Pseudorandomness via the discrete Fourier transform

Parikshit Gopalan
Microsoft Research

Daniel M. Kane
University of California, San Diego

Raghu Meka
University of California, Los Angeles

May 30, 2015

Abstract

We present a new approach to constructing unconditional pseudorandom generators against classes of functions that involve computing a linear function of the inputs. We give an explicit construction of a pseudorandom generator that fools the *discrete Fourier transforms* of linear functions with seed-length that is nearly logarithmic (up to polyloglog factors) in the input size and the desired error parameter. Our result gives a single pseudorandom generator that fools several important classes of tests computable in logspace that have been considered in the literature, including halfspaces (over general domains), modular tests and combinatorial shapes. For all these classes, our generator is the first that achieves near logarithmic seed-length in both the input length and the error parameter. Getting such a seed-length is a natural challenge in its own right, which needs to be overcome in order to derandomize RL — a central question in complexity theory.

Our construction combines ideas from a large body of prior work, ranging from a classical construction of [NN93] to the recent gradually increasing independence paradigm of [KMN11, CRSW13, GMR+12], while also introducing some novel analytic machinery which might find other applications.

1 Introduction

A central goal of computational complexity is to understand the power that randomness adds to efficient computation. The main questions in this area are whether $\mathsf{BPP} = \mathsf{P}$ and $\mathsf{RL} = \mathsf{L}$, which respectively assert that randomness can be eliminated from efficient computation, at the price of a polynomial slowdown in time, and a constant blowup in space. It is known that proving $\mathsf{BPP} = \mathsf{P}$ will imply strong circuit lower bounds that seem out of reach of current techniques. In contrast, proving $\mathsf{RL} = \mathsf{L}$, could well be within reach. Indeed, *bounded-space algorithms* are a natural computational model for which we know how to construct strong *pseudo-random generators*, PRGs, unconditionally.

Let RL denote the class of randomized algorithms with $O(\log n)$ work space which can access the random bits in a read-once pre-specified order. Nisan [Nis92] devised a PRG of seed length $O(\log^2(n/\varepsilon))$ that fools RL with error ε. This generator was subsequently used by Nisan [Nis94] to show that $\mathsf{RL} \subseteq \mathsf{SC}$ and by Saks and Zhou [SZ99] to prove that RL can be simulated in space $O(\log^{3/2} n)$. Constructing PRGs with the optimal $O(\log(n/\varepsilon))$ seed length for this class and showing that $\mathsf{RL} = \mathsf{L}$ is arguably the outstanding open problem in derandomization (which might not require a breakthrough in lower bounds). Despite much progress in this area [INW94, NZ96, RR99, Rei08, RTV06, BRRY14, BV10, KNP11, De11, GMR+12], there are few cases where we can improve on Nisan's twenty year old bound of $O(\log^2(n/\varepsilon))$ [Nis92].

1.1 Fourier shapes

A conceptual contribution of this work is to propose a class of functions in RL which we call *Fourier shapes* that unify and generalize the problem of fooling many natural classes of test functions that are computable in logspace and involve computing linear combinations of (functions of) the input variables. In the following, let $\mathbb{C}_1 = \{z : |z| \leq 1\}$ be the unit-disk in the complex plane.

Definition 1. *A (m, n)-*Fourier shape *$f : [m]^n \to \mathbb{C}_1$ is a function of the form $f(x_1, \ldots, x_n) = \prod_{j=1}^{n} f_j(x_j)$ where each $f_j : [m] \to \mathbb{C}_1$. We refer to m and n as the alphabet size and the dimension of the Fourier shape respectively.*

Clearly, (m, n)-Fourier shapes can be computed with $O(\log n)$ workspace, as long as the bit-complexity of $\log(f_j)$ is logarithmic for each j; a condition that can be enforced without loss of generality. Since our goal is to fool functions $f : \{0,1\}^n \to \{0,1\}$, it might be unclear why we should consider complex-valued functions (or larger domains). The answer comes from the discrete Fourier transform which maps integer random variables to $\mathbb{C}_1$. Concretely consider a Boolean function $f : \{0,1\}^n \to \{0,1\}$ of the form $f(x) = g(\sum_j w_j x_j)$ where $x \in \{0,1\}^n$, $w_j \in \mathbb{Z}$, and $g : \mathbb{Z} \to \{0,1\}$ is a *simple* function like a threshold or a mod function. To fool such a function f, it suffices to *fool* the linear function $w(x) = \sum_j w_j x_j$. A natural way to establish the closeness of distributions on the integers is via the discrete

Fourier transform. The discrete Fourier transform of $w(x)$ at $\alpha \in [0,1]$ is given by

$$\phi_\alpha(w(x)) = \exp(2\pi i \alpha \cdot w(x)) = \prod_{j=1}^{n} \exp(2\pi i \alpha w_j x_j)$$

which is a Fourier shape.

Allowing a non-binary alphabet m not only allows us to capture more general classes of functions (such as combinatorial shapes), it makes the class more robust. For instance, given a Fourier shape $f : \{0,1\}^n \to \mathbb{C}$, if we consider inputs bits in blocks of length $\log(m)$, then the resulting function is still a Fourier shape over a larger input domain $[m]$ (in dimension $n/\log(m)$). This allows certain compositions of PRGs and simplifies our construction even for the case $m = 2$.

1.1.1 PRGs for Fourier shapes and their applications.

A PRG is a function $\mathcal{G} : \{0,1\}^r \to [m]^n$. We refer to r as the seed-length of the generator. We say $\mathcal{G}$ is *explicit* if the output of $\mathcal{G}$ can be computed in time $\mathrm{poly}(n)$.[1]

Definition 2. *A* PRG $\mathcal{G} : \{0,1\}^r \to [m]^n$ *fools a class of functions* $\mathcal{F} = \{f : [m]^n \to \mathbb{C}\}$ *with error* ε *(or* ε*-fools* $\mathcal{F}$*) if for every* $f \in \mathcal{F}$*,*

$$\left| \mathop{\mathbb{E}}_{x \in_u [m]^n} [f(x)] - \mathop{\mathbb{E}}_{y \in_u \{0,1\}^r} [f(\mathcal{G}(y))] \right| < \varepsilon.$$

We motivate the problem of constructing PRGs for Fourier shapes by discussing how they capture a variety of well-studied classes like halfspaces (over general domains), combinatorial rectangles, modular tests and combinatorial shapes.

PRGs for halfspaces. Halfspaces are functions $h : \{0,1\}^n \to \{0,1\}$ that can be represented as

$$h(x) = \mathbb{1}^+(\langle w, x \rangle - \theta)$$

for some *weight* vector $w \in \mathbb{Z}^n$ and *threshold* $\theta \in \mathbb{Z}$ where $\mathbb{1}^+(a) = 1$ if $a \geq 0$ and 0 otherwise. Halfspaces are of central importance in computational complexity, learning theory and social choice. Lower bounds for halfspaces are trivial, whereas the problem of proving lower bounds against depth-2 TC_0 or halfspaces of halfspaces is a frontier open problem in computational complexity. The problem of constructing explicit PRGs that can fool halfspaces is a natural challenge that has seen a lot of exciting progress recently [DGJ+09, MZ13, Kan11b, Kan14, KM15]. The best known PRG construction for halfspaces is that of Meka and Zuckerman [MZ13] who gave a PRG with seed-length $O(\log n + \log^2(1/\varepsilon))$, which is $O(\log^2(n))$ for polynomially small error. They also showed

[1] Throughout, for a multi-set S, $x \in_u S$ denotes a uniformly random element of S.

that PRGs against RL with inverse polynomial error can be used to fool halfspaces, and thus constructing better PRGs for halfspaces is a necessary step towards progress for bounded-space algorithms. However, even for special cases of halfspaces (such as derandomizing the Chernoff bound), beating seed-length $O(\log^2(n))$ has proved difficult.

We show that a PRG for $(2, n)$-Fourier shapes with error ε/n^2 also fools halfspaces with error ε. In particular, PRGs fooling Fourier shapes with polynomially small error also fool halfspaces with small error.

PRGs for generalized halfspaces. PRGs for (m, n)-Fourier shapes give us PRGs for halfspaces not just for the uniform distribution over the hypercube, but for a large class of distributions that have been studied in the literature. We can derive these results in a unified manner by considering the class of *generalized halfspaces*.

Definition 3. *A generalized halfspace over $[m]^n$ is a function $g : [m]^n \to \{0, 1\}$ that can be represented as*

$$g(x) = \mathbb{1}^+\left(\sum_{j=1}^{n} g_j(x_j) - \theta\right).$$

where $g_j : [m] \to \mathbb{R}$ are arbitrary functions for $j \in [n]$ and $\theta \in \mathbb{R}$.

PRGs for (m, n)-Fourier shapes imply PRGs for generalized halfspaces. This in turn captures settings of fooling halfspaces with respect to the Gaussian distribution and the uniform distribution on the sphere [KRS12, MZ13, Kan14, KM15], and a large class of product distributions over $\mathbb{R}^n$ [GOWZ10].

Derandomizing the Chernoff-Hoeffding bound. A consequence of fooling generalized halfspaces is to *derandomize* Chernoff-Hoeffding type bounds for sums of independent random variables which are ubiquitous in the analysis of randomized algorithms. We state our result in the language of "randomness-efficient samplers" (cf. [Zuc97]). Let $X_1, \ldots, X_n$ be independent random variables over a domain $[m]$ and let $g_1, \ldots, g_n : [m] \to [-1, 1]$ be arbitrary bounded functions. The classical Chernoff-Hoeffding bounds [Hoe63] say that

$$\Pr\left[\left|\sum_{i=1}^{n} g_i(X_i) - \sum_{i=1}^{n} \mathbb{E}[g_i(X_i)]\right| \geq t\right] \leq 2\exp(-t^2/4n).$$

There has been a long line of work on showing sharp tail bounds for pseudorandom sequences starting from [SSS95] who showed that similar tail bounds hold under limited independence. But all previous constructions for the polynomial small error regime required seed-length $O(\log^2(n))$. PRGs for generalized halfspaces give Chernoff-Hoeffding tail bounds with polynomially small error, with seed-length $\tilde{O}(\log(n))$.

PRGs for modular tests. An important class of functions in L is that of *modular tests*, i.e., functions of the form $g : \{0,1\}^n \to \{0,1\}$, where $g(x) = \mathbb{1}(\sum_i a_i x_i \bmod m \in S)$, for $m \leq M$, coefficients $a_i \in \mathbb{Z}_m$ and $S \subseteq \mathbb{Z}_m$. Such a test is computable in L as long as $M \leq \mathrm{poly}(n)$. The case when $m = 2$ corresponds to small-bias spaces, for which optimal constructions were first given in the seminal work of Naor and Naor [NN93]. The case of arbitrary m was considered by [LRTV09] (see also [MZ09]), their generator gives seed-length $\tilde{O}(\log(n/\varepsilon) + \log^2(M))$. Thus for $M = \mathrm{poly}(n)$, their generator does not improve on Nisan's generator even for constant error ε. PRGs fooling $(2, n)$-Fourier shapes with polynomially small error fools modular tests.

PRGs for combinatorial shapes. *Combinatorial shapes* were introduced in the work of [GMRZ13] as a generalization of combinatorial rectangles and to address fooling linear sums in statistical distance. These are functions $f : [m]^n \to \{0,1\}$ of the form

$$f(x) = h\left(\sum_{i=1}^{n} g_i(x_i)\right)$$

for functions $g_i : [m] \to \{0,1\}$ and a function $h : \{0,\ldots,n\} \to \{0,1\}$. The best previous generators of [GMRZ13] and [De14] for combinatorial shapes achieve a seed-length of $O(\log(mn)+\log^2(1/\varepsilon))$, $O(\log m+\log(n/\varepsilon)^{3/2})$; in particular, the best previous seed-length for polynomially small error was $O(\log^{3/2}(n))$. PRGs for (m, n)-Fourier shapes with error ε/n imply PRGs for combinatorial shapes.

Combinatorial rectangles are a well-studied subset of combinatorial shapes [EGL+98, ASWZ96, LLSZ97, Lu02]. They are functions that can be written as $f(x) = \prod_j \mathbb{1}(x_j \in A_j)$ for some arbitrary subsets $A_j \subseteq [m]$. The best known PRG due to [GMR+12, GY14] gives a seed-length of $O(\log(mn/\varepsilon)\log\log(mn/\varepsilon))$. Combinatorial rectangles are special cases of Fourier shapes so our PRG for (m, n)-Fourier shapes also fools combinatorial rectangles, but requires a slightly longer seed. The *alphabet-reduction* step in our construction is inspired by the generator of [GMR+12, GY14].

1.1.2 Achieving optimal error dependence via Fourier shapes.

We note that having generators for Fourier shapes with seed-length $\tilde{O}(\log(n))$ even when ε is polynomially small is essential in our reductions: we sometimes need error $\varepsilon/\mathrm{poly}(n)$ for Fourier shapes in order to get ε error for our target class of functions. Once we have this, starting with ε a sufficiently small polynomial results in polynomially small error for the target class of functions.

We briefly explain why previous techniques based on *limit theorems* were unable to achieve polynomially small error with optimal seed-length, by considering the setting of halfspaces under the uniform distribution on $\{0,1\}^n$. Fooling halfspaces is equivalent to fooling all linear functions $L(x) = \sum_i w_i x_i$ in Kolmogorov or cdf distance. Previous work

on fooling halfspaces [DGJ+09, MZ13] relies on the Berry-Esséen theorem, a quantiative form of the central limit theorem, to show that the cdf of *regular* linear functions is close to that of the Gaussian distribution, both under the uniform distribution and under the pseudorandom distribution. However, even for the majority function (which is the most regular linear function), the discreteness of $\sum_i x_i$ means that the Kolmogorov distance from the Gaussian distribution is $1/\sqrt{n}$, even when x is uniformly random. Approaches that show closeness in cdf distance by comparison to the Gaussian distribution seem unlikely to give polynomially small error with optimal seed-length.

