Pseudorandom Generators for Group Products

(version with the geometric proof — 2nd draft)

Michal Koucký*, Prajakta Nimbhorkar† and Pavel Pudlák‡

February 25, 2011

Abstract

We prove that the pseudorandom generator introduced by Impagliazzo et al. in [INW94] with proper choice of parameters fools group products of a given finite group G. The seed length is $O(\log n(|G|^{O(1)} + \log \frac{1}{\delta}))$, where n is the length of the word and δ is the allowed error. The result implies that the pseudorandom generator with seed length $O(\log n(2^{O(w \log w)} + \log \frac{1}{\delta}))$ fools read-once permutation branching programs of width w. As an application of the pseudorandom generator one obtains small-bias spaces for products over all finite groups [MZ09].

*Institute of Mathematics, Academy of Sciences, Prague, e-mail: koucky@math.cas.cz. Partially supported by GA ČR P202/10/0854, project No. 1M0021620808 of MŠMT ČR, Institutional Research Plan No. AV0Z10190503 and grant IAA100190902 of GA AV ČR.

†Chennai Mathematical Institute, Chennai, India. e-mail: prajakta@cmi.ac.in. The work was done when the author was a student at The Institute of Mathematical Sciences, Chennai, India. Part of the work was done while visiting Institute of Mathematics, Academy of Sciences, Prague supported by project No. 1M0021620808 of MŠMT ČR.

‡Institute of Mathematics, Academy of Sciences, Prague and Institute of Theoretical Computer Science, Prague, e-mail: pudlak@math.cas.cz. Partially supported by Institutional Research Plan No. AV0Z10190503, project No. 1M0021620808 of MŠMT ČR and grant IAA100190902 of GA AV ČR.

1 Introduction

Our result is motivated by the problem of derandomizing space bounded computations. It is well-known that for the latter problem, it suffices to find efficient constructions of pseudorandom generators for polynomial size read-once branching programs. As this still seems to be too hard, researchers in computational complexity focused on special cases of this problem. In particular, the case of oblivious read-once constant width branching programs has been extensively studied. But even this special case is still open; so branching programs with further restriction have been studied. In this paper we present pseudorandom generators for *permutation* read-once constant width branching programs.

When working with permutation read-once constant width branching programs, it is more natural to recast the problem in terms of finite groups as has been done by Meka and Zuckerman [MZ09]. Let G be a finite group and $w = (g_1, g_2, \dots, g_n)$ a string of elements of G, which we will call a *group word.* The group word w determines a probability distribution Rnd^w on G by taking products of random substrings of w. The distribution is formally defined by setting the probability $\mathrm{Rnd}^w(g)$ of an element $g \in G$ to

$$\mathrm{Rnd}^w(g) = \frac{1}{2^n}|\{(x_1, \dots, x_n) \in \{0,1\}^n|\ g = g_1^{x_1} \dots g_n^{x_n}\}|.$$

The goal of derandomization is to replace the uniform distribution on the set $\{0,1\}^n$ by a distribution efficiently generated from r random bits, where $r = O(\log n)$, so that the resulting distribution is still very close to Rnd^w for any group word of length n. A pseudorandom generator is determined by an efficiently computable function $\Gamma : \{0,1\}^r \to \{0,1\}^n$. The elements of $\{0,1\}^r$ are called *seeds* and r is the *seed length.*

In general, a pseudorandom generator is used to approximate any polynomial time computable distribution. In this paper we are interested only in the distributions of the form above for a fixed finite group. Given such a function Γ, the corresponding distribution D^w_Γ is defined by

$$D^w_\Gamma(g) = \frac{1}{2^r}|\{y \in \{0,1\}^r|\ g = g_1^{\Gamma(y)_1} \cdots g_n^{\Gamma(y)_n}\}|.$$

($\Gamma(y)_i$ are the bits of the string $\Gamma(y) \in \{0,1\}^n$.) The goal is to find pseudorandom generators Γ such that D^w_Γ approximates very well the distribution Rnd^w for every w. It is well-known that for a random function Γ and $r = O(\log n)$, the distance between Rnd^w and D^w_Γ is at most $1/n^{O(1)}$. However, prior to our work no explicit constructions with logarithmic seed length had been known that would give D^w_Γ of distance δ for arbitrarily small positive constant δ. We analyze the Impagliazzo-Nisan-Wigderson generator (in the sequel abbreviated by 'INW generator') introduced in [INW94] and show that it gives pseudorandom generators such that $\|\mathrm{Rnd}^w - D^w_\Gamma\|_\infty \le \delta$, for arbitrary $\delta > 0$, where the seed length is $O(\log n \cdot (|G|^{O(1)} + \log 1/\delta))$.

Note that except for the dependence on δ this also solves the problem of finding pseudorandom generators for bounded with permutation branching programs, because a read-once permutation branching program of width k on n inputs can be described as a group word $g_1^{x_1} \dots g_n^{x_n}$, $g_i \in \mathcal{S}_k$, where $\mathcal{S}_k$ is the symmetric group on k elements. We will explain this connection in Section 1.3.

Our result also implies a polynomial-time construction of small-bias spaces for products over all finite groups. Previous constructions were known only for abelian groups [NN93, AGHP92, ?] and solvable groups [MZ09]. This follows from Theorem 1.1 of Meka and Zuckerman [MZ09].

1.1 Comparison with previous results

There has been a series of results concerning the power that randomness gives to space-bounded computation, and the simulation of randomized logspace machines by deterministic machines (c.f. [AKS87, BNS89, Nis92, Nis94, NZ96, SZ99]). In [Sav70], it was shown that a non-deterministic space S machine can be simulated by a deterministic machine that uses S^2 space, which implies $\mathsf{RL} \subseteq \mathsf{L}^2$. This was improved to $\mathsf{BPL} \in \mathsf{L}^{3/2}$ by [SZ99]. This is the best known bound for deterministic simulation of a randomized logspace machine.

Another approach to the same problem is to construct a pseudorandom generator with a short $(O(\log n))$ seed and to replace the random string of a randomized logspace machine by the output of a pseudorandom generator. For such a machine, a pseudorandom generator with $O(\log^2 n)$ due to [Nis92] is known. Other constructions with the same seed-length are known [NZ96, INW94, RR99].

As a logspace machine can be modelled as a branching program of width and length polynomial in n, the subsequent work has been focussed on designing pseudorandom generators for branching programs of constant width, which have length polynomial in n. Due to Barrington's theorem [Bar89], it is known that this class of branching programs is the same as the class NC^1, and in fact NC^1 can be simulated by a width 5 permutation branching program. The work on pseudorandom generators for bounded-width branching programs has been restricted to *read-once* branching programs. A general motivation for looking for pseudorandom generators that fool read once branching programs is that such generators would suffice to derandomize BPL. Unfortunately, it is not known that pseudorandom generators that fool read once bounded-width branching programs would suffice to derandomize RNC^1.

For width 2 branching programs, a generator having error δ is equivalent to an δ-biased space, which can be constructed with $O(\log n + \log \delta^{-1})$ seed-length [NN93, AGHP92]. Recently, a pseudorandom generator has been given by [BV10] for width w permutation read-once branching programs, which has seed-length $O((w^4 \log\log n + \log 1/\delta)\log n)$ and by [BRRY10] for width w *regular* read-once branching programs, which has seed-length $O((\log w + \log\log n + \log 1/\delta)\log n)$. Regular branching programs are more general than permutations branching programs. In [vv10], a hitting set of polynomial size has been given for width 3 read-once branching programs.

Pseudorandom generators with seed-length $O(\log n)$ for group products were previously known only for finite cyclic groups [LRTV09, MZ09]. Our result gives a generator for all finite groups. Our seed-length depends polynomially on the order of the group whereas the previously known generators for cyclic groups have a seed-length which depends logarithmically on the order of the group. Our result also implies that the INW generator with seed-length $O(\log n(\log 1/\delta + exp(w)))$ fools permutation programs of width w. The connection between group products and permutation branching programs will be explained shortly.

1.2 A brief outline of the proof

INW generator is based on recursive application of the following construction, called the *expander product* of two pseudorandom generators. This construction uses two pseudorandom generators $\Gamma_1, \Gamma_2 : \{0,1\}^r \to \{0,1\}^n$ and a 2^d-regular expander graph F with the vertex set $\{0,1\}^r$. It produces a pseudorandom generator $\Gamma_1 \otimes_F \Gamma_2 : \{0,1\}^{r+d} \to \{0,1\}^{2n}$. In the INW generator this construction is always applied with $\Gamma_1 = \Gamma_2$.

Starting with the trivial generator, the identity on $\{0,1\}^d$, and applying the expander product k times with expanders of degree 2^d, we obtain a pseudorandom generator $\Gamma : \{0,1\}^{d(k+1)} \to \{0,1\}^{d2^k}$. Thus if d is a constant, the seed length is logarithmic in the length of the output.

It is not difficult to prove, using the well known properties of expanders, that the $D^w_{\Gamma_1\otimes_F\Gamma_2}$ approximates $D^w_{\Gamma_1\times\Gamma_2}$, the distribution produced by sampling Γ_1 and Γ_2 independently. The latter distribution can also be described as the group G *convolution* of $D^{w_1}_{\Gamma_1}$ with $D^{w_2}_{\Gamma_2}$, which is denoted by $D^{w_1}_{\Gamma_1} * D^{w_2}_{\Gamma_2}$, where $w = w_1w_2$ and $|w_1| = |w_2|$. The advantage of this description is that it is an operation on *distributions*; we do not need to know the two generators.

The error of the approximation of $D^{w_1}_{\Gamma_1} * D^{w_2}_{\Gamma_2}$ by $D^w_{\Gamma_1\otimes_F\Gamma_2}$ is bounded by $O(\lambda(F))$, where $\lambda(F)$ denotes the second largest (in absolute value) eigenvalue of the normalized adjacency matrix of F. Since there are explicit constructions of expanders in which $\lambda(F)$ is an arbitrary small constant, the error can be set to be smaller than any fixed $\gamma > 0$.

Note that Rnd^w is the distribution that we obtain from the uniform distributions on $\{1_G, g_i\}$ by repeated applications of convolution. Thus one can study how the error develops with repeated application of the expander construction.

The fact that the expander product approximates the convolution with an arbitrary small positive error does not imply anything interesting. If in each step the error increases by a constant, then after a constant number of steps we do not have any control of it. Here comes a crucial observation: *the error does not increase always and sometimes it also decreases.* To see that this is possible, consider a model situation in which D_{Γ_1} is the uniform distribution on G. (We will omit the superscripts from now on.) Then $D_{\Gamma_1} * D_{\Gamma_2}$ is the uniform distribution. Hence $D_{\Gamma_1\otimes_F\Gamma_2}$ is γ-close to the uniform distribution. Note that this is regardless what is the distance of D_{Γ_2} from the distribution produced by random bits. This remains essentially true if we only assume that D_{Γ_1} is very close to the uniform distribution.

This suggests the following strategy: to prove that in each step of the construction

1. either $D_{\Gamma_1\otimes_F\Gamma_2}$ is closer to the uniform distribution than D_{Γ_1} and D_{Γ_2} by a constant additive term,

2. or the error does not increase.

Since case 1. can only occur a finite number of times, the accumulated error will be bounded by a constant depending on γ. This is not literally true, because the expander construction can always introduce an error, even if both D_{Γ_1} and D_{Γ_1} are the uniform distributions. So one must also use the fact mentioned above that the error decreases when one of the distributions is very close to the uniform distribution.

It is not difficult to formalize this intuition in the special case of groups of prime size—the groups without proper subgroups. It is substantially more difficult to prove our result in the case of general groups that have proper subgroups. The reason is that instead of convergence to the global uniform distribution, there can be convergence to a distribution that is uniform only on each coset of some subgroup H. (In fact it is more complicated as we have to consider double cosets determined by a pair of subgroups.) But when converging to a uniform distribution on cosets of H it can diverge from a uniform distribution on cosets of another subgroup J. The difficult part is to show that when we alternate the process of converging on cosets of different subgroups, the error will still be bounded by a constant.

The main technical tool that we design is our Approximate Convolution Theorem which states that in a formula consisting of convolutions and simple (*natural*) distributions one can replace the convolutions by functions computing the convolutions only approximately while keeping the resulting distribution close to the original one. The resulting error does not depend on the size or structure of the formula provided that each approximate convolution introduces only bounded error and satisfies certain technical conditions.

