Electronic Colloquium on Computational Complexity, Report No. 113 (2010)

Pseudorandom Generators for Group Products

preliminary version*

Michal Koucký†, Prajakta Nimbhorkar‡and Pavel Pudlák§

July 16, 2010

Abstract

We prove that the pseudorandom generator introduced in [INW94] fools group products of a given finite group. The seed length is $O(\log n \log \frac{1}{\varepsilon})$, where n the length of the word and ε is the precision. The result is equivalent to the statement that the pseudorandom generator fools read-once permutation branching programs of constant width.

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 solve the problem for *permutation* read-once constant width branching programs.

*We did not attempt to optimize the constants involved. We are also aware that several arguments need more detailed explanations. We plan to improve the presentation in subsequent revisions.

†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.

‡The Institute of Mathematical Sciences C. I. T. Campus, Taramani, Chennai-600113, e-mail: prajakta@imsc.res.in. 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.

ISSN 1433-8092

When working with permutation read-once constant width branching programs, it is more natural to recast the problem in terms of finite groups. Let G be a finite group and $w = (g_1, g_2, \ldots, g_n)$ 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

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

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_Γ^w is defined by

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

($\Gamma(y)_i$ are the bits of the string $\Gamma(y) \in \{0,1\}^n$.) The goal is to find pseudorandom generators Γ such that D_Γ^w 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_Γ^w 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_Γ^w of distance ε for arbitrarily small positive ε. 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 $\|\mathrm{Rnd}^w - D_\Gamma^w\|_\infty \leq \varepsilon$, for arbitrary constant $\varepsilon > 0$, where the seed length is $O(\log n \cdot (|G|^{O(1)} + \log 1/\varepsilon))$.

Note that 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 $g_1^{x_1} \ldots g_n^{x_n}$, $g_i \in \mathcal{S}_k$, where $\mathcal{S}_k$ is the symmetric group on k elements. We will explain this connection in Section 1.3.

1.1 Comparison with previous results

There has been a series of results concerning the power 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 due to [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 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 would suffice to derandomize RNC^1.

For width 2 branching programs, a generator having error ε is equivalent to an ε-biased space, which can be constructed with $O(\log n + \log \varepsilon^{-1})$ 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 $O((w^4 \log\log n + \log 1/\varepsilon) \log n)$ and by [BRRY10] for width w *regular* read-once branching programs, which has seed-length $O((\log w + \log\log n + \log 1/\varepsilon) \log n)$. Regular branching programs are more general than permutations branching programs. In [vv10], a construction of polynomial-time computable 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. The 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 $O(\log n(\log 1/\varepsilon + exp(w)))$ 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 $w = w_1 w_2$ 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 inform 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.[1] 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.

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 $g_1 \ldots g_k \in (\mathcal{S}_k)^n$, 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 pseudo-random generators that would approximate the probability that a random

[1] In fact it is more complicated. We have to consider double cosets determined by a pair of subgroups.

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 $1 \leq i, j \leq k$, 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, $1 \leq i \leq k$, 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.

1.4 Map of the paper

Section 2 contains notations and elementary preliminaries that are necessary for the rest of the paper. We rely heavily on not-quite-standard notation for several concepts, for which we do not know of any standard notation. Section 3 contains the description of the pseudorandom generator and the statement of our main result. This section uses very little of the non-standard notation and one can possibly read it with only intuitive understanding of γ-approximate convolution. The same goes also for most of next Section 4 which contains the proof that expander product can be viewed as a γ-approximate convolution. Section 5 contains the statements related to the main technical contribution of this paper which is Approximate Convolution Theorem. Rest of the paper is devoted to the proof of this theorem. Section 6 contains elementary facts about various concepts used later in the proofs. These facts do not form a cohesive reading and the reader may want to just quickly glance at the statements and return to the proofs only later. These facts are referred in Sections 7 and 8 which contain the proofs of the main technical tools.

2 Notations and Preliminaries

2.1 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 in-

dexes.) An all one vector is denoted by $\mathbb{1}$, 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 disjoint nonempty sets $S_1, S_2, \ldots, S_k \subseteq \{1, 2, \ldots, m\}$, $x^{\|(S_1,S_2,\ldots,S_k)}$ is the vector in $\mathbb{R}^m$ such that for all $i \in \{1, \ldots, m\}$:

1. if $i \in S_j$ for some $j \in \{1, \ldots, k\}$ then $x_i^{\|(S_1,S_2,\ldots,S_k)} = \frac{1}{|S_j|} \sum_{\ell \in S_j} x_\ell$, and
2. $x_i^{\|(S_1,S_2,\ldots,S_k)} = x_i$ otherwise.

Thus, $x^{\|(S_1,S_2,\ldots,S_k)}$ is constant on sets (of coordinates) $S_1, S_2, \ldots, S_k$. Furthermore, $x^{\perp(S_1,S_2,\ldots,S_k)} = x - x^{\|(S_1,S_2,\ldots,S_k)}$. Notice that $x^{\perp(S_1,S_2,\ldots,S_k)}$ is zero in coordinates not belonging to any of the S_i's. The following facts are easy to check:

Proposition 1 *Let $m \geq 1$ be an integer. For any $x \in \mathbb{R}^m$ and disjoint nonempty sets $S_1, S_2, \ldots, S_k \subseteq \{1, 2, \ldots, m\}$ it holds:*

$$\begin{aligned} x^{\perp(S_1,S_2,\ldots,S_k)} &\perp x^{\|(S_1,S_2,\ldots,S_k)}, \\ x^{\perp(S_1,S_2,\ldots,S_k)} &= \sum_{i=1}^{k} x^{\perp(S_i)}. \end{aligned}$$

