Mixing in non-quasirandom groups

W. T. Gowers

Emanuele Viola*

November 5, 2020

Abstract

We initiate a systematic study of mixing in non-quasirandom groups. Let A and B be two independent, high-entropy distributions over a group G. We show that the product distribution AB is statistically close to the distribution $F(AB)$ for several choices of G and F, including:

(1) G is the affine group of 2×2 matrices, and F sets the top-right matrix entry to a uniform value,

(2) G is the lamplighter group, that is the wreath product of $\mathbb{Z}_2$ and $\mathbb{Z}_n$, and F is multiplication by a certain subgroup,

(3) G is H^n where H is non-abelian, and F selects a uniform coordinate and takes a uniform conjugate of it.

The obtained bounds for (1) and (2) are tight.

This work is motivated by and applied to problems in communication complexity. We consider the 3-party communication problem of deciding if the product of three group elements multiplies to the identity. We prove lower bounds for the groups above, which are tight for the affine and the lamplighter groups.

*Email: viola@ccs.neu.edu. Supported by NSF grant CCF-1813930.

1 Introduction and our results

Computing the product of elements from a group is a fundamental problem in theoretical computer science that arises and has been studied in a variety of works including [KMR66, Mix89, BC92, IL95, BGKL03, PRS97, Amb96, AL00, Raz00, MV13, Mil14, GV19, Sha16], some of which are discussed more below. In this work we study this problem in the model of *communication complexity* [Yao79, KN97, RY19]. Previous work in this area [MV13, GV19] has found applications in cryptography, specifically to the construction of leakage-resilient circuits [MV13], and mathematics [Sha16].

We consider the following basic communication problem. Each of several parties receives an element from a finite group G. The parties need to decide if the product of their elements is equal to 1_G. They have access to public randomness, and can err with constant probability say $1/100$. For two parties, this is the *equality* problem (because $ab = 1_G$ iff $a = b^{-1}$) and can be solved with constant communication. Thus the first interesting case is for 3 parties.

Definition 1. We denote by $R_3(G)$ the randomized 3-party communication complexity of deciding if $abc = 1_G$, where the first party receives a, the second b, and the third c.

The simplest efficient protocol is over $G = \mathbb{Z}_2^n$. The parties use the public randomness to select a *linear hash function* $f_S : \mathbb{Z}_2^n \to \mathbb{Z}_2$ defined as $f_S(x) = \sum_{i \in S} x_i \mod 2$. The parties then send $f_S(a), f_S(b), f_S(c)$ and compute $f_S(a) + f_S(b) + f_S(c) = f_S(a+b+c)$. The latter is always 0 if $a+b+c = 0$, while it is 0 with probability $1/2$ over the choice of S if $a+b+c \neq 0$. By repeating the test a bounded number of times, one can make the failure probability less than 1%. This shows $R_3(\mathbb{Z}_2^n) = O(1)$. Throughout this paper $O(.)$ and $\Omega(.)$ denote absolute constants.

The communication is also constant over the cyclic group $\mathbb{Z}_n$ of integers modulo n: $R_3(\mathbb{Z}_n) = O(1)$ [Vio14]. But this is a bit more involved, because linear hash functions (with small range) do not exist. One can use instead a hash function which is *almost linear*. Such a hash function was analyzed in the work [DHKP97] and has found many other applications, for example to the study of the 3SUM problem [BDP08, Păt10].

The above raises the following natural question: For which groups G is $R_3(G)$ small?

It is fairly straightforward to prove lower bounds on $R_3(G)$ when G is *quasirandom* [Gow08], a type of group that is discussed more in detail below. Such lower bounds for $R_3(G)$ appear in the survey [Vio19] and also follow from the results in this paper (using what we later call the kernel method).

In this paper we prove lower bounds for groups to which the results for quasirandom groups do not apply. The groups we consider are natural, and they were considered before in the computer science literature, for example in the context of expander graphs [Wig10, LMR15, Zha17] and low-distortion embeddings [LNP09, ANV10]. We also complement the lower bounds with some new upper bounds. These results are closely related to the study of *mixing* in groups. We discuss these two perspectives in turn.

1.1 Communication complexity

To set the stage, we begin by discussing upper bounds on $R_3(G)$. We show that for any abelian group G we have $R_3(G) = O(1)$. This result generalizes the results for $\mathbb{Z}_2^n$ and $\mathbb{Z}_n$

mentioned earlier. More generally we can prove upper bounds for groups which contain large abelian subgroups, or that have *irreps* of bounded dimension. Here and throughout, *irrep* is short for *irreducible representation*. Representation theory plays a key role in this paper and is reviewed later.

Theorem 2. *We have the following upper bounds on $R_3(G)$:*
(1) Suppose G is abelian. Then $R_3(G) = O(1)$
(2) Suppose H is a subgroup of G. Then $R_3(G) \leq O(|G|/|H| + R_3(H))$.
(3) Suppose every irrep of G has dimension $\leq c$. Then $R_3(G) \leq c'$ where c' depends only on c.

Our main results are lower bounds. We show that for several groups that are "slightly less abelian" than those covered in Theorem 2 the value R_3 is large. First, we prove tight bounds for the *affine* group.

Definition 3. The *affine* group over the field $\mathbb{F}_q$ with q elements is denoted by $\mathrm{Aff}(q)$. This is the group of invertible affine transformations $x \to ax + b$ where $a, b \in \mathbb{F}_q$ and $a \neq 0$. Equivalently, it is the group of matrices $\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$ where $a \neq 0$. Note $|\mathrm{Aff}(q)| = q(q-1)$.

Theorem 4. $R_3(\mathrm{Aff}(q)) = \Theta(\log |\mathrm{Aff}(q)|)$.

The upper bound is trivial since for any group G the input length is $O(\log |G|)$.

Then we consider the so-called finite *lamplighter group*. This group is obtained from $\mathbb{Z}_2^n$ by adding a "shift" of the coordinates, formally by taking the *wreath product* $\wr$ of $\mathbb{Z}_2$ and $\mathbb{Z}_n$.

Definition 5. The finite *lamplighter* group is $L_n := \mathbb{Z}_2 \wr \mathbb{Z}_n$. Elements of L_n can be written as $(x_0, x_1, \ldots, x_{n-1}; s)$ where $x_i \in \mathbb{Z}_2$ and $s \in \mathbb{Z}_n$ and we have $(x_0, x_1, \ldots, x_{n-1}; s) \cdot (x'_0, x'_1, \ldots, x'_{n-1}; s') = (x_0 + x'_{0+s}, x_1 + x'_{1+s}, \ldots, x_{n-1} + x'_{n-1+s}; s + s')$ where addition is modulo n. For $(x; s) \in L_n$ we call x the $\mathbb{Z}_2^n$ part and s the $\mathbb{Z}_n$ part. Note $|L_n| = 2^n \cdot n$.

In other words, when multiplying $(x; s)$ and $(x'; s')$ we first shift x' by s, and then we sum component-wise. We prove a tight communication bound for $R_3(L_n)$.

Theorem 6. $R_3(L_n) = \Theta(\log \log |L_n|)$.

The upper bound is as follows. The parties can first communicate the $\mathbb{Z}_n$ parts. This takes $O(\log n) = O(\log \log |L_n|)$ communication. Then the parties can shift their $\mathbb{Z}_2^n$ parts privately, and finally use the constant-communication protocol for $\mathbb{Z}_2^n$.

We then move to groups of the form H^n. An interesting setting is when $|H|$ is small compared to n, say H has constant size.

Theorem 7. *Let H be a non-abelian group. Then $R_3(H^n) = \Omega(\log n)$.*

It is an interesting open question whether a bound of $\Omega(n)$ holds. We note that for the corresponding 4-party problem of deciding if $abcd = 1_G$ such an $\Omega(n)$ bound can be established by a reduction from lower bounds for *disjointness*. The proof proceeds by encoding the And of two bits by a group product of length four, see [Vio19]. However, those techniques do not seem to apply to the three-party problem, and appear unrelated to mixing.

1.2 Mixing in groups

At a high level, mixing refers to the general phenomenon that when we have several independent, high-entropy distributions over a group and we combine them in natural ways, for example by multiplying, the resulting random variable becomes closer to the uniform distribution, closer than the original distributions are. Our notion of (non) entropy of a distribution A is the *collision probability* $\mathbb{P}[A = A']$ where A and A' are independent and identically distributed. We define next a scaled version which is more convenient.