1.3 Permutation branching programs

A permutation branching program B of width k is a branching program with the following properties. The vertices of the program can be divided into levels $0, 1, \ldots, m$ such that every arrow goes from a level i to the level $i+1$ and the size of each level is k. It is oblivious, which means that at each level only one variable is queried. For each level i, the arrows labeled by 0 (respectively 1) define a one to one mapping onto the next level $i+1$. One of the initial (level 0) vertices is the input vertex. The terminal vertices (level m) are divided into accepting ones and rejecting ones. In this paper we are only interested in read once branching programs, which means that each variable x_i is used only on one level. Thus the length of B is equal to the number of variables and we can assume, that the variables are red in the order $x_1, x_2, \ldots, x_n$.

Assume that the vertices on each level are labeled by $1, \ldots, k$. Then the two one-to-one mappings between levels i and $i+1$ can be identified with permutations on a k elements set, in other words, with two elements of the symmetric group S_k. By relabelling nodes in each level of the branching program one can assume that the permutations corresponding to bit 0 can always be the identity mappings. Thus a permutation branching program of width k is determined by a group word $g_1 \ldots g_n \in (\mathcal{S}_k)^n$, except for the the choice of the accepting vertices (we can assume that the input vertex has label 1).

The derandomization problem for branching programs is to find pseudorandom generators that would approximate the probability that a random input is accepted. In the case of permutation branching programs we can look for pseudorandom generators satisfying the following property: for every pair of indices $1 \le i, j \le k$, they should approximate the probability that for a random input, starting in the i-th initial vertex we will end in the j-th terminal vertex. It is clear that this is equivalent to the original question.

Our result gives a pseudorandom generator that *for every fixed permutation* $\pi \in \mathcal{S}_k$ approximates the probability that for a random input and *every* i, $1 \le i \le k$, from an initial vertex labelled i, we reach a terminal vertex labelled $\pi(i)$. Again, it is not difficult to see that this problem is equivalent to the original problem (assuming we want logarithmic size seed and constant error). Here is a sketch of the proof of this equivalence. Our generator solves the previous problem, because the probability that from i we reach j is the sum of the probabilities of the permutations that map i to j. To prove the other direction of the equivalence, given a group word $g_1, \ldots, g_n \in G^n$, consider the permutation branching program of width $|G|$ in which the nonidentical mapping from the level i to the level $i+1$ is given by the action of g_i on G.

2 The pseudorandom generator and our result

As explained in Introduction, INW generator is obtained by recursively applying the expander product. Let us recall the relevant facts.

Recall that a (N, M, λ)-expander is an undirected M-regular multi-graph on N vertices whose second largest (in absolute value) eigenvalue of its normalized adjacency matrix is at most λ.

Let $\Gamma_1, \Gamma_2 : \{0,1\}^r \to \{0,1\}^n$ be two functions and F be a 2^d-regular multi-graph with vertex set $\{0,1\}^r$. (Think of Γ_1 and Γ_2 as pseudorandom generators and F as an expander.) Furthermore, let ν be a function that given an $y \in \{0,1\}^r$ and $z \in \{0,1\}^d$, gives a neighbor of y in F that is reached by the edge labeled z. Then the expander product of Γ_1 and Γ_2 is the function $\Gamma_1 \otimes_F \Gamma_2 : \{0,1\}^{r+d} \to \{0,1\}^{2n}$ defined by

$$(\Gamma_1 \otimes_F \Gamma_2)(y, z) = (\Gamma_1(y), \Gamma_2(\nu(y, z))).$$

Notice that given random $y \in \{0,1\}^r$ and $z \in \{0,1\}^d$, the pair $(y, \nu(y,z))$ is a random edge of the graph F.

When Γ_1 and Γ_2 are one-to-one functions (which is true in the case of the pseudorandom generators used in the construction of INW generator) we can also view the construction as follows. Take disjoint copies of the *ranges* of Γ_1 and Γ_2 and a bipartite expander on them. Then the range of $\Gamma_1 \otimes_F \Gamma_2$ will be the concatenation of the pairs of strings connected by an edge. (However, this view of the product has the drawback of problematic constructibility.)

For the construction of INW generator we need an explicitly constructible family of (N, M, λ)-expanders for an increasing sequence of N and a constant M that are powers of two. Such a sequence can be obtained from e.g. [GG81] or [RVW00] (we restate Lemma 5.1 from [RV05]).

Lemma 1 *There is a universal constant $c_0 > 0$ such that for every constant $0 < \lambda < 1$ and $d = c_0 \lceil \log 1/\lambda \rceil$, there exists a sequence F_m of $(2^{dm}, 2^d, \lambda)$-expanders, where $m = 1, 2, \ldots$. Neighbors in F_m are computable in space $O(m)$, i.e., given a vertex name $y \in \{0,1\}^{dm}$ and an edge label $z \in \{0,1\}^d$, we can compute $\nu(y,z)$ in space $O(dm)$ and time poly(dm).*

For $0 < \lambda < 1$ and an integer $n \geq 1$, (λ, n)-INW generator is obtained recursively as follows. We start by letting $\Gamma_0 : \{0,1\}^d \to \{0,1\}^d$ be the identity mapping. Then $\Gamma_{i+1} = \Gamma_i \otimes_{F_i} \Gamma_i$, where F_i is the $(2^{d(i+1)}, 2^d, \lambda)$-expander from the previous lemma. This gives (λ, n)-INW pseudorandom generator for every $n = d2^k$ where $k > 0$, namely $\Gamma_k : \{0,1\}^{d(k+1)} \to \{0,1\}^{d2^k}$. To obtain (λ, n)-INW pseudorandom generators for an arbitrary n, we take the smallest $n' = d2^k \geq n$ (which is less than $2n$ for all n large enough) and use only the first n output bits of the (λ, n')-INW generator. Hence, (λ, n)-INW generator giving n bits of output has seed length $O(\log n \cdot \log 1/\lambda)$.

One can easily verify that the output of the generator on a given seed can be computed in space linear in the seed length.

We make the following claim.

Theorem 2 (Main Theorem) *There is a constant $c > 0$ such that for any finite group G and $0 < \delta < 1$ if $\lambda = \delta / 2^{c|G|^{11}}$ then (λ, n)-INW generator Γ uses seeds of length $O(\log n \cdot (|G|^{11} + \log 1/\delta))$ to produce n bits such that for every $w \in G^n$,*

$$\|\mathrm{Rnd}^w - D_\Gamma^w\| \leq \delta.$$

The output of the generator is computable in space linear in the seed length.

We believe that the dependency on the size of the group in Main Theorem can be improved. To prove the theorem we need to introduce notation to prove auxiliary results.

3 Notations and Preliminaries

3.1 Notation on vectors

The i-th coordinate of any vector x can be referred to as either x_i or $x(i)$. (We use either of the two notations so to avoid confusion with double indexes.) An all one vector is denoted by $\vec{1}$, where its dimension is taken from the context. The *support of a vector* $x \in \mathbb{R}^I$ with coordinates labelled by elements of a set I is $\mathrm{supp}(x) = \{i \in I;\ x_i \neq 0\}$. For two real valued vectors x and y of the same dimension we define their inner product to be $\langle x, y \rangle = \sum x_i \cdot y_i$, where the sum is taken over all the coordinates of x and y, respectively. We say that x and y are *orthogonal* if $\langle x, y \rangle = 0$; we denote this by $x \perp y$. Notice, if $\mathrm{supp}(x) \cap \mathrm{supp}(y) = \emptyset$ then $x \perp y$. The ℓ_2-norm of a real valued vector x is defined as $\|x\| = \sqrt{\langle x, x\rangle}$; the ℓ_∞-norm is $\|x\|_\infty = \max_i |x_i|$.

Let $m \geq 1$ be an integer. For $x \in \mathbb{R}^m$ and a partition $\mathcal{S} = \{S_1, S_2, \dots, S_k\}$ of the set $\{1, 2, \dots, m\}$, $x^{\|\mathcal{S}}$ is the vector in $\mathbb{R}^m$ which is constant on the blocks of $\mathcal{S}$, (i.e., such that if $i, j \in S_\ell$ for some $\ell \in \{1, \dots, k\}$, then $x_i = x_j$). Further, $x^{\perp\mathcal{S}} = x - x^{\|\mathcal{S}}$.

We say that two partitions $(L_1, L_2, \dots, L_\ell)$ and $(K_1, K_2, \dots, K_k)$ of $\{1, 2, \dots, m\}$ are *connected* if their join is the whole set $\{1, 2, \dots, m\}$ (i.e., if for any $x, y \in S$ there exists a path of hyper-edges between x and y.)

In the proof of our result we will only use vectors indexed by elements of a fixed finite group G; so a vector x will be an element of $\mathbb{R}^G$.

3.2 Notation on groups and convolution

We consider finite groups, and probability distributions on them. The size of a group is considered to be a constant throughout the paper. Let G be a finite group. Denote the identity element of G as 1_G. Let $D \in \mathbb{R}^G$ be a probability distribution on G. For $g \in G$, $D(g)$ denotes the probability of picking g, if an element is chosen from G according to D. We also treat probability distributions over G as vectors in $\mathbb{R}^G$, indexed by elements of G. For a probability distribution D on G and a subset $S \subseteq G$, we define $D(S) = \sum_{g \in S} D(g)$.

For a group G and set $S \subseteq G$, the *subgroup generated by* S is the smallest subgroup of G containing S; we will denote it by $\langle S\rangle$. For a probability distribution $D \in \mathbb{R}^G$ we denote $\langle D\rangle = \langle \mathrm{supp}(D)\rangle$. For a group G, its subgroup $H \leq G$ and $g \in G$, the set $gH = \{gh;\ h \in H\}$ is called a *left coset of* H (or *left* H*-coset*) and the set $Hg = \{hg;\ h \in H\}$ is called a *right coset of* H (or *right* H*-coset*). A well known fact is that if H is a subgroup of G then G can be partitioned into left (and right) H-cosets and thus the size of H divides the size of G. For subgroups $L, K \leq G$ and an element $g \in G$, define $LgK = \{agb;\ a \in L, b \in K\}$. It is easy to verify that G can be partitioned into parts such that each part is of the form LgK for some $g \in G$. We will call such sets *double cosets.*

For a subgroup $H \leq G$, let $(L_1, L_2, \dots, L_k)$ be the partition of G into left H-cosets and $(K_1, K_2, \dots, K_k)$ be the partition of G into right H-cosets. For a vector $x \in \mathbb{R}^G$ we define

$$\begin{aligned} x^{\|H} &= x^{\|(L_1, L_2, \dots, L_k)} & \qquad x^{\perp H} &= x^{\perp(L_1, L_2, \dots, L_k)} \\ x^{H\|} &= x^{\|(K_1, K_2, \dots, K_k)} & \qquad x^{H\perp} &= x^{\perp(K_1, K_2, \dots, K_k)}. \end{aligned}$$

For $0 \leq \Delta \leq 1$, a subgroup $H \leq G$ and a probabilistic distribution $D \in \mathbb{R}^G$ we say that D is Δ*-uniform on left* H*-cosets* if $\|D^{\perp H}\|_\infty \leq \Delta$.

Definition 3 (Convolution of vectors:) *Given two vectors $u, v \in \mathbb{R}^G$, define the convolution of v and u, denoted by $v * u$, as follows:*

$$(v * u)_h = \sum_{g \in G} v_g \cdot u_{g^{-1}h}$$

Thus convolution of two probability distributions D_1, D_2 on G is another probability distribution D where $D(h) = \sum_{g \in G} D_1(g) \cdot D_2(g^{-1}h)$. Notice that convolution is a linear operation so $(u+v)*w = u*w + v*w$ and $u*(v+w) = u*v + u*w$.

For $g \in G$, if $g \neq 1_G$ then $[g]$ denotes the probability distribution where $[g](1_G) = [g](g) = 1/2$, otherwise $[g]$ denotes the probability distribution where $[g](1_G) = 1$.

Definition 4 *A distribution $D \in \mathbb{R}^G$ is* natural *if for some $g_1, \ldots, g_k \in G$,*

$$D(g) = \frac{1}{2^k} |\{x_1 \ldots x_k \in \{0,1\}^k | g = g_1^{x_1} \ldots g_k^{x_k}\}|$$

for all $g \in G$. Equivalently, D is natural, if

$$D = [g_1] * [g_2] * \cdots * [g_k]$$

for some elements $g_1, \ldots, g_k \in G$.