Remark 2 *The reader may find helpful the following explanation of the above notation. If one were to plot a graph of $x^{\|(S_1,S_2,\ldots,S_k)}(i)$ as a function of i, one would see it consists of several parallel (horizontal) lines as $x^{\|(S_1,S_2,\ldots,S_k)}$ is constant on each of the sets $S_1, \ldots, S_k$. Hence the notation $\|$. On the other hand $x^{\perp(S_1,S_2,\ldots,S_k)}$ is the component of x that sticks out in such a graph from the parallel lines. Hence the use of symbol $\perp$. (It also happens that $x^{\perp(S_1,S_2,\ldots,S_k)}$ is parallel to any $y^{\|(S_1,S_2,\ldots,S_k)}$.)*

We say that two partitions $(L_1, L_2, \ldots, L_\ell)$ and $(K_1, K_2, \ldots, K_k)$ of $\{1, 2, \ldots, m\}$ are *connected* if for any $x, y \in S$ there exists a *path of hyperedges* between x and y, i.e., a sequence of sets $C_1, C_2, \ldots, C_m$ where each $C_i \in \{L_1, L_2, \ldots, L_\ell, K_1, \ldots, K_k\}$, $x \in C_1$, $y \in C_m$ and $C_i \cap C_{i+1} \neq \emptyset$ for $i \in \{1, \ldots, m-1\}$.

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 $\mathbb{R}^G$.

2.2 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 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 $g \in G$.

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

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

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, $1_G \in \mathrm{supp}(D)$ and for every $g \in \mathrm{supp}(D)$, $D(g) \geq 2^{-|\mathrm{supp}(D)|}$. However, we will not need the latter fact. Clearly, convolution of two natural distributions is natural. 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, $\epsilon = D - R$, $\epsilon \perp \mathbb{1}$, and $\mathrm{supp}(\epsilon) \subseteq \mathrm{supp}(R)$. Observe that, 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 $D = R_1 * R_2 + \epsilon'$.

Our goal will be to approximate the actual convolution by an expander product thus, we consider also any γ-*approximate convolution* $*_\gamma$, defined below:

Definition 5 *For $0 < \gamma \in \mathbb{R}$, $*_\gamma : \mathbb{R}^G \times R^G \to \mathbb{R}^G$ is any operation that satisfies the following properties. For any two probability distributions $D_1, D_2 \in \mathbb{R}^G$:*

1. *$D_1 *_\gamma D_2$ is a probability distribution,*
2. *$\mathrm{supp}(\epsilon) \subseteq \mathrm{supp}(D_1 * D_2)$,*
3. *$||\epsilon^{\langle D_1 \rangle \|}|| = ||\epsilon^{\| \langle D_2 \rangle}|| = 0$, and*
4. *$||\epsilon^{\langle D_1 \rangle \perp}||, ||\epsilon^{\perp \langle D_2 \rangle}|| < \gamma$,*

where $\epsilon = D_1 * D_2 - D_1 *_\gamma D_2$.

The meaning of the third 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 other words, each right $\langle D_1 \rangle$-coset has the same probability mass in $D_1 * D_2$ and D, and similarly for left $\langle D_2 \rangle$-cosets.

Remark 6 *The reader may be puzzled by the fact that $*_\gamma$ is defined as a function (operation) on distributions. When considering pseudorandom generators it typically matters not only what* is *a particular distribution but also how it* was obtained *(as that may determine how such distributions will be composed by the generator). This is true for our construction as well. However, one can achieve similar effect by using different $*_\gamma$ operations at*

different places tailored to particular distributions since $*_\gamma$ is not defined uniquely. Thus for example in Approximate Convolution Theorem we allow different $*_\gamma$ at different places of the formula and the theorem remains true for arbitrary choice of these operations.

In order to shield the reader from particular technical details how a distribution was obtained in the technically most difficult part of the paper (Section 5 and related sections) we opted for this definition. Thus Section 5 can be read without having any particular pseudorandom generator in mind.

One could possibly think of $*_\gamma$ as a some kind of fuzzy *operation or ternary relation which satisfies the above condition.*

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 $\lambda(D) = \max\{\lambda^{\mathrm{R}}(D), \lambda^{\mathrm{L}}(D)\}$.

3 The pseudorandom generator

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

3.1 Expander product

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 $y' \in \{0,1\}^d$, gives a neighbor of y in F that is reached by the edge labeled y'. 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

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

Notice that given random $y \in \{0,1\}^r$ and $y' \in \{0,1\}^d$, the pair $(y, \nu(y, y'))$ 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.)

3.2 INW generator

For the construction of INW generator we need an explicitely 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 7 *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 $y' \in \{0,1\}^d$, we can compute $\nu(y, y')$ 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 8 (Main Theorem) *Let G be any finite group of size at least four and $0 < \delta < 1$ be arbitrary. Let $\lambda = \delta/(2^{c_1|G|^{12}} \cdot \sqrt{|G|})$, where c_1 is the constant from Theorem 10. Then (λ, n)-INW generator Γ uses seeds of length $O(\log n \cdot (|G|^{12} + \log 1/\delta))$ to produce n bits such that for every $w \in G^n$,*

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

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. Since every group of size three or less is a subgroup of some group of size six, the theorem is also applicable to such groups.

We provide the proof of Main Theorem next. 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. These two ingredients are formalized in the following two statements.

Lemma 9 *Let a word w over some group G be given. Let $w = w_1 w_2$, 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 $\lambda = \gamma/\sqrt{|G|}$. There is a γ-approximate convolution $*_\gamma : \mathbb{R}^G \times \mathbb{R}^G \to \mathbb{R}^G$ such that $D^{w_1}_{\Gamma_1} *_\gamma D^{w_2}_{\Gamma_2} = D^{w_1 w_2}_{\Gamma_1 \otimes_F \Gamma_2}$.*

This lemma is proven in Section 4. The following theorem is the main technical tool of this paper.