Definition 8. The *scaled collision probability* of a distribution A over G is $N(A) := |G|\mathbb{P}[A = A']$, where A and A' are independent and identically distributed. Equivalently, $N(A) = (|G| \|A\|_2)^2$ where $\|A\|_2$ is the L_2 norm $\sqrt{\mathbb{E}_x \mathbb{P}[A = x]^2}$.

To illustrate the normalization, note that $N(A) \leq |G|$ and it can be shown $N(A) \geq 1$. If A is uniform over a set of size $\delta|G|$ we have $N(A) = \delta^{-1}$. The uniform distribution has $\delta = 1$ and $N = 1$, the distribution concentrated on a single point has $\delta = 1/|G|$ and $N = |G|$. Distributions that are uniform on a constant fraction of the elements have $N \leq O(1)$; in the latter setting the main ideas in this paper are already at work, so one can focus on it while reading the paper.

To measure the distance between distributions we use *total variation distance.*

Definition 9. The total variation distance between distributions A and B is $\Delta(A, B) = \sum_x |\mathbb{P}[A = x] - \mathbb{P}[B = x]|$. Equivalently, $\Delta(A, B)$ is the ℓ_1 norm $\sum_x |f(x)|$ of the function $f(x) = \mathbb{P}[A = x] - \mathbb{P}[B = x]$.

We can now illustrate a basic result about mixing. Suppose that A and B are independent random variables over a group G such that $N(A)$ and $N(B)$ are $O(1)$. We would like to show that the random variable AB is close to the uniform distribution U over G. This is false for example over the group $\mathbb{Z}_2^n$. Indeed, A and B could each be the uniform distribution where the first coordinate is 0, and then AB would be the same as A, which has $\Delta(A, U) \geq \Omega(1)$.

Remarkably, however, for other groups one can show that $\Delta(AB, U)$ is small. We state this fundamental result next.

Theorem 10. *Let A and B be two independent random variables over G. We have*

$$\Delta(AB, U) \leq \sqrt{\frac{N(A)N(B)}{d}},$$

where d is the minimum dimension of a non-trivial irrep of G.

This theorem appears in equivalent form as Lemma 3.2 in [Gow08]. The formulation above appears in [BNP08]. Other proofs were discovered later, and the result is now considered folklore. The importance of this result is that for several groups the value d is large, and so the theorem shows that AB is close to U. In particular, for non-abelian *simple* groups we have that d grows with the size of the group, and for the *special-linear* group $SL_2(q)$ d is polynomial in the size of the group. For more discussion and pointers, we refer the reader to Section 13 in [Gow17] and to the original paper [Gow08]. The latter calls *quasirandom* the groups that have a large d.

In this work we consider several groups for which one cannot prove a good bound on $\Delta(AB, U)$ for every two independent distributions with small N. In particular, the group has an irrep of small dimension. The question arises of what type of mixing result, if any, makes sense.

Our approach to mixing Our approach is to show that even though $\Delta(AB, U)$ might be large, nevertheless AB acquires some "invariance property" of U which the distributions A and B in isolation may not have. One natural property of U is that it is invariant under multiplication by a fixed element: for any $y \in G$ we have that yU and U are the same distribution. So a first attempt is to say that G mixes if there exists a non-identity element y such that $\Delta(AB, yAB)$ is small, for any independent A and B with small N.

We show that this is indeed the case for the affine and the lamplighter group.

However, for groups like H^n this notion cannot be met: for any fixed $y \neq 1_G$, one can define A and B which fix one coordinate i where $y_i \neq 1_H$ and are uniform on the others; these distributions have small N but $\Delta(AB, yAB)$ is large. To overcome this obstacle we will use randomness in our definition of the invariance property.

In the special case that H does not have irreps of dimension one, we show that AB is almost invariant under selecting a uniform coordinate and replacing that coordinate with a uniform element. In other words, if Y is the uniform distribution over H^n obtained by setting a uniformly selected coordinate to a uniform element in H and the others to 1 then $\Delta(AB, YAB)$ is small. For general non-abelian H, which might have a unidimensional irrep, this does not work. For example, if $H = H' \times \mathbb{Z}_2$ we cannot change the $\mathbb{Z}_2$ part. Rather than replacing a coordinate with a uniform element, we take a uniform conjugate. That is, we show that $\Delta(AB, YABY^{-1})$ is small where Y is as before.

To capture these various possibilities, we say that the group mixes if there exists a distribution F on functions from G to G such that $\Delta(AB, F(AB))$ is small. For example, F could be the (fixed, deterministic) function $F(x) = yx$ corresponding to multiplication by a fixed element y. Over a group of the form H^n, F could be the random function $F(x) = YxY^{-1}$ which selects a uniform coordinate and takes a uniform conjugate of that coordinate.

Intuitively, in all these cases AB becomes somewhat uniform in the sense that it doesn't change much if we apply F to it. Of course for this to be of any use we need that $F(AB) \neq AB$ often. We have arrived to the following definition.

Definition 11. A group G is (ϵ, β)-mixing for (scaled collision probability) $N \leq \eta$ if there exists a distribution F on functions from G to G such that for every distributions A and B with $N(A), N(B) \leq \eta$ we have:
(1) $\Delta(AB, F(AB)) \leq \epsilon$, and
(2) $\mathbb{P}[F(AB) = AB] \leq \beta$.
We also say that G mixes *via* F.

Another important motivation for this definition is given by the following result which links our notion of mixing to communication lower bounds.

Lemma 12. *Suppose a group G is $(\epsilon, 0.99)$-mixing for $N \leq 1/\epsilon$. Then $R_3(G) \geq \Omega(\log(1/\epsilon))$.*

The communication lower bounds in the previous section are obtained by establishing mixing results and then using this Lemma 12. We also use this lemma in the contrapositive: by the communication upper bounds from Theorem 2 we obtain non-mixing results. As evident from the statement of the lemma, for the application to communication complexity the setting $\epsilon = \eta$ in Definition 11 suffices, but below we state the more general tradeoff.

The above definition of mixing can be considered "least-useful." It is a bare-minimum notion that in particular suffices for the communication lower bounds. It is also natural to try to prove a "most-general" mixing result by identifying F such that $F(x)$ has the largest possible entropy. In several cases, our results also identify such F. This also gives additional information in the communication lower bounds. As the proofs will show, the communication lower bounds will establish that the parties, on input $a, b, c \in G$, cannot distinguish $c = (ab)^{-1}$ from $c = F((ab)^{-1})$. Thus understanding via what functions F the group mixes is useful in understanding what information about the product abc the parties can compute.

We now state our mixing results. First we obtain a mixing result for the affine group.

Theorem 13. *The affine group* $\mathrm{Aff}(q)$ *is* $(O(s/\sqrt{q}), 0)$ *mixing for* $N \leq s$ *via*

$$F(x) := \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} \cdot x$$

for any $u \neq 0$.

The error parameter $O(s/\sqrt{q})$ is tight up to polynomials, as the size of the group is $q(q-1)$. Specifically, $\mathrm{Aff}(q)$ is not $(s/q^c, 0.99)$-mixing for $N \leq s$ for some constant c. This result also achieves a "most general" mixing in terms of F. Note that the matrices $\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}$ with $u \in \mathbb{F}_q$ form a subgroup H of $\mathrm{Aff}(q)$, in fact the additive group of $\mathbb{F}_q$. In particular the theorem gives $(O(s/\sqrt{q}), 1/q)$-mixing via $F(x) := Hx$, where Hx stands for multiplying x by a uniform element from H, and the $1/q$ is to account for the probability that $u = 0$. In turn, note that for any $a, b \in \mathbb{F}_q$ we have

$$H \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & U \\ 0 & 1 \end{pmatrix}$$

where U is the uniform distribution over $\mathbb{F}_q$. Thus, the theorem is saying that for any high-entropy distributions A and B, the distribution AB is close to the distribution obtained from AB by replacing the top-right entry with a uniform element in $\mathbb{F}_q$. This result is the strongest possible in the sense that the top-left entry of AB cannot be changed by F with noticeable probability. This is because that entry is the multiplicative group of $\mathbb{F}_q$, an abelian group which does not have mixing, as follows from Theorem 2 and Lemma 12.

Then we obtain a mixing result for the lamplighter group.

Theorem 14. *The lamplighter group* L_n *is* $(O(s/n^{1/4}), 0)$ *mixing for* $N \leq s$ *via*

$$F(x) := y \cdot x$$

where $y \in L_n$ *depends only on* n.

