_3, we can use (5.8) since

$$\begin{aligned}\Delta_3 &= \sum_{\substack{(x,y)\in\Theta_2,K',W' \\ K'\cap I(y)=\emptyset}} |f(y) - f(x_W^K)|^2 \\ &\le \sum_{\substack{x,K',W',j,g \\ K'\cap I(x_g^j)=\emptyset}} |f(x_g^j) - f(x_W^K)|^2.\end{aligned}$$

Hence,

$$\begin{aligned}&\sum_{\Theta_2} |f(x) - f(y)|^2 \\ &\le \frac{3\binom{n-m}{m}\binom{n}{m}}{M^2\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}} \left(\frac{\overline{m}m^2}{n} + 4\frac{\overline{m}m^2}{n} + \frac{(n-m)(1+m)^2}{n}\right) |G|^{2m+1} D^2 R_{\approx} \\ &\le \frac{3\left(1+2m+6m^2\right)\binom{n-m}{m}\binom{n}{m}}{M^2\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}} |G|^{2m+1} D^2 R_{\approx} \ .\end{aligned}$$

This and (5.20) yield

$$R_1 \le \left(5+6m+18m^2\right) \frac{\binom{n}{m}\binom{n-m}{m}}{M^2\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}} |G|^{2m+1} D^2 R_{\approx} \ . \qquad (5.22)$$

Using (5.12), (5.19) and (5.22), we obtain

$$\begin{aligned}\mathscr{E}_Q(\tilde{f},\tilde{f}) &= \frac{1}{2n|G|^{n+1}} (R_1 + 2R_2 + R_3) \\ &\le \frac{5(1+2m)^2\binom{n}{m}\binom{n-m}{m}}{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}} \frac{|G|^{2m} D^2}{n|G|^n M^2} R_{\approx} \\ &\le \frac{20(1+2m)^2\binom{n}{m}\binom{n-m}{m}}{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}} \frac{|\mathscr{X}||G|^{2m} D^2 n}{|\mathscr{Z}|M^2} \mathscr{E}_P(f,f)\end{aligned}$$

because $\mathscr{E}_P = \frac{1}{4n(n-1)} R_{\approx}$. This ends the proof of Proposition 5.2.

Proof of Corollary 5.3 and Theorem 3.2. From Lemma 5.1 and Proposition 5.2, the second largest eigenvalue of P is bounded by

$$\beta_1(P) \le 1 - \frac{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}M^2}{20(1+2m)^2\binom{n}{m}\binom{n-m}{m}|G|^{2m}D^2n^2}.$$

For the log-Sobolev constant, we get

$$\alpha(P) \ge \frac{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}M^2(|G|-2)}{20(1+2m)^2\binom{n}{m}\binom{n-m}{m}|G|^{2m+1}\log(|G|-1)D^2n^2} \quad .$$

These estimates and Theorem 2.3 give

$$\|H_t^x - \pi\|_{\mathrm{TV}} \le \mathrm{e}^{1-c}$$

for

$$t = \frac{20(1+2m)^2\binom{n}{m}\binom{n-m}{m}|G|^{2m+1}D^2n^2}{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}M^2(|G|-2)}[(\log(|G|-1))(\log\log|G|^n) + 4c]$$

with $c > 0$. This proves Theorem 3.2.

Remark. In the above proof, we have never used specifically the fact that m is the minimum size of a generating set in G. This leads to the following extensions of Corollary 5.3 and Theorem 3.2:

Theorem 5.5 *For any group G with $m(G) = m$, $\overline{m}(G) = \overline{m}$, any $m_* \ge m$, and $n \ge 2m_* + \overline{m}$, the chain P defined at* (3.1) *satisfies*

$$\beta_1(P) \le 1 - \frac{\binom{n-\overline{m}}{m_*}\binom{n-\overline{m}-m_*}{m_*}M_*^2}{20(1+2m_*)^2\binom{n}{m_*}\binom{n-m_*}{m_*}|G|^{2m_*}D^2n^2}$$

and

$$\alpha(P) \ge \frac{\binom{n-\overline{m}}{m_*}\binom{n-\overline{m}-m_*}{m_*}M_*^2(|G|-2)}{20(1+2m_*)^2\binom{n}{m_*}\binom{n-m_*}{m_*}|G|^{2m_*+1}\log(|G|-1)D^2n^2} \quad .$$

Further, for all $c > 0$, the semigroup $H_t = \mathrm{e}^{-t(I-P)}$ associated to the chain P at (3.1) *on generating n-tuples satisfies*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \le \mathrm{e}^{1-c}$$

for

$$t = \frac{20(1+2m_*)^2\binom{n}{m_*}\binom{n-m_*}{m_*}|G|^{2m_*+1}D^2n^2}{\binom{n-\overline{m}}{m_*}\binom{n-\overline{m}-m_*}{m_*}M_*^2(|G|-2)}$$

$$\times[(\log(|G|-1))(\log\log|G|^n)+4c] \ .$$

Here $M_* = f(m_*, G)$ is the number of distinct generating m_*-tuples and $D = D(G)$ is the maximal diameter of G.

This extension is interesting because it may happen that $M = f(m, G)$ is small or difficult to bound from below whereas $M_* = f(m_*, G) \approx |G|^{m_*}$ for certain $m_* > m$. For instance, this happens for cyclic groups of very composite order, see Section 6.C, Example 2.

6 Group combinatorics

This section discusses the four group theoretical parameters needed to apply Theorem 3.2. There is a growing literature on group combinatorics. See e.g., the surveys [5, 33]. We will use the following notation. Given a group G, the lower central series $G_1 = G \supset G_2 \supset \cdots \supset G_i \cdots$ is defined inductively by

$$G_i = [G_{i-1}, G]$$

where, for any two subgroups H, K, $[H, K]$ is the commutator group generated by $[h,k] = h^{-1}k^{-1}hk, h \in H, k \in K$. By definition, G is nilpotent of class c if $G_c \neq \{\mathrm{id}\}$ and $G_{c+1} = \{\mathrm{id}\}$.

6.A. The size of minimal generating sets

Let G be a finite group and $S \subset G$ be a set of generators. Say S is minimal if no smaller subset of S generates G. Recall that we have defined $\overline{m} = \overline{m}(G)$ to be the maximum size of a minimal generating set and $m = m(G)$ to be the minimum size of a generating set. If S is a generating set with $|S| = \overline{m}(G)$, then deleting successive elements of S results in a strictly decreasing sequence of subgroups. This shows that $\overline{m}(G)$ is bounded by the length of the longest chain of subgroups in G. If $|G| = \prod p^{a_p}$ is the factorization of the size of G into distinct prime powers, we see

$$\overline{m}(G) \leq \Omega(|G|) = \sum_{p \mid |G|} a_p. \tag{6.1}$$

In any solvable group, the length of the longest chain of subgroups is exactly $\Omega(|G|)$ but, in general, $\Omega(|G|)$ only gives an upper bound. Further, the length of the longest chain of subgroups is only an upper

bound for $\overline{m}$. There is a large literature on chains of subgroups in permutation groups and finite groups of Lie type. See [33].