We depart from the *derandomized limit theorem* approach taken by several previous works [DGJ+09, DKN10, GOWZ10, HKM12, GMRZ13, MZ13] and work directly with the Fourier transform. A crucial insight (that is formalized in Lemma 9.2) is that fooling the Fourier transform of linear forms to within polynomially small error implies polynomially small Kolmogorov distance.

1.2 Our results

Our main result is the following:

Theorem 1.1. *There is an explicit generator $\mathcal{G} : \{0,1\}^r \to [m]^n$ that fools all (m,n)-Fourier shapes with error ε, and has seed-length $r = O(\log(mn/\varepsilon) \cdot (\log\log(mn/\varepsilon))^2)$.*

We now state various corollaries of our main result starting with fooling halfspaces.

Corollary 1.2. *There is an explicit generator $\mathcal{G} : \{0,1\}^r \to \{0,1\}^n$ that fools halfspaces over $\{0,1\}^n$ under the uniform distribution with error ε, and has seed-length $r = O(\log(n/\varepsilon)(\log\log(n/\varepsilon))^2)$.*

The best previous generator due to [MZ13] had a seed-length of $O(\log n + \log^2(1/\varepsilon))$, which is $O(\log^2 n)$ for polynomially small error ε.

We also get a PRG with similar parameters for generalized halfspaces.

Corollary 1.3. *There is an explicit generator $\mathcal{G} : \{0,1\}^r \to [m]^n$ that ε-fools* generalized halfspaces *over $[m]^n$, and has seed-length $r = O(\log(mn/\varepsilon) \cdot (\log\log(mn/\varepsilon))^2)$.*

From this we can derive PRGs with seed-length $O(\log(n/\varepsilon)(\log\log(n/\varepsilon))^2)$ for fooling halfspaces with error ε under the Gaussian distribution and the uniform distribution on the sphere. Indeed, we get the following bound for arbitrary product distributions over $\mathbb{R}^n$, which depends on the 4^{th} moment of each co-ordinate.

Corollary 1.4. *Let X be a product distribution on $\mathbb{R}^n$ such that for all $i \in [n]$,*

$$\mathbb{E}[X_i] = 0, \mathbb{E}[X_i^2] = 1, \mathbb{E}[X_i^4] \leq C.$$

There exists an explicit generator $\mathcal{G} : \{0,1\}^r \to \mathbb{R}^n$ such that if $Y = \mathcal{G}(z)$, then for every halfspace $h : \mathbb{R}^n \to \{0,1\}$,

$$|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]| \leq \varepsilon.$$

The generator G has seed-length $r = O(\log(nC/\varepsilon)(\log\log(nC/\varepsilon))^2)$.

This improves on the result of [GOWZ10] who obtained seedlength $O(\log(nC/\varepsilon)\log(C/\varepsilon)$ for this setting via a suitable modification of the generator from [MZ13].

The next corollary is a near-optimal derandomization of the Chernoff-Hoeffding bounds. To get a similar guarantee, the best known seed-length that follows from previous work [SSS95, MZ13, GOWZ10] was $O(\log(mn) + \log^2(1/\varepsilon))$.

Corollary 1.5. *Let $X_1, \ldots, X_n$ be independent random variables over the domain $[m]$. Let $g_1, \ldots, g_n : [m] \to [-1, 1]$ be arbitrary bounded functions. There exists an explicit generator $G : \{0,1\}^r \to [m]^n$ such that if $(Y_1, \ldots, Y_n) = G(z)$ where $z \in_u \{0,1\}^r$, then Y_i is distributed identically to X_i and*

$$\Pr\left[\left|\sum_{i=1}^{n} g_i(Y_i) - \sum_{i=1}^{n} \mathbb{E}[g_i(Y_i)]\right| \geq t\right] \leq 2\exp(-t^2/2n) + \varepsilon.$$

G has seed-length $r = O(\log(mn/\varepsilon)(\log\log(mn/\varepsilon))^2)$.

We get the first generator for fooling modular tests whose dependence on the modulus M is near-logarithmic. The best previous generator from [LRTV09] had a seed-length of $\tilde{O}(\log(n/\varepsilon) + \log^2(M))$, which is $\tilde{O}(\log^2 n)$ for $M = \mathrm{poly}(n)$.

Corollary 1.6. *There is an explicit generator $\mathcal{G} : \{0,1\}^r \to \{0,1\}^n$ that fools all linear tests modulo m for all $m \leq M$ with error ε, and has seed-length $r = O(\log(Mn/\varepsilon) \cdot (\log\log(Mn/\varepsilon))^2)$.*

Finally, we get a generator with near-logarithmic seedlength for fooling combinatorial shapes. [GMRZ13] gave a PRG for combinatorial shapes with a seed-length of $O(\log(mn) + \log^2(1/\varepsilon))$. This was improved recently by De [De14] who gave a PRG with seed-length $O(\log m + \log(n/\varepsilon)^{3/2})$; in particular, the best previous seed-length for polynomially small error was $O((\log(n)^{3/2})$.

Corollary 1.7. *There is an explicit generator $\mathcal{G} : \{0,1\}^r \to [m]^n$ that fools (m,n)-combinatorial shapes to error ε and has seed-length $r = O(\log(mn/\varepsilon)(\log\log(mn/\varepsilon))^2)$.*

1.3 Other related work

Starting with the work of Diakonikolas et al. [DGJ+09], there has been a lot of interest in constructing PRGs for halfspaces and related classes such as intersections of halfspaces and polynomial threshold functions over the domain $\{\pm 1\}^n$ [DKN10, GOWZ10, HKM12, MZ13, Kan11b, Kan11a, Kan14]. Rabani and Shpilka [RS10] construct optimal hitting set generators for halfspaces over $\{\pm 1\}^n$; hitting set generators are weaker than PRGs.

Another line of work gives PRGs for halfspaces for the uniform distribution over the sphere (*spherical caps*) or the Gaussian distribution. For spherical caps, Karnin, Rabani

and Shpilka [KRS12] gave a PRG with a seed-length of $O(\log n + \log^2(1/\varepsilon))$. For the Gaussian distribution, [Kan14] gave a PRG which achieves a seed-length of $O(\log n + \log^{3/2}(1/\varepsilon))$. Recently, [KM15] gave the first PRGs for these settings with seedlength $O((\log(n/\varepsilon))(\log\log(n/\varepsilon)))$. Fooling halfspaces over the hypercube is known to be *harder* than the Gaussian setting or the uniform distribution on the sphere; hence our result gives a construction with similar parameters up to a $O(\log\log n)$ factor. At a high level, [KM15] also uses a iterative dimension reduction approach like in [KMN11, CRSW13, GMR+12]; however, the final construction and its analysis are significantly different from ours.

Gopalan *et al.* [GOWZ10] gave a generator fooling halfspaces under product distributions with bounded fourth moments, whose seed-length is $O(\log(n/\varepsilon)\log(1/\varepsilon))$.

The present work completely subsumes a manuscript of the authors which essentially solved the special-case of derandomizing Chernoff bounds and a special class of halfspaces [GKM14].

2 Proof overview

We describe our PRG for Fourier shapes as in Theorem 1.1. The various corollaries are derived from this Theorem using properties of the discrete Fourier transform of integer-valued random variables.

Let us first consider a very simple PRG: $O(1)$-wise independent distributions over $[m]^n$. At a glance, it appears to do very poorly as it is easy to express the parity of a subset of bits as a Fourier shape and parities are not fooled even by $(n-1)$-wise independence. The starting point for our construction is that bounded independence does fool a special but important class of Fourier shapes, namely those with polynomially small *total variance*.

For a complex valued random variable Z, define the variance of Z as

$$\sigma^2(Z) = \mathbb{E}\left[|Z - \mathbb{E}[Z]|^2\right] = \mathbb{E}[|Z|^2] - |\,\mathbb{E}[Z]|^2.$$

It is easy to verify that

$$\sigma^2(Z) + |\,\mathbb{E}[Z]|^2 = E[|Z|^2],$$

so that if Z takes values in $\mathbb{C}_1$, then

$$\sigma^2(Z) + |\,\mathbb{E}[Z]|^2 \leq 1.$$

The *total-variance* of a (m,n)-Fourier shape $f : [m]^n \to \mathbb{C}_1$ with $f(x) = \prod_{j=1}^n f_j(x_j)$ is defined as

$$\mathsf{Tvar}(f) = \sum_j \sigma^2(f_j(x_j)).$$

To gain some intuition for why this is a natural quantity, note that $\mathsf{Tvar}(f)$ gives an easy upper bound on the expectation of a Fourier shape:

$$\left|\mathop{\mathbb{E}}_{x\in[m]^n}[f(x)]\right| = \prod_j |\mathbb{E}[f_j(x_j)]| \leq \prod_j \sqrt{1-\sigma^2(f_j(x_j))} \leq \exp(-\mathsf{Tvar}(f)/2). \tag{1}$$

This inequality suggests a natural dichotomy for the task of fooling Fourier shapes. It suggests that high variance shapes where $\mathsf{Tvar}(f) \gg \log(1/\varepsilon)$ are *easy* in the sense that $\mathbb{E}[f] \ll \varepsilon$ is small for such Fourier shapes. So a PRG for such shapes only needs to ensure that $\mathbb{E}[f]$ is also sufficiently small under the pseudorandom output.

To complement the above, we show that if the total-variance $\mathsf{Tvar}(f)$ is very small, then generators based on limited independence do fairly well. Concretely, our main technical lemma says that limited independence fools products of bounded (complex-valued) random variables, provided that the sum of their variances is small.

Lemma 2.1. *Let $Y_1, \ldots, Y_n$ be k-wise independent random variables taking values in $\mathbb{C}_1$. Then,*

$$\left| \mathbb{E}[Y_1 \cdots Y_n] - \prod_{j=1}^{n} \mathbb{E}[Y_j] \right| \leq \exp(O(k)) \left(\frac{\sum_{j=1}^{n} \sigma^2(Y_j)}{\sqrt{k}} \right)^{\Omega(k)}.$$

We defer discussion of the proof to Section 2.4, and continue the description of our PRG construction. Recall that we are trying to fool a (m,n)-Fourier shape $f : [m]^n \to \mathbb{C}_1$ with $\mathsf{Tvar}(f) \leq O(\log(1/\varepsilon)$ to error $\varepsilon = \mathrm{poly}(1/nm)$. It is helpful to think of the desired error ε as being fixed at the beginning and staying unchanged through our iterations, while m and n change during the iterations. Generating k-wise independent distributions over $[m]^n$ takes $O(k \log(mn))$ random bits. Thus if we use $k = O(\log(1/\varepsilon))$-wise independence, we would achieve error ε, but with seed-length $O(\log(1/\varepsilon) \log(mn))$ rather than $O(\log(1/\varepsilon))$.

On the other hand, if $\mathsf{Tvar}(f) \leq 1/(mn)^c$ for a fixed constant c, then choosing $k = O(\log(1/\varepsilon)/(\log mn))$-wise independence is enough to get error ε while also achieving seed-length $O(k \log(mn)) = O(\log(1/\varepsilon))$ as desired. We exploit this observation by combining the use of limited independence with the recent *iterative-dimension-reduction* paradigm of [KMN11, CRSW13, GMR+12]. Our construction reduces the problem of fooling Fourier shapes with $\mathsf{Tvar}(f) \leq O(\log(1/\varepsilon))$ through a sequence of iterations to fooling Fourier shapes where the total variance is polynomially small in m, n in each iteration and then uses limited independence in each iteration.

To conclude our high-level description, our generator consists of three modular parts. The first is a generator for Fourier shapes with high total variance: $\mathsf{Tvar}(f) \geq \mathrm{poly}(\log(1/\varepsilon))$. We then give two reductions to handle low variance Fourier shapes: an alphabet-reduction step reduces the alphabet m down to $\sqrt{m}$ and leaves n unchanged, and a dimension-reduction step that reduces the dimension from n to $\sqrt{n}$ while possibly blowing up the alphabet to $\mathrm{poly}(1/\varepsilon)$. We describe each of these parts in more detail below.

2.1 Fooling high-variance Fourier shapes

We construct a PRG with seed-length $O(\log(mn/\varepsilon) \log\log(1/\varepsilon))$ which ε-fools (m,n)-Fourier shapes f when $\mathsf{Tvar}(f) \geq (\log(1/\varepsilon))^C$ for some sufficiently large constant C. We build the generator in two steps.

In the first step, we build a PRG with seed-length $O(\log(mn))$ which achieves constant error for (m,n)-Fourier shapes f with $\mathsf{Tvar}(f) \geq 1$. In the second step, we drive the error down to ε as follows. We hash the coordinates into roughly $(\log(1/\varepsilon))^{O(1)}$ buckets, so that for at least $\Omega(\log(1/\varepsilon))$ buckets, f restricted to the coordinates within the bucket has total-variance at least 1. We use the PRG with constant error within each bucket, while the seeds across buckets are recycled using a PRG for small-space algorithms. This construction is inspired by the construction of small-bias spaces due to Naor and Naor [NN93]; the difference being that we use generators for space bounded algorithms for amplification, as opposed to expander random walks as done in [NN93].

2.2 Alphabet-reduction

The next building block in our construction is *alphabet-reduction* which helps us assume without loss of generality that the alphabet-size m is polynomially bounded in terms of the dimension n. This is motivated by the construction of [GMR+12].

Concretely, we show that constructing an ε-PRG for (m,n)-Fourier shapes can be reduced to that of constructing an ε'-PRG for (n^4,n)-Fourier shapes for $\varepsilon' \approx \varepsilon/(\log m)$. The alphabet-reduction step consists of $(\log\log m)$ steps where in each step we reduce fooling (m,n)-Fourier shapes for $m > n^4$, to that of fooling $(\sqrt{m},n)$-Fourier shapes, at the cost of $O(\log(m/\varepsilon))$ random bits.