The error parameter $O(s/n^{1/4})$ is tight up to polynomials. As mentioned earlier, $R_3(L_n) = O(\log n)$ and hence for some constant c the group L_n is not $(1/n^c, 0.99)$-mixing for $N \leq n^c$ by Lemma 12.

As in Theorem 13, the group L_n also mixes via $F(x) = Hx$ where H is the uniform distribution over a subgroup. The definition of H depends on the prime factorization of n. The simplest case is when n is prime. In that case H is the subgroup $\{(z;0) : \sum_i z_i = 0 \mod 2\}$ and note that for any $(x;s) \in L_n$ we have

$$H(x;s) = (Z;s)$$

where Z is uniform over $\mathbb{Z}_2^n$ conditioned on $\sum_i Z_i = \sum_i x_i \mod 2$. Thus, the theorem for n prime is saying that for any high-entropy A and B, the distribution AB is close to the distribution obtained from AB by replacing the $\mathbb{Z}_2^n$ part x (i.e., $AB = (x;s)$) with a uniform element with the same parity as x. This result is strongest possible in the sense that $F(x;s)$ must preserve both the parity of x and the value s with high probability. One way to see this is to note that if F changes either the parity of x or s with high probability then the parties can in fact distinguish inputs of the form $a, b, (ab)^{-1}$ from those of the form $a, b, F((ab)^{-1})$. To do so, the parties can send the parities of the $\mathbb{Z}_2^n$ parts, and can use the efficient protocol for the $\mathbb{Z}_n$ part.

Then we consider direct-product groups H^n. We show that we have mixing for any non-abelian H. Mixing occurs via taking a random coordinate and computing a uniform conjugate of that coordinate.

Theorem 15. *Let H be a non-abelian group. The group H^n is $(O(s^{2/3}/n^{1/3}), 0.99)$ mixing for $N \leq s$ via*

$$F(x_1, x_2, \ldots, x_n) := (x_1, x_2, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n),$$

where $i \in \{1, 2, \ldots, n\}$ and $u \in H$ are uniform.

The error cannot be improved to $o(1/n)$ even for $N = |H|$, as A and B can just fix a coordinate. But an interesting question is whether the bound on N can be increased to exponential.

Under the stronger assumption that H does not have an irrep of dimension one we improve the bound in several respects, none of which affects the communication results. First, instead of taking a random conjugate of a coordinate we can simply set that coordinate to uniform. Second, we improve the error to about $1/\sqrt{n}$. And third, we show that the bound still holds if one distribution has exponential N (see the proof for this statement).

Theorem 16. *Let H be a group with no non-trivial irrep of dimension one. The group H^n is $(O(s\sqrt{\log(sn)}/\sqrt{n}), 1/|H|)$ mixing for $N \leq s$ via*

$$F(x_1, x_2, \ldots, x_n) := (x_1, x_2, \ldots, x_{i-1}, x_i u, x_{i+1}, \ldots, x_n),$$

where $i \in \{1, 2, \ldots, n\}$ and $u \in H$ are uniform.

The smallest group H with no non-trivial irrep of dimension one is the alternating group on five elements, of size 60.

1.3 Techniques for mixing results, and organization

Our main tool for the mixing results is *non-abelian Fourier analysis*, which we review in Section 2. We prove in Section 3 that (the probability mass function of) AB can be approximated by a function whose Fourier coefficients are *few* and have *small dimension*. Then we give different ways in which this fact can be exploited. First, we show that if the *intersection of the kernels of irreps of small dimension* is non-trivial, then we can take F to be multiplication by any non-identity element in that intersection. We call this method the *kernel* method, presented in Section 4. Using known facts about the representation theory of the affine group, Theorem 13 is proved in Section 4.1. For the lamplighter group we also use known facts about its representation theory, and we show that the small-dimensional representations lie, in a suitable sense, within a small-dimensional vector space. This is done in Section 4.2

Note that the kernel $K = \{k \in G : \rho(k) = I\}$ of an irrep ρ, where I is the identity matrix, is a *normal subgroup* of G. (The latter means that $g^{-1}kg \in K$ for every $k \in K$ and $g \in G$, which is true because $\rho(g^{-1}kg) = \rho(g^{-1})\rho(k)\rho(g) = \rho(g^{-1}g) = I$.) In particular, the intersection of kernels is also a normal subgroup, and it is in fact known that all normal subgroups arise in this way. Hence, the kernel method shows that $\Delta(AB, HAB)$ is small, where H is the uniform distribution over a normal subgroup. The applicability of the method hinges on our understanding of what normal subgroups arise when considering intersection of kernels of irreps of *bounded dimension*.

The kernel method cannot be applied to groups of the form H^n. For such groups, we use the fact that the irreps ρ of H^n are *tensor products* $\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$ of irreps ρ_i of H, and in particular the dimension of ρ is the product of the dimensions of the ρ_i. Then the key observation is that *low-dimensional irreps of H^n must be tensor products of mostly one-dimensional ρ_i*. And then we use the fact that unidimensional irreps are constant on conjugacy classes. In the special case that H does not have irreps of dimension one we can conclude the stronger fact that most ρ_i are trivial. And then we can get the refined result by extending a well-known Fourier expression for average sensitivity to the non-abelian setting. This appears in Section 5.

We briefly comment on how we prove the communication upper bounds (or equivalently the non-mixing results) in Theorem 2. The proof is in Section 6. Item (1) builds on the result for $\mathbb{Z}_n$ that we mentioned earlier and is obtained using the characterization of abelian groups, the Chinese remainder theorem, and hashing. Item (2) uses the random self-reducibility of the $abc = 1_G$ problem together with efficient protocols for disjointness. While (3) follows from (2) and (1) and a known characterization of groups whose irreps all have bounded dimension.

Finally, the proof of Lemma 12 is in Section 7.

1.4 Open problems

This work raises several interesting questions. First, can we characterize groups which admit non-trivial mixing? (We can define non-trivial as (ϵ, β)-mixing for $N = \omega(1)$ where ϵ and β are bounded constants.) We ask whether a group G has non-trivial mixing if and only if G has irreps of unbounded dimension. Note that we prove the "only if" direction in this work.

Can we prove this at least for some important classes of groups? Can we characterize the groups for which the kernel method suffices?

Another question is whether the bound on N in the H^n results can be improved to exponential, for both distributions. This points to the interesting question of discovering suitable generalizations of classical results in additive combinatorics, such as the Freiman-Ruzsa theorem, for groups of the form H^n.

It would also be interesting to study if the results in this paper can be extended to the *number-on-forehead* [CFL83] model. The study of group products in this model could lead to the solution of several outstanding problems. For example, it is conjectured in [GV19] that computing the product of many elements is hard even for more than logarithmically many parties (a well-known barrier, see e.g. [Vio17]). Moreover, the problem of computing the product of just three elements could also lead to stronger separations between deterministic and randomized communication. Specifically, it is pointed out in [Vio19] that the "corners" result in [Aus16] can be used to obtain a separation whose parameters match the state-of-the-art [BDPW10] but hold for a different function. And as remarked in [Aus16] stronger results could be within reach. For an exposition of the relevant result in [Aus16] see [Vio19]. Can the results for interleaved products in [GV19] or for "corners" in [Aus16] be suitably extended to other groups such as those in this paper? Those groups might be easier to understand than quasirandom groups, possibly leading to improved results.

2 Non-abelian Fourier analysis

The books by Serre [Ser77], Diaconis [Dia88], and Terras [Ter99] are good references for representation theory and non-abelian Fourier analysis. The Barbados notes [Wig10] and Section 13 of [Gow17] provide briefer introductions. The exposition in these sources is not always consistent, and often has different aims from ours. So let us give a quick account of the theory that is most relevant for this work.

Matrices. Let M be a square complex matrix. We denote by $\overline{M}$ the conjugate of M, by M^T the transpose of M, and by M^* the conjugate transpose $\overline{M^T}$ (aka adjoint, Hermitian conjugate, etc.). The matrix M is *unitary* if $M^{-1} = M^*$.

The *Hilbert-Schmidt operator* (or Frobenius norm) of M is

$$\|M\|_{HS}^2 := \sum_{i,j} |M_{i,j}|^2 = \mathrm{tr}(MM^*).$$

To verify the latter note that

$$\mathrm{tr}(MM^*) = \sum_i (MM^*)_{i,i} = \sum_i \sum_k M_{i,k} M^*_{k,i} = \sum_i \sum_k M_{i,k} \overline{M}_{i,k} = \sum_{i,k} |M_{i,k}|^2.$$

If $M = AB$ we have

$$\|M\|_{HS}^2 = \sum_{i,j} |\sum_k A_{i,k} B_{k,j}|^2 \leq \sum_{i,j} (\sum_k |A|_{i,k}^2)(\sum_k |B|_{k,j}^2) = \|A\|_{HS}^2 \, \|B\|_{HS}^2 \,, \qquad (1)$$

where the inequality is Cauchy-Schwarz.