One can show that for a natural distribution D, $1_G \in \text{supp}(D)$ and for every $g \in \text{supp}(D)$, $D(g) \geq 2^{-|\text{supp}(D)|}$. However, we will not need the latter fact. Clearly, convolution of two natural distributions is natural.

Definition 5 *For two probability distributions $D, R \in \mathbb{R}^G$ we say that $D = R + \epsilon$ is a* natural decomposition *of D if R is natural, $\epsilon = D - R$, $\epsilon \perp \vec{1}$, and* $\text{supp}(\epsilon) \subseteq \text{supp}(R)$.

Note that a probability distribution may have more than one natural decomposition.

Lemma 6 *For probability distributions $D_1, D_2, R_1, R_2 \in \mathbb{R}^G$, if $D_1 = R_1 + \epsilon_1$ and $D_2 = R_2 + \epsilon_2$ are natural decompositions, then $D = D_1 * D_2$ has a natural decomposition of the form $D = R_1 * R_2 + \epsilon'$.*

Proof: We have

$$D_1 * D_2 = R_1 * R_2 + (R_1 + \epsilon_1) * \epsilon_2 + \epsilon_1 * (R_2 + \epsilon_1).$$

One can easily show that in general, if $\text{supp}(v_i) \subseteq \text{supp}(R_i)$ and R_i are probability distributions, for $i = 1, 2$, then $\text{supp}(v_1 * v_2) \subseteq \text{supp}(R_1 * R_2)$. Thus in our case $\text{supp}((R_1 + \epsilon_1) * \epsilon_2 + \epsilon_1 * (R_2 + \epsilon_1)) \subseteq \text{supp}(R_1 * R_2)$. The fact that $(R_1 + \epsilon_1) * \epsilon_2 \perp \vec{1}$ and $\epsilon_1 * (R_2 + \epsilon_1) \perp \vec{1}$ follows from Lemma 13. □

We will approximate the convolution of two probability distributions D_1 and D_2 by a distribution D resulting from the expander construction. For the proof, we will only need the properties of D listed in the next definition.

Definition 7 *Let $0 < \gamma < 1$ and let D_1, D_2 and D be distributions. Let $\epsilon = D_1 * D_2 - D$. We say that the distribution D is a* γ-approximate convolution *of D_1 and D_2, if the vector ϵ satisfies the following conditions:*

1. $\text{supp}(\epsilon) \subseteq \text{supp}(D_1 * D_2)$,
2. $||\epsilon^{\langle D_1 \rangle \|}|| = ||\epsilon^{\| \langle D_2 \rangle}|| = 0$, *and*
3. $||\epsilon^{\langle D_1 \rangle \perp}||$, $||\epsilon^{\perp \langle D_2 \rangle}|| < \gamma$.

The meaning of the second condition is that the error ϵ redistributes the probability mass only within each right $\langle D_1 \rangle$-coset and left $\langle D_2 \rangle$-coset.

In the sequel we will denote a γ-approximate convolution of D_1 and D_2 by $D_1 *_\gamma D_2$. The reader should, however, keep in mind that this is only a convenient notation, *not* a uniquely defined operation on probability distributions.

For a probability distribution $D \in \mathbb{R}^G$ on a group G, we define

$$\lambda^{\mathrm{R}}(D) = \max \frac{\|x * D\|}{\|x\|},$$

where the maximum is over all vectors $x \in \mathbb{R}^G$ with $\|x^{\|\langle D\rangle}\| = 0$. Symmetrically we define $\lambda^{\mathrm{L}}(D) = \max \frac{\|D*x\|}{\|x\|}$. Let $\lambda(D) = \max\{\lambda^{\mathrm{R}}(D), \lambda^{\mathrm{L}}(D)\}$.

The following claim is an immediate consequence of definition of $\lambda(D)$.

Proposition 8 *Let G be a finite group. Let $\epsilon, D \in \mathbb{R}^G$, where D is a probability distribution. Then*

$$\|\epsilon^{\perp\langle D\rangle} * D\| \quad \leq \quad \lambda(D) \cdot \|\epsilon^{\perp\langle D\rangle}\|.$$

Similarly for right $\langle D\rangle$-cosets and convolution by D from left.

4 Basic properties of ℓ_2-norm, groups and convolution

In this section we review and establish some simple facts that will be needed for the proof of our main theorem. The reader may want to skip this section during the first reading and use it only later as a reference.

4.1 Facts on ℓ_2-norm

For $x \in \mathbb{R}^m$

$$\|x\| \leq \sqrt{m}\|x\|_\infty \quad \text{and} \quad \|x\|_\infty \leq \|x\|.$$

Note that the dimension m will be the size of the group G, which is a constant. Thus if the constant factor does not play role, one can use any of the standard norms. For us, it will be the most convenient to use the ℓ_2-norm.

We will need the following two lemmas that estimate the ℓ_2-norm when one of the components of the vector is changed. Recall that for $x \perp y$, $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

Lemma 9 *Let $0 < \delta, \epsilon < 1$ be reals and $x, x', y \in \mathbb{R}^m$ vectors satisfying $x \perp y$, $x' \perp y$, $\|x\| \geq \delta\|x+y\|$ and $\|x'\| \leq (1-\epsilon)\|x\|$. Then*

$$\|x' + y\| \leq (1 - \frac{\epsilon\delta^2}{2})\|x+y\|.$$

Proof: Since $x, x' \perp y$, we have

$$\begin{aligned} \|x'+y\|^2 &= \|x'\|^2 + \|y\|^2 \\ &\leq (1-\epsilon)^2\|x\|^2 + \|y\|^2 \\ &= \|x+y\|^2 - (1-(1-\epsilon)^2)\|x\|^2 \\ &\leq (1-(1-(1-\epsilon)^2)\delta^2)\|x+y\|^2. \end{aligned}$$

Thus

$$\|x'+y\| \leq \sqrt{1-(1-(1-\epsilon)^2)\delta^2}\|x+y\|$$

Since $(1-(1-\epsilon)^2)\delta^2 = (2\epsilon - \epsilon^2)\delta^2 \geq \epsilon\delta^2$, we have

$$\sqrt{1-(1-(1-\epsilon)^2)\delta^2} \leq \sqrt{1-\epsilon\delta^2} \quad \leq \quad 1 - \frac{\epsilon\delta^2}{2}.$$

□

The following fact formalizes an informal intuition that if ℓ_2-norm of a vector is large and we have two *independent* directions then the vector must be large in at least one of the two directions.

Lemma 10 *Let $m > 0$ be an integer. Let $\mathcal{L} = (L_1, L_2, \ldots, L_\ell)$ and $\mathcal{K} = (K_1, K_2, \ldots, K_k)$ be connected partitions of $\{1, 2, \ldots, m\}$. Let $\epsilon \in \mathbb{R}^m$ be such that $\epsilon \perp \vec{1}$. Let $\alpha, \beta \in \mathbb{R}$ satisfy $\alpha > 0$ and $\beta > 4\alpha\ell\sqrt{m}$. If $\|\epsilon\| \geq \beta$ and $\|\epsilon^{\perp\mathcal{L}}\| \leq \alpha$ then:*

$$\|\epsilon^{\perp\mathcal{K}}\| \geq \frac{\beta}{2\ell\sqrt{m}} - \alpha.$$

Proof: Since $\|\epsilon\| \geq \beta$, there exists a coordinate $max \in \{1, \ldots, m\}$ such that $|\epsilon_{max}| \geq \beta/\sqrt{m}$. W.l.o.g. $\epsilon_{max} > 0$ as we can consider $-\epsilon$ instead of ϵ. Since $\epsilon \perp \vec{1}$, there is also a coordinate $min \in \{1, \ldots, m\}$ with $\epsilon_{min} < 0$. Since $\|\epsilon^{\perp\mathcal{L}}\| \leq \alpha$, the absolute value of each coordinate of $\epsilon^{\perp\mathcal{L}}$ is at most α. Thus, the coordinates of $\epsilon^{\perp\mathcal{L}}$ corresponding to the elements of the same part L_i differ at most 2α. For each $i = 1, \ldots, \ell$ consider the interval $[\min_{j \in L_i} \epsilon_j, \max_{j \in L_i} \epsilon_j]$. The sum of their lengths is at most $2\alpha\ell$. Thus, there are $a, b \in \mathbb{R}$ such that $0 \leq a < b < \epsilon_{max}$, and

$$b - a \geq \frac{\epsilon_{max} - 2\alpha\ell}{\ell - 1} \geq \frac{\beta/(\sqrt{m}) - 2\alpha\ell}{\ell} \geq \beta/(\sqrt{m} \cdot \ell) - 2\alpha,$$

and no coordinate of ϵ has its value in the interval of (a, b). Consider $S = \{i \in \{1, \ldots, m\};\ \epsilon_i \leq a\}$. Clearly, $\emptyset \neq S \neq \{1, \ldots, m\}$ and S is a union of some L_i's as $b - a > 2\alpha$. Hence, from connectedness of $\mathcal{L}$ and $\mathcal{K}$ there is a part K_i with two elements s and t such that $\epsilon_s \leq a < b \leq \epsilon_t$. Thus, $\epsilon_t - \epsilon_s \geq b - a$. Hence

$$\|\epsilon^{\perp\mathcal{K}}\| \geq \|\epsilon^{\perp K_i}\| \geq \frac{\epsilon_t - \epsilon_s}{2} \geq \frac{b-a}{2} \geq \frac{\beta}{2\ell\sqrt{m}} - \alpha.$$

□

4.2 Facts on convolution

The next proposition is straightforward to prove so we leave the proof to an interested reader.

Proposition 11 *For a finite group G, let $x, D \in \mathbb{R}^G$, where D is a probability distribution. Then*

$$\begin{aligned}
(x * D)^{\|\langle D\rangle} &= x^{\|\langle D\rangle}, \\
(x * D)^{\perp\langle D\rangle} &= x^{\perp\langle D\rangle} * D, \\
(D * x)^{\langle D\rangle\|} &= x^{\langle D\rangle\|}, \\
(D * x)^{\langle D\rangle\perp} &= D * x^{\langle D\rangle\perp}.
\end{aligned}$$

The following is a consequence of the previous proposition.

Lemma 12 *Using the same notation as in Proposition 11*

$$\begin{aligned}\|x*D\|,\|D*x\| &\leq \|x\|,\\ \|(x*D)^{\perp\langle D\rangle}\| &\leq \|x^{\perp\langle D\rangle}\|,\\ \|(D*x)^{\langle D\rangle\perp}\| &\leq \|x^{\langle D\rangle\perp}\|.\end{aligned}$$

Proof:

We will prove the first part. The two remaining parts follow trivially from the first one. Let δ_g denote the vector such that $\delta_g(g) = D(g)$ and $\delta_g(h) = 0$ for $h \neq g$. Then $D = \sum_g \delta_g$. By linearity of convolution, we have

$$\|x*D\| = \|\sum_g x*\delta_g\| \leq \sum_g \|x*\delta_g\| = \sum_g D(g)\cdot\|x\| = \|x\|.$$

□

Lemma 13 *For a finite group G, let $x, y \in \mathbb{R}^G$ and $H \leq G$ be a subgroup of G. If $x \perp \vec{1}$ and* $\mathrm{supp}(x) \subseteq H$ *then $(y*x)^{\|H} = 0$ and $(x*y)^{H\|} = 0$.*

Proof: We prove $(x*y)^{H\|} = 0$, the other case is symmetric. For any $g \in G$ and any $b, b' \in H$,

$$\sum_{a\in Hg} y_{b^{-1}a} = \sum_{a\in Hg} y_{b'^{-1}a}.$$

Hence, by the definition of convolution and properties of x

$$\begin{aligned}\sum_{a\in Hg}(x*y)(a) &= \sum_{a\in Hg}\sum_{b\in G} x_b\cdot y_{b^{-1}a}\\ &= \sum_{a\in Hg}\sum_{b\in H} x_b\cdot y_{b^{-1}a}\\ &= \sum_{b\in H} x_b\cdot\sum_{a\in Hg} y_{b^{-1}a}\\ &= 0.\end{aligned}$$

The lemma follows. □

Proposition 14 *Let G be a finite group. Let $x, y \in \mathbb{R}^G$. Then $\|x*y\| \leq \sqrt{|G|}\cdot\|x\|\cdot\|y\|$.*

Proof: By the Cauchy-Schwarz inequality,

$$|(x*y)_h| = |\sum_g x_g y_{g^{-1}h}| \leq \sqrt{\sum_g x_g^2}\cdot\sqrt{\sum_g y_{g^{-1}h}^2} = \|x\|\cdot\|y\|,$$

for every $h \in G$. Hence $\|x*y\| \leq \sqrt{|G|}\cdot\|x\|\cdot\|y\|$.