We now describe a single step that reduces the alphabet from m to $\sqrt{m}$. Consider the following procedure for generating a uniformly random element in $[m]^n$:

- For $D \approx \sqrt{m}$, sample uniformly random subsets

$$S_1 = \{X[1,1], X[1,2], \ldots, X[D,1]\}, \ldots, S_n = \{X[1,n], X[2,n], \ldots, X[D,n]\} \subseteq [m].$$

- Sample $Y = (Y_1, \ldots, Y_n)$ uniformly at random from $[D]^n$.
- Output $(Z_1, \ldots, Z_n)$, where $Z_j = X[Y_j, j]$.

Our goal is to derandomize this procedure. The key observation is that once the subsets $S_1, \ldots, S_n$ are chosen, we are left with a (D,n)-Fourier shape as a function of Y. So the choice of Y can be derandomized using a PRG for Fourier shapes with alphabet $[D]$, and it suffices to derandomize the choice of the X's. A calculation shows that (because the Y's are uniformly random), derandomizing the choice of the X's reduces to that of fooling a Fourier shape of total-variance $1/m^{\Omega(1)}$. Lemma 2.1 implies that this can be done with limited independence.

2.3 Dimension-reduction for low-variance Fourier shapes

We show that constructing an ε-PRG for (n^4,n)-Fourier shapes f with $\mathsf{Tvar}(f) \leq \mathrm{poly}(\log(mn/\varepsilon))$ can be reduced to that of ε'-fooling $(\mathrm{poly}(n/\varepsilon), \sqrt{n})$-Fourier shapes for $\varepsilon' \approx \varepsilon/\log n$. Note

that here we decreased the dimension at the expense of increasing the alphabet-size. However, this can be fixed by employing another iteration of alphabet-reduction. This is the reason why considering (m, n)-Fourier shapes for arbitrary m helps us even if we were only trying to fool $(2, n)$-Fourier shapes. The dimension-reduction proceeds as follows:

1. We first hash the coordinates into roughly $\sqrt{n}$ buckets using a k-wise independent hash function $h \in_u \mathcal{H} = \{h : [n] \leftarrow [\sqrt{n}]\}$ for $k \approx O(\log(n/\varepsilon)/\log n)$. Note that this only requires $O(\log(n/\varepsilon))$ random bits.

2. For the coordinates within each bucket we use a k'-wise independent string in $[m]^n$ for $k' \approx O(\log(n/\varepsilon)/\log n)$. We use true independence across buckets. Note that this requires $\sqrt{n}$ independent seeds of length $r = O(\log(n/\varepsilon))$.

While the above process requires too many random bits by itself, it is easy to analyze. We then reduce the seed-length by observing that if we fix the hash function h, then what we are left with as a function of the seeds used for generating the symbols in each bucket is a $(2^r \leq \text{poly}(n/\varepsilon), \sqrt{n})$-Fourier shape. So rather than using independent seeds, we can use the output of a generator for such Fourier shapes.

The analysis of the above construction again relies on Lemma 2.1. The intuition is that since $\mathsf{Tvar}(f) \leq \text{poly}(\log(n/\varepsilon))$, and we are hashing into $\sqrt{n}$ buckets, for most hash functions h the Fourier shape restricted to each bucket has variance $O(1/n^c)$ for some fixed constant $c > 0$. By Lemma 2.1, limited independence fools such Fourier shapes.

2.4 Main Technical Lemma

The lemma can be seen as a generalization of a similar result proved for real-valued random variables in [GY14](who also have an additional restriction on the means of the random variables Y_j). However, the generalization to complex-valued variables is substantial and seems to require different proof techniques.

We first consider the case where the Y_j's not only have small total-variance, but also have small absolute deviation from their means. Concretely, let $Y_j = \mu_j(1 + Z_j)$ where $\mathbb{E}[Z_j] = 0$ and $|Z_j| \leq 1/2$. In this case, we do a variable change $W_j = \log(1 + Z_j)$ (taking the principal branch of the algorithm) to rewrite

$$\prod_j Y_j = \prod_j \mu_j(1 + Z_j) = \prod_j \mu_j \cdot \exp\left(\sum_j W_j\right).$$

We then argue that $\exp(\sum_j W_j)$ can be approximated by a polynomial $P(W_1, \ldots, W_n)$ of degree less than k with small expected error. The polynomial P is obtained by truncating the Taylor series expansion of the exp( ) function. Once, we have such a low-degree polynomial approximator, the claim follows as limited independence fools low-degree polynomials.

To handle the general case where Z_j's are not necessarily bounded, we use an inclusion-exclusion argument and exploit the fact that with high probability, not many of the Z_j's (say more than $k/2$) will deviate too much from their expectation. We leave the details to the actual proof.

3 Preliminaries

We start with some notation:

- For $v \in \mathbb{R}^n$ and a hash function $h : [n] \to [m]$, define

$$h(v) = \sum_{j=1}^{m} \|v_{|h^{-1}(j)}\|_2^4 \tag{2}$$

- $\mathbb{C}_1 = \{z : z \in \mathbb{C}, |z| \leq 1\}$ be the unit disk in the complex plane.
- For a complex valued random variable Z,

$$Var(Z) \equiv \sigma^2(Z) \equiv \mathbb{E}\left[|Z - \mathbb{E}[Z]|^2\right].$$

- Unless otherwise stated c, C denote universal constants.
- Throughout we assume that n is sufficiently large and that $\delta, \varepsilon > 0$ are sufficiently small.
- For positive functions f, g, h we write $f = g + O(h)$ when $|f - g| = O(h)$.
- For a integer-valued random variable Z, its Fourier transform is given as follows: for $\alpha \in [0, 1]$, $\hat{Z}(\alpha) = \mathbb{E}[\exp(2\pi i \alpha Z)]$. Further, given the Fourier coefficients $\hat{Z}(\alpha)$, one can compute the probability density function of Z as follows: for any integer j,

$$\Pr[Z = j] = \int_0^1 \exp(2\pi i j \alpha) \hat{Z}(\alpha)\, d\alpha.$$

Definition. *For $n, m, \delta > 0$ we say that a family of hash functions $\mathcal{H} = \{h : [n] \to [m]\}$ is δ-biased if for any $r \leq n$ distinct indices $i_1, i_2, \ldots, i_r \in [n]$ and $j_1, \ldots, j_r \in [m]$,*

$$\Pr_{h \in_u \mathcal{H}}[h(i_1) = j_1 \wedge h(i_2) = j_2 \wedge \cdots \wedge h(i_r) = j_r] = \frac{1}{m^r} \pm \delta.$$

We say that such a family is k-wise independent if the above holds with $\delta = 0$ for all $r \leq k$.

We say that a distribution over $\{\pm 1\}^n$ is δ-biased or k-wise independent if the corresponding family of functions $h : [n] \to [2]$ is.

Such families of functions can be generated efficiently using small seeds.

Fact 3.1. *For $n, m, k, \delta > 0$, there exist explicit δ-biased families of hash functions $\mathcal{H} = \{h : [n] \to [m]\}$ that can be generated efficiently from a seed of length $s = O(\log(n/\delta))$. There are also, explicit k-wise independent families that can be generated efficiently from a seed of length $s = O(k \log(nm))$.*

Taking the pointwise sum of such generators modulo m gives a family of hash functions that is both δ-biased and k-wise independent generated from a seed of length $s = O(\log(n/\delta) + k \log(nm))$.

3.1 Basic Results

We start with the simple observation that to δ-fool an (m, n)-Fourier shape f, we can assume the functions in f have bit-precision $2\log_2(n/\delta)$. This observation will be useful when we use PRGs for small-space machines to fool Fourier shapes in certain parameter regimes.

Lemma 3.2. *If a* PRG $\mathcal{G} : \{0,1\}^r \to [m]^n$ *δ-fools (m, n)-Fourier shapes $f = \prod_j f_j$ when $\log(f_j)$'s have bit precision $2\log_2(n/\delta)$, then $\mathcal{G}$ fools all (m, n)-Fourier shapes with error at most 2δ.*

Proof. Consider an arbitrary (m, n)-Fourier shape $f : [m]^n \to \mathbb{C}_1$ with $f = \prod_j f_j$. Let $\tilde{f}_j : [m] \to \mathbb{C}_1$ be obtained by truncating the $\log(f_j)$'s to $2\log_2(n/\delta)$ bits. Then, $|f_j(x_j) - \tilde{f}_j(x_j)| \leq \delta/n$ for all $x_j \in [m]$. Therefore, if we define $\tilde{f} = \prod_j \tilde{f}_j$, then for any $x \in [m]^n$, (as the f_j's and $\tilde{f}_j$'s are in $\mathbb{C}_1$)

$$\left|f(x) - \tilde{f}(x)\right| = \left|\prod_j f_j(x_j) - \prod_j \tilde{f}(x_j)\right| \leq \sum_j \left|f_j(x_j) - \tilde{f}(x_j)\right| \leq \delta.$$

The claim now follows as the above inequality holds point-wise and by assumption, $\mathcal{G}$ δ-fools $\tilde{f}$. □

We collect some known results about pseudorandomness and prove some other technical results that will be used later.

We shall use PRGs for small-space machines or read-once branching programs (ROBP) of Nisan [Nis92], [NZ96] and Impagliazzo, Nisan and Wigderson [INW94]. We extend the usual definitions of read-once branching programs to compute complex-valued functions; the results of [Nis92], [NZ96], [INW94] apply to this extended model readily[2].

Definition 4 ((S, D, T)-ROBP)**.** *An (S, D, T)-ROBP M is a layered directed graph with $T + 1$ layers and 2^S vertices per layer with the following properties.*

[2]This is because these results in fact give guarantees in terms of statistical distance.

- *The first layer has a single* start *node and the vertices in the last layer are labeled by complex numbers from* $\mathbb{C}_1$.
- *A vertex* v *in layer* i, $0 \le i < T$ *has* 2^D *edges to layer* $i+1$ *each labeled with an element of* $\{0,1\}^D$.

A graph M *as above naturally defines a function* $M : \left(\{0,1\}^D\right)^T \to \mathbb{C}_1$ *where on input* $(z_1, \ldots, z_T) \in \left(\{0,1\}^D\right)^T$ *one traverses the edges of the graph according to the labels* $z_1, \ldots, z_T$ *and outputs the label of the final vertex reached.*

Theorem 3.3 ([Nis92], [INW94])**.** *There exists an explicit* PRG $\mathcal{G}^{INW} : \{0,1\}^r \to \left(\{0,1\}^D\right)^T$ *which* ε*-fools* (S, D, T)*-branching programs and has seed-length* $r = O(D + S \log T + \log(T/\delta) \cdot (\log T))$.

Theorem 3.4 ([NZ96])**.** *For all* $C > 1$ *and* $0 < c < 1$, *there exists an explicit* PRG $\mathcal{G}^{NZ} : \{0,1\}^r \to \left(\{0,1\}^D\right)^T$ *which* ε*-fools* (S, S, S^C)*-branching programs for* $\varepsilon = 2^{-\log^{1-c} S}$ *and has seed-length* $r = O(S)$.

The next two lemmas quantify load-balancing properties of δ-biased hash functions in terms of the ℓ_p-norms of vectors. Proofs can be found in Appendix A.

Lemma 3.5. *Let* $p \ge 2$ *be an integer. Let* $v \in \mathbb{R}^n$ *and* $\mathcal{H} = \{h : [n] \to [m]\}$ *be either a* δ*-biased hash family for* $\delta > 0$ *or a* p*-wise independent family for* $\delta = 0$. *Then*

$$\mathbb{E}[h(v)^p] \le O(p)^{2p} \left(\frac{\|v\|_2^4}{m}\right)^p + O(p)^{2p} \|v\|_4^{4p} + m^p \|v\|_2^{4p} \delta.$$

Lemma 3.6. *For all* $v \in \mathbb{R}_+^n$, *let* $p \ge 2$ *be even and* $\mathcal{H} = \{h : [n] \to [m]\}$ *a* p*-wise independent family, and* $j \in [m]$,

$$\Pr\left[\left|\left\|v_{|h^{-1}(j)}\right\|_1 - \|v\|_1 / m\right| \ge t\right] \le \frac{O(p)^{p/2} \|v\|_2^p}{t^p}.$$

4 Fooling products of low-variance random variables

We now show one of our main technical claims that products of complex-valued random variables are fooled by limited independence if the sum of variances of the random variables is small. The lemma is essentially equivalent to saying that limited independence fools low-variance Fourier shapes.

Lemma 4.1. *Let* $Y_1, \ldots, Y_n$ *be* k*-wise independent random variables taking values in* $\mathbb{C}_1$. *Then,*

$$\left|\mathbb{E}[Y_1 \cdots Y_n] - \prod_{j=1}^n \mathbb{E}[Y_j]\right| \le \exp(O(k)) \cdot \left(\frac{\sum_j \sigma^2(Y_j)}{k}\right)^{\Omega(k)}.$$

More concretely, let $X_1, \ldots, X_n$ be independent random variables taking values in $\mathbb{C}_1$. Let $\sigma_i^2 = \mathrm{Var}(X_i)$ and $\sum_{i=1}^n \sigma_i^2 \leq \sigma^2$. Let k be a positive even integer and let $Y_1, \ldots, Y_n$ be a Ck-wise independent family of random variables with each Y_i distributed identically to X_i. Then, we will show that for C a sufficiently big constant,

$$|\mathbb{E}[Y_1 \cdots Y_n] - \mathbb{E}[X_1 \cdots X_n]| = \exp(O(k)) \cdot (\sigma/\sqrt{k})^k. \tag{3}$$

We start with the following standard bound on moments of bounded random variables whose proof is deferred to appendix B.