Representation theory. Let G be a group. A *representation* ρ of G with dimension d maps elements of G to $d \times d$ unitary, complex matrices so that $\rho(xy) = \rho(x)\rho(y)$. Thus ρ is a homomorphism from G to the group of linear transformations of the vector space $\mathbb{C}^d$. We denote by d_ρ the dimension of ρ.

If there is a non-trivial subspace W of $\mathbb{C}^d$ that is invariant under ρ (that is $\rho(x)W \subseteq W$ for every $x \in G$) then ρ is *reducible*; otherwise it is *irreducible.* Irreducible representations are abbreviated *irreps* and play a critical role in Fourier analysis. We denote by $\widehat{G}$ a complete set of inequivalent irreducible representations of G.

Fourier analysis Let $f : G \to \mathbb{C}$. We use the L norms: $\|f\|_2^2 = \mathbb{E}_x |f(x)|^2$. Note that

$$N(p) = (|G| \, \|p\|_2)^2.$$

The ρ *Fourier coefficient* of f is

$$\widehat{f}(\rho) := \mathbb{E}_x f(x)\overline{\rho(x)}.$$

The Fourier inversion formula is then

$$f(x) = \sum_\rho d_\rho \mathrm{tr}(\widehat{f}(\rho)\rho(x)^T),$$

where tr is the trace and ρ ranges over $\widehat{G}$, here and below, unless specified otherwise.

We define the *convolution* as follows (which is off by a factor of $|G|$ from some texts):

$$p * q(x) := \sum_y p(y)q(y^{-1}x).$$

Note that if p and q are distributions then $p * q$ is the distribution obtained by sampling x from p, y from q, and then outputting xy.

We note that under this normalization we have

$$\widehat{p * q}(\alpha) = |G|\hat{p}(\alpha)\hat{q}(\alpha).$$

Parseval's equality is

$$\mathbb{E} f(x)\overline{g(x)} = \sum_\rho d_\rho tr(\hat{f}(\rho)\hat{g}(\rho)^*).$$

In case $f = g$ this becomes

$$\mathbb{E}|f(x)|^2 = \sum_\rho d_\rho tr(\hat{f}(\rho)\hat{f}(\rho)^*) = \sum_\rho d_\rho \left\|\hat{f}(\rho)\right\|_{HS}^2.$$

3 Fourier truncation

Let p and q be distributions over a group G. We show that $p * q$ is well-approximated by a function with few Fourier coefficients. Specifically we just take certain "heavy" coefficients.

Before stating the lemma for arbitrary groups it is instructive to see it in the basic setting of $\mathbb{Z}_2^n$. Here the irreps are $\rho(x) = \chi_\alpha(x) = (-1)^{\sum_i \alpha_i x_i}$ where $\alpha, x \in \{0,1\}^n$.

Lemma 17. *Let p and q be distributions over $G = \mathbb{Z}_2^n$. Let $R = \{\alpha : \hat{p}^2(\alpha) \geq \theta^2/|G|^2\}$. We have $\Delta(p * q, \sum_{\alpha \in R} \widehat{p * q}(\alpha)\chi_\alpha) \leq \sqrt{N(q)}\theta$. Moreover $|R| \leq N(p)/\theta^2$.*

To make sense of the parameters note that if $N(p) = O(1)$ then the bound on Δ is $O(\theta)$ and that on $|R|$ is $O(1/\theta)^2$.

Once we have this statement, the verification is straightforward.

Proof. The bound on $|R|$ follows from Parseval because $\|p\|_2^2 = \sum_\alpha \hat{p}^2(\alpha) \geq |R|\theta^2/|G|^2$.

For the bound on Δ let $f := \sum_{\alpha \notin R} \widehat{p * q}(\alpha)\chi_\alpha$. We have

$$\sum_x f(x) = |G|\mathbb{E}_x f(x) \leq |G|\sqrt{\mathbb{E}_x f(x)^2} = |G|\sqrt{\sum_{\alpha \notin R} \hat{f}(\alpha)^2} = |G|\sqrt{\sum_{\alpha \notin R} p \hat{*} q(\alpha)^2}.$$

Now we use the fact that $p \hat{*} q(\alpha) = |G|\hat{p}(\alpha)\hat{q}(\alpha)$, and the bound on $\hat{p}$ to get

$$\sum_x |f(x)| \leq |G|^2 \sqrt{\sum_{\alpha \notin R} \hat{p}(\alpha)^2 \hat{q}(\alpha)^2} \leq |G|\theta \sqrt{\sum_{\alpha \notin R} \hat{q}(\alpha)^2} \leq |G|\theta \sqrt{\sum_\alpha \hat{q}(\alpha)^2} \leq |G|\theta \|q\|_2 .$$

The last inequality is Parseval's. □

One can expect that this proof generalizes to any group, and indeed it does.

Lemma 18. *Let p and q be distributions over a group G. Let $R = \{\rho : \|\hat{p}(\rho)\|_{HS}^2 \geq \theta^2/|G|^2\}$. Let $f := \sum_{\rho \in R} \mathrm{tr}(\widehat{p * q}(\rho)\rho(.)^T)$. We have $\Delta(p * q, f) \leq \sqrt{N(q)}\theta$. Moreover $\sum_{\rho \in R} d_\rho \leq N(p)/\theta^2$.*

Proof. The bound on $\sum_{\rho \in R} d_\rho$ follows from Parseval's fact that $\|p\|_2^2 = \sum_\rho d_\rho \|\hat{p}(\rho)\|_{HS}^2$. The latter is $\geq \sum_{\rho \in R} d_\rho \theta^2/|G|^2$, and the bound follows.

For the bound on Δ let $e(x) := p * q(x) - f(x) = \sum_{\rho \notin R} \mathrm{tr}(\widehat{p * q}(\rho)\rho(x)^T)$. We seek to bound $\sum_x |e(x)|$. First we bound the L_2 norm. By Parseval we have

$$\|e\|_2^2 = \sum_{\rho \notin R} d_\rho \|\hat{e}(\rho)\|_{HS}^2 .$$

Now we use the fact that $\hat{e}(\rho) = \widehat{p * q}(\rho) = |G|\hat{p}(\rho)\hat{q}(\rho)$. Then we use Equation (1) in the preliminaries to obtain

$$\|\hat{e}(\rho)\|_{HS}^2 = |G|^2 \|p(\rho)\hat{q}(\rho)\|_{HS}^2 \leq |G|^2 \|\hat{p}(\rho)\|_{HS}^2 \|\hat{q}(\rho)\|_{HS}^2 .$$

For every $\rho \notin R$ we have $\|\hat{p}(\rho)\|_{HS}^2 \leq \theta^2/|G|^2$, and so $\|\hat{e}(\rho)\|_{HS}^2 \leq \theta^2 \|\hat{q}(\rho)\|_{HS}^2$. This gives

$$\|e\|_2^2 \leq \theta^2 \sum_\rho d_\rho \|\hat{q}(\rho)\|_{HS}^2 = \theta^2 \|q\|_2^2 ,$$

using Parseval once again.

Hence we can bound

$$\sum_x |e(x)| \leq |G| \|e\|_2 \leq \theta |G| \|q\|_2 = \theta\sqrt{N(q)}.$$

□

4 The kernel method

In this section we first develop the kernel method in the next lemma, and then we apply it to the affine and lamplighter groups.