□

Lemma 15 *Let G be a finite group. Let $0 < \gamma < 1$ and $R_1, R_2 \in \mathbb{R}^G$ be probability distributions. Let $\epsilon_1, \epsilon_2 \in \mathbb{R}^G$ be such that $R_1 + \epsilon_1$ and $R_2 + \epsilon_2$ are also probability distributions. Then*

$$\|(R_1 + \epsilon_1) *_\gamma (R_2 + \epsilon_2) - R_1 * R_2\| \leq \|\epsilon_1\| + \|\epsilon_2\| + \gamma.$$

Proof: By linearity

$$(R_1 + \epsilon_1) * (R_2 + \epsilon_2) = R_1 * R_2 + (R_1 + \epsilon_1) * \epsilon_2 + \epsilon_1 * R_2.$$

By Lemma 12

$$\|(R_1 + \epsilon_1) * \epsilon_2\| \leq \|\epsilon_2\|$$

and

$$\|\epsilon_1 * R_2\| \leq \|\epsilon_1\|.$$

By the triangle inequality

$$\|(R_1 + \epsilon_1) * (R_2 + \epsilon_2) - R_1 * R_2\| \leq \|\epsilon_1\| + \|\epsilon_2\|$$

The lemma follows from properties of $*_\gamma$. □

Lemma 16 *Let G be a finite group. Let $0 < \Delta, \gamma < 1$ and $R_1, R_2 \in \mathbb{R}^G$ be probability distributions such that $\langle R_1 \rangle = \langle R_2 \rangle = H$, R_1 is Δ-uniform on left H-cosets, and R_2 is Δ-uniform on right H-cosets. Let $\epsilon_1, \epsilon_2 \in \mathbb{R}^G$ be orthogonal to $\vec{1}$ and* $\mathrm{supp}(\epsilon_1), \mathrm{supp}(\epsilon_2) \subseteq H$*. Then*

$$\begin{aligned}\|(R_1 + \epsilon_1) *_\gamma (R_2 + \epsilon_2) - R_1 * R_2\| \quad &\leq \quad |G| \cdot \Delta \cdot (\|\epsilon_1\| + \|\epsilon_2\|) \\ &\qquad + \sqrt{|G|} \cdot \|\epsilon_1\| \cdot \|\epsilon_2\| + \gamma.\end{aligned}$$

Proof: By linearity

$$(R_1 + \epsilon_1) * (R_2 + \epsilon_2) = R_1 * R_2 + R_1 * \epsilon_2 + \epsilon_1 * R_2 + \epsilon_1 * \epsilon_2.$$

Since $\mathrm{supp}(\epsilon_2) \subseteq H$ and $\epsilon_2 \perp \vec{1}$, $R_1^{H\|} * \epsilon_2$ is the zero vector. Hence,

$$R_1 * \epsilon_2 = (R_1^{H\|} + R_1^{H\perp}) * \epsilon_2 = R_1^{H\perp} * \epsilon_2.$$

Since R_1 is Δ-uniform on left H-cosets, each coordinate of $R_1^{H\perp}$ is at most Δ in absolute value, hence $\|R_1^{H\perp}\| \leq \sqrt{|G|} \cdot \Delta$. Thus by Proposition 14,

$$\|R_1 * \epsilon_2\| = \|R_1^{H\perp} * \epsilon_2\| \leq |G| \cdot \Delta \cdot \|\epsilon_2\|.$$

Similarly,

$$\|\epsilon_1 * R_2\| \leq |G| \cdot \Delta \cdot \|\epsilon_1\|.$$

From these inequalities and the triangle inequality

$$\|(R_1 + \epsilon_1) * (R_2 + \epsilon_2) - R_1 * R_2\| \leq |G| \cdot \Delta \cdot (\|\epsilon_1\| + \|\epsilon_2\|) + \sqrt{|G|} \cdot \|\epsilon_1\| \cdot \|\epsilon_2\|.$$

The lemma follows from properties of $*_\gamma$. □

4.3 Facts on natural distributions

Lemma 17 *Let D be a natural probability distribution on a finite group G. For every subgroup $H \leq G$, if* $\operatorname{supp}(D) \setminus H \neq \emptyset$ *then $D(S) \leq 1/2$ for all left and right H-cosets S.*

Proof: We prove it for right cosets, the case of left cosets is symmetric. Let $D = [g_1] * [g_2] * \cdots * [g_n]$. Take the smallest k such that $D_k = [g_1] * [g_2] * \cdots * [g_k]$ is not contained in H. Then clearly $D_k(Hg_k) = D_k(H) = 1/2$ so no right H-coset has probability more than 1/2. Furthermore by induction on $\ell > k$, for any right H-coset S, $(D_{\ell-1} * [g_\ell])(S) = \frac{1}{2}D_{\ell-1}(S) + \frac{1}{2}D_{\ell-1}(Sg_\ell^{-1}) \leq 1/2$. □

Lemma 18 *Let D be a natural probability distribution on a finite group G. For every subgroup $H \leq G$, if the support of D is not in H, then there exists two right H-cosets $S_1 \neq S_2$ such that $D(S_1) \geq |H|/|G|$ and $D(S_2) \geq |H|/2(|G| - |H|)$. Symmetrically for left H-cosets.*

Proof: Take S_1 to be the right H-coset with the largest probability and S_2 the right H-coset with the second largest probability. □

Lemma 19 *Let D be a natural probability distribution on a finite group G. Suppose that the support of D generates G. Then there exists an element $a \in G$ and a set $K \subseteq G$ such that $D(a) \geq 1/2|G|$, $D(g) \geq 1/2|G|$ for every $g \in K$, and Ka^{-1} generates G.*

Proof: Assume $|G| > 1$ otherwise the claim is trivial. By Lemma 18 applied on $H = \{1_G\}$, there exist two elements $g \neq g'$ such that $D(g), D(g') \geq 1/2|G|$. Let b_1 be one of them that is not equal to 1_G and let a be the other. Now define inductively a sequence of elements $b_1, b_2, \ldots$ such that $D(b_i) \geq 1/2|G|$, for $i \geq 1$, and $b_1a^{-1}, \ldots, b_ka^{-1}$ span subgroups of increasing size. Suppose we already have $b_1, \ldots, b_k$ and $b_1a^{-1}, \ldots, b_ka^{-1}$ span a proper subgroup B_k. Since $D(B_ka) \leq 1/2$ by Lemma 17, there exists an element $b_{k+1} \notin B_ka$ that has probability $\geq \frac{1}{2}|G|$. Since $b_{k+1} \notin B_ka$, we have $b_{k+1}a^{-1} \notin B_k$, hence $b_1a^{-1}, \ldots, b_ka^{-1}, b_{k+1}a^{-1}$ span a larger subgroup. Since G is finite, we eventually get a set K with the properties required by the lemma. □

The following lemma bounds $\lambda(D)$ for natural distributions D.

Lemma 20 *Let $D \in \mathbb{R}^G$ be a natural probability distribution on a finite group G. Then*

$$\lambda(D) \leq 1 - 1/c_G,$$

where $c_G = 16|G|^4$.

We use the same technique that is used to estimate the second largest (in absolute value) eigenvalue of graphs (cf. [Lov93]) to prove the lemma.

Proof: We prove that $\lambda^{\mathrm{R}}(D) \leq 1 - 1/c_G$, the case for $\lambda^{\mathrm{L}}(D)$ is symmetric. Let $|G| = m$. Let us assume first that $\langle D \rangle = G$. Then for any $x \in \mathbb{R}^G$, $\|x^{\|\langle D \rangle}\| = 0$ if and only if $x \perp \vec{1}$. Hence, let $x \in \mathbb{R}^G$ be such that $x \perp \vec{1}$, $\|x\| = 1$ and $\|x * D\|$ be maximal possible. Since $\{x \in \mathbb{R}^G;\ x \perp \vec{1}\ \&\ \|x\| = 1\}$ forms a compact space such x exists. Clearly, $\lambda^{\mathrm{R}}(D) = \|x * D\|$.

Let $\tilde{D}$ be the $m \times m$ matrix indexed by elements of G and defined by $\tilde{D}(g,h) = D(g^{-1}h)$, for all $g, h \in G$. Clearly, $\tilde{D}$ is doubly-stochastic as well as $\tilde{D}\tilde{D}^T$. Moreover by definition of convolution, $x * D = x\tilde{D}$. Thus, $\|x * D\|^2 = x\tilde{D}\tilde{D}^T x^T = \lambda^{\mathrm{R}}(D)^2$. Hence

$$\begin{aligned}1-(\lambda^{\mathrm{R}}(D))^2 &= x(I-\tilde{D}\tilde{D}^T)x^T\\ &= \sum_{i\in G} x_i^2 - \sum_{i,j\in G} x_i x_j (\tilde{D}\tilde{D}^T)_{i,j}\\ &= \frac{1}{2}\sum_{i\in G} x_i^2 \sum_{j\in G}(\tilde{D}\tilde{D}^T)_{i,j} + \frac{1}{2}\sum_{j\in G} x_j^2 \sum_{i\in G}(\tilde{D}\tilde{D}^T)_{i,j} - \sum_{i,j\in G} x_i x_j (\tilde{D}\tilde{D}^T)_{i,j}\\ &= \sum_{i,j\in G} \frac{(\tilde{D}\tilde{D}^T)_{i,j}}{2}(x_i-x_j)^2\ .\end{aligned}$$

Since, the right hand side is non-negative, $\lambda^{\mathrm{R}}(D) \leq 1$.

As $\|x\| = 1$, there is a coordinate g_+ such that $|x_{g_+}| \geq \frac{1}{\sqrt{m}}$. Without loss of generality, let $x_{g_+} \geq \frac{1}{\sqrt{m}}$ where $g_+ \in G$. Since $x \perp \vec{1}$, there is another coordinate $x_{g_-} < 0$. Let a and K be the element and the set from Lemma 19. Since Ka^{-1} generates G, there exists $g_0, \ldots, g_\ell \in G$, $\ell \leq m$, such that $g_0 = g_+$, $g_\ell = g_-$, and $g_k^{-1} g_{k+1} \in Ka^{-1}$ for $0 \leq k < \ell$. (Take $h_1, h_2, \ldots, h_\ell \in Ka^{-1}$ such that $g_+^{-1} g_- = h_1 h_2 \cdots h_\ell$ and set inductively for $k = 0, \cdots, \ell-1$, $g_{k+1} = g_k h_{k+1}$.) We will show that

$$(\tilde{D}\tilde{D}^T)_{g_k, g_{k+1}} \geq \frac{1}{4m^2}. \tag{1}$$

By definition

$$(\tilde{D}\tilde{D}^T)_{g_k,g_{k+1}} = \sum_{s=1}^{k} (\tilde{D})_{g_k,s} (\tilde{D}_1^T)_{s,g_{k+1}} = \sum_{s=1}^{k} D(g_k^{-1}s) D(g_{k+1}^{-1}s).$$

We will lower-bound it by the term in which $s = g_{k+1}a$. Since $g_k^{-1}s = g_k^{-1}g_{k+1}a$ and $g_k^{-1}g_{k+1} \in Ka^{-1}$, we have $g_k^{-1}s \in K$, whence $D(g_k^{-1}s) \geq 1/2m$. Further, $g_{k+1}^{-1}s = a$, hence $D(g_{k+1}^{-1}s) \geq 1/2m$. Thus we get (1).