Lemma 4.2. *Let $Z_1, \ldots, Z_n \in \mathbb{C}$ be random variables with $\mathbb{E}[Z_i] = 0$, $\|Z_i\|_\infty < B$ and $\sum_i Var(Z_i) \leq \sigma^2$. Then, for all even positive integers k,*

$$\mathbb{E}\left[\left|\sum_i Z_i\right|^k\right] \leq 2^{O(k)}(\sigma\sqrt{k} + Bk)^k.$$

We also use some elementary properties of the (complex-valued) log and exponential functions:

Lemma 4.3. *1. For $z \in \mathbb{C}$ with $|z| \leq 1/2$, $|\log(1+z)| \leq 2|z|$, where we take the principle branch of the logarithm.*

2. For $w \in \mathbb{C}$ and $k > 0$,

$$\left|\exp(w) - \sum_{j=0}^{k-1} w^k/k!\right| \leq O(1)\frac{|w|^k}{k!} \cdot \max(1, \exp(\Re(w))).$$

3. For a random variable $Z \in \mathbb{C}$ with $|Z|_\infty \leq 1/2$, $\mathbb{E}[Z] = 0$, and $W = \log(1+Z)$ the principle branch of the logarithm function (phase between $(-\pi, \pi)$), $Var(W) \leq 4Var(Z)$.

4. For any complex-valued random variable $W \in \mathbb{C}$, $|\exp(\mathbb{E}[W])| \leq \mathbb{E}[|\exp(W)|]$.

Proof. Claims (1), (2) follow from the Taylor series expansions for the complex-valued log and exponential functions.

For (3), note that $Var(W) \leq \mathbb{E}[|W|^2] \leq 4\,\mathbb{E}[|Z|^2] = 4Var(Z)$.

For (4), note that $|\exp(\mathbb{E}[W])| = |\exp(E[\Re(W)])|$ and similarly $|\exp(W)| = |\exp(\Re(W))|$. The statement now follows from Jensen's inequality applied to the random variable $\Re(W)$. □

We prove Lemma 4.1 or equivalently, Equation (3) by proving a sequence of increasingly stronger claims. We begin by proving that Equation (3) holds if X_j's have small absolute deviation, i.e., lie in a disk of small radius about a fixed point.

Lemma 4.4. *Let X_i and Y_i be as above. Furthermore, assume that $Y_i = \mu_i(1+Z_i)$ for complex numbers $\mu_i = \mathbb{E}[Y_i]$ and random variables Z_i so that with probability 1, $|Z_i| \le B \le 1/2$ for all i. Let $\tilde{\sigma}_i^2 = Var(Z_i)$, and $\tilde{\sigma}^2 = \sum_{i=1}^n \tilde{\sigma}_i^2$. Then we have that*

$$|\mathbb{E}[X_1 \cdots X_n] - \mathbb{E}[Y_1 \cdots Y_n]| = \exp(O(k)) \cdot (\tilde{\sigma}/k^{1/2} + B)^k.$$

Proof. Let $W_j = \log(1+Z_j)$, taking the principle branch of the logarithm function and let $W_j' = W_j - \mathbb{E}[W_j]$. Then, by Lemma 4.3 (1), (3), $|W_j| \le 2|Z_j| \le 2B$, so that $|W_i|' \le 4B$ and $\mathrm{Var}(W_j') = O(\tilde{\sigma}_j^2)$. Finally, let $W = \sum_{j=1}^n W_j'$.

Now, by Lemma 4.3 (3)

$$\begin{aligned}\prod_{i=1}^n Y_i &= \prod_{i=1}^n (\mu_i \exp(\mathbb{E}[W_i])) \exp(W) \\ &= \prod_{i=1}^n (\mu_i \exp(\mathbb{E}[W_i])) \left(\sum_{\ell=0}^{k-1} \frac{W^\ell}{\ell!} + O(1) \cdot \left(\frac{|W|^k}{k!} \right) \cdot \max(1, \exp(\Re(W))) \right).\end{aligned}$$

Note that the expectation of the ℓ^{th} powers of W are fooled by the k-wise independence of the Y's for $\ell < k$. Therefore the difference in the expectations between the product of Y's and the product of X's is at most

$$O(1) \cdot \prod_{i=1}^n (\mu_i \exp(\mathbb{E}[W_i])) \, \mathbb{E}\left[\left(\frac{|W|^k}{k!} \right) \cdot \max(1, \exp(\Re(W))) \right] \tag{4}$$

Now, by Lemma 4.3 (4),

$$\left| \prod_{i=1}^n \mu_i \exp(\mathbb{E}[W_i]) \right| = \left| \prod_{i=1}^n \mu_i \cdot \exp\left(\mathbb{E}\left[\sum_i W_i \right] \right) \right| \le$$
$$\left| \prod_{i=1}^n \mu_i \cdot \mathbb{E}\left[\exp\left(\sum_i W_i \right) \right] \right| = \mathbb{E}\left[\left| \prod_{i=1}^n \mu_i \exp(W_i) \right| \right] = \mathbb{E}\left[\left| \prod_{i=1}^n Y_i \right| \right] \le 1.$$

Further,

$$\left| \prod_{i=1}^n \mu_i \exp(\mathbb{E}[W_i]) \right| \cdot \exp(\Re(W)) = \left| \prod_{i=1}^n \mu_i \exp(\mathbb{E}[W_i]) \cdot \exp(W) \right| = \left| \prod_{i=1}^n Y_i \right| \le 1.$$

Therefore, by Lemma 4.2, the expression in (4) is at most

$$O(1)\, \mathbb{E}\left[\frac{|W|^k}{k!} \right] \le 2^{O(k)} \cdot \left(\frac{\tilde{\sigma}\sqrt{k} + Bk}{k} \right)^k = 2^{O(k)} \cdot (\tilde{\sigma}/k^{1/2} + B)^k.$$

□

Next, we relax the conditions to handle the case where we only require the means of the X_j's be far from zero.

Lemma 4.5. *Let X_i and Y_i be as in Equation (3). Let $\mu_i = \mathbb{E}[X_i]$. If $|\mu_i| \geq (\sigma/\sqrt{k})^{1/3}$ for all i, then Equation (3) holds.*

Proof. We assume throughout that $\sigma/\sqrt{k}$ is less than a sufficiently small constant; otherwise, there is nothing to prove. Further, note that there can be at most k different indices $j \in [n]$ where $\sigma_j \geq \sigma/\sqrt{k}$. As even after conditioning on the values of the corresponding Y's, the remaining Y_j's are $(C-1)k$-independent, it suffices to prove the lemma when $\sigma_j \leq \sigma/\sqrt{k}$ for all j.

To apply Lemma 4.4, we consider a truncation of our random variables: define

$$\tilde{Y}_i = \begin{cases} Y_i & \text{if } |Y_i - \mu_i| \leq (\sigma/\sqrt{k})^{2/3} \\ \mu_i & \text{else} \end{cases}$$

We claim that the variables $\tilde{Y}_i$ satisfy the conditions of Lemma 4.4. Let $\tilde{\mu}_i = \mathbb{E}[\tilde{Y}_i]$. Note that by Chebyshev bound, $\Pr(\tilde{Y}_i \neq Y_i) \leq \sigma_i^2(\sigma/\sqrt{k})^{-4/3} \leq (\sigma/\sqrt{k})^{2/3}$. Therefore, $|\mu_i - \tilde{\mu}_i| \leq (\sigma/\sqrt{k})^{2/3}$, so that $|\tilde{\mu}_i| \geq (1/2)|\mu_i|$. Furthermore, letting $\tilde{Y}_i = \tilde{\mu}_i(1+Z_i)$, we have that

$$\mathbb{E}[Z_i] = 0, \quad \|Z_i\|_\infty \leq 2(\sigma/\sqrt{k})^{1/3}, \quad \mathrm{Var}(Z_i) \leq 4\sigma_i^2(\sigma/\sqrt{k})^{-2/3}, \quad \sum_i \mathrm{Var}(Z_i) \leq 4\sigma_i^2(\sigma/\sqrt{k})^{-2/3}. \tag{5}$$

Finally, note that

$$\prod_{i=1}^n Y_i = \prod_{i=1}^n (Y_i - \tilde{Y}_i + \tilde{Y}_i) = \sum_{S \subseteq [n]} \prod_{i \in S} (Y_i - \tilde{Y}_i) \prod_{i \notin S} \tilde{Y}_i.$$

We truncate the above expansion to only include terms corresponding to sets S with $|S| < m$ for $m = O(k)$ to be chosen later. Let

$$P_m(Y_1, \ldots, Y_n) = \sum_{S \subseteq [n], |S| < m} \prod_{i \in S} (Y_i - \tilde{Y}_i) \prod_{i \notin S} (\tilde{Y}_i),$$

and let N equal the number of i so that $Y_i \neq \tilde{Y}_i$. We claim that

$$\left| \prod_{j=1}^n Y_j - P_m(Y_1, \ldots, Y_n) \right| \leq 2^m \binom{N}{m}.$$

The above clearly holds when $N < m$, since in this case for any S of size at least m we have $\prod_{i \in S}(Y_i - \tilde{Y}_i) = 0$. On the other hand for $N \geq m$ we note that there

are at most $\sum_{\ell=0}^{m-1}\binom{N}{\ell} \leq 2^m\binom{N}{m}$ subsets S for which this product is non-zero. Hence, $|P_m(Y_1,\ldots,Y_n)| < 2^m\binom{N}{m}$.

We now argue that Ck-wise independence fools the individual terms of P_m when $m = O(k)$. This is because, the Y_j for $j \in S$ are independent and conditioned on their values, the remaining $\tilde{Y}_j$ for $j \notin S$ are still $C'k$-wise independent for some sufficiently large constant C'. Therefore, applying Lemma 4.4 with parameters as given by Equation (5), Ck-wise independence fools $P_m(Y_1,\ldots,Y_n)$ up to error

$$\sum_{S\subseteq[n],|S|<m}\prod_{i\in S}\left|\mathbb{E}[Y_i-\tilde{Y}_i]\right|\cdot 2^{O(k)}\left(\frac{\tilde{\sigma}^2}{\sqrt{k}}\left(\frac{\sigma}{\sqrt{k}}\right)^{-2/3}+\left(\frac{\sigma}{\sqrt{k}}\right)^{1/3}\right)^{3k},$$

where

$$\left(\frac{\tilde{\sigma}^2}{\sqrt{k}}\left(\frac{\sigma}{\sqrt{k}}\right)^{-2/3}+\left(\frac{\sigma}{\sqrt{k}}\right)^{1/3}\right)^{3k} = O(\sigma/\sqrt{k})^k.$$

Therefore, P_m is fooled to error

$$\sum_{\ell=0}^{m-1}\mathbb{E}\left[\binom{N}{\ell}\right]2^{O(k)}\cdot(\sigma/\sqrt{k})^k.$$

Note that the expectation above is the same as what it would be if the Y_i's were fully independent, in which case it is at most

$$\mathbb{E}[2^N] = \prod_{i=1}^{n}(1+\Pr(Y_i\neq\tilde{Y}_i)) \leq \exp\left(\sum_{i=1}^{n}\Pr(Y_i\neq\tilde{Y}_i)\right) = \exp\left(O\left(\sum_{i=1}^{n}\sigma_i^2(\sigma/\sqrt{k})^{-4/3}\right)\right) = \exp(O(\sigma^{2/3}k^{2/3})) = \exp(O(k)).$$

Therefore, Ck-wise independence fools P_m to error $2^{O(k)}\cdot(\sigma/\sqrt{k})^k$.

On the other hand, the expectation of $\binom{N}{m}$ is

$$\begin{aligned}\sum_{S\subseteq[n],|S|=m}\prod_{i\in S}\Pr(Y_i\neq\tilde{Y}_i) &\leq \frac{\left(\sum_{i=1}^{n}\Pr(Y_i\neq\tilde{Y}_i)\right)^m}{m!}\\ &\leq \frac{\left(\sum_{i=1}^{n}\sigma_i^2(\sigma/\sqrt{k})^{-4/3}\right)^m}{m!}\\ &\leq O\left((\sigma^2/m)(\sigma^2/k)^{-2/3}\right)^m.\end{aligned}$$

Taking $m = 3k/2$ yields a final error of $\exp(O(k))\cdot(\sigma/\sqrt{k})^k$. This completes our proof. □

Finally, we can extend our proof to cover the general case.

Proof of Lemma 4.1. Note that it suffices to prove that Equation (3) holds. As before, it suffices to assume that $\sigma/\sqrt{k} \ll 1$ and that $\sigma_i \leq \sigma/\sqrt{k}$ for all i.

Let m be the number of i so that $|\mathbb{E}[Y_i]| \leq (\sigma/\sqrt{k})^{1/3}$. Assume that the Y's with *small* expectation are $Y_1, \ldots, Y_m$. We break into cases based upon the size of m.

On the one hand if $m \leq 6k$, we note that for C sufficiently large, the values of $Y_1, \ldots, Y_m$ are independent of each other, and even after conditioning on them, the remaining Y_i's are still $C'k$-wise independent. Thus, applying Lemma 4.5 to the expectation of the product of the remaining Y_i we find that the difference between the expectation of the product of X's and product of Y's is as desired.