Example 6.A.a: Cyclic groups. Take G to be the cyclic group $\mathbb{Z}/r\mathbb{Z}$ where $r = \prod_1^k p_i^{a_i}$ is a factorization of r into distinct prime powers ($a_i \neq 0$). Then (6.1) gives $\overline{m} \leq \sum_1^k a_i$ whereas, of course, $\overline{m} = k$. Here, $m = 1$.

Example 6.A.b: The symmetric group. Take G to be the symmetric group S_d. Then, for primes $p \leq d$, $a_p = [d/p] + [d/p^2] + \cdots \leq d/[p(1 - 1/p)] \leq 2d/p$. Thus,

$$\overline{m}(S_d) \leq 2d \sum_{p \leq d} \frac{1}{p} \sim 2d \log\log d \ .$$

In fact, using results of Babai [3] and Cameron et al. [8], $\overline{m}(S_d) \leq 2d$. This follows from an exact formula for the length of the longest chain of subgroup in S_d. The longest chain only gives an upper bound on $\overline{m}$. Indeed, preliminary computations based on the classification of simple finite groups seems to indicate that $\overline{m}(S_d) = d - 1$. The classical generating set $S = \{(1,2), (2,3), \ldots, (d-1, d)\}$ shows $\overline{m}(S_d) \geq d - 1$. Observe also that $m(S_d) = 2$ (e.g., the transposition $(1,2)$ and the long cycle $(1, \ldots, d)$ generate).

Example 6.A.c: *p* groups. Let $|G| = p^a$ for some prime p. For such groups, $\overline{m}(G)$ can be explicitly determined. We need a bit of elementary group theory connected with the Burnside basis theorem. Suzuki [35], p. 93, is a splendid reference. Let $\Phi(G) = \Phi$ be the Frattini subgroup, i.e., the intersection of the subgroups of order p^{a-1}. This is a normal subgroup and the Burnside basis theorem says the quotient G/Φ is isomorphic to a vector space over the field $\mathbb{Z}/p\mathbb{Z}$. If $|G/\Phi| = p^b$, Burnside's theorem says S generates G if and only if the images of S generate G/Φ as a linear space. Thus, G can be generated by b generators. Further, if S is a generating set with $|S| \geq b$, the images of S in G/Φ generate and so some subset of size b in S generates G. Thus we have proved

$$\text{for a } p- \text{ group } G \text{ with } |G/\Phi| = p^b, \quad \overline{m}(G) = m(G) = b \ . \tag{6.2}$$

For instance, let G be the group of upper-triangular $n \times n$ matrices with ones on the diagonal and entries mod p, p prime. Thus $|G| = p^{n(n-1)/2}$. It is well known that $\Phi(G)$ is the subgroup with zeros just above the diagonal. Thus, $|G/\Phi| = p^{n-1}$ and $\overline{m}(G) = n - 1$. This

is a good deal smaller than the bound (6.1). A simple set of generators is s_i with ones on the diagonal, a one in position $(i, i+1)$ and zeros elsewhere, $1 \le i \le n-1$. For example, when $n = 3$, G is the Heisenberg group with entries mod p. The two generators are

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} .$$

For many further examples, see [19], Section 5.C.

Isaacs [31] gives some bounds for $\overline{m}(G)$ for p-groups. Among other things, he shows that, if $p \ge 3$ and G is non Abelian, $\overline{m}(G) \le f - p + 3$ with f the dimension of a faithful characteristic zero representation. Observe that in the case of the Heisenberg group, $f = p$ is the smallest possible degree of such a representation and Isaacs' bound reads $\overline{m}(G) \le 3$ whereas the right answer is 2. There are slight variants where Isaac's bound is sharp.

Example 6.A.d: Nilpotent groups. If G is nilpotent, G is the direct product of its Sylow p-groups: $G = \prod_1^k \mathbf{S}(p_i)$ where the p_i's are the distinct primes that divide $|G|$. Clearly, $\overline{m}(G) = \sum_1^k \overline{m}(\mathbf{S}(p_i))$ whereas $m(G) = \max_i m(\mathbf{S}(p_i))$.

Example 6.A.e: Metacyclic groups. If p, q are primes such that q divides $p-1$, there exists exactly one non-Abelian group of order pq. It is a semidirect product $H(p,q) = \mathbb{Z}/q\mathbb{Z} \propto \mathbb{Z}/p\mathbb{Z}$. These have $m(G) = \overline{m}(G) = 2$. See [5].

6.B. The number of generating tuples

Let $f(k, G)$ be the number of ordered k-tuples that generate G. We are interested in this number for at least two reasons:

First, when n is large enough, namely $n \ge m(G) + \overline{m}(G)$, Lemma 3.1 shows that our walk is irreducible on the set of all generating n-tuples. Thus, $f(n, G) = |\mathscr{X}|$ is the size of the natural state space $\mathscr{X}$ of the chain P at (3.1).

Second, our main result, Theorem 3.2, involves the quantity

$$M = M(G) = f(m(G), G) \ .$$

In order to apply Theorem 3.2, it is crucial to have good lower bounds on this M, if possible of the type $M \ge c|G|^{m(G)}$ where c does not depend on G. Similarly, Theorem 5.4 requires lower bounds on

$M_* = f(m_*, G)$ for some fixed $m_* \geq m$. A simple yet useful observation for our purpose is that

$$\text{the ratio} \quad f(k,G)/|G|^k \quad \text{is an increasing function of } k \ . \tag{6.3}$$

In this section we show how to use Möbius inversion on the subgroups of G to give a formula for the size of the state space. This is work of P. Hall (1936) [27]. A clear elementary exposition appears in Constantine [10]. A recent survey is in [29]. The role of Möbius inversion comes from the observation that, with obvious notation,

$$|G|^k = \sum_{\{\mathrm{id}\} \subseteq H \subseteq G} f(k,H) \ .$$

Thus,

$$f(k,G) = \sum_{\{\mathrm{id}\} \subseteq H \subseteq G} |H|^k \mu(H,G) \tag{6.4}$$

with $\mu(H,G)$ the Möbius function of the interval $[H,G]$. This function is known for several classes of groups. We describe results for nilpotent groups. Such a group is the direct product of its Sylow p-groups. Now, if $G = \prod_1^\ell G_i$ where $|G_i| = p_i^{a_i}$ are distinct prime powers, any subgroup H of G has the form $H = \prod_1^\ell H_i$ with H_i a subgroup of G_i. Further, the partial order $(H_i)_1^\ell \leq (H_i')_1^\ell \iff H_i \subseteq H_i', \ 1 \leq i \leq \ell$ coincides with the inclusion ordering on subgroups of G. It is standard that the Möbius function factors

$$\mu(H,G) = \prod_1^\ell \mu(H_i, G_i) \ .$$

Hence, $f(k,G)$ also factors and

$$f(k,G) = \prod_1^\ell f(k,G_i) \ .$$

To treat nilpotent groups it is thus enough to determine the Möbius functions of p-groups with p a prime. For p-groups, $\mu(H,G)$ is zero unless H contains the Frattini subgroup $\Phi(G)$. If H contains $\Phi(G)$ and has index p^ℓ in G, then

$$\mu(H,G) = (-1)^\ell p^{\binom{\ell}{2}} \ .$$

The determination of Möbius functions for other classes of groups is not simple. P. Hall [27] treats $\mathrm{PSL}_2(\mathbb{Z}/p\mathbb{Z})$ and Gaschütz [25] treats solvable groups.