Using this estimate and the expression for $1-(\lambda^{\mathrm{R}}(D))^2$ derived above, we get

$$\begin{aligned}1-(\lambda^{\mathrm{R}}(D))^2 &\geq \frac{1}{8m^2}\sum_{j=0}^{\ell-1}(x_{g_j} - x_{g_{j+1}})^2\\ &\geq \frac{1}{8m^2\ell}\Big(\sum_{j=0}^{\ell-1}(x_{g_j} - x_{g_{j+1}})\Big)^2 \quad \text{(by Cauchy-Schwarz inequality)}\\ &\geq \frac{1}{8m^2\ell}(x_{g_+} - x_{g_-})^2\\ &\geq \frac{1}{8m^4}\end{aligned}$$

As $\lambda^{\mathrm{R}}(D) \leq 1$, we have

$$1-\lambda^{\mathrm{R}}(D) \geq \frac{1}{16m^4}$$

Consider the case when $\langle D\rangle = H \lneq G$. Let $k = |H|$. If $\tilde{D}$ is the same matrix as above then $D_{g,h} > 0$ implies that $g^{-1}h \in H$ so $h \in gH$. Thus $D_{g,h} > 0$ implies that h is in the left coset gH and $hH = gH$. One can easily verify that rows and columns of $\tilde{D}$ can be reordered so that $\tilde{D} = I_{m/k} \otimes \tilde{D}_H$, where $\tilde{D}_H$ is the $k \times k$ matrix defined by $\tilde{D}_H(h_1, h_2) = D_{h_1^{-1}h_2}$ for $h_1, h_2 \in H$, and $I_{m/k}$ is the identity matrix of rank m/k. Consider any vector x such that $\|x^{\|\langle D\rangle}\| = 0$. If $(A_1, \ldots, A_{\ell_A})$ is the partition of G into left H-cosets, then $x = \sum_{i=1}^{\ell_A} x^{\perp(A_i)}$. As $\tilde{D} = I_{m/k} \otimes \tilde{D}_H$, for any left coset A_i, $x^{\perp(A_i)} * D = (x * D)^{\perp(A_i)}$. Also $x^{\perp(A_i)} \perp \vec{1}$ so $\|x^{\perp(A_i)} * D\| \leq \lambda^{\mathrm{R}}(\tilde{D}_H)\|x^{\perp(A_i)}\|$. Now

$$\|x * D\|^2 \;=\; \|\sum_{i=1}^{\ell_A} x^{\perp(A_i)} * D\|^2 \leq \lambda^{\mathrm{R}}(\tilde{D}_H)^2 \cdot \sum_{i=1}^{\ell_A} \|x^{\perp(A_i)}\|^2 = \lambda^{\mathrm{R}}(\tilde{D}_H)^2 \cdot \|x\|^2$$

and the lemma follows. □

5 Expander product well approximates convolution

In this section we will estimate the error introduced by the expander product of two pseudorandom generators, the basic step of INW generator, and prove Lemma 23. Similar bounds were proven in [INW94]. Rather than adapting their results, we will give a direct proof based on Expander Mixing Lemma.

Lemma 21 *Let a word w over some group G be given. Let $w = w_1 w_2$, with $|w_1| = |w_2| = n$. Let $\Gamma_1, \Gamma_2 : \{0,1\}^r \to \{0,1\}^n$ be two functions and let F be an $(2^r, 2^d, \lambda)$-expander. Then*

$$\|D^{w_1}_{\Gamma_1} * D^{w_2}_{\Gamma_2} - D^{w}_{\Gamma_1 \otimes_F \Gamma_2}\| \leq \lambda\sqrt{|G|}.$$

The proof of Lemma 21 uses *Expander Mixing Lemma*, stated below (see e.g. [AS92] Corollary 2.5).

Lemma 22 (Expander Mixing Lemma) *Let $F = (V, E)$ be a (N, M, λ)-expander. For any two subsets $S \subseteq U$, $T \subseteq V$, let $e(S,T)$ denote the number of edges between S and T. Then*

$$|e(S,T) - \frac{M \cdot |S| \cdot |T|}{N}| \leq \lambda M \sqrt{|S| \cdot |T|}.$$

Note that we do not require the sets S and T to be disjoint.

Proof of Lemma 21. Let $w_1 = g_1 \ldots g_n$, $w_2 = h_1 \ldots h_n$, $N = 2^r$ and $M = 2^d$. For $g \in G$ put

$$\begin{aligned} U_g &= \{y \in \{0,1\}^r |\; g_1^{\Gamma_1(y)_1} \cdots g_n^{\Gamma_1(y)_n} = g\}, \\ V_g &= \{y \in \{0,1\}^r |\; h_1^{\Gamma_2(y)_1} \cdots h_n^{\Gamma_2(y)_n} = g\}. \end{aligned}$$

Then $\{U_g\}_{g\in G}$ and $\{V_g\}_{g\in G}$ are partitions of $\{0,1\}^r$. Using the expander mixing lemma, we have

$$\left| e(U_g, V_h) - \frac{M \cdot |U_g| \cdot |V_h|}{N} \right| \leq \lambda M \sqrt{|U_g| \cdot |V_h|} \tag{2}$$

for all $g, h \in G$. Dividing by MN, we get (from now on we are omitting the superscripts w_1, w_2 and w)

$$\begin{aligned} \left| \frac{e(U_g, V_h)}{MN} - \frac{|U_g|}{N} \cdot \frac{|V_h|}{N} \right| &\leq \lambda \sqrt{\frac{|U_g|}{N} \cdot \frac{|V_h|}{N}} \\ \therefore \left| \frac{e(U_g, V_h)}{MN} - D_{\Gamma_1}(g) D_{\Gamma_2}(h) \right| &\leq \lambda \sqrt{D_{\Gamma_1}(g) D_{\Gamma_2}(h)}. \end{aligned}$$

Therefore for each $k \in G$, we have

$$\begin{aligned}
\Big|\frac{1}{MN}\sum_{\substack{g,h:\\gh=k}} e(U_g, V_h) - \sum_{\substack{g,h:\\gh=k}} D_{\Gamma_1}(g)D_{\Gamma_2}(h)\Big| &\leq \lambda \sum_{\substack{g,h:\\gh=k}} \sqrt{D_{\Gamma_1}(g)D_{\Gamma_2}(h)} \\
\therefore |D_{\Gamma_1\otimes_F\Gamma_2}(k) - D_{\Gamma_1} * D_{\Gamma_2}(k)| &\leq \lambda \sum_{\substack{g,h:\\gh=k}} \sqrt{D_{\Gamma_1}(g)D_{\Gamma_2}(h)}
\end{aligned}$$

Squaring and summing over all $k \in G$, we get

$$\begin{aligned}
\|D_{\Gamma_1\otimes_F\Gamma_2} - D_{\Gamma_1} * D_{\Gamma_2}\|^2 &\leq \lambda^2 \sum_{k\in G} \Big(\sum_{\substack{g,h:\\gh=k}} \sqrt{D_{\Gamma_1}(g)D_{\Gamma_2}(h)}\Big)^2 \\
&\leq \lambda^2 \sum_{k\in G} (\|\sqrt{D_{\Gamma_1}}\|^2 \|\sqrt{D_{\Gamma_2}}\|^2) \\
&= \lambda^2 |G|,
\end{aligned}$$

where $\sqrt{D}$ is the vector with entries equal to the square roots of the entries of D, and the last inequality follows from Cauchy-Schwarz inequality. □

Lemma 23 *Let a word w over some group G be given. Let $w = w_1w_2$, with $|w_1| = |w_2| = n$. Let $0 < \gamma < 1$ be given. Let $\Gamma_1, \Gamma_2 : \{0,1\}^r \to \{0,1\}^n$ be two functions and let F be an $(2^r, 2^d, \lambda)$-expander, where $\lambda = \gamma/\sqrt{|G|}$. The distribution given by $D^{w_1w_2}_{\Gamma_1\otimes_F\Gamma_2}$ is a γ-approximate convolution of $D^{w_1}_{\Gamma_1}$ and $D^{w_2}_{\Gamma_2}$.*

Proof of Lemma 23. Define $D^{w_1}_{\Gamma_1} *_\gamma D^{w_2}_{\Gamma_2} = D^{w_1w_2}_{\Gamma_1\otimes_F\Gamma_2}$. Clearly, we only have to verify the latter three conditions of Definition 7 concerning $\epsilon = D^{w_1}_{\Gamma_1} * D^{w_2}_{\Gamma_2} - D^{w_1w_2}_{\Gamma_1\otimes_F\Gamma_2}$. Let $N = 2^r$ and $M = 2^d$. (We drop the superscripts of D for the rest of the proof).

1. The support of $D_{\Gamma_1} * D_{\Gamma_2}$ is the set of elements of the form $gg^{-1}h$ such that $g \in supp(D_{\Gamma_1})$ and $g^{-1}h \in supp(D_{\Gamma_2})$. It follows from the definition that only such elements are in $supp(D_{\Gamma_1\otimes_F\Gamma_2})$. Hence also $\mathrm{supp}(\epsilon) \subseteq supp(D_{\Gamma_1} * D_{\Gamma_2})$.

2. Let A be a right coset of $\langle D_{\Gamma_1}\rangle$. Let B be the elements $y \in \{0,1\}^r$ such that $w_{2,1}^{\Gamma_2(y)_1} w_{2,2}^{\Gamma_2(y)_2} \cdots w_{2,n}^{\Gamma_2(y)_n} \in A$. The weight of A in D_{Γ_2} is $|B|/N$. The weight of A in $D_{\Gamma_1\otimes_F\Gamma_2}$ is $e(\{0,1\}^r, B)/MN$. Since F is M-regular, $e(\{0,1\}^r, B)/MN = |B|/N$. Hence $D^{\langle D_{\Gamma_1}\rangle\|}_{\Gamma_1\otimes_F\Gamma_2} = D^{\langle D_{\Gamma_1}\rangle\|}_{\Gamma_2}$, which means $\epsilon^{\langle D_{\Gamma_1}\rangle\|} = \vec{0}$. The other case follows by symmetry.

3. This follows from 2. and the lemma above, because $\epsilon = \epsilon^{\langle D_{\Gamma_i}\rangle\|} + \epsilon^{\langle D_{\Gamma_i}\rangle\perp}$, for $i = 1, 2$. □

6 Proof of the main theorem

In this section we will prove a more general Approximate Convolution Theorem and show that our main theorem is an easy consequence. The Approximate Convolution Theorem shows that an arbitrary convolution of natural distributions can be well approximated by γ-approximate convolutions.

For this section we will fix a finite group G of size at least 4. Since every group of size three or less is a subgroup of some group of size six, our theorems are applicable to such groups as well. We define two parameters Δ and τ depending only on the group:

$$\Delta = \tfrac{1}{16|G|^2} \qquad\qquad \tau = \tfrac{\Delta^2}{2\cdot c_G \cdot \sqrt{|G|}} = \tfrac{1}{8192\cdot|G|^{8.5}}.$$

Here c_G is the constant from Lemma 20 below. Let $T = \lceil 1/\tau \rceil$.

Theorem 24 (Approximate Convolution Theorem) *There is a universal constant $c_1 > 0$ such that for every finite group G and $0 < \gamma < 1$ the following holds.*

Let F be a formula consisting of convolutions $*$ and natural probability distributions on G. Let F' be obtained from F by replacing the convolutions by γ-approximate convolutions. If R denotes the distribution computed by F, and D denotes the distribution computed by F' then

$$\|D - R\| \leq \gamma 2^{c_1|G|^{11}}.$$

We would like to draw attention of the reader to the remarkable fact that the conclusion of this theorem does not depend in any way on the size or structure of the formula F.

To prove the theorem we will classify probability distributions according to their closeness to the uniform distribution. Notice that for every probability distribution R on G, its norm is bounded by $1/\sqrt{|G|} \leq \|R\| \leq 1$ and R is the uniform distribution if and only if $\|R\| = 1/\sqrt{|G|}$. This motivates the following definition.

Definition 25 *For a probability distribution $R \in \mathbb{R}^G$, we say that the* rank of R is i *(*$\mathrm{rank}(R) = i$*) if*

$$i\tau \leq 1 - \|R\| < (i+1)\tau.$$

The rank of R corresponds to its distance from the uniform distribution: the higher the rank the closer the distribution is to uniform. The rank is in the range from 0 to T. Next lemma summarizes some properties of rank.