For $m \geq 6k$ we note that

$$\left|\mathbb{E}\left[\prod_{i=1}^{n} X_i\right]\right| = \prod_{i=1}^{n} |\mathbb{E}[Y_i]| \leq (\sigma/\sqrt{k})^{m/3}.$$

Therefore, it suffices to show that

$$\left|\mathbb{E}\left[\prod_{i=1}^{n} Y_i\right]\right| = O(\sigma/\sqrt{k})^{k}.$$

Notice that so long as at least $3k$ of $Y_1, \ldots, Y_m$ have absolute value less than $2(\sigma/\sqrt{k})^{1/3}$, then

$$\left|\prod_{i=1}^{n} Y_i\right| = O(\sigma/\sqrt{k})^{k}.$$

Therefore, it suffices to show that this occurs except with probability at most $O(\sigma/\sqrt{k})^{k}$. Let N be the number of $1 \leq i \leq m$ so that $|Y_i| \geq 2(\sigma/\sqrt{k})^{1/3}$. Note that

$$\mathbb{E}[N] = \sum_{i=1}^{m} \Pr(|Y_i| \geq 2(\sigma/\sqrt{k})^{1/3}) \leq \sum_{i=1}^{m} \sigma_i^2 (\sigma/\sqrt{k})^{-2/3} \leq \sigma^2 (\sigma^2/k)^{-1/3}.$$

On the other hand, we have that

$$\begin{aligned}
\Pr(N \geq 3k) &\leq \mathbb{E}\left[\binom{N}{3k}\right] \\
&= \sum_{S\subseteq[m],|S|=3k} \prod_{i\in S} \Pr(|Y_i| \geq 2(\sigma/\sqrt{k})^{1/3}) \\
&\leq \frac{\left(\sum_{i=1}^{m} \Pr(|Y_i| \geq 2(\sigma/\sqrt{k})^{1/3})\right)^{3k}}{(3k)!} \\
&= \frac{\mathbb{E}[N]^{3k}}{(3k)!} \\
&\leq O((\sigma^2/k)^{2/3})^{3k} \\
&\leq O(\sigma/\sqrt{k})^{k}.
\end{aligned}$$

This completes the proof. □

5 A Generator for high-variance Fourier shapes

In this section, we construct a generator that fools Fourier shapes with high variance.

Theorem 5.1. *There exists a constant $C > 0$, such that for all $\delta > 0$, there exists an explicit generator $\mathcal{G}_\ell : \{0,1\}^{r_\ell} \to [m]^n$ with seed-length $r_\ell = O(\log(mn/\delta)\log\log(1/\delta))$ such that for all Fourier shapes $f : [m]^n \to \mathbb{C}_1$ with $\mathsf{Tvar}(f) \geq C\log^5(1/\delta)$, we have*

$$\left| \mathop{\mathbb{E}}_{z\sim\{0,1\}^{r_\ell}}[f(\mathcal{G}_\ell(z))] - \mathop{\mathbb{E}}_{X\in_u[m]^n}[f(X)] \right| < \delta.$$

We start with the simple but crucial observation that Fourier shapes with large variance have small expectation.

Lemma 5.2. *For any Fourier shape $f : [m]^n \to \mathbb{C}_1$, we have*

$$\left| \mathop{\mathbb{E}}_{X\in_u[m]^n}[f(X)] \right| \leq \exp(-\mathsf{Tvar}(f)/2). \tag{6}$$

Proof. Let $f(x) = \prod_j f_j(x_j)$. Since $f_j(x) \in \mathbb{C}_1$, we have $|f_j(x)| \leq 1$. Let $\mu_j = \mathbb{E}_{X_j\in[m]}[f_j(X_j)]$. For $X \in_u [m]^n$,

$$\sigma_j^2 = \mathbb{E}[|f_j(X_j) - \mu_j|^2] = \mathbb{E}[|f_j(X_j)|^2] - |\mu_j|^2 \leq 1 - |\mu_j|^2.$$

Hence

$$\left|\mathop{\mathbb{E}}_{X}[f(X)]\right| = \prod_{j=1}^{n} |\mu_j| \leq \prod_{j=1}^{n} (1-\sigma_j^2)^{1/2}$$
$$\leq \exp(-\sum_{j=1}^{n} \sigma_j^2/2) \leq \exp(-\mathsf{Tvar}(f)/2).$$

□

We build the generator in two steps. We first build a generator with seed-length $O(\log n)$ which achieves constant error for all f with $\mathsf{Tvar}(f) \geq 1$. In the second step, we reduce the error down to δ. This construction is inspired by a construction of Naor and Naor [NN93] of small-bias spaces.

5.1 A generator with constant error

Our goal in this subsection is get a generator with constant error for Fourier shapes where $\mathsf{Tvar}(f) = \Omega(1)$. We start by showing that when $\mathsf{Tvar}(f) = \Theta(1)$ (instead of just $\Omega(1)$), $O(1)$-wise independence is enough to fool f.

Lemma 5.3. *For all constants $0 < c_1 < c_2$, there exist $p \in \mathbb{Z}_+$ and $0 < c' < 1$ such that the following holds. For any (m,n)-Fourier shape, f with $\mathsf{Tvar}(f) \in [c_1, c_2]$, and $Z \sim [m]^n$ $2p$-wise independent,*

$$\left|\mathop{\mathbb{E}}_{Z}[f(Z)]\right| < c'.$$

Proof. Let $f = \prod_j f_j$, $X \in_u [m]^n$. Now, by Lemma 4.1 applied to $Y_j = f_j(Z_j)$, we have,

$$|\mathbb{E}[f(Z)] - \mathbb{E}[f(X)]| \leq \exp(O(p))(\mathsf{Tvar}(f)/\sqrt{p})^{\Omega(p)} = \exp(O(p))(c_2/\sqrt{p})^{\Omega(p)}.$$

Note that by taking p to be a sufficiently large constant compared to c_2, we can make the last bound arbitrary small.

On the other hand, by Equation (6),

$$|\mathbb{E}[f(X)]| \leq \exp(-\mathsf{Tvar}(f)/2) \leq \exp(-c_1/2).$$

Therefore,

$$|\mathbb{E}[f(Z)]| \leq \exp(-c_1/2) + \exp(O(p))(c_2/\sqrt{p})^{\Omega(p)} < c'$$

for p sufficiently large constant and some constant $0 < c' < 1$. □

We reduce the general case of $\mathsf{Tvar}(f) \in [1,n]$ to the case above where $\mathsf{Tvar}(f) = \Theta(1)$ by using the Valiant-Vazirani technique of sub-sampling. For $B \subseteq [n]$ let $\mathsf{Tvar}(f_B) = \sum_{i \in B} \sigma_i^2$. If we sample a random subset $B \subseteq [n]$ with $|B| \approx n/\mathsf{Tvar}(f)$ in a pairwise

independent manner, we will get $\mathsf{Tvar}(f_B) = \Theta(1)$ with $\Omega(1)$ probability. Since we do not know $\mathsf{Tvar}(f)$, we sample $\log(n)$ subsets whose cardinalities are geometrically increasing; one of them is likely to satisfy the desired bound.

We set up some notation that will be used in the remainder of this section.

- Assume n is a power of 2, and set $T = \log_2(n) - 1$. Let $\Pi \subseteq \mathbb{S}_n$ be a family of pairwise independent permutations so that $\pi \in_u \Pi$ can be sampled efficiently with $O(\log n)$ random bits. For $0 \le j \le T$, let $B_j = \{\pi(i) : i \in \{2^j, \ldots, 2^{j+1} - 1\}\}$ be the 2^j co-ordinates that land in the j^{th} bucket.
- For $v \in \mathbb{R}^n$, let $v^j = v_{B_j}$ denote the projection of v onto coordinates in bucket j. Similarly, for $x \in [m]^n$, let x^j denote the projection of x to the co-ordinates in B_j.
- Fix an (m, n)-Fourier shape $f : [m]^n \to \mathbb{C}_1$ with $f(x) = \prod_i f_i(x_i)$. Define $f^j : [m]^{B_j} \to \mathbb{C}_1$ as $f^j(x^j) = \prod_{i \in B_j} f_i(x_i)$.

Lemma 5.4. *Let $v \in \mathbb{R}^n$ with $\|v\|_2^2 \in [1, n]$, $\|v\|_\infty \le 1$ and $t \in [\log_2 n]$ be such that $n/2^{t+1} \le \|v\|_2^2 \le n/2^t$. Then,*

$$\Pr_{\pi \in_u \Pi}\left[\left\|v^t\right\|_2^2 \in [1/6, 4/3]\right] \ge 7/16.$$

The proof of this lemma is standard and is deferred to Appendix C.

This naturally suggests using an $O(1)$-wise independent distribution within each bucket. But using independent strings across the $\log(n)$ buckets would require a seed of length $O(\log(mn) \cdot (\log n))$. We analyze our generator assuming independence across distinct buckets, but then recycle the seeds using PRGs for space bounded computation to keep the seed-length down to $O(\log(mn))$ (rather than $O(\log^2(n))$).

We now prove the main claim of this subsection.

Lemma 5.5. *There exists an explicit generator $\mathcal{G}_1 : \{0,1\}^r \to [m]^n$ with $r = O(\log(mn))$ such that for all Fourier shapes $f : [m]^n \to \mathbb{C}_1$ with $\mathsf{Tvar}(f) \ge 1$, we have*

$$\left|\mathop{\mathbb{E}}_{z \sim \{0,1\}^r}[f(\mathcal{G}_1(z))]\right| \le c.$$

for some constant $0 < c < 1$.

Proof. Let $\pi \in_u \Pi$ and let $Z^j \sim [m]^{2^j}$ be an independent p-wise independent string for a parameter $p = O(1)$ to be chosen later. Define

$$\mathcal{G}_1'(\pi, Z^0, \ldots, Z^T) = Y, \quad \text{where } Y_{B_j} = Z^j \text{ for } j \in \{0, \ldots, T\}.$$

In other words, the generator applies the string Z^j to the coordinates in bucket B_j.

Observe that $f(Y) = \prod_{j=0}^{\log(n)-1} f^j(Z^j)$. Since the Z^j's are independent of each other

$$|\mathbb{E}[f(Y)]| = \left| \prod_{j=0}^{\log(n)-1} \mathbb{E}[f^j(Z^j)] \right| \leq \left| \mathbb{E}[f^t(Z^t)] \right|,$$

for any $t \leq T$. Applying Lemma 5.3 to $v = (\sigma_1(f_1), \ldots, \sigma_n(f_j))$, we get that for some $t \leq T$, $\mathsf{Tvar}(f^t) = \left\|v^t\right\|_2^2 \in [1/6, 4/3]$ with probability at least $7/16$. Conditioned on this event, Lemma 5.3 implies that for p a sufficiently large constant, there exists a constant $c' < 1$ so that $\left|\mathbb{E}[f^t(Z^t)]\right| < c'$. Therefore, overall we get

$$|\mathbb{E}[f(Y)]| \leq \left|\mathbb{E}[f^t(Z^t)]\right| \leq \frac{9}{16} + \frac{7c'}{16} = c'' < 1.$$

We next improve the seed-length of $\mathcal{G}_1'$ using the PRG for ROBPs of Theorem 3.4. To this end, note that by Lemma 3.2 we can assume that every $\log(f_i(x_i))$, and hence every $\log(f^j(x^j))$, has bit precision at most $O(\log n)$ bits (since our goal is to get error $\delta = O(1)$). Further, each Z^j can be generated efficiently with $O(\log(mn))$ random bits.

Thus, for a fixed permutation π, the computation of $f(\mathcal{G}'(\pi, Z^1, \ldots, Z^T))$ can be done by a (S, D, T)-ROBP where S, T are $O(\log n)$ and $D = O(\log(mn))$: for $j \in \{1, \ldots, T\}$, the ROBP computes $f^j(Z^j)$ and multiplies it to the product computed so far, which can be done using $O(\log n)$ bits of space. Let $\mathcal{G}^{NZ} : \{0,1\}^r \to \left(\{0,1\}^D\right)^T$ be the generator in Theorem 3.4 fooling (S, D, T)-ROBPs as above with error $\delta < (1 - c'')/2$. $\mathcal{G}^{NZ}$ has seedlength $O(\log(mn))$. Let

$$\mathcal{G}_1(\pi, z) = \mathcal{G}_1'(\pi, \mathcal{G}^{NZ}(z)).$$

It follows that $|\mathbb{E}[f(\mathcal{G}_1(\pi, z))]| < c$ for some constant $c < 1$. Finally, the seed-length of $\mathcal{G}_1$ is $O(\log(mn))$ as π can be sampled with $O(\log n)$ random bits and the seed-length of $\mathcal{G}^{NZ}$ is $O(\log(mn))$. The lemma is now proved. □

5.2 Reducing the error

We now amplify the error to prove Theorem 5.1. The starting point for the construction is the observation that for $X \in_u [m]^n$, $|\mathbb{E}[f(X)]| \leq \exp(-\mathsf{Tvar}(f/2)) \leq \delta$ once $\mathsf{Tvar}(f) \gg \log(1/\delta)$. Therefore, it suffices to design a generator so that $\mathbb{E}[f] \ll \delta$, when $\mathsf{Tvar}(f)$ is sufficiently large.

Our generator will partition $[n]$ into $m = O((\log(1/\delta))^5)$ buckets $B_1, \ldots, B_m$, using a family of hash functions with the following *spreading* property:

Definition 5. *A family of hash functions $\mathcal{H} = \{h : [n] \to [m]\}$ is said to be (B, ℓ, δ)-spreading if for all $v \in [0,1]^n$ with $\|v\|_2^2 \geq B$,*

$$\Pr_{h \in_u \mathcal{H}}[|\{j \in [m] : \left\|v_{h^{-1}(j)}\right\|_2^2 \geq B/2m\}| \geq \ell] \geq 1 - \delta.$$

Using the notation from the last subsection, we write $f(x) = \prod_{j=1}^{m} f^j(x^j)$ where $f^j(x^j) = \prod_{i \in B_j} f_i(x_i)$. If $\mathsf{Tvar}(f)$ is sufficiently large, then the spreading property guarantees that for at least $\Omega(\log(1/\delta))$ of the buckets B_j, $\mathsf{Tvar}(f^j) \geq 1$. If we now generate $X \in [m]^n$ by setting X_{B_j} to be an independent instantiation of the generator $\mathcal{G}_1$ from Lemma 5.3, then we get $\mathbb{E}[f(X)] \ll \delta$. As in the proof of Lemma 5.5, we keep the seed-length down to $\tilde{O}(\log(n/\delta))$ by recycling the seeds for the buckets using a PRG for small-space machines.