Example 6.B.a: *p*-groups. Start with $G = (\mathbb{Z}/p\mathbb{Z})^\ell$, p prime. Here $\Phi(G) = \{\mathrm{id}\}$. The number of subgroups of index p^i is equal to the number $N(\ell, \ell - i)$ of linear subspaces of dimension $\ell - i$ in $(\mathbb{Z}/p\mathbb{Z})^\ell$. It is well known that the numbers $N(\ell, v)$ are given by the p-binomial coefficients

$$N(\ell, v) = N(\ell, \ell - v) = \binom{\ell}{v}_p = \frac{(p^\ell - 1)\cdots(p^{\ell-v+1} - 1)}{(p^v - 1)\cdots(p - 1)} \quad .$$

Thus,

$$f(k, (\mathbb{Z}/p\mathbb{Z})^\ell) = \sum_{i=1}^{\ell} (-1)^i \binom{\ell}{\ell - i}_p p^{k(\ell-i)+\binom{i}{2}}.$$

Lemma 6.1 *For $k \geq \ell$,*

$$f(k, (\mathbb{Z}/p\mathbb{Z})^\ell) \geq p^{\ell k}\left(1 - \frac{p^\ell - 1}{p - 1}p^{-k}\right) \quad .$$

Further, for $p = 2$ and $k \geq \ell$, $f(k, (\mathbb{Z}/2\mathbb{Z})^\ell) \geq \frac{1}{4}2^{\ell k}$.

Proof. For $i \geq 3$, we have

$$\frac{\binom{\ell}{\ell-i}_p p^{k(\ell-i)+\binom{i}{2}}}{\binom{\ell}{\ell-i+1}_p p^{k(\ell-i+1)+\binom{i-1}{2}}} = \frac{(p^{\ell-i+1} - 1)p^{-k+i-1}}{p^i - 1} \leq 1$$

Thus, $i \mapsto \binom{\ell}{\ell-i}_p p^{k(\ell-i)+\binom{i}{2}}$ is a decreasing function of $i \geq 2$. It follows that

$$f(k, (\mathbb{Z}/p\mathbb{Z})^\ell) \geq p^{\ell k} - \frac{p^\ell - 1}{p - 1}p^{(\ell-1)k} = p^{\ell k}\left(1 - \frac{p^\ell - 1}{p - 1}p^{-k}\right) \quad .$$

When $p = 2$ and $\ell = k$, this bound is poor but there is another way to bound $f(k, (\mathbb{Z}/p\mathbb{Z})^\ell)$. Restricting ourselves to $k = \ell$ for simplicity, its is easy to see that

$$f(\ell, (\mathbb{Z}/p\mathbb{Z})^\ell) = (p^\ell - 1)(p^\ell - p)\cdots(p^\ell - p^{\ell-1})$$

where each factor represents the number of vectors in $(\mathbb{Z}/p\mathbb{Z})^\ell$ that are linearly independent of the previously chosen vectors. Thus, using (6.3),

$$f(k, (\mathbb{Z}/p\mathbb{Z})^\ell) \geq p^{\ell k} \times \prod_1^\ell (1 - p^{-i}) \geq p^{\ell k}\mathrm{e}^{-1/(p-1)} \quad .$$

This proves the last assertion of the lemma.

Let now G be finite p-group with Frattini subgroup $\Phi = \Phi(G)$. Since a set S of k elements of G generates G if and only if the projection of S to the Frattini quotient G/Φ generates G/Φ, we have:

$$f(k, G) = |\Phi|^k f(k, G/\Phi) \ .$$

Since $G/\Phi = (\mathbb{Z}/p\mathbb{Z})^{m(G)}$ is an elementary Abelian p-group, Lemma 6.1 yields the following.

Lemma 6.2 *Let G be a finite p-group. Put $m = m(G)$. For $k \geq m$, the number $f(k, G)$ of k-tuples that generate G satisfies*

$$f(k, G) \geq |G|^k \left(1 - \frac{p^m - 1}{p - 1} p^{-k}\right) \ .$$

Further, for $p = 2$ and $k \geq m$, $f(k, G) \geq \frac{1}{4}|G|^k$.

Example 6.B.b: Nilpotent groups. The results obtained for p-groups and the factorisation property let us now compute $f(k, G)$ for any nilpotent group G. Indeed, we simply have to write G as the product of its Sylow p-groups $G = \prod_{p||G|} \mathbf{S}(p)$ where p runs over all primes that divide $|G|$. Here, we recall that $m = m(G) = \max_p m(p)$ where $m(p) = m(\mathbf{S}(p))$. For any $k \geq m$,

$$f(k, G) = \prod_{p||G|} f(k, \mathbf{S}(p)) \ .$$

In particular,

Lemma 6.3 *Let G be a nilpotent group with Sylow decomposition $G = \prod_{p||G|} \mathbf{S}(p)$. Set $m = m(G)$ and $m(p) = m(\mathbf{S}(p))$. For $k \geq m$,*

$$f(k, G) \geq \frac{1}{4}|G|^k \prod_{\substack{p||G| \\ p \neq 2}} \left(1 - \frac{p^{m(p)} - 1}{p - 1} p^{-k}\right) \ .$$

If $2 \nmid |G|$ the factor $1/4$ can be removed.

Remark. For the cyclic group $G = \mathbb{Z}/r\mathbb{Z}$ with $r = \prod_{p|r} p^{a(p)}$, $m = m(p) = 1$ and $f(k, G) \geq \frac{1}{8} r^k$ if $k \geq 2$ whereas

$$f(1, G) = r \prod_{p|r} (1 - 1/p) \ .$$

Here, $r \mapsto f(1, \mathbb{Z}/r\mathbb{Z}) = \phi(r)$ is the classic Euler function. This shows that there are cases where $|G|^{-1} f(1, G) \to 0$ as r tends to infinity. However, we have

$$\liminf_{r\to\infty} \frac{f(1, \mathbb{Z}/r\mathbb{Z}) \log\log r}{r} = e^{-\gamma}$$

where γ is the Euler constant (See [28], pg. 267).