Theorem 10 *There is a universal constant $c_1 > 0$ such that for any group G of size at least four and any $0 < \gamma < 1$ 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

$$\|D - R\| \leq \gamma 2^{c_1 |G|^{12}}.$$

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. This theorem is an immediate consequence of Theorem 13 stated in Section 5.

We are ready to prove Main Theorem.

Proof of Main Theorem. Let $\gamma = \delta/2^{c_1|G|^{12}}$ and $\lambda = \gamma/\sqrt{|G|}$. 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 $\mathrm{Rnd}^{w'}$. Observe, $\mathrm{Rnd}^{w'} = D^{w'}_{\Gamma_0}$. Using convolution form a balanced formula F out of $\mathrm{Rnd}^{w'}$, 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 $D^{w_1 w_2}_{\Gamma_i \otimes_F \Gamma_i} = D^{w_1 w_2}_{\Gamma_{i+1}}$ 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 9. 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 10

$$\|D - R\| \leq \gamma 2^{c_1|G|^{12}} = \delta.$$

□

4 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 9. Similar bounds were proven in [INW94]. Rather than adapting their results, we will give a direct proof based on Expander Mixing Lemma.

Lemma 11 *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 11 uses *Expander Mixing Lemma*, stated below (see e.g. [AS92] Corollary 2.5).

Lemma 12 (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 11. Let $w_1 = g_1 \dots g_n$, $w_2 = h_1 \dots h_n$, $N = 2^r$ and $M = 2^d$. For $g \in G$ put

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

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

$$\left| e(U_g, V_h) - \frac{M \cdot |U_g| \cdot |V_h|}{N} \right| \leq \lambda M \sqrt{|U_g| \cdot |V_h|} \tag{1}$$

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)

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

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

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

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. □

This allows us to prove Lemma 9.

Proof of Lemma 9. Define $D^{w_1}_{\Gamma_1} *_\gamma D^{w_2}_{\Gamma_2} = D^{w_1 w_2}_{\Gamma_1 \otimes_F \Gamma_2}$. For other $D_1, D_2 \in \mathbb{R}^G$, where $D_1 \neq D^{w_1}_{\Gamma_1}$ or $D_2 \neq D^{w_2}_{\Gamma_2}$, we can extend $*_\gamma$ almost arbitrarily so we define $D_1 *_\gamma D_2 = D_1 * D_2$. Clearly, we only have to verify the latter three conditions of Definition 5 concerning $\epsilon = D^{w_1}_{\Gamma_1} * D^{w_2}_{\Gamma_2} - D^{w_1 w_2}_{\Gamma_1 \otimes_F \Gamma_2}$. 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 $i = 1, 2$. □

5 Main Approximate Convolution Theorem

In this section we establish our Approximate Convolution Theorem stated below which shows that an arbitrary convolution of natural distributions can be well approximated by γ-approximate convolutions. For this and following sections we will fix a finite group G of size at least 4. We define two parameters Δ and τ depending on the group:

$$\begin{aligned} \Delta &= \frac{1}{16|G|^2} \\ \tau &= \frac{\Delta^2}{2 \cdot c_G \cdot \sqrt{|G|}} = \frac{1}{8192 \cdot |G|^{8.5}}. \end{aligned}$$

Here c_G is the constant from Lemma 31. Let $T = \lceil 1/\tau \rceil$.

We state Approximate Convolution Theorem next.

Theorem 13 (Approximate Convolution Theorem) *Let $0 < \gamma < 1$ be real. Let $E_0 = 0$ and $E_i = 2h \cdot (E_{i-1} + \gamma)$, for $i > 0$, where h is the constant from Lemma 16. Assume that γ is small so that $E_T < 1/8|G|$.*

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 $\|D - R\| \leq E_T$.

Solving the recurrence $E_i = 2h \cdot (E_{i-1} + \gamma)$ in Theorem 13 gives the error $E_T \leq (2h)^{1+T}\gamma$.

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 14 *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. (Section 7 contains its proof.)

Lemma 15 *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* $\mathrm{rank}(R_2) < \mathrm{rank}(R_1 * R_2)$*.*

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 16 (Key Convergence Lemma) *Let* $0 < e_1, \gamma < 1$ *be reals. Let* $D_0, D_1, D_2, \ldots, D_t, R_0, \ldots, R_t$ *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*

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

where $h = (600^{|G|} \cdot |G|^{3|G|} \cdot c_G)^{|G|} \cdot 1200^{|G|} \cdot |G|^{4|G|}$.

The proof of Key Convergence Lemma is in Section 8.

Proof of Approximate Convolution Theorem. 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, $\|R_u - D_u\| \leq E_{\mathrm{rank}(R_u)}$. 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 25). Consider a node u of degree two in such a tree with children v and w. Distribution $D_u = R_u + \epsilon_u$ is the γ-approximate convolution of two distributions $D_v = R_v + \epsilon_v$ and $D_w = R_w + \epsilon_w$, where $\mathrm{rank}(R_v) = \mathrm{rank}(R_w) = \mathrm{rank}(R_u)$. (All the decompositions are natural.)

By Lemma 15, 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 26 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 $E_i/4 + h(E_{i-1} + \gamma) \leq E_i$. The claim follows. □

6 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 our main Approximate Convolution Theorem. The reader may want to skip this section during the first reading and use it only later as a reference.