We start by showing that the desired hash functions can be generated from a small-bias family of hash functions. We show that it satisfies the conditions of the lemma by standard moment bounds. The proof is in Appendix C

Lemma 5.6. *For all constants C_1, there exist constants C_2, C_3 such that following holds. For all $\delta \geq 0$, there exists an explicit hash family $\mathcal{H} = \{h : [n] \to [T]\}$, where $T = C_2 \log^5(1/\delta))$ which is $(C_3 \log^5(1/\delta), C_1 \log(1/\delta), \delta)$-spreading and $h \in_u \mathcal{H}$ can be sampled efficiently with $O(\log(n/\delta))$ bits.*

We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1. Let $\ell = C \log(1/\delta)$ for some constant to be chosen later and let $\mathcal{H} = \{h : [n] \to [T]\}$ be a (B, ℓ, δ)-spreading family as in Lemma 5.6 above for $B = \Theta(\log^5(1/\delta))$ and $T = \Theta(\log^5(1/\delta))$. Let $\mathcal{G}_1 : \{0,1\}^{r_1} \to [m]^n$ be the generator in Lemma 5.3. Define a new generator $\mathcal{G}'_\ell : \mathcal{H} \times (\{0,1\}^{r'})^T \to [m]^n$ as:

$$\mathcal{G}'_\ell(h, z^1, \ldots, z^T) = X, \quad \text{where } X_{h^{-1}(j)} = \mathcal{G}_1(z^j) \text{ for } j \in [T].$$

Let $f : [m]^n \to \mathbb{C}_1$ with $\mathsf{Tvar}(f) \geq \max(2T, B)$. For $h \in \mathcal{H}$, let $I = \{j : \mathsf{Tvar} f^j \geq 1\}$. For any fixed $h \in \mathcal{H}$, as the z^j's are independent of each other,

$$|\mathbb{E}[f(X)]| = \prod_{j=1}^{m} \left|\mathbb{E}[f^j(\mathcal{G}_1(z^j))]\right| \leq \prod_{j \in I} \left|\mathbb{E}[f^j(\mathcal{G}_1(z^j))]\right| \leq c^{|I|},$$

where $c < 1$ is the constant from Lemma 5.3. By the spreading property of $\mathcal{H}$, with probability at least $1 - \delta$, $|I| \geq C \log(1/\delta)$. Therefore, for C sufficiently large,

$$|\mathbb{E}[f(X)]| \leq \delta + c^{C \log(1/\delta)} < 2\delta.$$

As in Lemma 5.5, we recycle the seeds for the various buckets using the PRGs for ROBPs. By Lemma 3.2, we may assume that f^j has bit precision at most $O(\log(n/\delta))$ bits. Further note that

$$f(\mathcal{G}'_\ell(h, z^1, \ldots, z^m)) = \prod_{j=1}^{m} f^j(\mathcal{G}_1(z^j)).$$

For a fixed hash function $h \in \mathcal{H}$, this can be computed by a (S, D, T)-ROBP where $S = O(\log(n/\delta))$ and $D = O(\log(mn))$, corresponding to the various possible seeds for $\mathcal{G}_1$. Let $\mathcal{G}^{INW} : \{0,1\}^r \to \left(\{0,1\}^D\right)^T$ be a generator fooling (S, D, T)-ROBPs as in Theorem 3.3 with error δ and define

$$\mathcal{G}_\ell(h, z) = \mathcal{G}'_\ell(h, \mathcal{G}^{INW}(z)).$$

The seed-length is dominated by the seed-length of $\mathcal{G}^{INW}$, which is

$$O(\log(mn/\delta) \log T) = O(\log(mn/\delta) \log\log(1/\delta)).$$

It follows that $|\mathbb{E}[f(\mathcal{G}_\ell(h, z))]| < 3\delta$, whereas for a truly random $Y \in_u [m]^n$,

$$|\mathbb{E}[f(Y)]| \leq \exp(-\mathsf{Tvar}(f)/2) < \delta.$$

The theorem now follows. □

6 Alphabet reduction for Fourier shapes

In this section, we describe our alphabet-reduction procedure, which reduces the general problem of constructing an ε-PRG for (m, n)-Fourier shapes where m could be much larger than n, to that of constructing an $\varepsilon/\log(m)$-PRG for (n^4, n)-Fourier shapes. This reduction is composed of $O(\log\log m)$ steps where in each step we reduce fooling (m, n)-Fourier shapes to fooling $(\sqrt{m}, n)$-Fourier shapes. Each of these steps in turn will cost $O(\log(m/\varepsilon))$ random bits, so that the overall cost is $O(\log(m/\varepsilon) \cdot (\log\log m))$. Concretely, we show the following:

Theorem 6.1. *Let $n, \delta > 0$ and suppose that for some $r' = r'(n, \delta')$, for all $m' \leq n^4$ there exists an explicit generator $\mathcal{G}_{m'} : \{0,1\}^{r_1} \to [m']^n$ which δ'-fools (m', n)-Fourier shapes. For all m, there exists an explicit generator $\mathcal{G}_m : \{0,1\}^r \to [m]^n$ which $(\delta' + \delta)$-fools (m, n)-Fourier shapes with seed-length $r = r' + O(\log(m/\delta)\log\log(m))$.*

Proof. We prove the claim by showing that for $m > n^4$, we can reduce $(\delta+\delta')$-fooling (m, n)-Fourier shapes to that of δ'-fooling $(\sqrt{m}, n)$-Fourier shapes with $O(\log(m/\delta))$ additional random bits. The theorem follows by applying the claim $\log\log(m)$ until the alphabet size drops below n^4 when we can use $\mathcal{G}_{m'}$. This costs a total of $r' + O(\log(m/\delta)\log\log(m))$ random bits, and gives error $\delta' + \log\log(m)\delta$. The claim follows by replacing δ with $\delta/\log\log(m)$.

Thus, suppose that $m > n^4$ and for $D = \lfloor\sqrt{m}\rfloor$, we have a generator $\mathcal{G}_D : \{0,1\}^{r_D} \to [D]^n$ which δ'-fools (D, n)-Fourier shapes. The generator $\mathcal{G}_m$ works as follows:

1. Generate a matrix $X \in [m]^{D \times n}$ where
   - Each column of X is from a pairwise independent distribution over $[m]^D$.
   - The different columns are k-wise independent for $k = C\log(1/\delta)/\log(m)$ for some sufficiently large constant C.

2. Generate $Y = (Y_1, \ldots, Y_n) = \mathcal{G}_D(z) \in [D]^n$ for $z \in_u \{0,1\}^{r_D}$.

3. $\mathcal{G}_m$ outputs $Z = (Z_1, \ldots, Z_n) \in [m]^n$ where $Z_j = X[Y_j, j]$ for $j \in [n]$.

Each column of X can be generated using a seed of length $2 \log m$. By using seeds for various columns that are k-wise independent, generating X requires seedlength $O(k \log m) = O(\log(1/\delta))$ (as $m > n^2$), while the number of bits needed to generate Z is $r_D + O(\log(1/\delta))$.

Fix an (m,n)-Fourier shape $f : [m]^n \to \mathbb{C}_1$, $f(z) = \prod_j f_j(z_j)$. For $x \in [m]^{D \times n}$, define a (D,n)-Fourier shape $f^x : [D]^n \to \mathbb{C}_1$ by:

$$f^x(y_1, \ldots, y_n) = \prod_{j=1}^{n} f_j(x[y_j, j]).$$

Note that $f(Z) = f^X(Y)$.

Let X', Y' be random variables distributed uniformly over $[m]^{D \times n}$ and $[D]^n$ respectively. Let $Z'_j = X'[Y'_j, j]$ for $j \in [n]$, so that Z' is uniform over $[m]^n$ and $f(Z') = f^{X'}(Y')$. Our goal is to show that $f(Z')$ and $f(Z)$ are close in expectation. We do this by replacing X' and Y' by X and Y respectively.

That we can replace Y' with Y follows from the pseudorandomness of $\mathcal{G}_D$. For any fixed $x \in [m]^n$, as $\mathcal{G}_D$ fools (D,n)-Fourier shapes,

$$\left| \mathop{\mathbb{E}}_{Y = \mathcal{G}_D(z)} [f^x(Y)] - \mathop{\mathbb{E}}_{Y' \in_u [D]^n} [f^x(Y')] \right| \leq \delta'. \tag{7}$$

We now show that for truly random Y', one can replace X by X'. Note that

$$\mathop{\mathbb{E}}_{Y' \in_u [D]^n} [f^x(Y')] = \prod_{j=1}^{n} \left(\frac{1}{D} \cdot \left(\sum_{\ell=1}^{D} f_j(x[\ell, j]) \right) \right) \equiv B_f(x). \tag{8}$$

where we define the *bias-function* $B_f : [m]^{D \times n} \to \mathbb{C}_1$ as above. We claim that X fools B_f:

$$\left| \mathbb{E}[B_f(X)] - \mathbb{E}[B_f(X')] \right| \leq \delta. \tag{9}$$

For $j \in [n]$, let

$$A_j = \frac{1}{D} \left(\sum_{\ell=1}^{D} f_j(X[\ell, j]) \right), \; A'_j = \frac{1}{D} \left(\sum_{\ell=1}^{D} f_j(X'[\ell, j]) \right)$$

so that

$$B_f(X) = \prod_{j=1}^{n} A_j, \; B_f(X') = \prod_{j=1}^{n} A'_j.$$

Since $f_j(X[\ell, j]) \in \mathbb{C}_1$ for $\ell \in [D]$, it follows that $A_j, A'_j \in \mathbb{C}_1$. Since the $f_j(X[\ell, j])$s are pairwise independent variables,

$$\mathbb{E}[A_j] = \mathbb{E}[A'_j], \ \mathrm{Var}[A_j] = \mathrm{Var}[A'_j].$$

Note that

$$\mathbb{E}[B_f(X')] = \mathbb{E}[\prod_{i=1}^{n} A'_j] = \prod_{j=1}^{n} \mathbb{E}[A'_j] = \prod_{j=1}^{n} \mathbb{E}[A_j], \ \mathbb{E}[B_f(X)] = \mathbb{E}[A_1 \cdots A_n]. \tag{10}$$

The random variables $A_1, \ldots, A_n$ are k-wise independent. Further, we have

$$\mathrm{Var}(A_j) = \frac{1}{D^2} \sum_{\ell=1}^{D} \mathrm{Var}(f_j(X[\ell, j])) = \frac{\sigma^2(f_j)}{D} \leq \frac{1}{D}.$$

Therefore, by Lemma 4.1,

$$\left| \mathbb{E}[A_1 \cdots A_n] - \prod_{j=1}^{n} \mathbb{E}[A_j] \right| \leq \left(\frac{n}{D}\right)^{\Omega(k)} \leq m^{-\Omega(k)} \leq \delta \tag{11}$$

where the second to last inequality follows becase $n \leq m^{1/4}$ and $D \geq \sqrt{m}/2$, and the last holds for $k = C \log(1/\delta)/\log(m)$ for a sufficiently big constant C. Equation 9 now follows from Equations (11) and (10).

Finally,

$$\begin{aligned}
\left|\mathbb{E}[f(Z)] - \mathbb{E}[f(Z')]\right| &= \left|\mathbb{E}[f^X(Y)] - \mathbb{E}[f^{X'}(Y')]\right| \\
&= \left|\mathbb{E}[f^X(Y')] - \mathbb{E}[f^{X'}(Y')]\right| + \delta' && \text{Equation (7)} \\
&= \left|\mathbb{E}[B_f(X)] - \mathbb{E}[B_f(X')]\right| + \delta' && \text{Equation (8)} \\
&\leq \delta + \delta'. && \text{Equation (9)}
\end{aligned}$$

Hence the theorem is proved. □

7 Dimension reduction for low-variance Fourier shapes

We next describe our *dimension reduction* step for low-variance Fourier shapes. We start with an (m, n)-Fourier shape where $m \leq n^4$ and $\mathsf{Tvar}(f) \leq \log(n/\delta)^c$. We show how one can reduce the dimension to $t = \sqrt{n}$, at a price of a blowup in the alphabet size m' which now becomes $(n/\delta)^c$ for some (large) constant c.

Theorem 7.1. *Let $\delta > 0$, $n > 0$ and $t = \lceil \sqrt{n} \rceil$. There is a constant c and $m' \leq (n/\delta)^c$ such that the following holds: if there exists an explicit* PRG *$\mathcal{G}' : \{0,1\}^{r'} \to [m']^t$ with seed-length $r' = r'(n, \delta')$ which δ'-fools (m', t)-Fourier shapes, then there exists an explicit generator $\mathcal{G} : \{0,1\}^r \to [m]^n$ with seed-length $r = r' + O(\log(n/\delta))$ which $(\delta + \delta')$-fools (m, n)-Fourier shapes f with $m \leq n^4$ and $\mathsf{Tvar}(f) \leq n^{1/9}$.*

We first set up some notation. Assume that we have fixed a hash function $h : [n] \to [t]$. For $x \in [m]^n$ and $j \in [t]$, let x^j denote the projection of x onto co-ordinates in $h^{-1}(j)$. For an (m, n)-Fourier shape $f : [m]^n \to \mathbb{C}_1$ with $f = \prod_{i=1}^n f_i$, let

$$f^j(x^j) = \prod_{i:h(i)=j} f_i(x_i)$$

$$\text{so that} \quad f(x) = \prod_{j=1}^{t} f^j(x^j).$$

We start by constructing an easy to analyze generator G_1 which hashes co-ordinates into buckets using k-wise independence and then uses independent k-wise independent strings within a bucket. Let

$$k = C\frac{\log(n/\delta)}{\log(n)} \tag{12}$$

where C will is a sufficiently large constant. Let $\mathcal{H} : \{[n] \to t\}$ be a k-wise independent family of hash functions. Let $G_0 : \{0,1\}^{r_0} \to [m]^n$ be a k-wise independent generator over $[m]^n$. Define a new generator $G_1 : \mathcal{H} \times (\{0,1\}^{r_0})^t \to [m]^n$ as:

$$G_1(h, z_1, \ldots, z_t) = Z, \text{where } Z^j = G_0(z_j) \ \forall \ j \in [t]. \tag{13}$$

We argue that G_1 fools (m, n)-Fourier shapes with small total variance as in the theorem. Our analysis proceeds as follows:

- With high probability over $h \in_u \mathcal{H}$, each of the f^j's has low variance except for a few heavy co-ordinates (roughly $\mathsf{Tvar}(f)/t$ after dropping $k/2$ heavy coordinates).
- Within each bin we have k-wise independence, whereas the distributions across bins are independent. So even conditioned on the heavy co-ordinates in a bin, the remaining distribution in the bin is $k/2$-wise independent. Hence each f^j is fooled by Lemma 4.1.