Lemma 19. *Suppose that an element $y \in G$ belongs to the kernel of every irrep of G of dimension $\leq t$. Then G is $(2s/\sqrt{t}, 0)$-mixing for $N \leq s$ via $F(x) := y \cdot x$.*

Proof. We prove it for $F(x) = y^{-1}x$, which is the same since each kernel is a subgroup. Let p and q be two distributions with $N(p)$ and $N(q) \leq s$. For a function f denote by f_y the function $f_y(x) := f(yx)$. We need to bound $\Delta(p * q, (p * q)_y)$. We apply Lemma 18 with $\theta := \sqrt{N(p)/t}$. Let f be the corresponding function. Then we can bound

$$\begin{aligned}\Delta(p * q, (p * q)_y) &\leq \Delta(p * q, f) + \Delta(f, f_y) + \Delta(f_y, (p * q)_y) \\ &\leq 2\Delta(p * q, f) + \Delta(f, f_y).\end{aligned}$$

By Lemma 18 the first term is at most $2\sqrt{N(q)}\theta = 2\sqrt{N(q)N(p)/t} = 2s/\sqrt{t}$. To bound the second term we use the fact that f only has representations ρ in R, where R is as in Lemma 18. Because $\sum_{\rho \in R} d_\rho \leq N(p)/\theta^2 = t$, every value d_ρ with $\rho \in R$ is at most t. By Fourier inversion we have:

$$\Delta(f, f_y) = \sum_x |f(x) - f(yx)| = \sum_x \left| \sum_{\rho \in R} d_\rho \left(\mathrm{tr} \hat{f}(\rho)\rho(x)^T - tr\hat{f}(\rho)\rho(yx)^T \right) \right|.$$

Because ρ is a representation we have $\rho(yx) = \rho(y) \cdot \rho(x)$, and by the assumption on y we have $\rho(y) = I$, and so $\rho(yx) = \rho(x)$. Hence each term in the inner sum is 0, and $\Delta(f, f_y) = 0$. □

4.1 Affine group

In this section we prove Theorem 13, the mixing result for the affine group $\mathrm{Aff}(q)$. We use the kernel method, so we want to find a non-identity element y that belongs to the kernel of low-dimensional representations of $\mathrm{Aff}(q)$.

The irreps of $Aff(q)$ are given explicitly for example in [Ter99].

Lemma 20. *[Ter99], Page 273. The group $\mathrm{Aff}(q)$ has $q-1$ irreps of dimension 1 of the form $\rho \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = \chi(a)$, where χ is a character of the multiplicative group of $\mathbb{F}_q$, and one irrep of dimension $q-1$. This includes all the irreps.*

Because $\chi(1) = 1$, we have that for any $b \in \mathbb{F}_q$ the element $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$ belongs to the kernel of any representation of dimension $\leq q-2$. By Lemma 19, $\mathrm{Aff}(q)$ is $(2s/\sqrt{q-2}, 0)$-mixing for $N \leq s$.

4.2 Lamplighter group

In this section we prove Theorem 14, the mixing result for the lamplighter group L_n. We use the kernel method, so we want to find a non-identity element y that belongs to the kernel of low-dimensional representations of L_n.

The representation theory of L_n is given in [TFF09] (see also the book [TFF14]). To state their result we need some notation. Every string $\theta \in \{0,1\}^n$ has a *period* $t(\theta)$ which is the smallest integer t such that $\theta_i = \theta_{i+t}$ for every i, where the $+$ works modulo n. We denote by Θ_t a set of representatives for the orbits under $\mathbb{Z}_n$ of the strings with period t. In other words, every string with period t is in Θ_t up to a shift.

For example, for $n = 4$ we have the following representatives

$$\begin{aligned}\Theta_1 &= \{0000, 1111\},\\ \Theta_2 &= \{1010\},\\ \Theta_4 &= \{0001, 0011, 0111\}.\end{aligned}$$

For $s \in \mathbb{Z}_n$ and $\theta \in \mathbb{Z}_2^n$ we denote by $s\theta$ the vector θ shifted by s, that is $(s\theta)_i = \theta_{i-s}$.

Lemma 21. *[TFF09] For every t, every $\theta \in \Theta_t$, and every $r \in \{0, 1, \ldots, n/t - 1\}$ the group L_n has a $t \times t$ irrep ρ defined as:*

$$\rho(x;k)_{s,j} = \begin{cases} 0 & \text{if } t \text{ does not divide } k + j - s \\ (-1)^{\langle s\theta, x\rangle} \cdot e^{2\pi \boldsymbol{i} r(k+j-s)/n} & \text{otherwise,} \end{cases}$$

where $s, j \in \{0, 1, \ldots, t-1\}$, $\boldsymbol{i} = \sqrt{-1}$, and $\langle a, b\rangle = \sum_i a_i b_i$. This includes all the irreps.

We shall only consider $k = 0$ in which case the matrix $\rho(x;k)$ is diagonal (since t can only divide $j - s$ if $j = s$) and we have

$$\rho(x;0)_{s,s} = (-1)^{\langle s\theta, x\rangle} \cdot e^{2\pi \boldsymbol{i} r 0/n} = (-1)^{\langle s\theta, x\rangle}.$$

Note that these matrices do not depend on r anymore.

Now we investigate the kernels of these matrices. Fix t. Note that each vector $s\theta$ where $\theta \in \Theta_t$ has period t. The key to obtaining tight bounds is the following observation.

Claim 22. The vectors in $\mathbb{Z}_2^n$ with period t are contained in a vector space of dimension t.

Proof. Any vector of period t is in the span of the t vectors $0^i 10^{t-1} 10^{t-1} \cdots 10^{t-1-i}$ for $i = 0, 1, \ldots, t-1$. □

Hence, the set of vectors whose period is $\leq t$ is contained in a vector space of dimension $\leq \sum_{i \leq t} i < t^2$. Therefore, as long as $t \leq \sqrt{n}$ there exists a vector y' which is orthogonal to any vector with period $\leq t$. This implies that $y = (y'; 0)$ is in the intersection of the kernels of the representations of dimension $\leq t$, because $\rho(y';0)_{s,s} = (-1)^0 = 1$, and so $\rho(y';0) = I$. Appealing to Lemma 19 concludes the proof.

The case n prime Because the period of a string divides n, if n is prime the period is either $t = n$ or $t = 1$. The strings of period 1 are the all-zero and the all-one string. Thus, for any vector y' with parity 0, we have that $y = (y'; 0)$ is in the intersection of the kernels of the representations of dimension $< n$.

5 Product groups

In this section we prove theorems 15 and 16: mixing results for groups of the form H^n. We begin with the proof of Theorem 15 and present the proof of the other theorem in subsection 5.1.

The next key lemma shows that every low-dimensional irrep of H^n can be written as a product of unidimensional irreps of H and one irrep of H^n that depends on few coordinates.

Lemma 23. *Let ρ be an irrep of H^n of dimension d_ρ. Then there are a set $S \subseteq \{1, 2, \ldots, n\}$ with $|\bar{S}| \leq \log_2 d_\rho$, unidimensional irreps ρ_i for $i \in S$ of H, and an irrep ρ^* of $H^{|\bar{S}|}$ such that*

$$\rho(x_1, x_2, \ldots, x_n) = \rho^*(x_{\bar{S}}) \prod_{i \in S} \rho_i(x_i).$$

Proof. We use the following fact from representation theory. Any irrep of H^n is the tensor product of n irreps of H. This can be found as Theorem 10 in Section 3.2 in [Ser77] or as Theorem 9 in [Dia88]. In particular, the dimensions multiply.

Let $\rho = \otimes_{i=1}^n \rho_i$ where the ρ_i are irreps of H. We define S to be the set of indices i such that the dimension of ρ_i is ≤ 1. Because the dimensions multiply, we have

$$2^{|\bar{S}|} \leq d_\rho.$$

Letting ρ^* be the tensor product of ρ_i with $i \in \bar{S}$ concludes the proof. □

Now we use the above lemma in conjunction with Fourier truncation to claim that $p * q$ is close to a function that does not change when we conjugate most coordinates.

Lemma 24. *Let p and q be distributions on H^n. For any θ there is a function f such that $\Delta(p * q, f) \leq \sqrt{N(q)}\theta$ and for all but $N(p)/\theta^2$ coordinates i, and every u,*

$$f(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n).$$

Proof. We use Lemma 18 and let $f := \sum_{\rho \in R} \operatorname{tr}(\widehat{p * q}(\rho)\rho(.)^T)$. By Lemma 23, each ρ can be written as $\rho^*(x_{\bar{S}}) \prod_{i \in S} \rho_i(x_i)$ where $|\bar{S}| \leq \log_2 d_\rho$. Note that the ρ_i are constant on conjugacy classes since they map to $\mathbb{C}$: $\rho_i(u^{-1}x_i u) = \rho_i(u^{-1})\rho_i(x_i)\rho_i(u) = \rho_i(x_i)$.

Summing all the $\rho \in R$, the number of coordinates that change with conjugation is $\leq \sum_{\rho \in R} \log_2 d_\rho \leq \sum_{\rho \in R} d_\rho$. By Lemma 18 the latter quantity is at most $N(p)/\theta^2$. □

We have not used the logarithmic dependence between the dimension and the number of coordinates. A linear dependence would have been enough. We will exploit the logarithmic dependence below to give a refined bound.