Using the results presented above, one sees that there exists an explicit constant $\varepsilon > 0$ such that, for any nilpotent group G,

$$f(m(G), G) \geq \varepsilon \frac{|G|^{m(G)}}{\log\log|G|}$$

whereas

$$f(k, G) \geq \tfrac{1}{8}|G|^k \text{ if } k > m(G) \ .$$

Example 6.B.c: The alternating and symmetric groups. It is known [1, 5] that the probability that a randomly selected pair of permutations in S_d generates A_d or S_d is $1 - 1/d + O(1/d^2)$. Thus

$$|A_d|^{-2} f(2, A_d) \to 1 \text{ as } d \to \infty$$

and

$$|S_d|^{-2} f(2, S_d) \to 3/4 \text{ as } d \to \infty \ .$$

This shows that for $G = A_d$ or S_d there exists a constant c independent of d such that, for all $k \geq 2$,

$$f(k, G) \geq c|G|^k.$$

Example 6.B.d: $G = \mathrm{PSL}_2(\mathbb{Z}/p\mathbb{Z})$, p prime. For this group, $m = 2$ whereas $\overline{m}$ seems unknown. P. Hall [27] studies $f(k, G)$, $k \geq m = 2$. He obtained manageable formulas which show that $f(k, G)$ is of the same order as $|G|^k = [\frac{1}{2}p(p^2 - 1)]^k$ when $k \geq 2$. Actually, $\lim_{k\to\infty} f(k, G)/|G|^k = 1$ and $\lim_{|G|\to\infty} f(k, G)/|G|^k = 1$, $k \geq 2$.

6.C. The maximum diameter

The maximum diameter is a difficult quantity to bound except in a few special circumstances. The difficulty comes from the fact that, in

general, we do not know or understand most generating sets of a group. We start with a difficult result of Babai and Seress concerning permutation groups. See [5, 6, 7].

Example 6.C.a: Permutation groups. For the alternating group A_d, Babai and Seress [6] prove that

$$D(A_d) \leq (1 + o(1))\mathrm{e}^{\sqrt{d \log d}} \ .$$

They show in [7] that this result extends to any permutation group of degree d. In particular,

$$D(S_d) \leq (1 + o(1))\mathrm{e}^{\sqrt{d \log d}} \ .$$

Further, they show that any transitive permutation group G of degree d satisfies

$$D(G) \leq \mathrm{e}^{c(\log d)^3} D(A_d)$$

for some $c > 0$. This last result is important because Babai and Seress [6] conjecture that, for every simple group G,

$$D(G) = (\log |G|)^{O(1)} \ . \tag{6.5}$$

In the case of A_d or S_d this amounts to conjecturing that there exists a independent of d such that

$$D(A_d), D(S_d) \leq d^a \ . \tag{6.6}$$

Indeed, for all we know at present writing it is possible that $a = 2$.

Example 6.C.b: Metacyclic groups. For the metacyclic groups $H(p, q) = (\mathbb{Z}/q\mathbb{Z}) \propto (\mathbb{Z}/p\mathbb{Z})$ with q, p primes and $q|(p-1)$ of example 6.A.e, Babai et al. [5] state that, for fixed q,

$$D(H(p, q)) \leq O\left(p^{1/(q-1)}\right)$$

whereas for $q \geq p^{\varepsilon}$ with $\varepsilon > 1/2$,

$$D(H(p, q)) \leq O(q) \ .$$

In both cases the estimate is optimal.

Example 6.C.c: *p*-groups and nilpotent groups. To state our bound on the maximum diameter requires some classical notation.

Let G be a nilpotent group of class c with lower central series $G_1 = G \supset G_2 \supset \cdots \supset G_c \supset \{\mathrm{id}\}$. The lower central series of G/G_i is ([35],II, pg. 13) $G/G_i \supset G_2/G_i \cdots G_i/G_i = \{\mathrm{id}\}$.

Simple commutators in a set $\{g_1, \ldots, g_r\}$ are defined inductively as follows. The simple commutators of length 1 are the g_i's. Simple commutators of length ℓ in the g_i's are commutators of the form $c = [c', g]$ with c' a simple commutator of length $\ell - 1$ and $g \in \{g_1, \ldots, g_r\}$. Thus, a simple commutator c of length ℓ has form

$$c = [[\cdots [x_1, x_2], \ldots,]x_\ell]$$

with $x_i \in \{g_1, \ldots, g_r\}$. There are r^ℓ possible simple commutators of length ℓ (of course some of these may be equal). Written out as a word in g_i, g_i^{-1} such a commutator involves $2^\ell + 2^{\ell-1} - 2$ elements. If G is generated by $\{g_1, \ldots, g_r\}$ then G_i/G_{i+1} is generated by the simple commutators of length i mod (G_{i+1}) ([26], Theorem 10.2.3). The following theorem bounds the diameter of a nilpotent group in terms of the exponent of $G/[G, G]$. The corollaries that follow give bounds on the maximum diameter.

Theorem 6.4 *Let G be a nilpotent group with class c. Let $\{g_1, \ldots, g_r\}$ be a minimal set of generators of G. Then the diameter γ of G in these generators satisfies*

$$\tfrac{1}{2}[\exp(G/[G, G]) - 1] \leq \gamma \leq (2r)^{c+1} \exp(G/[G, G])$$

where $\exp(G/[G, G])$ *is the exponent of* $G/[G, G]$.

Example: The Heisenberg group $\mathrm{mod}(k)$ has class 2 and $\exp(G/[G, G]) = k$. Theorem 6.5 shows that any set of two generators has diameter essentially k, uniformly in k.

Corollary 6.5 *Let G be a p-group with class c and Frattini rank b* (i.e, $|G/\Phi| = p^b$). *Then the maximum diameter $D(G)$ satisfies*

$$\tfrac{1}{2}[\exp(G/[G, G]) - 1] \leq D(G) \leq (2b)^{c+1} \exp(G/[G, G]) \ .$$

Proof of Theorem 6.4. The theorem holds trivially for Abelian groups (class $c = 1$). For clarity, we first prove the case $c = 2$. Let $e = \exp(G/[G, G])$. Then for any $g \in G$, $g^e \in [G, G]$. Hence, any g can be written

$$g = g_1^{a_1} \cdots g_r^{a_r} w$$

with $0 \le a_i < e$ and $w \in [G, G]$. Now, $[G, G]$ is generated by $\{[g_i, g_j]; i \ne j\}$, so the class 2 case is proved if $\exp([G, G]) \le e$. For this, recall that in general, if x, y are such that $[x, y] \in Z(G)$, then $[x^u, y^v] = [x, y]^{uv}$. This follows from the formula $[xy, z] = y^{-1}[x, z]y[y, z]$. In particular, $[x^e, y] = [x, y]^e$ for all $x, y \in G$ because $c = 2$ implies $[G, G] \subset Z(G)$. Now $x^e \in [G, G] \subset Z(G)$, so $[x^e, y] = [x, y]^e = \mathrm{id}$.

Inductively, suppose we have shown that, for any group G of class $c - 1$ generated by $\{x_1, \ldots, x_r\}$, any element is expressible as

$$c_1^{a_1} \cdots c_\ell^{a_\ell}$$

with c_i simple commutators in $\{x_1, \ldots, x_r\}$ and $0 \le a_i < \exp(G/[G, G])$.