However, the seed-length of G_1 is prohibitively large: since we use independent seeds across the various buckets, the resulting seed-length is $O(\sqrt{n}\log(n/\delta))$. The crucial observation is that we can *recycle* the seeds for various buckets using a generator that fools (m', t)-Fourier shapes with $m' = 2^{r_0} = \text{poly}(n/\delta)$ and $t = O(\sqrt{n})$. Given such a generator

$\mathcal{G}' : \{0,1\}^{r'} \to [m']^t$ which δ-fools (m', t)-Fourier shapes, our final generator for small-variance Fourier shapes is $\mathcal{G}_s : \mathcal{H} \times \{0,1\}^{r'} \to [m]^n$ is defined as

$$\mathcal{G}(h, w) = G_1(h, \mathcal{G}'(w)). \tag{14}$$

It is worth mentioning that even though the original Fourier shape $f : [m]^n \to \mathbb{C}_1$ has low total variance, the generator $\mathcal{G}'$ needs to fool all (m', t)-Fourier shapes, not just those with low variance.

7.1 Analysis of the dimension-reduction step

For $\alpha > 0$, to be chosen later, let $L = \{j \in [n] : \sigma^2(f_j) \geq \alpha\}$ denote the α-large indices and $S = [n] \setminus L$ denote the small indices. We call a hash function $h \in \mathcal{H}$ (α, β)-*good* if the following two conditions hold for every bin $h^{-1}(j)$ where $j \in [t]$:

1. The bin does not have too many large indices: $|h^{-1}(j) \cap L| \leq k/2$.
2. The small indices in the bin have small total variance:

$$\sum_{\ell \notin L : h(\ell) = j} \sigma^2(f_\ell) \leq \beta.$$

Using standard moment bounds for k-wise independent hash functions one can show that $h \in_u \mathcal{H}$ is (α, β)-good with probability at least $1 - n^{-\Omega(k)}$ for $\alpha = n^{-\Omega(1)}$ and $\beta = n^{-\Omega(1)}$. We defer the proof of the following Lemma to Appendix D.

Lemma 7.2. *Let* $\mathsf{Tvar}(f) \leq n^{1/9}$ *and let* $\mathcal{H} = \{h : [n] \to [t]\}$ *be a* k*-wise independent family of hash functions for* $t = \Theta(\sqrt{n})$*. Then* $h \in \mathcal{H}$ *is* $(n^{-1/3}, n^{-1/36})$*-good with probability* $1 - O(k)^{k/2} n^{-\Omega(k)}$*.*

We next argue that if $h \in \mathcal{H}$ is (α, β)-good then, k-wise independence is sufficient to fool f^j for each $j \in [t]$.

Lemma 7.3. *Let* $h \in \mathcal{H}$ *be* (α, β)*-good, and let* $j \in [t]$*. For* $Z' \sim [m]^n$ k*-wise independent, and* $Z'' \in_u [m]^n$*,*

$$\left|\mathbb{E}[f^j(Z')] - \mathbb{E}[f^j(Z'')]\right| \leq \exp(O(k)) \cdot \beta^{\Omega(k)}.$$

Proof. Fix $j \in [t]$. By relabelling coordinates, let us assume that $h^{-1}(j) = \{1, \ldots, n_j\}$ and $L \cap h^{-1}(j) = \{1, \ldots, r\}$, where $r \leq k/2$. As Z' is k-wise independent, $(Z'_1, \ldots, Z'_r)$ is uniformly distributed over $[m]^r$. We couple Z' and Z'' by taking $Z'_i = Z''_i$ for $i \leq r$. Even after conditioning on these values, $Z'_{r+1}, \ldots, Z'_{n_j}$ are $k/2$-wise independent.

Let $Y_\ell = f_\ell(Z'_\ell)$ for $\ell \in \{r+1, \ldots, n_j\}$. As h is (α, β)-good,

$$\sum_{\ell=r+1}^{n/t} \sigma^2(Y_\ell) \leq \beta.$$

Therefore, by Lemma 4.1,

$$\left|\mathbb{E}\left[\prod_{\ell=r+1}^{n_j} Y_\ell\right] - \prod_{\ell=r+1}^{n_j} \mathbb{E}[Y_\ell]\right| \le \exp(O(k)) \cdot \beta^{\Omega(k)}. \tag{15}$$

But since $Z'_\ell = Z'_\ell$ for $\ell \le r$, we have

$$\begin{aligned}
\mathbb{E}[f^j(Z')] &= \prod_{\ell=1}^{r} \mathbb{E}[f^\ell(Z'_\ell)]\, \mathbb{E}[\prod_{\ell=r+1}^{n} Y_\ell], \\
\mathbb{E}[f^j(Z'')] &= \prod_{\ell=1}^{r} \mathbb{E}[f^\ell(Z'_\ell)] \prod_{\ell=r+1}^{n_j} \mathbb{E}[Y_\ell], \\
\left|\mathbb{E}[f^j(Z')] - \mathbb{E}[f^j(Z'')]\right| &= \left|\prod_{\ell=1}^{r} \mathbb{E}[f^\ell(Z'_\ell)]\, \mathbb{E}[\prod_{\ell=r+1}^{n} Y_\ell] - \prod_{\ell=1}^{r} \mathbb{E}[f^\ell(Z'_\ell)] \prod_{\ell=r+1}^{n_j} \mathbb{E}[Y_\ell]\right| \\
&\le \left|\mathbb{E}[\prod_{\ell=r+1}^{n_j} Y_\ell] - \prod_{\ell=r+1}^{n_j} \mathbb{E}[Y_\ell]\right| && \text{Since } |f^\ell(Z'_\ell)| \le 1 \\
&\le \exp(O(k)) \cdot \beta^{\Omega(k)}. && \text{Equation (15)}
\end{aligned}$$

□

We use these lemmas to prove Theorem 7.1.

Proof of Theorem 7.1. Let $f : [m]^n \to \mathbb{C}_1$ be a Fourier shape with $\mathsf{Tvar}(f) \le n^{1/9}$. Let G_1 be the generator in Equation (13) with parameters as above. We condition on $h \in_u \mathcal{H}$ being $(n^{-1/3}, n^{-1/36})$-good; by Lemma 7.2 this only adds an additional $O(k)^{k/2} n^{-\Omega(k)}$ to the error. We fix such a good hash function h.

Recall that $G_1(h, z_1, \ldots, z_t) = Z$ where $Z^j = G_0(z_j)$ for $j \in [t]$. Since the z_js are independent, so are the Z^j's. Hence,

$$\mathbb{E}[f(G_1(h, z^1, \ldots, z^t))] = \prod_{j=1}^{t} \mathop{\mathbb{E}}_{h} \left[f^j(Z^j)\right].$$

By Lemma 7.3, for $(n^{-1/3}, n^{-1/36})$-good h, if $Y \in_u [m]^n$, then

$$\begin{aligned}
&\left|\prod_{j=1}^{t} \mathbb{E}\left[f^j(Z^j)\right] - \prod_{j=1}^{t} \mathbb{E}\left[f^j(Y^j)\right]\right| \\
&\leq \sum_{r=0}^{t-1} \left|\prod_{j=1}^{r} \mathbb{E}\left[f^j(Y^j)\right] \prod_{j=r+1}^{t} \mathbb{E}\left[f^j(Z^j)\right] - \prod_{j=1}^{r+1} \mathbb{E}\left[f^j(Y^j)\right] \prod_{j=r+2}^{t} \mathbb{E}\left[f^j(Z^j)\right]\right| \\
&\leq \sum_{r=0}^{t-1} \left|\mathbb{E}\left[f^{r+1}(Z^{r+1})\right] - \left[f^{r+1}(Y^{r+1})\right]\right| \\
&\leq \exp(O(k)) \cdot O(tn^{-k/36}).
\end{aligned}$$

Combining the above equations we get that for $Y \in_u [m]^n$,

$$|\mathbb{E}[f(G_1(h, z_1, \ldots, z_t))] - \mathbb{E}[f(Y)]| \leq O(k)^{k/2} n^{-\Omega(k)} + \exp(O(k)) \cdot O(tn^{-k/36}) \leq \delta \quad (16)$$

where the last inequality holds by taking C in Equation (12) to be a sufficiently large constant.

We next derandomize the choice of the z^j's by using a PRG for appropriate Fourier shapes. Let r_0 be the seed-length of the generator G_0 obtained by setting $k = C \log(n/\delta)/(\log n)$ as above, and let c be such that $r_0 \leq c \log(n/\delta)$. Let

$$m' = 2^{r_0} \leq \left(\frac{n}{\delta}\right)^c$$

and identify $[m']$ with $\{0,1\}^{r_0}$. Given a hash function $h \in \mathcal{H}$, let us define $\bar{f}^j : [m'] \to \mathbb{C}_1$ for $j \in [t]$ and $\bar{f} : [m']^n \to \mathbb{C}_1$ as

$$\bar{f}^j(z_j) = f^j(G_0(z_j)), \;\; \bar{f}(z) = \prod_{i=1}^{t} \bar{f}^j(z_j)$$

respectively. Observe that $\bar{f}$ is a Fourier shape, and

$$f(G_1(h, z_1, \ldots, z_t)) = \prod_{j=1}^{t} f^j(G_0(z_j)) = \bar{f}(z).$$

By assumption, we have an explicit generator $\mathcal{G}' : \{0,1\}^{r'} \to [m']^t$ which δ'-fools (m', t)-Fourier shapes. We claim that $\mathcal{G} : \mathcal{H} \times \{0,1\}^{r'} \to [m]^n$ defined as

$$\mathcal{G}_s(h, w) = G_1(h, \mathcal{G}'(w))$$

$(\delta' + \delta)$ fools small-variance (m, n)-Fourier shapes.

Since $\mathcal{G}'$ fools (m', t)-Fourier shapes,

$$|\mathbb{E}[f(\mathcal{G}(h, w))] - \mathbb{E}[f(G_1(h, z_1, \ldots, z_t))]| \leq \delta'.$$

By Equation (16), whenever $\mathsf{Tvar}(f) \leq \log(n/\delta)^C$,

$$\left|\mathbb{E}[f(G_1(h, z_1, \ldots, z_t))] - \mathbb{E}[f(Z')]\right| \leq \delta.$$

Combining these equations,

$$\left|\mathbb{E}[f(\mathcal{G}(h, w))] - \mathbb{E}[f(Z')]\right| \leq \delta' + \delta.$$

The seed-length required for $\mathcal{G}_s$ is $O(\log(n/\delta))$ for h and r' for w. □

8 Putting things together

We put the pieces together and prove our main theorem, Theorem 1.1. We show the following lemma which allows simultaneous reduction in both the alphabet and the dimension, going from fooling (m, n)-Fourier shapes to fooling $(n^2, \lceil\sqrt{n}\rceil)$-Fourier shapes.

Lemma 8.1. *Let $\delta > 0$, $n > \log^C(1/\delta)$ for some sufficiently large constant C, and $t = \lceil\sqrt{n}\rceil$. If there exists an explicit PRG $\mathcal{G}'' : \{0,1\}^{r''} \to [m'']^t$ with seed-length $r'' = r''(n, \delta)$ which δ-fools (m'', t)-Fourier shapes for all $m'' \leq n^2$, then there exists an explicit generator $\mathcal{G} : \{0,1\}^r \to [m]^n$ with seed-length $r = r'' + O(\log(mn/\delta)\log\log(mn))$ which 4δ-fools (m, n)-Fourier shapes.*[3]

Proof. Let r'' be the seed-length required for $\mathcal{G}''$ to have error δ. Let $m' \leq (n/\delta)^c$ be as in the statement of 7.1. Applying Theorem 6.1 to $\mathcal{G}''$, we get a generator $\mathcal{G}'$ with seedlength $r'' + O(\log(n/\delta)\log\log(n/\delta))$ that $\delta' = 2\delta$-fools $(m', \sqrt{n})$ Fourier shapes. Invoking Theorem 7.1 with $\mathcal{G}'$, we get an explicit generator $\mathcal{G}_s : \{0,1\}^{r_s} \to [m]^n$ which 3δ fools (m, n)-Fourier shapes $f : [m]^n \to \mathbb{C}_1$ with $\mathsf{Tvar}(f) \leq n^{1/9}$ and $m \leq n^4$, with seed-length

$$r_s = r'' + O(\log(n/\delta)\log\log(n/\delta)).$$

For $m \leq n^4$, let $\mathcal{G}_\ell : \{0,1\}^{r_\ell} \to [m]^n$ be a generator for large Fourier shapes as in Theorem 5.1, which δ-fools (m, n)-Fourier shapes $f : [m]^n \to \mathbb{C}_1$ with $\mathsf{Tvar}(f) \geq C\log^5(1/\delta)$. Since $m \leq n^4$, this generator requires seed-length

$$r_\ell = O(\log(n/\delta)\log\log(1/\delta)).$$

[3] Comparing this to Theorem 7.1, the main difference is that we do not assume that $\mathsf{Tvar}(f)$ is small. Further, the generator $\mathcal{G}''$ for small dimensions requires $m'' \leq n^2$, and our goal is to fool Fourier shapes in n dimensions with arbitrary alphabet size m.