6.1 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 17 *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*

$$\|x + y\| - \|x' + y\| \geq \frac{\epsilon\delta^2}{2}\|x + y\|.$$

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

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

Therefore

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

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

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

□

Lemma 18 *Let $0 < \mu$ be a real and $x, x', y \in \mathbb{R}^m$ vectors satisfying $x \perp y$, $x' \perp y$, and $\|x'\| \leq \|x\| + \mu$. Then*

$$\|x' + y\| \leq \|x + y\| + \mu.$$

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

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

The lemma follows. □

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 19 *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 $\epsilon \perp \mathbb{1}$. Let $\alpha, \beta \in \mathbb{R}$ satisfy $\alpha > 0$ and $\beta > 4\alpha\ell\sqrt{m}$. If $\|\epsilon\| \geq \beta$ and $\|\epsilon^{\perp\mathcal{L}}\| < \alpha$ 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 $\epsilon \perp \mathbb{1}$, there is also a coordinate $min \in \{1, \ldots, m\}$ with $\epsilon_{min} < 0$. Since $\|\epsilon^{\perp\mathcal{L}}\| < \alpha$, the absolute value of each coordinate of $\epsilon^{\perp\mathcal{L}}$ is smaller than α. Thus, the coordinates of $\epsilon^{\perp\mathcal{L}}$ corresponding to the elements of the same part L_i differ by less than 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

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

□

6.2 Facts on convolution

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

Proposition 20 *For a finite group G, let $x, D \in \mathbb{R}^G$, where D is a probability distribution. Let $(L_1, L_2, \ldots, L_\ell)$ be distinct left $\langle D\rangle$-cosets and $(K_1, K_2, \ldots, K_k)$ be distinct right $\langle D\rangle$-cosets of G. Then*

$$\begin{aligned}
(x * D)^{\|\langle D\rangle} &= x^{\|\langle D\rangle}, \\
(x * D)^{\perp(L_1,\ldots,L_\ell)} &= x^{\perp(L_1,\ldots,L_\ell)} * D, \\
(D * x)^{\langle D\rangle\|} &= x^{\langle D\rangle\|}, \\
(D * x)^{\perp(K_1,\ldots,K_k)} &= D * x^{\perp(K_1,\ldots,K_k)}.
\end{aligned}$$

The following is a consequence of the previous proposition.

Lemma 21 *Using the same notation as in Proposition 20*

$$\begin{aligned}
\|x * D\|, \|D * x\| &\leq \|x\|, \\
\|(x * D)^{\perp(L_1,\ldots,L_\ell)}\| &\leq \|x^{\perp(L_1,\ldots,L_\ell)}\|, \\
\|(D * x)^{\perp(K_1,\ldots,K_k)}\| &\leq \|x^{\perp(K_1,\ldots,K_k)}\|.
\end{aligned}$$

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 22 *For a finite group G, let $x, y \in \mathbb{R}^G$ and $H \leq G$ be a subgroup of G. If $x \perp \mathbb{1}$ 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. □

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

Proposition 23 *Let G be a finite group. Let $\epsilon, D \in \mathbb{R}^G$, where D is a probability distribution. Let $(A_1, A_2, \ldots, A_\ell)$ be different left $\langle D \rangle$-cosets. Then*

$$\|\epsilon^{\perp(A_1, A_2, \ldots, A_\ell)} * D\| \leq \lambda(D) \cdot \|\epsilon^{\perp(A_1, A_2, \ldots, A_\ell)}\|.$$

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

Proposition 24 *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 25 *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 21

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

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 26 *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 $\mathbb{1}$ and $\mathrm{supp}(\epsilon_1), \mathrm{supp}(\epsilon_2) \subseteq H$. Then*

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

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 $\epsilon_2 \perp \mathbb{1}$, $R_1^{H\|} * \epsilon_2$ 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 24,

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

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 $*_\gamma$. □

6.2.1 Facts on groups

Lemma 27 *Let G be a finite group, $L, K, H \leq G$ be subgroups of G, and $g \in G$. The following hold.*

1. *The set $\langle L \cup H \rangle gK$ can be partitioned into right H-cosets, that we call $A_1, \ldots, A_\ell$.*
2. *The set $\langle L \cup H \rangle gK$ can be partitioned into sets $B_1, \ldots, B_k$, each B_i of the form $Lg'K$, for some $g' \in G$.*
3. *The two partitions $(A_1, \ldots, A_\ell)$ and $(B_1, \ldots, B_k)$ are connected partitions of $\langle L \cup H \rangle gK$.*

Symmetrically for $Lg \langle K \cup H \rangle$ and its decomposition into left H-cosets.

Proof:

Part 1. Notice that $\langle L \cup H \rangle gK$ is a disjoint union of the right cosets of the form $\langle L \cup H \rangle gk$ for $k \in K$. Since H is a subgroup of $\langle L \cup H \rangle$, every right coset of $\langle L \cup H \rangle$ can be partitioned into right cosets of H.

Part 2. Let there be an $Lg'K$ such that $Lg'K \cap \langle L \cup H \rangle gK \neq \emptyset$ and $Lg'K \nsubseteq \langle L \cup H \rangle gK$. Let $\ell_1 g' k_1 \in \langle L \cup H \rangle gK$ where $\ell_1 \in L, k_1 \in K$. Therefore $\ell_1 g' k_1 = \ell_2 g k_2$ for some $\ell_2 \in \langle L \cup H \rangle, k_2 \in K$. Let $\ell_3 g' k_3 \notin \langle L \cup H \rangle gK$.

But then

$$\ell_3 g' k_3 = \ell_3 \ell_1^{-1} \ell_1 g' k_1 k_1^{-1} k_3 = \ell_3 \ell_1^{-1} \ell_2 g k_2 k_1^{-1} k_3 \in \langle L \cup H \rangle gK$$

which contradicts the assumption that $\ell_3 g' k_3 \notin \langle L \cup H \rangle gK$.