Let G be a group of class c with lower central series $G = G_1 \supset G_2 \supset \cdots \supset G_c \supset \{\mathrm{id}\}$. Then $G_c \subset Z(G)$ ([26], 10.2.1) and G/G_c have class $c - 1$. Set $\exp(G/[G, G]) = e$. If G is generated by $\{g_1, \ldots, g_r\}$, then G/G_c is generated by $\bar{g}_i = g_i \bmod(G_c)$. Thus, with obvious notation, any $\bar{g} \in G/G_c$ can be written

$$\bar{g} = \bar{c}_1^{a_1} \cdots \bar{c}_\ell^{a_\ell}$$

with $\bar{c}_i$ simple commutators in $\{\bar{g}_1, \ldots, \bar{g}_r\}$ and

$$0 \le a_i < \exp((G/G_c)/[G/G_c, G/G_c]) = \exp(G/[G, G]) = e \ .$$

Here we have used the fact that

$$(G/G_c)/[G/G_c, G/G_c] = (G/G_c)/([G, G]/G_c) = G/[G, G] \ .$$

Now, if $\bar{g}_i$ is chosen in G as $\bar{g}_i = g_i w_i$, $w_i \in G_c$, any $g \in G$ can be written as

$$g = c_1^{a_1} \cdots c_\ell^{a_\ell} w, \quad w \in G_c \ .$$

Furthermore $w = z_1^{u_1} \cdots z_k^{u_k}$ with z_i simple commutators of length c and u_i non-negative integers. For any such $z = [x, y]$ with x a simple commutator of length $c - 1$ and $y \in \{g_1, \ldots, g_r\}$, $z^e = [x, y]^e = [x^e, y] = 1$. Hence, we can assume that $0 \le u_i < e$. This shows that the inductive assumption passes from class $c - 1$ to class c. It also proves the claimed upper bound: there are at most $r + r^2 + \cdots + r^c$ simple commutators and any of them involves at most 2^{c+1} generators.

For the lower bound, check the bound in $G/[G, G]$.

Remark. We have chosen to work with simple commutators for clarity. The proof goes through as stated with what Marshall Hall

[26], Chapter 11, calls basic commutators. There are fewer of these so the constant in the upper bound can be slightly improved.

It may happen that the automorphism group of a p-group G acts transitively on minimal generating sets. In this case, the diameter is the same for any minimal generating set. This happens for the Heisenberg group mod (p). We now give another example.

Example 6.C.d: The Burnside group $B(3, r)$. The group $B(3,r)$ is the largest group generated by r elements and whose exponent is 3. It is known that this group is finite, nilpotent of class 3 and has order $3^{t(r)}$ with $t(r) = r + \binom{r}{2} + \binom{r}{3}$. Here, $m = \overline{m} = r$ and the maximum diameter D is bounded above by r^3. This can easily be shown by using the collection formula (e.g. [26], pg 178–182) and the fact that the automorphism group of $B(3,r)$ acts transitively on the sets of all generating r-tuples.

7 Examples of walks on generating sets

This section applies the tools developed above to some examples. We treat fixed G with n large, cyclic groups, p-groups of small class, the Burnside group $B(3,r)$, metacyclic groups and the symmetric group. Throughout, for G a finite group, we write m for the minimum size of a generating set and $\overline{m}$ for the maximum size of a minimal generating set (Section 6.A). We write M for the number of generating m-tuples in G and D for the maximum diameter of G (Section 6.C). Finally π is the uniform distribution on the set of all generating n-tuples.

Example 1: Fixed G, large n. As a first result, our estimates show that for any fixed finite G and all sufficiently large n, order $n^2 \log n$ steps suffice for convergence.

Theorem 7.1 *There is an explicit function $C(N) > 0$ such that, for any finite group G of order at most N, for any $n \geq 2m + \overline{m}$, for any $c > 0$ and $\ell = C(N)n^2(\log n + c)$, the Markov chain P defined at* (3.1) *satisfies*

$$\|P_x^\ell - \pi\|_{\mathrm{TV}} \leq 2\mathrm{e}^{-c}$$

for any starting state x.

Proof. Theorem 2.2 shows that, for every x,

$$\|P_x^\ell - \pi\|_{\mathrm{TV}} \leq 2\mathrm{e}^{-c} \quad \text{for } \ell \geq 1 + \frac{c}{1-\beta} + \frac{1}{4\alpha}\log\log\frac{1}{\pi(x)}$$

with $\beta = \max\{\beta_1, |\beta_{\min}|\}$ and α the log-Sobolev constant. Proposition 4.2 bounds

$$\beta_{\min} \geq -1 + \frac{n - \overline{m}}{n^2|G|(2D+1)} \quad .$$

Corollary 5.3 shows that

$$\beta - 1 \leq 1 - \frac{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}M^2}{20(1+2m)^2\binom{n}{m}\binom{n-m}{m}|G|^{2m}D^2n^2} \quad .$$

These bounds show that there is $C_1(N)$ such that, if G has order at most N,

$$\beta \leq 1 - \frac{C_1(N)}{n^2} \quad .$$

Corollary 5.3 further bounds

$$\alpha(P) \geq \frac{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}M^2(|G|-2)}{20(1+2m)^2\binom{n}{m}\binom{n-m}{m}|G|^{2m+1}\log(|G|-1)D^2n^2} \geq \frac{C_2(N)}{n^2} \quad .$$

Here we have used the very crude bounds $m, \overline{m}, D \leq |G|$, $M \geq 1$. Also $\pi(x) \geq \frac{1}{|G|^n}$ so $\log\log\frac{1}{\pi(x)} \leq C_3(N)\log n$. Combining bounds completes the proof.

Remarks. 1. A crude bound on $C(N)$ is $(20N)^{3N+6}$. Theorem 3.2 gives a more precise result.

2. The best lower bound we know is $c(G)n\log n$. Indeed, this many steps are required to have a good chance of hitting every coordinate (from the classical coupon collectors result). We conjecture that this is the right answer. The only case where this has been proved is for $G = \mathbb{Z}/2\mathbb{Z}$ [12].

3. The above result can be refined in special cases to give more explicit constants. We now turn to these developments.

Example 2: The cyclic group $G = \mathbb{Z}/r\mathbb{Z}$. Roughly we show that order $(n^2\log n)r^2(\log r)^a$ steps suffice for convergence (for some a).

Theorem 7.2 *Let $G = \mathbb{Z}/r\mathbb{Z}$ with $r = \prod_1^k p_i^{a_i}$ the prime decomposition of r.*

1. *For $n \geq 2k$* and $c > 0$, *there is a constant A such that*,

$$\|P_x^\ell - \pi\|_{\mathrm{TV}} \leq 2\mathrm{e}^{-c} \quad \textit{for} \quad \ell \geq A(nr)^2[(\log r)(\log\log r^n) + c] \quad .$$

2. *For $n = k + 2$ and $c > 0$, working in continuous time, there is a constant A such that*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \le \mathrm{e}^{1-c} \quad \textit{for} \quad t \ge Ar^2(\log r)^4(\log\log r)^2 \\ ((\log r)(\log\log r) + c) \ .$$

Proof. For $G = \mathbb{Z}/r\mathbb{Z}$, $m = 1$, $\overline{m} = k$, and

$$M = r\prod_1^k \left(1 - p_i^{-1}\right) \ .$$

As shown in Section 6.A,

$$\liminf_{|G|\to\infty} M/|G| = 0$$

but there exists a constant c_1 such that

$$M \ge \frac{c_1|G|}{\log\log|G|} \ . \tag{7.1}$$

For part (1), we use Theorem 5.5 with $m_* = 2$. Lemma 6.3 shows that $M_* = f(2, G)$ satisfies $M_* \ge c_2 r^2$ for some constant c_2. The maximum diameter is bounded by $r/2$. Using these estimates in Theorem 5.5 yields

$$\beta_1 \le 1 - \frac{C_1}{(nr)^2} \quad \text{and} \quad \alpha \ge \frac{C_2}{(nr)^2\log r}$$

with constants C_1, C_2. Proposition 4.2 bounds the least eigenvalue $\beta_{\min}$ from below by $-1 + \frac{n-k}{n^2r^2}$. Hence, Theorem 2.2 yields the claim in part (1).