The following definition corresponds to the effect of applying F.

Definition 25. Let p be the probability mass function of a distribution on H^n. We define $c(p)$ to be the probability mass function of the distribution which samples $(x_1, x_2, \ldots, x_n)$ from p and then picks a uniform i and replaces x_i with a uniform conjugate $u^{-1}x_i u$ of x_i.

We have $c(p)(x_1, x_2, \ldots, x_n) = \mathbb{E}_{i,u} p(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n)$. We aim to bound $\Delta(p * q, c(p * q))$.

Lemma 26. *Let f be the function from Lemma 24. Then*

$$\Delta(f, c(f)) \leq \left(\frac{N(p)}{\theta^2 n}\right) \cdot 2(\sqrt{N(q)}\theta + 1).$$

Proof. We have

$$\Delta(f, c(f)) = \sum_x |\mathbb{E}_{i,u\in H}(f(x) - f(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n))|.$$

By the triangle inequality this is at most

$$\mathbb{E}_{i,u\in H} \sum_x |f(x) - f(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n)|.$$

For all but $N(p)/\theta^2$ coordinates i the inner sum is 0. For any other i, by the triangle inequality the inner sum is at most

$$\sum_x |f(x)| + \mathbb{E}_{u\in H} \sum_x |f(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n)| = 2\sum_x |f(x)|.$$

From Lemma 18 we know that $\Delta(f, p * q) \leq \sqrt{N(q)}\theta$. Hence we have

$$\sum_x |f(x)| = \sum_x |f(x) - p * q(x) + p * q(x)| \leq \Delta(f, p * q) + \sum_x |p * q(x)| \leq \sqrt{N(q)}\theta + 1,$$

because $p * q$ is a probability distribution. The result follows. □

We also note that c does not increase distance.

Lemma 27. *Let $f, g : G \to \mathbb{C}$ be any functions. Then $\Delta(c(f), c(g)) \leq \Delta(f, g)$.*

Proof. By the triangle inequality:

$$\begin{aligned}\Delta(c(f), c(g)) &= \sum_x |\mathbb{E}_{i,u}(f(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n) - g(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n))| \\ &\leq \mathbb{E}_{i,u} \sum_x |f(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n) - g(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n)|.\end{aligned}$$

Fix i to maximize the expectation, and the result follows. □

So we can bound the distance as follows.

$$\begin{aligned}\Delta(p * q, c(p * q)) &\leq \Delta(p * q, f) + \Delta(f, c(f)) + \Delta(c(f), c(p * q)) \\ &\leq 2\Delta(p * q, f) + \Delta(f, c(f)) \\ &\leq 2\sqrt{N(q)}\theta + O(N(p)/\theta^2 n)(\sqrt{N(q)}\theta + 1).\end{aligned}$$

Setting $\theta^3 = \frac{N(p)}{n\sqrt{N(q)}}$ we obtain distance

$$O(N(p)N(q)/n)^{1/3}.$$

It remains to show that $F(AB) \neq AB$ often. First we note that $N(p * q) \leq N(p)$. To see this, let A, A' be distributed according to p and let B and B' be distributed according to q. Then

$$N(p * q) = |G|\mathbb{P}[AB = A'B'].$$

We can fix the outcomes of B and B' to maximize this probability. Hence there is an element $g \in G$ such that

$$N(p * q) \leq |G|\mathbb{P}[A = gA'].$$

The latter probability equals

$$\sum_x p(x)p(g^{-1}x) \leq \sqrt{\sum_x p^2(x)}\sqrt{\sum_x p^2(g^{-1}x)} = \sqrt{\sum_x p^2(x)}\sqrt{\sum_x p^2(x)} = \mathbb{P}[A = A'],$$

where the inequality is Cauchy-Schwarz. Hence $N(p * q) \leq |G|\mathbb{P}[A = A'] = N(p)$.

Now that we have a bound on $N(p * q)$ we can apply the following lemma.

Lemma 28. *There is $\alpha > 0$ such that the following holds:*

Let H be a non-abelian group. Let p be a distribution over H^n with $N(p) \leq (1 + \alpha)^n$. With probability ≥ 0.01 over x sampled from p, i uniform in $\{1, 2, \ldots, n\}$ and u uniform in H we have

$$(x_1, \ldots, x_{i-1}, u^{-1}x_i u, x_{i+1}, \ldots, x_n) \neq (x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n).$$

For the proof we need some basic concepts from group theory. For an element $h \in H$ the *centralizer* of h is the set $\{x : hx = xh\}$ of elements that commute with h. The *center* of H is the set $Z(H)$ of elements that commute with every element from H, that is the set of elements whose centralizer is H. Both the centralizer of h and the center of H are subgroups of H. To prove the lemma, first we show that with high probability over x, many coordinates do not belong to the center of H. When we take a uniform conjugate of that coordinate, the probability that we get the same element is at most the size of the centralizer. Since the latter is a non-trivial subgroup of H, by Lagrange's theorem that probability is at most $1/2$. Details follow.

Proof. Let Z be the center of H. Since Z is a subgroup of H, the size of Z divides the size of H. Since H is not abelian we have $|Z| \leq |H|/2$. So the complement $\bar{Z}$ of the center is at least half the group. Let $S \subseteq H^n$ be the set of tuples with at most $n/3$ coordinates in $\bar{Z}$. By a Chernoff bound we have

$$|S| \leq (c|H|)^n,$$

for some $c \leq 1 - \Omega(1)$.

The collision probability of any distribution q over S is at least $1/|S|$. This is because $\sum_{x \in S} q(x)^2 \geq |S|(\mathbb{E}_{x \in S} q(x))^2 = 1/|S|$ by Cauchy-Schwarz.

Now suppose that a sample from p lands in S with probability β. Then the collision probability of p would be at least β^2 times the probability of collision conditioned on landing in S, which is at least $1/S$ by what we just said. Hence the collision probability would be $\beta^2/|S| \geq \beta^2/(c|H|)^n$. Hence $N(p) \geq \beta^2 c^{-n}$. Under the assumption on $N(p)$ for a small enough α we have, say, $\beta \leq 1/2$.

For any sample that is not in S we have probability at least $1/3$ over i that $x_i \in \bar{Z}$ and so the centralizer of x_i is a strict subgroup of H. By Lagrange's theorem its size divides $|H|$. Hence its size is at most $|H|/2$. Hence the probability over u that $u^{-1}x_iu = x_i$ is at most $1/2$. □

5.1 No unidimensional rep case

Using some of the ideas in the previous proof, in this subsection we prove Theorem 16, a refined bound under the stronger assumption that H does not have a uni-dimensional representation. For slight convenience, if $f : H^n \to \mathbb{C}$ we write f_i for the function $f_i(x_1, \ldots, x_i \ldots x_n, y) = f(x_1, \ldots, y, \ldots x_n)$. Let v be the distribution over H^n where a uniformly selected coordinate is set to a uniform element and the others are set to 1. Our goal is to bound $\Delta(p * q, p * q * v)$. In the next lemmas we give a convenient expression for this quantity. Note that $p * v(x) = \sum_y p(y)v(y^{-1}x) = \mathbb{E}_{i,h}p_i(x, h)$.

Lemma 29. *Let $f : G \to \mathbb{C}$ be a function. Then $\Delta^2(f, f*v) \leq |G|^2(\mathbb{E}_x|f(x)|^2 - \mathbb{E}_{i,x}|\mathbb{E}_{y\in H}f_i(x, y)|^2)$.*

Proof. The LHS equals

$$\begin{aligned}
&\left(\sum_{x\in G} |f(x) - \mathbb{E}_{i,y\in H}f_i(x, y)|\right)^2 \\
&\leq \left(\sum_{x\in G} \mathbb{E}_i|f(x) - \mathbb{E}_{y\in H}f_i(x, y)|\right)^2 \\
&\leq |G|^2\mathbb{E}_{x,i}|f(x) - \mathbb{E}_{y\in H}f_i(x, y)|^2 \\
&= |G|^2\mathbb{E}_{x,i}\left(|f(x)|^2 - f(x)\overline{\mathbb{E}_{y\in H}f_i(x, y)} - \overline{f(x)}\mathbb{E}_{y\in H}f_i(x, y) + |\mathbb{E}_{y\in H}f_i(x, y)|^2\right) \\
&= |G|^2(\mathbb{E}_{x,i}|f(x)|^2 - 2\mathbb{E}_{x,i}|\mathbb{E}_{y\in H}f_i(x, y)|^2 + \mathbb{E}_{x,i}|\mathbb{E}_{y\in H}f_i(x, y)|^2) \\
&= |G|^2(\mathbb{E}_{x,i}|f(x)|^2 - \mathbb{E}_{x,i}|\mathbb{E}_{y\in H}f_i(x, y)|^2).
\end{aligned}$$

□

For $\rho = \otimes\rho_i$ we denote by $\#trivial(\rho)$ the number of trivial ρ_i (that is those equal to 1) and by $\#non - trivial$ n minus that number. The following is a non-abelian version of a well-known result in abelian Fourier analysis (Theorem 2.38 in [O'D14]).