Part 3. Let us assume that $k, \ell \geq 2$ as otherwise the claim is trivial. Clearly it suffices to show that one can find a *path of hyper-edges* consisting of A_i's and B_j's between g and any other element $g' \in \langle L \cup H \rangle gK$. Let $g' = \ell_m \ell_{m-1} \cdots \ell_2 \ell_1 g k_0$, where $\ell_1, \ell_2, \ldots, \ell_m \in L \cup H$ and $k_0 \in K$. (Since $L, H \leq G$ and G is finite such elements exist.) We show by induction on $i = 0, 1, \ldots, m$ that $\ell_i \ell_{i-1} \cdots \ell_1 g k_0$ is connected to g by a path of hyper-edges.

Base case. Clearly, g and gk_0 are both from LgK so this case is trivial.

Induction step. Let us assume that there is a path of hyper-edges between g and $\ell_{i-1} \cdots \ell_1 g k_0$. We will extend it into a path from g to $\ell_i \ell_{i-1} \cdots \ell_1 g k_0$. There are two cases. Consider the case when $\ell_i \in H$. Since $\langle L \cup H \rangle gK$ decomposes into right H-cosets and $\ell_{i-1} \cdots \ell_1 g k_0 \in \langle L \cup H \rangle gK$, $\ell_{i-1} \cdots \ell_1 g k_0$ is in a right H-coset A_j for some j. But then also $\ell_i \ell_{i-1} \cdots \ell_1 g k_0 \in A_j$ as right H-cosets are closed under left multiplication by elements of H. So extend the path from g to $\ell_{i-1} \cdots \ell_1 g k_0$ by A_j to get a path from g to $\ell_i \cdots \ell_1 g k_0$.

The other case of $\ell_i \in L$ is similar. Again, $\ell_{i-1} \cdots \ell_1 g k_0$ is in $B_j = Lg_j K$ for some j and $g_j \in G$ as $\langle L \cup H \rangle gK$ decomposes into B_j's. Since B_j is

closed under multiplication by elements of L from left, $\ell_i \cdots \ell_1 g k_0 \in B_j$. Thus again, g is connected to $\ell_i \cdots \ell_1 g k_0$. The lemma follows. □

6.2.2 Facts on natural distributions

Lemma 28 *Let D be a natural probability distribution on a finite group G. For every subgroup $H \leq G$, if $\mathrm{supp}(D) \setminus H \neq \emptyset$ then $D(S) \leq 1/2$ for every left and right H-coset 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, $(D_{\ell-1} * [g_\ell])(S) = \frac{1}{2} D_{\ell-1}(S) + \frac{1}{2} D_{\ell-1}(S g_\ell^{-1}) \leq 1/2$. □

Lemma 29 *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 30 *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 29 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 $b_1 a^{-1}, \ldots, b_k a^{-1}$ span subgroups of increasing size. Suppose we already have $b_1, \ldots, b_k$ and $b_1 a^{-1}, \ldots, b_k a^{-1}$ span a proper subgroup B_k. Since $D(B_k a) \leq 1/2$ by Lemma 28, there exists an element $b_{k+1} \notin B_k a$ that has probability $\geq \frac{1}{2}|G|$. Since $b_{k+1} \notin B_k a$, we have $b_{k+1} a^{-1} \notin B_k$, hence $b_1 a^{-1}, \ldots, b_k a^{-1}, b_{k+1} a^{-1}$ 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 31 *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 $x \perp \mathbb{1}$. Hence, let $x \in \mathbb{R}^G$ be such that $x \perp \mathbb{1}$, $\|x\| = 1$ and $\|x * D\|$ be maximal possible. Since $\{x \in \mathbb{R}^G;\ x \perp \mathbb{1}\ \&\ \|x\| = 1\}$ 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

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

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 $x \perp \mathbb{1}$, there is another coordinate $x_{g_-} < 0$. Let a and K be the element and the set from Lemma 30. 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

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

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 (2).

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 $x^{\perp(A_i)}\perp\mathbb{1}$ so $\|x^{\perp(A_i)}*D\|\leq\lambda^{\mathrm{R}}(\tilde{D}_H)\|x^{\perp(A_i)}\|$. Now

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

and the lemma follows. □

7 Proof of Lemma 15

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 28, 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,

$$\begin{aligned} R_1 * R_2 &= (R_1^{\| H} + R_1^{\perp H}) * R_2 \\ &= R_1^{\| H} + R_1^{\perp H} * R_2 \end{aligned}$$

and, by Lemma 31,

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

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 17,

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

□

8 Proof of Key Convergence Lemma

The Key Convergence Lemma is an immediate corollary to the following lemma after observing that one can add convolutions with the trivial distribution fully concentrated on 1_G arbitrarily. Such a distribution acts as the identity for convolution.

Lemma 32 *Let $0 < e_1, \gamma < 1$ be reals and $t > 0$ be an integer. Let $D_0, R_0 \in \mathbb{R}^G$ be probability distributions with their natural decomposition $D_0 = R_0 + \epsilon_0$. For $j \in \{L, K\}, i \in \{1, \ldots, t\}$, let $D_i^j, R_i^j \in \mathbb{R}^G$ be probability distributions with their natural decomposition $D_i^j = R_i^j + \epsilon_i^j$ where $\|\epsilon_i^j\| \leq e_1$. Let*

$$\begin{aligned} \hat{D}_0 &= D_0 & \hat{R}_0 &= R_0 \\ \hat{D}_i &= (D_i^L *_\gamma \hat{D}_{i-1}) *_\gamma D_i^K & \hat{R}_i &= (R_i^L * \hat{R}_{i-1}) * R_i^K, \\ \epsilon_i &= \hat{D}_i - \hat{R}_i \end{aligned}$$

where $i = 1, \ldots, t$.