For part (2), use Theorem 3.2, the lower bound (7.1) and the obvious estimate $2^k \le r$.

Remark. We conjecture that $(\log r)^a n\log n$ steps are enough for convergence in both cases (for some $a > 0$). This conjecture is proved in [20] when n is fixed and r is a prime. This is the only case where the kind of dramatic speedup reported by Celler et al. has been proved.

Example 3: *p*-groups. Our results are fairly sharp when applied to p-groups of bounded class and number of generators. We have developed things in continuous time.

Theorem 7.3 *Let G be a p-group with Frattini-rank and class both bounded by N. There is a function $A(N) > 0$ such that for all $n \geq 3N$ and $c > 0$, the chain* (3.1) *satisfies*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \leq \mathrm{e}^{1-c} \quad \text{for} \quad t \geq A(N) n^2 p^{2\omega} [(\log p)(\log\log p) + c]$$

where $p^\omega = \exp(G/[G,G])$.

Proof. Theorem 6.4 shows that the maximum diameter is of order p^ω up to constant multiples depending on N. For p-groups, $m = \overline{m} = b$. Corollary 5.3 shows that there are $A_1(N), A_2(N)$ such that

$$\beta_1(P) \leq 1 - \frac{A_1(N)}{n^2 p^{2\omega}} \quad \text{and} \quad \alpha(P) \geq \frac{A_2(N)}{n^2 p^{2\omega} \log p} \ .$$

The bound for α uses $|G| \leq p^{A_3(N)}$ for some constant $A_3(N)$ (of order $N^{O(N)}$). Using these ingredients in Theorem 3.2 completes the proof.

The following result gives a bound for general p-groups in terms of the maximum diameter D. This implies Corollary 3.4 above. The proof is the same as for Theorem 7.3.

Theorem 7.4 *For any p-group of Frattini-rank b and all $n \geq 3b$, $c > 0$, the chain* (3.1) *satisfies*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \leq \mathrm{e}^{1-c} \quad \textit{for} \ \ t \geq 320(1+2b)^{b+2}(nD)^2 \\ ((\log|G|)(\log\log|G|^n) + 4c) \ .$$

Remarks. 1. Specialize to the Heisenberg group mod p (say p is an odd prime). Then, $\omega = 1$, $b = m = \overline{m} = 2$ and D is of order p. The theorems show that order $(pn)^2$ steps suffice (up to logarithmic factors). Any element has order p which is odd, so that we can use Proposition 4.3 to bound the least eigenvalue from below by $-1 + \frac{2}{np^2}$. This can be used to show

$$\|P_x^\ell - \pi\|_{\mathrm{TV}} \leq 2\mathrm{e}^{-c} \quad \text{for} \quad \ell \geq 10^4 (np)^2 \left[(\log p^3)(\log\log p^{3n}) + 4c\right]$$

for any $c > 0$.

2. Similar results can be obtained for nilpotent groups.

3. In all cases above we conjecture that for $n \geq 3b$, order $(\log|G|)^a \, n \log n$ steps suffice for convergence with small universal a.

Example 4: The Burnside group $B(3, r)$. Let $G = B(3,r)$ be the Burnside group; this is generated by r elements and all elements in G

have order 3. Every r-generator exponent 3 group is a homomorphic image of $B(3,r)$. M. Hall [26] contains a clear description showing that $B(3,r)$ is finite, nilpotent of class 3 and has order $3^{t(r)}$ with $t(r) = r + \binom{r}{2} + \binom{r}{3}$.

Theorem 7.5 *Let $G = B(3,r)$. For $n = r^2, r \geq 3$ and $c > 0$, the chain* (3.1) *satisfies*

$$\|P_x^\ell - \pi\|_{\mathrm{TV}} \leq 2\mathrm{e}^{-c} \quad \text{for} \quad \ell \geq Ar^{15}(\log r + c)$$

with an explicit constant A.

Proof. Here, $m = \overline{m} = r$ and the maximum diameter D is bounded by r^3. By Lemma 6.2 the number M of generating r-tuples is bounded below by $M \geq \frac{1}{2}|G|^r$. Applying Proposition 4.3, we find that the least eigenvalue is bounded by $-1 + \frac{2}{9n}$. Further, for any $n \geq 3r$, Corollary 5.3 yields

$$\beta_1(P) \leq 1 - \frac{720\binom{n-2r}{r}}{\binom{n}{r} r^8 n^2} \quad \text{and} \quad \alpha(P) \geq \frac{720\binom{n-2r}{r}}{\binom{n}{r} r^{11} n^2}.$$

If $n = 3r$ these bounds are exponentially bad in r. For $n = r^2$, $r \geq 3$, they give

$$\beta_1(P) \leq 1 - \frac{1}{7200r^{12}} \quad \text{and} \quad \alpha(P) \geq \frac{1}{7200r^{15}} \quad .$$

These use $\binom{r^2-2r}{r} \Big/ \binom{r^2}{r} \geq \left[1 - \frac{2r-1}{r^2-r+1}\right]^r \geq \mathrm{e}^{-2} \geq 10^{-1}$ which in turn follows from $\log(1-u) \geq -2u$ for $0 < u < 5/7$). Using these ingredients in Theorem 2.2 completes the proof.

Remark. It is straightforward to show that at least $r^5/\log r$ steps are necessary for convergence.

Example 5: Metacyclic groups. For p, q primes satisfying $q|(p-1)$, let $G = H(p,q)$ be the unique non-Abelian group of order pq. The group $H(p,q)$ is the semidirect product of $\mathbb{Z}/p\mathbb{Z}$ by $\mathbb{Z}/q\mathbb{Z}$. Here, all minimal generating sets have two elements so that $m = \overline{m} = 2$. Further, there exists a numerical constant c such that $M \geq c|G|^2$. For q fixed and p large, the maximum diameter is of order $p^{1/(q-1)}$. See Examples 6.A.e, 6.C.b. Corollary 5.2 yields

$$\beta_1(P) \leq 1 - \frac{A}{n^2 p^{2/(q-1)}}.$$

Theorem 3.2 shows that a running time of order $n^2 p^{2/(q-1)}[\log p][\log\log p^n]$ suffices to reach uniformity.