Lemma 30. *Let $f : H^n \to \mathbb{C}$ be a function. We have*

$$\begin{aligned}
\mathbb{E}_{x,i}|\mathbb{E}_{y\in H}f_i(x, y)|^2 &= \frac{1}{n}\sum_\rho d_\rho \#trivial(\rho)\left\|\hat{f}(\rho)\right\|_{HS}^2 \\
&= \mathbb{E}|f(x)^2| - \frac{1}{n}\sum_\rho d_\rho \#non - trivial(\rho)\left\|\hat{f}(\rho)\right\|_{HS}^2.
\end{aligned}$$

Proof. Define the function $f_{-i} : H^{n-1} \to \mathbb{C}$ as

$$f_{-i}(x_1, x_2, \ldots, x_{n-1}) := \mathbb{E}_{y\in H}f(x_1, \ldots, x_{i-1}, y, x_i, \ldots, x_{n-1}).$$

Note that for any i we have

$$\mathbb{E}_x|\mathbb{E}_{y\in H}f_i(x,y)|^2 = \mathbb{E}_x|f_{-i}(x)|^2 = \sum_{\rho=\rho_1\otimes\ldots\otimes\rho_{n-1}} d_\rho \left\|\widehat{f_{-i}}(\rho)\right\|_{HS}^2$$

by Parseval. For $\rho = \rho_1\otimes\ldots\otimes\rho_{n-1}$ write ρ_{+i} for $\rho = \rho_1\otimes\cdots\rho_{i-1}\otimes 1\otimes\rho_i\otimes\ldots\otimes\rho_{n-1}$. We have

$$\widehat{f_{-i}}(\rho) = \mathbb{E}_{x\in H^{n-1}}f_{-i}(x)\overline{\rho(x)} = \mathbb{E}_{x\in H^{n-1},y\in H}f_i(x,y)\overline{\rho(x)} = \mathbb{E}_{x\in H^n}f(x)\overline{\rho_{+i}(x)} = \hat{f}(\rho_{+i}).$$

Averaging over i corresponds to summing over all representations in $\widehat{H^n}$ and multiplying each by the number of trivial components, and dividing everything by n.

This proves the first equality, and the second follows by Parseval. □

Let now f and R be given by Lemma 18. Putting the above results together we obtain

$$\Delta(f, f*v)^2 \leq |G|^2\frac{1}{n}\sum_{\rho\in R} d_\rho \#non-trivial(\rho)\left\|\hat{f}(\rho)\right\|_{HS}^2.$$

Reasoning as in the proof of Theorem 15, we know that $d_\rho \leq N(p)/\theta^2$ and so

$$\#non-trivial(\rho) \leq \log_2(N(p)/\theta^2)$$

because we are assuming that non-trivial irreps have dimension ≥ 2, and dimensions multiply in tensor products of representations. Hence we get

$$\Delta(f, f*v)^2 \leq |G|^2\frac{1}{n}\log(N(p)/\theta^2)\sum_\rho \left\|\hat{f}(\rho)\right\|_{HS}^2 \leq \frac{1}{n}\log(N(p)/\theta^2)N(p*q) \leq \frac{1}{n}\log(N(p)/\theta^2)N(q).$$

Here we are using the definition of f, Parseval, and the fact $N(p*q)\leq N(q)$.

Again as before, we get the bound

$$\begin{aligned}\Delta(p*q, p*q*v) &\leq 2\Delta(p*q, f) + \Delta(f, f*v)\\ &\leq 2N(q)\theta + \sqrt{\frac{1}{n}\log(N(p)/\theta^2)N(q)}.\end{aligned}$$

Setting say $\theta = 1/n$ gives the desired distance.

It remains to argue that $F(AB)\neq AB$ often. This probability is exactly $1/|H|$ (the chance of getting 1_G in v).

Finally, it is apparent from the bound on Δ that it remains non-trivial even for exponential $N(p)$.

6 Upper bounds on $R_3(G)$

In this section we prove Theorem 2, one item at the time.

Lemma 31. *Let G be a finite abelian group. Then $R_3(G) = O(1)$.*

Proof. For the proof we establish a series of simple reductions which reduce the problem in groups of a type (i) to the problem in groups of type $(i+1)$, where the last type is that of cyclic groups $\mathbb{Z}_m$ which is solved in Theorem 15 in [Vio14]. Next are the types. Throughout, the p_i are prime numbers and the the e_i are positive integers.

(1) abelian
(2) $\times_i \mathbb{Z}_{p_i^{e_i}}$
(3) $\times_i \mathbb{Z}_{p_i^{e_i}}$ where $p_i = p_j \Rightarrow e_i = e_j$
(4) $\times_i \mathbb{Z}_{p_i^{e_i}}$ where the p_i are distinct
(5) $\mathbb{Z}_m$.

The reduction from (1) to (2) is the fundamental theorem of abelian groups.

To reduce (2) to (3): suppose the group has factors $\mathbb{Z}_{p^e}$ and $\mathbb{Z}_{p^{e'}}$ where $e' > e$. We can replace the first factor with $\mathbb{Z}_{p^{e'}}$: On a given input, the parties privately multiply their input in the first factor by $p^{e'-e}$. We can repeat this argument until (3) holds.

To reduce (3) to (4): suppose the group has a factor $(\mathbb{Z}_{p^e})^k$. We can replace it with a single $\mathbb{Z}_{p^e}$: Using public randomness, the parties pick a uniform subset S of $\{1, 2, \ldots, k\}$. On a given input, each party privately replaces its input $a_1, a_2, \ldots, a_k$ in the factor $(\mathbb{Z}_{p^e})^k$ with $\sum_{i \in S} a_i$. We can repeat this argument until (4) holds. If the parties' inputs sum to 0, this will continue to hold. Otherwise, with probability $\geq 1/2$ the new inputs will not sum to 0 either.

The reduction from (4) to (5) is the Chinese Remainder Theorem. □

Lemma 32. *If H is a subgroup of G then $R_3(G) \leq O(|G|/|H| + R_3(H))$.*

The proof of the lemma uses *random self-reducibility* and efficient protocols for *disjointness on small sets* [HW07], [BCK+14].

Proof. By the *random self-reducibility* of the problem, given inputs $a, b, c \in G$, the parties can use public randomness to sample uniformly distributed ABC such that $abc = ABC$. Specifically, the parties may pick r_1 and r_2 uniformly from G and privately compute their new inputs as $A = ar_1^{-1}, B = r_1 b r_2^{-1}, C = r_2 c$.

Naively, we could proceed as follows. If both $A \in H$ and $B \in H$, then the parties can execute the protocol for H. So they only have to keep sampling until that happens. This would take $\Omega(s^2)$ trials and communication in expectation, where $s := |G|/|H|$.

To obtain the stronger bound, the parties will repeat the above sampling t independent times, obtaining A_i, B_i, C_i using elements $r_{i,1}, r_{i,2}$ for $i = 1, 2, \ldots t$. It suffices for the parties to identify an index i^* such that $A_{i^*} \in H$ and $B_{i^*} \in H$ and then they can run the protocol for H as before. By picking $t = O(s)^2$ the chance that such an index does not exist is at most $1/10$.

Now the first party can compute a vector $\alpha \in \{0,1\}^t$ where $\alpha_i = 1$ iff $A_i \in H$, and the second party can compute the vector β corresponding to whether $B_i \in H$.

All the computation up until this point was private and required no communication. Now however the parties need communication to find i^*.

The expected Hamming weight of α is $t/s = O(s)$, and by Markov's inequality we have that it is at most $O(s)$ except with probability $1/10$.

It remains to find the index i^* where $\alpha_i = \beta_i = 1$. This can be solved with communication $O(s)$ [HW07], [BCK+14]. (The latter paper explicitly proves this.) □

For the last item, it is known [IP64] that if all the irreducible representations of G have dimension $\leq d$ then G has an abelian subgroup of index $f(d)$.