Lemma 26 *The following hold:*

1. *For any two probability distributions $R_1, R_2 \in \mathbb{R}^G$,* $\mathrm{rank}(R_1), \mathrm{rank}(R_2) \leq \mathrm{rank}(R_1 * R_2)$.
2. *For any two natural probability distributions $R_1, R_2 \in \mathbb{R}^G$, if $\langle R_1 \rangle \neq \langle R_2 \rangle$ and R_1 is Δ-uniform on left $\langle R_2 \rangle$-cosets then* $\mathrm{rank}(R_2) < \mathrm{rank}(R_1)$. *Similarly, if $\langle R_1 \rangle \neq \langle R_2 \rangle$ and R_2 is Δ-uniform on right $\langle R_1 \rangle$-cosets then* $\mathrm{rank}(R_1) < \mathrm{rank}(R_2)$.
3. *For any two natural probability distributions $R_1, R_2 \in \mathbb{R}^G$, if R_1 is not Δ-uniform on left $\langle R_2 \rangle$-cosets then* $\mathrm{rank}(R_1) < \mathrm{rank}(R_1 * R_2)$. *Similarly, if R_2 is not Δ-uniform on right $\langle R_1 \rangle$-cosets then* $\mathrm{rank}(R_2) < \mathrm{rank}(R_1 * R_2)$.

Proof: The first part of the lemma follows trivially from properties of convolution. In both remaining parts we only consider the case of the left cosets as the case of right cosets is symmetric.

Part 2. Let $H = \langle R_2 \rangle$ and $\ell = |H|$. First we show that $H \subseteq \langle R_1 \rangle$. Since R_1 is Δ-uniform on left H-cosets, coordinates of R_1 corresponding to the same left H-coset differ by at most $2\Delta < 1/2|G|$. Clearly, some left H-coset gH must contain at least $\ell/|G|$ of probability mass under R_1. Thus, all coordinates from gH of R_1 have probability at least $\frac{1}{G} - \frac{1}{2|G|} > 0$. Hence, $gH \subseteq \mathrm{supp}(R_1)$. Since $g \in gH$, $g^{-1} \in \langle R_1 \rangle$ and $H = g^{-1}gH \subseteq \langle R_1 \rangle$. From the assumption of the lemma, $H \subsetneq \langle R_1 \rangle$ and $\ell \leq |G|/2$.

By Lemma 17, each left H-coset contains at most 1/2 of the total probability mass. By Δ-uniformity, no coordinate of R_1 can have value larger than $\frac{1}{2\ell} + \Delta$. By convexity of the squaring function (i.e., for any $0 \leq c \leq b \leq a$, $a^2 + b^2 \leq (a+c)^2 + (b-c)^2$), the norm $\|R_1\|$ is maximized when R_1 is concentrated on the fewest possible coordinates. Thus, concentrating R_1 to at most 2ℓ coordinates each of size at most $\frac{1}{2\ell} + \Delta$ can only increase ℓ_2-norm of R_1 (hence, decrease its rank). Thus,

$$\|R_1\| \leq \sqrt{2\ell \cdot \left(\frac{1}{2\ell} + \Delta\right)^2} \leq \sqrt{2\ell \cdot \left(\frac{1}{2\ell} + \frac{1}{16\ell}\right)^2} = \sqrt{2\ell \cdot \left(\frac{9}{16\ell}\right)^2} = \frac{1}{\sqrt{\ell}} \cdot \frac{9\sqrt{2}}{16} < \frac{1}{\sqrt{\ell}} \cdot \frac{4}{5}.$$

However, $\|R_2\| \geq \frac{1}{\sqrt{\ell}}$, since $\mathrm{supp}(R_2) \subseteq H$ and ℓ_2-norm is minimal when the probability is spread uniformly over $\mathrm{supp}(R_2)$. Thus, $\|R_2\| - \|R_1\| \geq \frac{1}{5\sqrt{\ell}} \geq \tau$.

Part 3. Let $H = \langle R_2 \rangle$. By our assumption, $R_1^{\perp H}$ contains a coordinate of absolute value $> \Delta$. Hence, $\|R_1^{\perp H}\| > \Delta$. Furthermore,

$$R_1 * R_2 = (R_1^{\| H} + R_1^{\perp H}) * R_2 = R_1^{\| H} + R_1^{\perp H} * R_2$$

and, by Lemma 20,

$$\|R_1^{\perp H} * R_2\| \quad \leq \quad \lambda(R_2) \cdot \|R_1^{\perp H}\| \leq \left(1 - \frac{1}{c_G}\right) \cdot \|R_1^{\perp H}\|.$$

Clearly, $\|R_1\| \geq 1/\sqrt{|G|}$. Since $\|R_1^{\perp H}\| \geq \Delta \geq \Delta \cdot \|R_1^{\| H} + R_1^{\perp H}\|$, by Lemma 9,

$$\|R_1\| - \|R_1 * R_2\| \geq \|R_1^{\| H} + R_1^{\perp H}\| - \|R_1^{\| H} + R_1^{\perp H} * R_2\| \quad \geq \quad \frac{\Delta^2}{2 \cdot c_G \cdot \sqrt{|G|}}.$$

□

In the proof of Approximate Convolution Theorem we will trade rank for error. We will allow the error of our approximate distribution grow with its rank. Thus we will also need a lemma which will bound the error introduced by γ-approximate convolutions when there is a long chain of such convolutions that is applied on a distribution without increasing its rank.

We are interested in an approximate convolution of distributions D_i that approximate some natural distributions R_i up-to error ϵ_i. We assume that each $D_i = R_i + \epsilon_i$ is a natural decomposition. We want to bound the increase in the error if we convolve many such distributions. The following lemma bounds the increase in the error.

Lemma 27 (Key Convergence Lemma) *Let $0 < e_1, \gamma < 1$ be reals. Let $D_0, D_1, \ldots, D_t$, $R_0, R_1, \ldots, R_t$ be probability distributions on G, where $D_i = R_i + \epsilon_i$ is a natural decomposition, for $i \in \{0, \ldots, t\}$. Let $\|\epsilon_i\| \leq e_1$ for $i > 0$. Let D be obtained by iteratively convolving D_0 with $D_1, D_2, \ldots, D_t$ where each of the convolutions is some γ-approximate convolution either from left or from right. Let R be obtained by the same sequence of convolutions (but exact) of R_0 with $R_1, R_2, \ldots, R_t$. Then $\epsilon = D - R$ satisfies*

$$\|\epsilon\| \leq \|\epsilon_0\| + (e_1 + \gamma)h,$$

where $h = |G|^{O(|G|^2)}$ depends only on G.

The proof of Key Convergence Lemma is in the next section.

Proof of Approximate Convolution Theorem.
Let $E_0 = 0$ and $E_i = 2h \cdot (E_{i-1} + \gamma)$, for $i > 0$. Solving the recurrence gives $E_T \leq (2h)^{1+T}\gamma$. Assume w.l.o.g. that γ is small so that $E_T < 1/8|G|$. We will show $\|D - R\| \leq E_T$.

Look on F as a tree and assign to each node of the tree the rank of the distribution computed by the subformula rooted at the node. Assign the same rank to nodes of F'. We denote the distribution computed by a node u in F by R_u and the corresponding node in F' by D_u. We claim that for any node u of F, $\|R_u - D_u\| \leq E_{\text{rank}(R_u)}$. In the rest of the proof we call the size of the difference the *error*.

Remove all the edges between nodes of different ranks in F'. Hence we obtain a forest each consisting of nodes of the same rank. We prove the claim by induction on the rank of nodes in a tree. (We describe the induction somewhat informally. The interested reader can easily formalize it.) The base case is trivial as leaves of the original formula have zero error. Consider a tree of nodes of some rank i. Leaves in such a tree have either zero error as they are leaves of F' or have error bounded by $2E_{i-1} + \gamma \leq E_i/4$ since they are obtained by a γ-approximate convolution of nodes of rank less than i (by induction hypothesis and Lemma 15). Consider a node u of degree two in such a tree with children v and w. Distribution $D_u = R_u + \epsilon_u$ is a γ-approximate convolution of two distributions $D_v = R_v + \epsilon_v$ and $D_w = R_w + \epsilon_w$, where $\text{rank}(R_v) = \text{rank}(R_w) = \text{rank}(R_u)$. (All the decompositions are natural.)

By Lemma 26, R_v is Δ-uniform on left $\langle R_w \rangle$-cosets and R_w is Δ-uniform on right $\langle R_v \rangle$-cosets, so $\langle R_v \rangle = \langle R_w \rangle$. Thus by Lemma 16 and the choice of γ and Δ, the size of the error $\|\epsilon_u\| \leq 2 \cdot |G| \cdot \Delta \cdot E_i + \sqrt{|G|} \cdot E_i^2 + \gamma \leq E_i/4$.

The remaining nodes are nodes of degree one and form possibly several paths, each path starting either in a leaf or a node of degree two. Hence each path starts in a node with error $\leq E_i/4$. Each node along the path represents a γ-approximate convolution of the distribution of the start node with a distribution of rank less than i, so of error at most E_{i-1}. Thus the Key Convergence Lemma applies and each node along the path has error bounded by $E_i/4 + h(E_{i-1} + \gamma) \leq E_i$. Thus we have

$$\|D - R\| \leq E_T \leq (2h)^{1+T}\gamma = |G|^{O(|G|^{10.5})} = 2^{O(|G|^{11})}.$$

□

We are ready to prove Main Theorem. The proof uses two key ingredients. The first ingredient shows that expander product approximates convolution of any two probability distributions well. Then the second ingredient shows that if we take any formula consisting of convolutions of natural distributions and we substitute the convolutions by approximate convolutions the ℓ_2-distance between the distributions computed by the two formulas can be bounded. Since INW generator is constructed recursively using the expander product, the distribution it induces can be thought of as a distribution obtained by a formula consisting of approximate convolutions of natural distributions.

Proof of Main Theorem. Let $\gamma = \delta/2^{c_1|G|^{11}}$ and $\lambda = \gamma/\sqrt{|G|}$, where c_1 is the constant from Theorem 24. Hence $\lambda = \delta/2^{c|G|^{11}}$ for some constant $c > c_1$.

Consider the (λ, n)-INW generator. Let $\Gamma_0, \Gamma_1, \ldots, \Gamma_k$ be the functions used to construct the generator, where $\Gamma_0 : \{0,1\}^d \to \{0,1\}^d$. Pad w by 1_G at the right end so that it would be of length $d2^k$. Break w into consecutive blocks of d elements and for each block w' compute $\text{Rnd}^{w'}$. Observe, $\text{Rnd}^{w'} = D_{\Gamma_0}^{w'}$. Using convolution form a balanced formula F out of $\text{Rnd}^{w'}$, for all the blocks w', so that F evaluates to Rnd^{w}. Hence, F is a full binary tree of depth k with each internal node being a convolution and each leaf being one of the $\mathrm{Rnd}^{w'}$. Notice, the structure of the formula corresponds to the structure of $(\lambda, d2^k)$-INW generator.

Thus, from leaves towards the root of F, inductively replace each convolution by some γ-approximate convolution which correctly computes the distribution $D^{w_1w_2}_{\Gamma_i \otimes_F \Gamma_i} = D^{w_1w_2}_{\Gamma_{i+1}}$ when applied to the distributions computed by the operands of the convolution, i.e., distributions $D^{w_1}_{\Gamma_i}$ and $D^{w_2}_{\Gamma_i}$ for some subwords w_1 and w_2 of w. Such a γ-approximate convolution exists by Lemma 23. The new formula F' obtained by replacing all the convolutions in F by their $*_\gamma$-approximate convolutions clearly computes $D^w_{\Gamma_k}$. Thus, by Theorem 24

$$\|D - R\| \leq \gamma 2^{c_1|G|^{11}} = \delta.$$

□

7 Proof of the Key Convergence Lemma

Lemma 28 *Let P and R be probability distributions on the group G and η and ϵ be vectors such that $P + \eta$ and $R + \epsilon$ are natural decomposition (of two probability distributions). Let $\gamma > 0$. Then*

$$(P + \eta) *_\gamma (R + \epsilon) - P * R = \eta^{\|\langle R\rangle} + \bar{\eta},$$

where

$$\|\bar{\eta}\| \leq \lambda(R)\|\eta^{\perp\langle R\rangle}\| + \|\epsilon\| + \gamma,$$

and $\bar{\eta}^{\|\langle R\rangle} = 0$.

We will also use the dual version of this lemma.