For $p^\epsilon < q$, $\epsilon > 1/2$, the diameter D is $O(q)$. Hence

$$\beta_1(P) \leq 1 - \frac{A}{(nq)^2}$$

and Theorem 3.2 shows that a time of order $(nq)^2[\log q][\log\log q^n]$ suffices. We conjecture that, for $G = H(p,q)$, a running time of order $(\log|G|)^a n \log n$ suffices for convergence for a universal constant a.

Example 6: The symmetric group. For the symmetric group S_d on d letters we have the following.

Theorem 7.6 *Let $G = S_d$. For $n \geq \overline{m} + 4$ and $c > 0$, the semigroup $H_t = \mathrm{e}^{-t(I-P)}$ associated to the chain P at* (3.1) *satisfies*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \leq \mathrm{e}^{1-c}$$

for

$$t = \frac{An^6 D^2}{(n-\overline{m})^4}[(\log(d!))(\log\log(d!)^n) + c] \ .$$

In particular, for $n \geq 3d$ *and* $c > 0$,

$$\|H_t^x - \pi\|_{\mathrm{TV}} \leq \mathrm{e}^{1-c}$$

for $t = A(nD)^2[(d\log d)(\log n) + c]$. Here D is the maximum diameter of S_d and A is a constant independent of n and d.

Proof. It is plain that $m(S_d) = 2$. Further (see Example 6.B.c), almost 3/4 of the pairs in $S_d \times S_d$ generate S_d. Thus, $M \geq a(d!)^2$ for some numerical constant a. For $n \geq \overline{m} + 4$, Theorem 3.2 gives the announced result. One also knows (see Example 6.A.b) that $d - 1 \leq \overline{m} \leq 2d$, hence the result for $n \geq 3d$.

Remark. In the special case where $n = 3d$, Corollary 3.5 follows. In this case, a running time of order $d^3 D^2 (\log d)^2$ suffices. Using $D = O(\mathrm{e}^{\sqrt{n\log n}})$ (cf., 6.C.a) gives a running time subexponential in d. It is conjectured that $D = O(d^A)$ for some A. On this conjecture, our estimates give a polynomial bound on the running time. It is easy to show that order d^2 steps are necessary in this case. Roughly, this analysis also works for classical finite simple groups.

8 Remarks on non-reversible versions of the walk

The chain proposed by Celler et al. [9] in their algorithm for generating random elements of a finite group G is not exactly the chain P at (3.1) studied in the previous sections but a close cousin that we will call $\widetilde{P}$. To describe $\widetilde{P}$, consider again the set of all generating n-tuples of group elements. Let S be a generating set with $|S| < n$. Start the walk by labeling the first $|S|$ coordinates with the elements of S and label all the remaining coordinates with the identity. The basic step of their walk is as follows: Pick a pair of coordinates (u, v) uniformly at random and multiply the group element at u by the element at v, either on the right or on the left, each with probability $1/2$.

To be more precise, denote by $g(v)$ the vth coordinate. The chains P and $\widetilde{P}$ differ in the following way: Once the ordered pair (u, v) of coordinates has been chosen uniformly at random the mechanism for P is

$$\text{replace } g(u) \text{ by } \begin{cases} \text{either} & g(u)g(v) \\ \text{or} & g(u)g(v)^{-1} \end{cases} \text{ each with equal probabilty}$$

whereas the mechanism for $\widetilde{P}$ is

$$\text{replace } g(u) \text{ by } \begin{cases} \text{either} & g(u)g(v) \\ \text{or} & g(v)g(u) \end{cases} \text{ each with equal probabilty } .$$

From a practical point of view the algorithm $\widetilde{P}$ has the important advantage that it does not require computing $g(v)^{-1}$. From a theoretical view-point the main difference between these two chains is that P is reversible whereas $\widetilde{P}$ is not if $G \neq (\mathbb{Z}/2\mathbb{Z})^k$. Both chains have the uniform distribution as their stationary measure because they are doubly stochastic (see below).

To describe formally the chain $\widetilde{P}$, we introduce some notation. For $x, y \in \mathscr{X} = G^n$, write

$$x \simeq y \text{ if } \begin{cases} x \text{ and } y \text{ differ exactly in one coordinate, say } x_i \neq y_i, \\ \text{and there exists } j \neq i \text{ such that } y_i = x_i x_j \text{ or } x_j x_i \ . \end{cases}$$

If $x \simeq y$ with $x_i \neq y_i$, let

$$\begin{aligned} \widetilde{N}(x, y) \ &= (\text{the number of } j \text{ such that } x_i^{-1} y_i = x_j) \\ &\quad + (\text{the number of } j \text{ such that } \ y_i x_i^{-1} = x_j) \ . \end{aligned}$$

Finally, let $N(x)$ be the number of coordinates equal to the identity. Then

$$\widetilde{P}(x,y) = \begin{cases} 0 & \text{if } x \not\simeq y \text{ and } x \neq y \\ \frac{\tilde{N}(x,y)}{2n(n-1)} & \text{if } x \simeq y \\ \frac{N(x)}{n} & \text{if } x = y \ . \end{cases} \tag{8.2}$$

It is useful to write $\widetilde{P}$ as the sum of two pieces, a right piece R and a left piece L defined as follows. The basic steps for the chains R, L are the same as before except that, after picking (u, v) uniformly at random, we replace $g(u)$ by $g(u)g(v)$ for the chain R and by $g(v)g(u)$ for the chain L. Thus,

$$\widetilde{P} = \tfrac{1}{2}(R + L) \ . \tag{8.3}$$

Finally, consider the chains R^*, L^* where, after picking (u, v) uniformly at random, we replace $g(u)$ by $g(u)g(v)^{-1}$ for the chain R^* and by $g(v)^{-1}g(u)$ for the chain L^*.

It is easy to verify that R, L are doubly stochastic with adjoint (i.e., transpose) R^*, L^* ("adjoint" refers to the space $\ell^2(\mathscr{X})$ whereas "transpose" refers to the matrices). This of course implies that $\widetilde{P}$ is doubly stochastic with adjoint $\widetilde{P}^* = \frac{1}{2}(R^* + L^*)$.

Further, $P = \frac{1}{2}(R + R^*)$. There is obviously a "left" version of P which is equal to $\frac{1}{2}(L + L^*)$. Lemma 3.1 generalizes (with the same proof) as follows.

Lemma 8.1 *Let G be a finite group and $n \geq m(G) + \overline{m}(G)$. Then, the chains*

$$R, L, R^*, L^*, \widetilde{P}, P$$

are all irreducible, doubly stochastic, aperiodic Markov chains on the set $\mathscr{X} \subset G^n$ of n-tuples $(x_1, \ldots, x_n)$ which generate G. They all have the uniform distribution $\pi(x) = |\mathscr{X}|^{-1}$ as stationary probability.

We show below that all our other results also extend to the chains $R, L, R^*, L^*, \widetilde{P}$. In particular, we have

Theorem 8.2 *Let K denote any one of the chains $R, L, R^*, L^*, \widetilde{P}$. For any group G, all $n \geq 2m(G) + \overline{m}(G)$ and all $c > 0$, the semigroup $H_t = \mathrm{e}^{-t(I-K)}$ associated to the chain K on generating n-tuples satisfies*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \leq \mathrm{e}^{1-c}$$

for

$$t = \frac{40(1+2m)^2 \binom{n}{m}\binom{n-m}{m}|G|^{2m+1}D^2n^2}{\binom{n-\overline{m}}{m}\binom{n-\overline{m}-m}{m}M^2(|G|-2)}[(\log(|G|-1))(\log\log|G|^n)+2c] \ .$$

Here $M = M(G)$ is the number of distinct generating m-tuples and $D = D(G)$ is the maximum diameter of G.