Let $L_i = \langle R_i^L \rangle$ and $K_i = \langle R_i^K \rangle$, for $i \in \{1, \ldots, t\}$. Let $L = \langle \bigcup_{i=1}^t L_i \rangle$ and $K = \langle \bigcup_{i=1}^t K_i \rangle$. For any $g \in G$,

$$\|\epsilon_t^{\perp(LgK)}\| \leq \max\left(h_{|LgK|}, (1 - d_{|LgK|}) \cdot \|\epsilon_0^{\perp(LgK)}\|\right)$$

and

$$\sum_{a \in LgK} \epsilon_t(a) = \sum_{a \in LgK} \epsilon_0(a),$$

where d_i and h_i satisfy:

$$\begin{aligned} h_0 &= e_1 + \gamma & d_0 &= \min(1/2, 1/c_G) \\ h_i &= \tfrac{1200 \cdot i^4}{d_{i-1}} \cdot h_{i-1} & d_i &= \frac{d_{i-1}}{600 \cdot i^3} \end{aligned}$$

In the lemma we allow each of the γ-approximate convolution to be different, although it has to satisfy the requirements on γ-approximate convolution.

Proof: We prove the lemma by induction on the size of LgH. W.l.o.g. we assume that for each odd $i \in \{1, \ldots, t\}$, $D_i^L(1_G) = R_i^L(1_G) = D_i^K(1_G) = R_i^K(1_G) = 1$ and for each even i, $D_i^L(1_G) = R_i^L(1_G) = 1$ or $D_i^K(1_G) = R_i^K(1_G) = 1$, since a distribution fully concentrated on 1_G acts as identity so it can be inserted into the chain of convolutions arbitrarily.

Base case: If $|LgK| = 1$ then $\epsilon_t^{\perp(LgK)} = 0$ and there is nothing to prove. Notice that $D_i^j(1_G) = R_i^j(1_G) = 1$ for all i and j.

Induction step: Let $m = |LgK| > 1$. We start by considering a special case. Define $K_\times = \left\langle \bigcup_{i=1}^{t-1} K_i \right\rangle$ and $L_\times = \left\langle \bigcup_{i=1}^{t-1} L_i \right\rangle$. We consider the case when $L_\times g K_\times \subsetneq LgK$. By our assumption either R_t^L or R_t^K is identity operation so there are two symmetric possibilities, either $LgK_\times \subsetneq LgK$ or $L_\times gK \subsetneq LgK$. We will analyze the former case, the analysis of the latter case is symmetric.

So consider, $LgK_\times \subsetneq LgK$. By Lemma 27, LgK can be partitioned into parts $(A_1, A_2, \ldots, A_{\ell_A})$, where each $A_i = Lg_iK_\times$ for some $g_i \in G$, and parts $(B_1, B_2, \ldots, B_{\ell_B})$, where each B_i is a left K_t-coset.

Denote

$$\begin{aligned}\widetilde{\epsilon_0} &= \epsilon_0^{\perp(LgK)},\\ \widetilde{\epsilon_\times} &= (\epsilon_{t-1})^{\perp(LgK)}, \text{ and}\\ \widetilde{\epsilon_t} &= \epsilon_t^{\perp(LgK)}.\end{aligned}$$

Since $D_t^L(1_G) = R_t^L(1_G) = 1$ by our assumption, $\hat{D}_t = \hat{D}_{t-1} * D_t^K$ and $\hat{R}_t = \hat{R}_{t-1} * R_t^K$.

Let $\beta = \|\widetilde{\epsilon_0}\|$ and $\alpha = \beta/(10 \cdot m^{3/2})$. Depending on the size of α we distinguish several cases.

Case 1) $\alpha > 60 \cdot m^{3/2}(m \cdot h_{m-1} + e_1 + \gamma)/d_{m-1}$. In this case we claim that

$$\|\widetilde{\epsilon_t}\| \leq (1 - 2 \cdot d_m) \cdot \|\widetilde{\epsilon_0}\|$$

Sub-case 1a) $\|\widetilde{\epsilon_0}^{\perp(A_1,A_2,\ldots,A_{\ell_A})}\| \geq \alpha$. Then there exists $i_0 \in \{1,\ldots,\ell_A\}$ such that $\|\widetilde{\epsilon_0}^{\perp(A_{i_0})}\| \geq \alpha/\sqrt{\ell_A} \geq 2h_{m-1}$. By induction hypothesis applied to $\hat{D}_{t-1}$, $\hat{R}_{t-1}$ and each $A_i = Lg_iK_\times$,

$$\begin{aligned}\|\widetilde{\epsilon_\times}^{\perp(A_{i_0})}\| &\leq (1 - d_{m-1})\|\widetilde{\epsilon_0}^{\perp(A_{i_0})}\|, \text{ and}\\ \|\widetilde{\epsilon_\times}^{\perp(A_i)}\| &\leq \|\widetilde{\epsilon_0}^{\perp(A_i)}\| + h_{m-1}\end{aligned}$$

for all $i \neq i_0$. Also

$$\begin{aligned}\widetilde{\epsilon_0} &= \widetilde{\epsilon_0}^{\|(A_1,A_2,\ldots,A_{\ell_A})} + \widetilde{\epsilon_0}^{\perp(A_1)} + \cdots + \widetilde{\epsilon_0}^{\perp(A_{\ell_A})}, \text{ and}\\ \widetilde{\epsilon_\times} &= \widetilde{\epsilon_\times}^{\|(A_1,A_2,\ldots,A_{\ell_A})} + \widetilde{\epsilon_\times}^{\perp(A_1)} + \cdots + \widetilde{\epsilon_\times}^{\perp(A_{\ell_A})},\end{aligned}$$

where the vectors on each right hand side are all pair-wise orthogonal, and $\widetilde{\epsilon_0}^{\|(A_1,A_2,\ldots,A_{\ell_A})} = \widetilde{\epsilon_\times}^{\|(A_1,A_2,\ldots,A_{\ell_A})}$ by the induction hypothesis. By Lemma 17,

$$\|\widetilde{\epsilon_0}\| - \|\widetilde{\epsilon_0} - \widetilde{\epsilon_0}^{\perp(A_{i_0})} + \widetilde{\epsilon_\times}^{\perp(A_{i_0})}\| \geq \frac{d_{m-1}}{200 \cdot m^3} \cdot \|\widetilde{\epsilon_0}\|.$$