Proof:

$$(P + \eta) *_\gamma (R + \epsilon) - P * R = \eta^{\|\langle R\rangle} + \eta^{\perp\langle R\rangle} * R + (P + \eta) * \epsilon + v,$$

where $v^{\|\langle R\rangle} = \vec{0}$ and $\|v\| \leq \gamma$ by Definition 7 (of the γ-approximate convolution). By Lemma 13, $((P + \eta) * \epsilon)^{\|R} = \vec{0}$, and by Lemma 12, $\|(P + \eta) * \epsilon\| \leq \|\epsilon\|$. By Proposition 11,

$$(\eta^{\perp\langle R\rangle} * R)^{\|R} = (\eta^{\perp\langle R\rangle})^{\|R} = \vec{0}.$$

We have $\|\eta^{\perp\langle R\rangle} * R\| \leq \lambda(R)\|\eta^{\perp\langle R\rangle}\|$ by the definition of $\lambda(R)$. Thus

$$\bar{\eta} := \eta^{\perp\langle R\rangle} * R + (P + \eta) * \epsilon + v$$

satisfies the conditions of the lemma. □

We need to introduce some more notation. Let a finitely dimensional real vector space be given. $\bar{0}$ will denote the trivial subspace $\{\vec{0}\}$. For a subspace V, we denote by $V^\perp$ the orthogonal complement of V. For two subspaces U, V, we will denote by $V^{\perp U} = V \cap U^\perp = \{v \in V;\ v \perp U\}$.

For two subspaces U and V that are not comparable by inclusion, we define the angle between them, $\angle(U, V)$, as follows. Let $W = U \cap V$. Then

$$\angle(U, V) = \min_{u \in U^{\perp W}, v \in V^{\perp W}, u, v \neq \vec{0}} \angle(u, v).$$

Similarly we define $\angle(u, V)$ for a vector $u \neq \vec{0}$ and a subspace V.

Given a subspace U, every vector v has a unique representation of the form $v = v^{\parallel U} + v^{\perp U}$ such that $v^{\parallel U} \in U$ and $v^{\perp U} \in U^{\perp}$. Clearly, $v^{\parallel U}$ is the projection of v to U.

Note that

$$\phi = \angle(u, V) \ \Leftrightarrow\ 0 \leq \phi \leq \pi/2 \text{ and } \sin\phi = \frac{\|u^{\perp V}\|}{\|u\|}. \tag{3}$$

For a partition $\mathcal{L}$ of the set $\{1, 2, \dots, m\}$, we shall denote by

$$U_{\mathcal{L}} := \{v \in \mathbb{R}^m;\ v^{\parallel\mathcal{L}} = v\},$$

the subspace of $\mathbb{R}^m$ of vectors v such that v is constant on every block of the partition $\mathcal{L}$.

Lemma 29 *Let $\mathcal{K}$ and $\mathcal{L}$ be incomparable partitions of the set $\{1, 2, \dots, m\}$ and let $\phi = \angle(U_{\mathcal{K}}, U_{\mathcal{L}})$. Then*

$$\sin(\phi/2) \geq \frac{1}{4m^{3/2}}.$$

We will apply this lemma to spaces defined by cosets and *double* cosets. We have to consider double cosets because in the following lemma we have the condition that $\mathcal{U}$ is closed under intersections. The intersection of a space defined by left cosets with a space defined by right cosets produces, in general, a space defined by double cosets.

Proof: We will first consider the case when the partitions $\mathcal{K}$ and $\mathcal{L}$ are connected. Then $U_{\mathcal{K}} \cap U_{\mathcal{L}} = \{a\vec{1};\ a \in \mathbb{R}\}$. Let ℓ be the number of blocks in the partition $\mathcal{L}$. Let $\phi = \angle(U_{\mathcal{K}}, U_{\mathcal{L}})$ and let $\kappa \subseteq U_{\mathcal{K}}$ and $\lambda \subseteq U_{\mathcal{L}}$ be lines orthogonal to $\vec{1}$ that span the angle ϕ. Let $\epsilon \perp \vec{1}$ be the unit vector such that $\angle(\kappa, \epsilon) = \angle(\epsilon, \lambda) = \phi/2$. Then $\sin(\phi/2) = \|\epsilon^{\perp\mathcal{K}}\| = \|\epsilon^{\perp\mathcal{L}}\|$.

Suppose that $\sin(\phi/2) < \frac{1}{4\ell\sqrt{m}}$. Set $\alpha = \sin(\phi/2)$ and $\beta = 1$. Then the assumptions of Lemma 10 are satisfied and we get $\|\epsilon^{\perp\mathcal{L}}\| > \frac{1}{4\ell\sqrt{m}}$. But this is a contradiction, hence $\sin(\phi/2) \geq \frac{1}{4\ell\sqrt{m}} \geq \frac{1}{4m^{3/2}}$ as required.

Now suppose that the partitions $\mathcal{K}$ and $\mathcal{L}$ are not connected. Let $\mathcal{J} = \{B_1, \dots, B_s\}$ be the join of $\mathcal{K}$ and $\mathcal{L}$, i.e., the finest partition whose blocks are unions of blocks of $\mathcal{K}$, respectively of $\mathcal{L}$. Then for every i, the restrictions of $\mathcal{K}$ and $\mathcal{L}$ to B_i are connected. The space $\mathbb{R}^m$ has the orthogonal decomposition $V_1 \oplus \dots \oplus V_s$, where V_i is the space of vectors that are zero everywhere except for B_i. The lemma in its general form is now a corollary of the special case above and the claim below.

Claim. Suppose that a given vector space has orthogonal decomposition $V_1 \oplus \dots \oplus V_s$. Let $u_i, v_i \in V_i$ be vectors with angles at least α, $0 < \alpha \leq \pi/2$ for all i. Then the the vectors $u = \sum_i u_i$ and $v = \sum_i v_i$ also have angle at least α.

Proof of Claim.

The condition on angles is equivalent to $\frac{\langle u_i, v_i\rangle}{\|u_i\|\|v_i\|} \leq \cos\alpha$. We need to show it also for u and v.

$$\frac{\langle u, v\rangle}{\|u\|\|v\|} = \frac{\sum\langle u_i, v_i\rangle}{\sqrt{\sum\|u_i\|^2}\sqrt{\sum\|u_i\|^2}} \leq \cos\alpha\frac{\sum\|u_i\|\|v_i\|}{\sqrt{\sum\|u_i\|^2}\sqrt{\sum\|u_i\|^2}} \leq \cos\alpha,$$

using Cauchy-Schwartz in the last inequality. □

In order to apply the claim we only need to realize what vectors $v \in U_{\mathcal{K}}$ and $u \in U_{\mathcal{L}}$ orthogonal to the intersection $U_{\mathcal{K}} \cap U_{\mathcal{L}} = U_{\mathcal{J}}$ are. Such a vector v (respectively u) restricted to a block B_i is constant on blocks of $\mathcal{K}$ (respectively $\mathcal{L}$) and orthogonal to the vector of 1s on B_i for every i. Hence the projections of the vectors u and v on subspaces V_i are precisely those to which the first part of the proof applies.

□

Lemma 30 *Let $0 < \lambda < 1, c > 0, 0 < \phi \leq \pi/2$ and $d \geq 0$ be constants. Let $\mathcal{U}$ be a set of subspaces such that*

1. $\dim U \leq d$ *for all* $U \in \mathcal{U}$,
2. $\mathcal{U}$ *is closed under intersections,*
3. $\angle(U, V) \geq \phi$ *for all* $U, V \in \mathcal{U}$ *incomparable by inclusion.*

Let $U_1, U_2, \ldots, U_t$ be a sequence of subspaces from $\mathcal{U}$. Let $v_0, v_1, v_2, \ldots, v_t$ be a sequence of vectors satisfying the following conditions.

1. $v_0 = \vec{0}$,
2. $v_{i+1} = v_i^{\|U_{i+1}} + u_i$,

   where $u_i \perp U_{i+1}$ *and* $\|u_i\| \leq \lambda \|v_i^{\perp U_{i+1}}\| + c$.

Then for all i,

$$\|v_i\| \leq f_{\lambda,c,\phi}(d) = O\left(\frac{c}{(1-\lambda)^{d+1}((\sin(\phi/2)^2/2)^{d(d+1)/2}}\right). \tag{4}$$

The key fact is that the bound does not depend on the length of the sequence.

Let us look more closely at one step in the sequence. Let us change notation and write that v is changed into v', and let the subspace be U. The condition is

$$v' = v^{\|U} + u, \quad \text{where} \quad u \perp U,\ \|u\| \leq \lambda \|v^{\perp U}\| + c.$$

First we change it to

$$v' = v^{\|U} + u' + z, \quad \text{where} \quad u', z \perp U,\ \|u'\| \leq \lambda \|v^{\perp U}\|,\ \|z\| \leq c.$$

Further, we can view u as obtained by first *shrinking* $v^{\perp U}$ to $\lambda' v^{\perp U}$, for some $0 \leq \lambda' \leq \lambda$, and then *rotating* to u inside of $U^{\perp}$. Note that this is a *linear* transformation. Thus we can assume that u' is a result of applying a linear transformation $L_{U,\lambda}$ to $v^{\perp U}$, a transformation that shrinks all vectors in $U^{\perp}$ by a factor at least λ. Further, we can extend $L_{U,\lambda}$ to the entire space by defining it to be the identity on U.

Hence we can assume that the sequence $v_1, \ldots, v_t$ is given by the recursion

2$'$. $v_{i+1} = L^{i+1}_{U_{i+1},\lambda}(v_i) + u_i$,

   where $u_i \perp U_{i+1}$, $\|u_i\| \leq c$,

   and $L^{i+1}_{U_{i+1},\lambda}$ is identity on U_{i+1}, preserves $U_{i+1}^{\perp}$ and shrinks all vectors $v \in U_{i+1}^{\perp}$ by a factor at least λ.

In the application the vectors are actually given by such linear transformations. In the sequel we will omit the upper indices in order to simplify notation. Instead we keep the lower indices that determine the properties of L that we need.

We will use the following simple facts:

- $\|L_{U,\lambda}(v)\| \leq \|v\|$ for all vectors v;

- if $U = \bar{0}$, then $L_{U,\lambda}$ shrinks all vectors;
- if $W \subseteq U$ is a subspace, then $L_{U,\lambda}$ preserves $U^{\perp W}$.

Lemma 31 *Suppose $v \notin U$ and $\angle(v, U) \geq \psi$. Then*

$$\|L_{U,\lambda}(v)\| \leq (1 - \epsilon\frac{(\sin\psi)^2}{2})\|v\|,$$

where $\epsilon = 1 - \lambda$.

Proof: This is precisely our Lemma 9 with $\delta = \sin\psi$, $x = v^{\perp U}$, $y = v^{\| U}$, and $x' = L_{U,\lambda}(v^{\perp U})$. □

Lemma 32 *Suppose that $U_1 \cap \dots \cap U_t = \bar{0}$, $\dim U_i \leq d$ for all i and assume the lower bound ϕ on the angles. Let $\delta = (\sin\frac{\phi}{2})^2/2$. Then*

$$\|L_{U_t,\lambda} \dots L_{U_1,\lambda}(v)\| \leq (1 - \epsilon\delta^d)\|v\|. \tag{5}$$

Proof: We will use induction on the dimension of subspaces d.

$\boxed{d = 0.}$
Then all $U_i = \bar{0}$. Hence already $\|L_{U_1,\lambda}(v)\| \leq \lambda\|v\| = (1 - \epsilon\delta^0)\|v\|$.

$\boxed{d \mapsto d + 1.}$
W.l.o.g. we can assume that $U_1 \cap \dots \cap U_{t'} \neq \bar{0}$ for every $t' < t$. If $U_t = \bar{0}$, then

$$\|L_{U_t,\lambda} \dots L_{U_1,\lambda}(v)\| \leq \lambda\|L_{U_{t-1},\lambda} \dots L_{U_1,\lambda}(v)\| \leq \lambda\|v\| \leq (1 - \epsilon\delta^d)\|v\|.$$

Otherwise $t \geq 2$. Let $W = U_2 \cap \dots \cap U_t$. Then $U_1 \cap W = \bar{0}$ and $\angle(U_1, W) \geq \phi$. (Here we are using the fact that $\mathcal{U}$ is closed under intersections.) Consider two cases.

1. If $\angle(v, U_1) \geq \phi/2$, then we get, by Lemma 31,

$$\|L_{U_t,\lambda} \dots L_{U_1,\lambda}(v)\| \leq \|L_{U_1,\lambda}(v)\| \leq (1 - \epsilon\delta)\|v\| \leq (1 - \epsilon\delta^d)\|v\|.$$

2. Otherwise $\angle(v, U_1) < \phi/2$. Let $w = L_{U_1,\lambda}(v)$. Since $L_{U_1,\lambda}$ preserves $v^{\| U_1}$ and contracts $v^{\perp U_1}$, also $\angle(w, U_1) < \phi/2$. This can be seen from the following inequalities.

$$\tan(\angle(w, U_1)) = \frac{\|w^{\perp W}\|}{\|w^{\| W}\|} = \frac{\|L(v^{\perp U_1})\|}{\|v^{\| W}\|} \leq \frac{\|v^{\perp W}\|}{\|v^{\| W}\|} = \tan(\angle(v, U_1)) \leq \tan(\phi/2).$$

Hence $\angle(w, W) \geq \phi/2$.