Proof. First, we observe that the definitions (2.2)–(2.4) make perfect sense for non-reversible chains. This allows us to extend the definitions of $\beta_1(P)$ and $\alpha(P)$ to non-reversible chains (note that $\beta_1(P)$ is not, in general, an eigenvalue of P). Now, there is a version of Theorem 2.3 for non-reversible chains. Namely, (see [18], Theorem 3.7).

Theorem 8.3 *Let P be a finite Markov chain with stationary measure π. Then, for all $c > 0$,*

$$\|H_t^x - \pi\|_{\mathrm{TV}} \le \mathrm{e}^{1-c} \quad \text{for } t = \frac{c}{1-\beta_1} + \frac{1}{2\alpha}\log\log\frac{1}{\pi(x)}.$$

To finish the proof when $K = R$ or R^*, we only have to observe that

$$\mathscr{E}_R = \mathscr{E}_{R^*} = \mathscr{E}_P \ .$$

This follows readily from the definition

$$\mathscr{E}_R(f,f) = \frac{1}{2}\sum_{x,y}|f(x)-f(y)|^2R(x,y)\pi(x)$$

because $R^*(x,y)\pi(x) = R(y,x)\pi(y)$ and $P(x,y) = \frac{1}{2}(R(x,y)+R^*(x,y))$. Of course, all the results obtained for P also holds for the left version of P and the above analysis works for L and L^*. Finally, the desired result for $\widetilde{P} = \frac{1}{2}(R+L)$ follows.

Similar results can be obtained in **discrete time** for the chains $K_+ = \frac{1}{2}(I+K)$ where K is one of the chains R, L, R^*, L^* or $\widetilde{P}$ and I is the identity matrix. For this, use Theorem 3.7' of [18].

References

[1] L. Babai: The probability of generating the symmetric group. J. Comb. Theory. A., **52**, 148–153 (1989)

[2] L. Babai: Local expansion of vertex-transitive graphs and random generation in finite groups. Proc. 23rd ACM STOC. 164–174 (1991)

[3] L. Babai: On the length of subgroup chains in the symmetric group Comm. Alg., **14**, 1729–1736 (1986)
[4] L. Babai: Randomization in Group Algorithms: Conceptual Questions. Preprint (1996)
[5] L. Babai., G. Hetyei., W.M. Kantor., A. Lubotzky., Á. Seress: On the diameter of finite groups. Foundations of Computer Science, 31st Annual Symposium, IEEE, 857–865 (1990)
[6] L. Babai., Á. Seress: On the diameter of Cayley graphs of the symmetric group. J. Comb. Theory, A, **49**, 175–179 (1988)
[7] L. Babai., and Á. Seress: On the diameter of permutation groups. European J. Comb. **13**, 231–243 (1992)
[8] P., Cameron., R. Solomon., A. Turull: Chains of subgroups in symmetric groups. Jour. Alg. **127**, 340–352 (1989)
[9] F. Celler., C. Leedham-Green., S. Murray., A. Niemeyer., E. O'Brien: Generating random elements of a finite group. Communications in Algebra, **23**, 4931–4948 (1995)
[10] G. Constantine: Combinatorial theory and statistical design. Wiley, New York (1987)
[11] F. Chung., R. Graham: Random walks on generating sets for finite groups. Electronic Journal of Combinatorics, **2**, no. R7 (1997)
[12] F. Chung., R. Graham: Stratified Random walks on an n-cube. Random Structures and Algorithms, **1**, 199–222 (1997)
[13] P. Diaconis: Group representations in probability and statistics. IMS, Hayward (1986)
[14] P. Diaconis., R.L. Graham R.L: The graph of generating sets of an Abelian group. Preprint, Dept. of Math. Harvard University (1995)
[15] P. Diaconis., S. Holmes: Statistical testing of computer algorithms for generating random elements of finite groups (1996)
[16] P. Diaconis., L. Saloff-Coste: Comparison theorems for reversible Markov chains. Ann. Appl. Prob.**3**, 696–730 (1993)
[17] P. Diaconis., L. Saloff-Coste: Comparison techniques for random walk on finite groups. Ann. Prob. **21**, 2131–2156 (1993)
[18] P. Diaconis., L. Saloff-Coste: Logarithmic Sobolev inequalities and finite Markov chains. Ann. Appl. Prob. **6**, 695–750 (1996)
[19] P. Diaconis., L. Saloff-Coste: Moderate growth and random walk on finite groups. G.A.F.A., **4**, 1–36 (1992)
[20] P. Diaconis., L. Saloff-Coste: Walks on generating sets of Abelian groups. Prob. Th. Rel. Fields.**105**, 393–421 (1996)
[21] P. Diaconis., L. Saloff-Coste: Random walks on finite groups: a survey of analytical techniques. In Probability on groups and related structures XI, Heyer (ed). World Scientific (1995)
[22] P. Diaconis., M. Sahshahani: The subgroup algorithm for generating uniform random variables. Probl. in Engin. Info. Sci., **1**, 15–32 (1987)
[23] P. Diaconis.,D. Stroock: Geometric bounds for eigenvalues for Markov chains. Ann. Appl. Prob. **1**, 36–61 (1991)
[24] L. Finkelstein., W. Kantor: Groups and Computation. Amer. Math. Soc. Providence (1993)
[25] W. Gaschütz: Die Eulersche funktion endlicher auflösbarer gruppen. Illinois J. Math., **3**, 469–476 (1959)

[26] M. Hall: The theory of groups. (Sec. Ed., 1976) Chelsea Publishing Company. New York (1959)
[27] P. Hall: The Eulerian functions of a group. Quart. J. Math. **7**, 134–151 (1936)
[28] G. Hardy., E. Wright: An Introduction to the theory of numbers. (Fifth ed.), Oxford University Press (1938)
[29] T. Hawkes., M. Issacs and M. Özaydin: On the Möbius function of a finite group. Rocky Mountain J. Math. **19**, 1003–1034 (1989)
[30] D. Holt and S. Rees: An implementation of the Neumann-Praeger algorithm for the recognition of special linear groups. Jour. Exper. Math. **1**, 237–242 (1992)
[31] M. Issacs: The number of generators of a linear p-group. Can. J. Math. **24**, 851–858 (1972)
[32] P. Neumann., S. Praeger: A recognition algorithm for special linear groups. Proc. London Math. Soc. **65**, 555–603 (1992)
[33] L. Pyber: Asymptotic results for permutation groups. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. **11**, 197–219 (1993)
[34] A. Sinclair: Algorithms for random generation and counting: a Markov chain approach. Birkhäuser, Boston (1993)
[35] M. Suzuki: Group theory I,II. Springer, New York (1982)