Furthermore, by the repeated use of Lemma 18,

$$\|\widetilde{\epsilon_0}\| - \|\widetilde{\epsilon_\times}\| \geq \frac{d_{m-1}}{200 \cdot m^3} \cdot \|\widetilde{\epsilon_0}\| - (\ell_A - 1)h_{m-1}.$$

From the properties of $*_\gamma$, for some $\epsilon_\gamma \in \mathbb{R}^G$, where $\|\epsilon_\gamma\| \leq \gamma$ and $\epsilon_\gamma^{\|K_t} = 0$, it holds

$$\epsilon_t = \epsilon_{t-1} * D_t^K + \hat{R}_{t-1} * \epsilon_t^K + \epsilon_\gamma.$$

By Proposition 20 and Lemma 21,

$$\widetilde{\epsilon_t} = \widetilde{\epsilon_\times} * D_t^K + (\hat{R}_{t-1} * \epsilon_t^K)^{\perp(LgK)} + \epsilon_\gamma^{\perp(LgK)}.$$

Since $\|\widetilde{\epsilon_\times} * D_t^K\| \le \|\widetilde{\epsilon_\times}\|$ and $\|\hat{R}_{t-1} * \epsilon_t^K\| \le \|\epsilon_t^K\|$,

$$\begin{aligned}\|\widetilde{\epsilon_t}\| &\le \|\widetilde{\epsilon_0}\| - \frac{d_{m-1}}{200 \cdot m^3} \cdot \|\widetilde{\epsilon_0}\| + (\ell_A - 1) \cdot h_{m-1} + \|\hat{R}_{t-1} * \epsilon_t^K\| + \|\epsilon_\gamma\| \\ &\le \|\widetilde{\epsilon_0}\| - \frac{d_{m-1}}{200 \cdot m^3} \cdot \|\widetilde{\epsilon_0}\| + m \cdot h_{m-1} + e_1 + \gamma.\end{aligned}$$

Since $\beta \ge 600 \cdot m^3(m \cdot h_{m-1} + e_1 + \gamma)/d_{m-1}$,

$$\|\widetilde{\epsilon_t}\| \le (1 - \frac{2}{3} \cdot \frac{d_{m-1}}{200 \cdot m^3}) \cdot \|\widetilde{\epsilon_0}\|$$

Sub-case 1b) $\|\widetilde{\epsilon_0}^{\perp(A_1,A_2,\ldots,A_{\ell_A})}\| < \alpha$. Then $\|\widetilde{\epsilon_0}^{\|(A_1,A_2,\ldots,A_{\ell_A})}\| > \sqrt{\beta^2 - \alpha^2} > \beta - \alpha$, since $0 < \alpha < \beta$. By induction hypothesis, $\widetilde{\epsilon_0}^{\|(A_1,A_2,\ldots,A_{\ell_A})} = \widetilde{\epsilon_\times}^{\|(A_1,A_2,\ldots,A_{\ell_A})}$ and $\|\widetilde{\epsilon_\times}^{\perp(A_1,A_2,\ldots,A_{\ell_A})}\| \le \alpha + \ell_A h_{m-1} \le (3/2)\alpha$. Thus, $\|\widetilde{\epsilon_0}\|/2 \le \beta - \alpha \le \|\widetilde{\epsilon_\times}\| \le \beta + mh_{m-1}$. In particular, $\|\widetilde{\epsilon_\times}\| \le \beta + (1/2)\alpha$ and $\|\widetilde{\epsilon_0}\|/2 \le \|\widetilde{\epsilon_\times}\|$. By Lemma 19,

$$\begin{aligned}\|\widetilde{\epsilon_\times}^{\perp(B_1,B_2,\ldots,B_{\ell_B})}\| &\ge \frac{\beta - \alpha}{2 \cdot m^{3/2}} - (3/2)\alpha \\ &\ge 3\alpha.\end{aligned}$$

By Proposition 23,

$$\|\widetilde{\epsilon_\times}^{\perp(B_1,B_2,\ldots,B_{\ell_B})} * R_t^K\| \le \lambda(R_t^K) \cdot \|\widetilde{\epsilon_\times}^{\perp(B_1,B_2,\ldots,B_{\ell_B})}\|.$$

Hence by Lemma 17,

$$\begin{aligned}\|\widetilde{\epsilon_\times} * R_t^K\| &\le \left(1 - \frac{1}{2} \cdot \left(\frac{3\alpha}{\beta + (1/2)\alpha}\right)^2 \cdot (1 - \lambda(R_t^K))\right) \cdot \|\widetilde{\epsilon_\times}\| \\ &\le \left(1 - \frac{1}{32 \cdot m^3} \cdot \frac{1}{c_G}\right) \cdot \|\widetilde{\epsilon_\times}\| \\ &\le \|\widetilde{\epsilon_\times}\| - \frac{1}{64 \cdot m^3} \cdot \frac{1}{c_G} \cdot \|\widetilde{\epsilon_0}\| \\ &\le \|\widetilde{\epsilon_0}\| + mh_{m-1} - \frac{1}{64 \cdot m^3} \cdot \frac{1}{c_G} \cdot \|\widetilde{\epsilon_0}\|.\end{aligned}$$

Since,

$$\epsilon_t = \epsilon_{t-1} * R_t^K + \hat{D}_{t-1} * \epsilon_t^K + \epsilon_\gamma.$$