We will apply the induction hypothesis to the vector $w^{\perp W}$ and the subspaces $U_2^{\perp W}, \dots, U_t^{\perp W}$ which have dimensions $\leq d$. This gives us

$$\|L_{U_t,\lambda} \dots L_{U_2,\lambda}(w^{\perp W})\| \leq (1 - \epsilon\delta^d)\|w^{\perp W}\|.$$

Let u be the vector on the left hand side. Then, by Lemma 31 (with $\psi = \phi/2$),

$$\|L_{U_t,\lambda} \dots L_{U_2,\lambda}(w)\| = \|w^{\| W} + u\| = L_{W,\lambda'}(w) \leq (1 - \epsilon'\delta)\|w\|,$$

where $\epsilon' = \epsilon\delta^d$ and $\lambda' = 1 - \epsilon'$. Substituting for ϵ' and using the fact that $\|w\| \leq \|v\|$ we obtain the inequality (5) with $d := d + 1$.

□

Proof of Lemma 30. By induction.

$\boxed{d = 0.}$ $f_{\lambda,c,\phi}(0) = \frac{c}{1-\lambda}$ (using the formula for the sum of a geometric series).

$\boxed{d \mapsto d + 1.}$ Let $k_0 = 0$. Let k_{j+1} be the least index $> k_j$ such that $U_{k_j+1} \cap \cdots \cap U_{k_{j+1}} = \bar{0}$. Thus $U_1, \dots, U_t$ is divided into segments, each having empty intersection except possibly for the last one.

(a) Let j be an arbitrary number such that $U_{k_j+1} \cap \cdots \cap U_{k_{j+1}} = \bar{0}$. Consider two cases.

1. $k_{j+1} = k_j + 1$. Then $U_{k_{j+1}} = \bar{0}$, whence $\|v_{k_{j+1}}\| \leq \lambda \|v_{k_j+1}\| + c$.
2. $k_{j+1} > k_j + 1$. Then $W = U_{k_j+1} \cap \cdots \cap U_{k_{j+1}-1} \neq \bar{0}$.

Let $t = k_{j+1} - k_j$. Define decompositions of the vectors $v_{k_j+i} = x_i + y_i$, for $i = 0, 1, 2, \dots, t$ as follows.

$x_0 = v_{k_j}$, $y_0 = \vec{0}$;

$x_{i+1} = L_{U_{k_j+i+1},\lambda}(x_i)$,
$y_{i+1} = L_{U_{k_j+i+1},\lambda}(y_i) + w_{i+1}$,

where the linear operation and the vector w_{i+1} are those used in the sequence of vectors v_k (i.e., $v_{k_j+i+1} = L_{U_{k_j+i+1},\lambda}(v_{k_j+i}) + w_{i+1}$). By Lemma 32,

$$\|x_t\| \leq \lambda' \|v_{k_j}\|,$$

where $\lambda' = (1 - \epsilon\delta^{d+1})$. Since $W \neq \bar{0}$, we can apply the induction assumption to $U^{\perp W}_{k_j+1}, \dots, U^{\perp W}_{k_{j+1}-1}$. This gives us

$$\|y_{t-1}\| \leq f_{\lambda,c,\phi}(d),$$

whence

$$\|y_t\| \leq \|L_{U_{k_{j+1}-1}}(y_{t-1}) + w_t\| \leq \|y_{t-1}\| + \|w_t\| \leq f_{\lambda,c,\phi}(d) + c.$$

Let us denote by $c' = f_{\lambda,c,\phi}(d) + c$. Then we have

$$\|v_{k_{j+1}}\| \leq \lambda' \|v_{k_j}\| + c'.$$

As $\lambda \leq \lambda'$ and $c \leq c'$, this inequality holds true in both cases.

(b) Let now r be such that $k_r + 1, \dots, n$ is the last segment. Then the same argument gives us

$$\|y_t\| \leq f_{\lambda,c,\phi}(d) + c.$$

But if $U_{k_r+1} \cap \cdots \cap U_n \neq \bar{0}$ we cannot use Lemma 32 to bound $\|x_t\|$ as above, so we only use $\|x_t\| \leq \|x_0\| = \|v_{k_r}\|$. Thus we get

$$\|v_t\| \leq \|v_{k_r}\| + c'.$$

Now we can estimate $\|v_t\|$. Using part (a) we get

$$\|v_{k_r}\| \leq \frac{c'}{1-\lambda'},$$

again using the sum of a geometric series. From (b) we then get

$$\|v_t\| \leq \frac{c'}{1-\lambda'} + c'.$$

This leads to the following recursion:

$f_{\lambda,c,\phi}(0) = \frac{c}{\epsilon}$;

$f_{\lambda,c,\phi}(d+1) = \frac{f_{\lambda,c,\phi}(d)+c}{\epsilon\delta^{d+1}} + f_{\lambda,c,\phi}(d)$.

Write the right hand side as

$$\frac{f_{\lambda,c,\phi}(d)}{\epsilon\delta^{d+1}} + \frac{c}{\epsilon\delta^{d+1}} + f_{\lambda,c,\phi}(d).$$

Assuming that $f_{\lambda,c,\phi}(d) \geq \frac{c}{\epsilon^{d+1}\delta^{d(d+1)}}$ the sum of the two last terms is exponentially smaller than the first term. Whence the recursion can be solved by a function that satisfies

$$f_{\lambda,c,\phi}(d) = O(\frac{c}{\epsilon^{d+1}\delta^{d(d+1)}}).$$

□

Proof of the Key Convergence Lemma 27. Let constants $0 < e_1, \gamma < 1$ and probability distributions $D_0, D_1, \ldots, D_t$, $R_0, R_1, \ldots, R_t$ be given. Recall that we are assuming that $D_i = R_i + \epsilon_i$ is a natural decomposition, for $i \in \{0, \ldots, t\}$ and $\|\epsilon_i\| \leq e_1$ for $i > 0$.

Further we can w.l.o.g. assume that $\epsilon_0 = \vec{0}$, because if we decompose the errors into a part resulting from ϵ_0 and the rest, then the first part will not increase in the process, because $\|x * D\|, \|D * x\| \leq \|x\|$ (by Lemma 12) .

Let $C_0 := D_0$ and $C_{i+1} := C_i *_\gamma D_{i+1}$, respectively $C_{i+1} := D_{i+1} *_\gamma C_i$ for $i = 0, \ldots, t-1$. Let $P_i + \eta_i = C_i$ be natural decompositions. In order to apply Lemmas 29 and 30, we define

$$U_i := \{v;\ v^{\|\langle R_i\rangle} = v\},\ \text{if}\ C_{i+1} = C_i *_\gamma D_i, \qquad \text{and} \qquad U_i := \{v;\ v^{\langle R_i\rangle\|} = v\},\ \text{if}\ C_{i+1} = D_i *_\gamma C_i.$$

Let $\mathcal{U}$ be the set of all possible intersections of spaces U_i. They are spaces of vectors that are constant on blocks of some partitions (given by cosets or double cosets), thus we can apply the lower bound on the angle from Lemma 29. Let $v_i := \eta_i$ and let $m = d = |G|$, $c = e_1 + \gamma$, $\lambda = 1 - 1/c_G$. We shall verify the conditions of Lemma 30.

By Lemma 28, assuming that the multiplication in the step $i+1$ is from the right, we have the decomposition

$$v_{i+1} = \eta_{i+1} = \eta_i^{\|\langle R_{i+1}\rangle} + \bar{\eta} = v_i^{\|\langle R_{i+1}\rangle} + \bar{\eta},$$

where

$$\|\bar{\eta}\| \leq \lambda(R_{i+1})\|\eta_i^{\perp\langle R_{i+1}\rangle}\| + \|\epsilon_i\| + \gamma \leq \lambda\|\eta_i^{\perp\langle R_{i+1}\rangle}\| + c = \lambda\|v_i^{\perp\langle R_{i+1}\rangle}\| + c$$

and $\bar{\eta}^{\|\langle R_{i+1}\rangle} = 0$. Clearly, $v_i^{\|\langle R_{i+1}\rangle} = v_i^{\|U_{i+1}}$, and $\bar{\eta}^{\|\langle R_{i+1}\rangle} = 0$ is equivalent to $\bar{\eta} \perp U_{i+1}$. Similarly, if the multiplication in the step $i+1$ is from the left, $v_i^{\langle R_{i+1}\rangle\|} = v_i^{\|U_{i+1}}$, and $\bar{\eta}^{\langle R_{i+1}\rangle\|} = 0$ is equivalent to $\bar{\eta} \perp U_{i+1}$.

It remains to substitute the bounds on $\lambda = 1 - 1/c_G$ from Lemma 20 and $\sin(\phi/2)$ from Lemma 29 into the inequality (4) of Lemma 30, which we leave to the reader.

□

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple construction of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992.

[AKS87] M. Ajtai, J. Komlos, and E. Szemeredi. Deterministic simulation in logspace. In *Proceedings of the nineteenth annual ACM symposium on Theory of Computing (STOC)*, pages 132–140, 1987.

[AS92] Noga Alon and Joel Spencer. *The Probabilistic Method.* John Wiley, 1992.

[Bar89] David A. Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *Journal of Computer and System Sciences*, 38(1):150 – 164, 1989.

[BNS89] László Babai, Noam Nisan, and Mario Szegedy. Multiparty protocols and logspace-hard pseudorandom sequences (extended abstract). In *Proceedings of the twenty first annual ACM Symposium on Theory of Computing (STOC)*, pages 1–11, 1989.

[BRRY10] Mark Braverman, Anup Rao, Ran Raz, and Amir Yehudayoff. Pseudorandom generators for regular branching programs. In *Proceedings of the fifty first annual symposium on Foundations of Computer Science (FOCS) (To appear)*, 2010.

[BV10] Joshua Brody and Elad Verbin. The coin problem, and pseudorandomness for branching programs. In *Proceedings of the fifty first annual symposium on Foundations of Computer Science (FOCS) (To appear)*, 2010.

[GG81] O. Gabber and Z. Galil. Explicit constructions of linear-sized superconcentrators. *Journal of Computer and System Sciences*, 22(3):407–420, 1981.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the 26th annual ACM Symposium on Theory of Computing (STOC)*, pages 356–364, 1994.

[Lov93] László Lovász. *Combinatorial Problems and Exercises.* Akadémiai Kiadó, Budapest, 1993.

[LRTV09] Shachar Lovett, Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Pseudorandom bit generators that fool modular sums. In *APPROX-RANDOM*, pages 615–630, 2009.

[MZ09] Raghu Meka and David Zuckerman. Small-bias spaces for group products. In *APPROX-RANDOM*, pages 658–672, 2009.

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computations. *Combinatorica*, 12(4):449–461, 1992.

[Nis94] Noam Nisan. $\mathsf{RL} \subseteq \mathsf{SC}$. *Computational Complexity*, 4(1):1–11, 1994.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[RR99] Ran Raz and Omer Reingold. On recycling the randomness of states in space bounded computation. In *Proceedings of the thirty first annual ACM Symposium on Theory of Computing (STOC)*, pages 159–168, 1999.

[RV05] Eyal Rozenman and Salil P. Vadhan. Derandomized squaring of graphs. In *APPROX-RANDOM*, pages 436–447, 2005.

[RVW00] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Annals of Mathematics*, pages 157–187, 2000.

[Sav70] Walter J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *Journal of Computer and System Sciences*, 4(2):177–192, 1970.

[SZ99] Michael E. Saks and Shiyu Zhou. Bp $_{\mathrm{h}}$space(s) $\subseteq$ dspace(s$^{3/2}$). *Journal of Computer and System Sciences*, 58(2):376–403, 1999.

[vv10] Jiří Šíma and Stanislav Žák. A polynomial time construction of a hitting set for read-once branching programs of width 3. Technical Report 088, Electronic Colloquium on Computational Complexity (ECCC), 2010.