7 Proof of Lemma 12

This proof is similar to a proof for quasirandom groups that appears in [Vio19]. First, by repeating the protocol we can assume that we have a protocol P^{rep} whose error probability is less than 0.001. Now consider the distributions $D_0 = (a, b, (ab)^{-1})$ and $D_1 = (a, b, F((ab)^{-1}))$ where a, b are uniform and G mixes via F. By assumption we have $\mathbb{E}[P^{rep}(D_0) = 1] \geq 0.999$ and $\mathbb{E}[P^{rep}(D_1) = 1] \leq 0.001 + 0.99 < 0.999$. Hence there is a fixing of the randomness to the protocol yielding a deterministic protocol P using c bits of communication such that

$$\mathbb{P}[P(D_0)] - \mathbb{P}[P(D_1)] \geq \Omega(1).$$

By a standard argument, which can be found in [KN97] for two parties and generalizes to more parties, P can be written as the sum of 2^c "rectangles" $R_i = A_i \times B_i \times C_i$. That is, $P(x, y, z) = \sum_{i \leq 2^c} A_i(x)B_i(y)C_i(z)$ where A_i is the characteristic function of the set A_i and the same for B_i and C_i.

For any rectangle R_i where either A_i or B_i has density $\leq \epsilon$ we have $\Delta(R_i(D_0), R_i(D_1)) \leq \epsilon$ just because a and b are uniform.

Consider a rectangle where A_i and B_i have both density $\geq \epsilon$. Note that the distribution of a conditioned on $a \in A_i$ has $N \leq 1/\epsilon$, and the same for b. It follows from the mixing property that for any such rectangle we have $\Delta(R_i(D_0), R_i(D_1)) \leq \epsilon$.

Hence we obtain

$$|\mathbb{P}[P(D_0)] - \mathbb{P}[P(D_1)]| \leq O(2^c\epsilon).$$

This contradicts the previous displayed equation unless $c = \Omega(\log(1/\epsilon))$.

Acknowledgment Emanuele Viola is grateful to Peter Ivanov for stimulating discussions.

References

[AL00] Andris Ambainis and Satyanarayana V. Lokam. Improved upper bounds on the simultaneous messages complexity of the generalized addressing function. In *Latin American Symposium on Theoretical Informatics (LATIN)*, pages 207–216, 2000.

[Amb96] Andris Ambainis. Upper bounds on multiparty communication complexity of shifts. In *Symp. on Theoretical Aspects of Computer Science (STACS)*, pages 631–642, 1996.

[ANV10] Tim Austin, Assaf Naor, and Alain Valette. The euclidean distortion of the lamplighter group. *Discret. Comput. Geom.*, 44(1):55–74, 2010.

[Aus16] Tim Austin. Ajtai-Szemerédi theorems over quasirandom groups. In *Recent trends in combinatorics*, volume 159 of *IMA Vol. Math. Appl.*, pages 453–484. Springer, [Cham], 2016.

[BC92] Michael Ben-Or and Richard Cleve. Computing algebraic formulas using a constant number of registers. *SIAM J. on Computing*, 21(1):54–58, 1992.

[BCK+14] Joshua Brody, Amit Chakrabarti, Ranganath Kondapally, David P. Woodruff, and Grigory Yaroslavtsev. Beyond set disjointness: the communication complexity of finding the intersection. In Magnús M. Halldórsson and Shlomi Dolev, editors, *ACM Symposium on Principles of Distributed Computing, PODC '14, Paris, France, July 15-18, 2014*, pages 106–113. ACM, 2014.

[BDP08] Ilya Baran, Erik D. Demaine, and Mihai Pătraşcu. Subquadratic algorithms for 3sum. *Algorithmica*, 50(4):584–596, 2008.

[BDPW10] Paul Beame, Matei David, Toniann Pitassi, and Philipp Woelfel. Separating deterministic from randomized multiparty communication complexity. *Theory of Computing*, 6(1):201–225, 2010.

[BGKL03] László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication complexity of simultaneous messages. *SIAM J. on Computing*, 33(1):137–166, 2003.

[BNP08] László Babai, Nikolay Nikolov, and László Pyber. Product growth and mixing in finite groups. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 248–257, 2008.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *15th ACM Symp. on the Theory of Computing (STOC)*, pages 94–99, 1983.

[DHKP97] Martin Dietzfelbinger, Torben Hagerup, Jyrki Katajainen, and Martti Penttonen. A reliable randomized algorithm for the closest-pair problem. *J. Algorithms*, 25(1):19–51, 1997.

[Dia88] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988.

[Gow08] W. T. Gowers. Quasirandom groups. *Combinatorics, Probability & Computing*, 17(3):363–387, 2008.

[Gow17] W. T. Gowers. Generalizations of Fourier analysis, and how to apply them. *Bull. Amer. Math. Soc. (N.S.)*, 54(1):1–44, 2017.

[GV19] W. T. Gowers and Emanuele Viola. Interleaved group products. *SIAM J. on Computing*, 48(3):554–580, 2019. Special issue of FOCS 2016.

[HW07] Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, 3(1):211–219, 2007.

[IL95] Neil Immerman and Susan Landau. The complexity of iterated multiplication. *Inf. Comput.*, 116(1):103–116, 1995.

[IP64] I. M. Isaacs and D. S. Passman. Groups with representations of bounded degree. *Canadian J. Math.*, 16:299–309, 1964.

[KMR66] Kenneth Krohn, W. D. Maurer, and John Rhodes. Realizing complex Boolean functions with simple groups. *Information and Control*, 9:190–195, 1966.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[LMR15] Shachar Lovett, Cristopher Moore, and Alexander Russell. Group representations that resist random sampling. *Random Struct. Algorithms*, 47(3):605–614,

2015.

[LNP09] James R. Lee, Assaf Naor, and Yuval Peres. Trees and Markov convexity. *Geom. Funct. Anal.*, 18(5):1609–1659, 2009.

[Mil14] Eric Miles. Iterated group products and leakage resilience against NC^1. In *ACM Innovations in Theoretical Computer Science conf. (ITCS)*, 2014.

[Mix89] David A. Mix Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *J. of Computer and System Sciences*, 38(1):150–164, 1989.

[MV13] Eric Miles and Emanuele Viola. Shielding circuits with groups. In *ACM Symp. on the Theory of Computing (STOC)*, 2013.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[Păt10] Mihai Pătraşcu. Towards polynomial lower bounds for dynamic problems. In *ACM Symp. on the Theory of Computing (STOC)*, pages 603–610, 2010.

[PRS97] Pavel Pudlák, Vojtěch Rödl, and Jiří Sgall. Boolean circuits, tensor ranks, and communication complexity. *SIAM J. on Computing*, 26(3):605–633, 1997.

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[RY19] Anup Rao and Amir Yehudayoff. *Communication complexity*. 2019. https://homes.cs.washington.edu/ anuprao/pubs/book.pdf.

[Ser77] Jean Pierre Serre. *Linear Representations of Finite Groups*. Springer, 1977.

[Sha16] Aner Shalev. Mixing, communication complexity and conjectures of Gowers and Viola. *Combinatorics, Probability and Computing*, pages 1–13, 6 2016. arXiv:1601.00795.

[Ter99] Audrey Terras. *Fourier analysis on finite groups and applications*, volume 43 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999.

[TFF09] Ceccherini-Silberstein T., Scarabotti F., and Tolli F. Representation theory of wreath products of finite groups. *Journal of Mathematical Sciences*, 156(1), 2009.

[TFF14] Ceccherini-Silberstein T., Scarabotti F., and Tolli F. *Representation Theory and Harmonic Analysis of Wreath Products of Finite Groups (London Mathematical Society Lecture Note Series)*. Cambridge University Press, 2014.

[Vio14] Emanuele Viola. The communication complexity of addition. *Combinatorica*, pages 1–45, 2014.

[Vio17] Emanuele Viola. Challenges in computational lower bounds. *SIGACT News, Open Problems Column*, 48(1), 2017.

[Vio19] Emanuele Viola. Non-abelian combinatorics and communication complexity. *SIGACT News, Complexity Theory Column*, 50(3), 2019.

[Wig10] Avi Wigderson. Representation theory of finite groups, and applications. Available at http://www.math.ias.edu/∼avi/TALKS/, 2010.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *11th ACM Symp. on the Theory of Computing (STOC)*, pages 209–213, 1979.

[Zha17] Yufei Zhao. Group representations that resist worst-case sampling, 2017.