By Proposition 20 and Lemma 21,

$$\widetilde{\epsilon_t} = \widetilde{\epsilon_\times} * R_t^K + (\hat{D}_{t-1} * \epsilon_t^K)^{\perp(LgK)} + \epsilon_\gamma^{\perp(LgK)}.$$

Since $\beta \ge 3 \cdot 64 \cdot m^3(m \cdot h_{m-1} + e_1 + \gamma)/d_{m-1}$,

$$\begin{aligned}\|\widetilde{\epsilon_t}\| &\le \|\widetilde{\epsilon_0}\| + m \cdot h_{m-1} - \frac{1}{64 \cdot m^3} \cdot \frac{1}{c_G} \cdot \|\widetilde{\epsilon_0}\| + e_1 + \gamma \\ &\le (1 - \frac{2}{3} \cdot \frac{1}{64 \cdot m^3} \cdot \frac{1}{c_G})\|\widetilde{\epsilon_0}\|.\end{aligned}$$

Case 2) $\alpha \leq 60 \cdot m^{3/2}(m \cdot h_{m-1} + e_1 + \gamma)/d_{m-1}$. In this case we claim that

$$\|\widetilde{\epsilon_t}\| \quad \leq \quad \|\widetilde{\epsilon_0}\| + mh_{m-1} + e_1 + \gamma.$$

Again, by induction hypothesis applied to $\hat{D}_{t-1}$, $\hat{R}_{t-1}$ and each $A_i = Lg_iK_\times$

$$\|\widetilde{\epsilon_\times}\| \leq \|\widetilde{\epsilon_0}\| + (\ell_A - 1)h_{m-1}.$$

Furthermore,

$$\|\widetilde{\epsilon_t}\| \leq \|\widetilde{\epsilon_\times}\| + e_1 + \gamma.$$

Thus, this case follows.

Cases 1 and 2 put together give that

$$\begin{array}{lcl} \|\widetilde{\epsilon_0}\| > 600 \cdot m^3 \cdot d_{m-1}^{-1} \cdot (m \cdot h_{m-1} + e_1 + \gamma) & \Rightarrow & \|\widetilde{\epsilon_t}\| \leq (1 - 2 \cdot d_m) \cdot \|\widetilde{\epsilon_0}\| \\ \|\widetilde{\epsilon_0}\| \leq 600 \cdot m^3 \cdot d_{m-1}^{-1} \cdot (m \cdot h_{m-1} + e_1 + \gamma) & \Rightarrow & \|\widetilde{\epsilon_t}\| \leq \|\widetilde{\epsilon_0}\| + mh_{m-1} + e_1 + \gamma. \end{array}$$

This is true when $L_\times gK_\times \subsetneq LgK$. In general case we partition the chain of convolutions into blocks where each block except possibly for the last one satisfy:

1. $\langle \bigcup L_i \rangle\, g\, \langle \bigcup K_i \rangle \subsetneq LgK$, where the union is taken over all the groups but the last one involved in the block, and
2. $\langle \bigcup L_i \rangle\, g\, \langle \bigcup K_i \rangle = LgK$, where the union is taken over all the groups involved in the block.

The last block of convolutions may not satisfy the second property in which case we know by the induction hypothesis that the convolutions of the last block increase ℓ_2-norm of the error at most by an additive term $m \cdot h_{m-1}$.

Thus we see that if $\|\widetilde{\epsilon_0}\|$ is large enough, the ℓ_2-norm of the error will shrink by a factor of at least $(1 - 2 \cdot d_m)$ in each block. Once the ℓ_2-norm of the error is below the threshold $600 \cdot m^3 \cdot d_{m-1}^{-1} \cdot (m \cdot h_{m-1} + e_1 + \gamma)$ it may increase in a given block by at most $mh_{m-1} + e_1 + \gamma$ but to no more than $600 \cdot m^3 \cdot d_{m-1}^{-1} \cdot (m \cdot h_{m-1} + e_1 + \gamma) + mh_{m-1} + e_1 + \gamma$. The last incomplete block may increase the error by the additional term $m \cdot h_{m-1}$.

Overall we see that in the general case

$$\begin{array}{lll} \|\widetilde{\epsilon_t}\| \leq \max & (& (1 - 2 \cdot d_m) \cdot \|\widetilde{\epsilon_0}\| + m \cdot h_{m-1}, \\ & & 600 \cdot m^3 \cdot d_{m-1}^{-1} \cdot (m \cdot h_{m-1} + e_1 + \gamma) + 2m \cdot h_{m-1} + e_1 + \gamma) \end{array}$$

The choice of h_m shows that

$$\begin{array}{rcl} h_m & = & \dfrac{1200 \cdot m^4}{d_{m-1}} \cdot h_{m-1} \\ & \geq & 600 \cdot m^3 \cdot d_{m-1}^{-1} \cdot (m \cdot h_{m-1} + e_1 + \gamma) + 2m \cdot h_{m-1} + e_1 + \gamma. \end{array}$$

Furthermore, if $h_m - m \cdot h_{m-1} \leq \|\widetilde{\epsilon_0}\|$ then $d_m \cdot \|\widetilde{\epsilon_0}\| > m \cdot h_{m-1}$. Hence,

$$\|\widetilde{\epsilon_t}\| \leq \max((1 - d_m) \cdot \|\widetilde{\epsilon_0}\|, h_m).$$

The preservation of $\sum_{a \in LgK} \epsilon_t(a)$ follows from the fact that each convolution acts on cosets that always partition LgK and from Lemma 22. □

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. $\mathrm{Bp}_{\mathrm{h}}\mathrm{space}(s) \subseteq \mathrm{dspace}(s^{3/2})$. *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.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de