Define the generator

$$\mathcal{G}_{\ell\oplus s}(w_1, w_2) = \mathcal{G}_\ell(w_\ell) \oplus \mathcal{G}_s(w_s)$$

where the seeds $w_\ell \in \{0,1\}^{r_\ell}$ and $w_s \in \{0,1\}^{r_s}$ are chosen independently and $\oplus$ is interpreted as the sum mod m. Note that the total seed-length is

$$r_\ell + r_s = r'' + O(\log(n/\delta)\log\log(n/\delta)).$$

We now analyze $\mathcal{G}_{\ell\oplus s}$. Let $Y = \mathcal{G}_\ell(w_1)$ and $Z = \mathcal{G}_\ell(w_2)$ and let $X \in_u [m]^n$. Fix an (m,n)-Fourier shape $f : [m]^n \to \mathbb{C}_1$. We consider two cases based on $\mathsf{Tvar}(f)$:

Case 1: $\mathsf{Tvar}(f) \geq C\log(1/\delta)^5$**.** For any $z \in [m]^n$, define a new Fourier shape $f_z(y) = f(y \oplus z)$. Then, for any fixed z, Y δ-fools f_z as $\mathsf{Tvar}(f_z) = \mathsf{Tvar}(f) \geq C\log(1/\delta)^5$. Therefore,

$$|\mathbb{E}[f(Y \oplus Z)] - \mathbb{E}[f(X)]| \leq \mathop{\mathbb{E}}_{Z} |\mathbb{E}[f_Z(Y)] - \mathbb{E}[f(X)]| \leq \delta.$$

Case 2: $\mathsf{Tvar}(f) \leq n^{1/9}$**.** Consider a fixing y of Y and define $f_y(Z) = f(y \oplus Z)$. Then, for any fixed y, Z 3δ-fools f_y as $\mathsf{Tvar}(f_y) \leq n^{1/9}$. Therefore,

$$|\mathbb{E}[f(Y \oplus Z)] - \mathbb{E}[f(X)]| \leq \mathop{\mathbb{E}}_{Y} |\mathbb{E}[f_Y(Z)] - \mathbb{E}[f(X)]| \leq 3\delta.$$

In either case, we have

$$|\mathbb{E}[f(Y \oplus Z)] - \mathbb{E}[f(X)]| \leq 3\delta.$$

Finally, for arbitrary m, by applying Theorem 6.1 to $\mathcal{G}_{\ell\oplus s}$, we get a generator $\mathcal{G} : \{0,1\}^r \to [m]^n$ that 4δ fools (m,n)-Fourier shapes with seed-length

$$O(\log(m/\delta)\log\log(m)) + r_\ell + r_s = r'' + O(\log(nm/\delta)\log\log(nm/\delta)).$$

□

We prove Theorem 1.1 by repeated applications of this lemma.

Proof of Theorem 1.1. Assume that the final error desired is δ'. Let $\delta = \delta'/4\log\log(n)$. Applying Lemma 8.1, by using $O(\log(mn/\delta')\log\log(mn/\delta'))$ random bits we reduce fooling (m,n)-Fourier shapes to fooling $(m', \lceil\sqrt{n}\rceil)$-Fourier shapes for $m' \leq n^2$.

We now apply the lemma $O(\log\log n)$ times to reduce to the case of fooling $(\log^C(1/\delta), \log^C(1/\delta))$-Fourier shapes. This can be done by noting that by Lemma 3.2 it suffices to fool Fourier shapes with $\log(f_i)$ having $O(\log(1/\delta))$ bits of precision. Such Fourier shapes can be computed by width-$O(\log(1/\delta))$ ROBPs, and thus using the generator from Theorem 3.3, we can fool this case with seed length $O(\log(1/\delta)\log\log(1/\delta))$ bits. Since each step requires $O(\log(n/\delta)\log\log(n/\delta)$ random bits, the overall seedlength is bounded by

$$O(\log(mn/\delta)\log\log(mn/\delta) + O(\log(n/\delta)(\log\log(n/\delta))^2).$$

□

9 Applications of PRGs for Fourier shapes

In this Section, we show how Theorem 1.1 implies near optimal PRGs for halfspaces, modular tests and combinatorial shapes. We first prove two technical lemmas relating closeness between Fourier transforms of integer valued random variables to closeness under other metrics. We define the Fourier distance, statistical distance and Kolmogorov distance between two integer-valued random variables respectively as

$$d_{FT}(Z_1, Z_2) = \max_{\alpha \in [0,1]} \left| \mathbb{E}[\exp(2\pi i \alpha Z_1)] - \mathbb{E}[\exp(2\pi i \alpha Z_2)] \right|, \tag{17}$$

$$d_{TV}(Z_1, Z_2) = \frac{1}{2} \sum_{j \in \mathbb{Z}} |\Pr(Z_1 = j) - \Pr(Z_2 = j)|, \tag{18}$$

$$d_K(Z_1, Z_2) = \max_{k \in \mathbb{Z}} (|\Pr(Z_1 \leq k) - \Pr(Z_2 \leq k)|) \tag{19}$$

The first standard claim relates closeness in statistical distance and Fourier distance for bounded integer valued random variables.

Lemma 9.1. *Let Z_1, Z_2 be two integer-valued random variables supported on $[0, N]$. Then,*

$$d_{TV}(Z_1, Z_2) \leq O(\sqrt{N}) \cdot d_{FT}(Z_1, Z_2).$$

Proof. Note that the distribution $Z_1 - Z_2$ is supported on at most $4N+1$ points. Therefore,

$$d_{TV}(Z_1, Z_2) = \|Z_1 - Z_2\|_1 \leq \sqrt{4N+1} \|Z_1 - Z_2\|_2.$$

On the other hand, the Plancherel identity implies that

$$\|Z_1 - Z_2\|_2 \leq d_{FT}(Z_1, Z_2).$$

This completes the proof. □

The second claim relates closeness in Kolmogorov distance to closeness in Fourier distance. The key is that unlike in Lemma 9.1, the dependence on N is logarithmic. This difference is crucial to fooling halfspaces with polynomially small error (since there N can exponential in the dimension n).

Lemma 9.2. *Let Z_1, Z_2 be two integer-valued random variables supported on $[-N, N]$. Then,*

$$d_K(Z_1, Z_2) \leq O(\log(N) \cdot d_{FT}(Z_1, Z_2)).$$

Proof. By definition we have that

$$d_K(Z_1, Z_2) = \max_{-N \leq k \leq N} (|\Pr(Z_1 \leq k) - \Pr(Z_2 \leq k)|).$$

We note that

$$\begin{aligned}\Pr(Z_i \le k) &= \sum_{j=-N}^{k} \Pr(Z_i = j) \\ &= \sum_{j=-N}^{k} \int_0^1 \exp(-2\pi i j\alpha)\, \mathbb{E}[\exp(2\pi i \alpha Z_i)] d\alpha \\ &= \int_0^1 s(k,N,\alpha)\, \mathbb{E}[\exp(2\pi i \alpha Z_i)] d\alpha\end{aligned}$$

where

$$s(k,N,\alpha) = \sum_{j=-N}^{k} \exp(-2\pi i j\alpha).$$

It is clear that $|s(k,N,\alpha)| \le 2N$. Further,

$$|s(k,N,\alpha)| = \left|\frac{\exp(-2\pi i k\alpha)(\exp(2\pi i(N+k+1)\alpha)-1)}{\exp(2\pi i\alpha)-1}\right| \le \frac{1}{|\exp(2\pi i\alpha)-1|} \le O(\frac{1}{[\alpha]})$$

where $[\alpha]$ is the distance between α and the nearest integer. Therefore, we have

$$\begin{aligned}|\Pr(Z_1 \le k) - \Pr(Z_2 \le k)| &\le \int_0^1 |s(k,N,\alpha)|\, |\mathbb{E}[\exp(2\pi i\alpha Z_1)] - \mathbb{E}[\exp(2\pi i\alpha Z_1)]|\, d\alpha \\ &\le \int_0^1 O\left(\min\left(N, \frac{1}{[\alpha]}\right)\right) d_{FT}(Z_1,Z_2) d\alpha \\ &= O(d_{FT}(Z_1,Z_2)) \left(\int_0^{1/N} N d\alpha + \int_{1/N}^{1/2} \frac{d\alpha}{\alpha} + \int_{1/2}^{1-1/N} \frac{d\alpha}{1-\alpha} + \int_{1-1/N}^{1} N d\alpha\right) \\ &= O(d_{FT}(Z_1,Z_2)\log(N)).\end{aligned}$$

□

9.1 Corollaries of the main result

We combine Lemma 9.2 with Theorem 1.1 to derive Corollary 1.2, which gives PRGs for halfspaces with polynomially small error from PRGs for $(2,n)$-Fourier shapes.

Proof of Corollary 1.2. Let $\mathcal{G} : \{0,1\}^r \to \{\pm 1\}^n$ be a PRG which δ-fools $(2,n)$-Fourier shapes (here we identify [2] with $\{\pm 1\}$ arbitrarily). We claim that $\mathcal{G}$ also fools all halfspaces with error at most $\varepsilon = O(n\log(n)\delta)$.

Let $h : \{\pm 1\}^n \to \{\pm 1\}$ be a halfspace given by $h(x) = \mathbb{1}^+(\langle w, x\rangle - \theta)$. It is well known that we can assume the weights and the threshold θ to be integers bounded in the range

$[-N, N]$ for $N = 2^{O(n \log n)}$ (cf. [LC67]). Let $X \in_u \{\pm 1\}^n$ and $Y = \mathcal{G}(y)$ for $y \in_u \{0,1\}^r$ and $Z_1 = \langle w, X \rangle$, $Z_2 = \langle w, Y \rangle$. Note that Z_1, Z_2 are bounded in the range $[-n \cdot N, n \cdot N]$.

We first claim that

$$d_{FT}(Z_1, Z_2) \leq \delta.$$

For $\alpha \in [0,1]$, we define $f_\alpha : \{\pm 1\}^n \to \mathbb{C}_1$ as

$$f_\alpha(x) = \exp(2\pi i \alpha \langle w, x \rangle) = \prod_{j=1}^{n} \exp(2\pi i \alpha w_j x_j) \tag{20}$$

then f_α is a $(2, n)$-Fourier shape. Hence,

$$|\,\mathbb{E}[f_\alpha(X)] - \mathbb{E}[f_\alpha(Y)]| \leq \delta.$$

That $d_{FT}(Z_1, Z_2) \leq \delta$ now follows from the definition of Fourier distance, and the fact that $\mathbb{E}[f_\alpha(X)]$ and $\mathbb{E}[f_\alpha(Y)]$ are the Fourier transforms of X and Y at α respectively.

Therefore, by Lemma 9.2 applied to Z_1, Z_2, $d_K(Z_1, Z_2) \leq O(n \log n)\delta$. Finally, note that

$$|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]| \leq d_K(\langle w, X \rangle, \langle w, Y \rangle) = d_K(Z_1, Z_2) \leq O(n \log n)\delta.$$

The corollary now follows by picking a generator as in Theorem 1.1 for $m = 2$ with error $\delta = \varepsilon/(Cn \log n)$ for sufficiently big C. □

To prove Corollary 1.3, we need the following lemma about generalized halfspaces.

Lemma 9.3. *In Definition 3, we may assume that each $g_i(j)$ is an integer of absolute value $(mn)^{O(mn)}$.*

Proof. Let $g : [m]^n \to \{0,1\}$ be a generalized halfspace where the g_is are arbitrary. Embed $[m]^n$ into $\{0,1\}^{mn}$ by sending each $x_i \in [m]$ to $(y_{i,1}, \ldots, y_{i,m})$ where $y_{i,j} = 1$ if $x_i = j$ and $y_{i,j} = 0$ otherwise. Note that

$$\sum_{i=1}^{n} g_i(x_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} g_i(j) y_{i,j}$$

However, the halfspace

$$\sum_{i=1}^{n} \sum_{j=1}^{m} g_i(j) y_{i,j} \geq \theta$$

over the domain $\{0,1\}^{mn}$ has a representation where the weights $g'_i(j)$ and θ' are integers of size at most $(mn)^{O(mn)}$. Hence we can replace each $g_i(j)$ in the defintion of g with $g'_i(j)$ without changing its value at any point in $[m]^n$. □

We now prove Corollary 1.3 giving PRGs for generalized halfspaces over $[m]^n$.

Proof of Corollary 1.3. Letting $X \in_u [m]^n$ and letting X' be obtained from a PRG for (m,n)-Fourier shapes with error at most ε, we let $Z_1 = \sum_i g_i(X_i)$ and $Z_2 = \sum_i g_i(X'_i)$. By Lemma 9.2 that $d_K(Z_1, Z_2) \leq O(\varepsilon nm \log(nm))$. Picking ε sufficiently small gives our generator for generalized halfspaces. □

Next we use Corollary 1.3 to get PRGs fooling halfspaces under general product distributions. From the definition of generalized halfspaces, it follows that if $\mathcal{D}$ is a discrete product distribution on $\mathbb{R}^n$ where each co-ordinate can be sampled using $\log(m)$ bits, then fooling halfspaces under $\mathcal{D}$ reduces to fooling generalized halfspaces over $[m]^n$ for some suitable choice of g_i. In fact [GOWZ10] showed that fooling such distributions is in fact sufficient to sufficient to fool continuous product distributions with bounded moments. The following is a restatement of [GOWZ10, Lemma 6.1].

Lemma 9.4. *Let X be a product distribution on $\mathbb{R}^n$ such that for all $i \in [n]$,*

$$\mathbb{E}[X_i] = 0, \mathbb{E}[X_i^2] = 1, \mathbb{E}[X_i^4] \leq C.$$

Then there exists a discrete product distribution Y such that for every halfspace h,

$$|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]| \leq \varepsilon.$$

Further, each Y_i can be sampled using $\log(n, 1/\varepsilon, C)$ random bits.

Note that the first and second moment conditions on X can be obtained for any product distribution by an affine transformation. Hence we get Corollary 1.4 from combining Lemma 9.4 with Corollary 1.3. In particular, there exist generators that fool all halfspaces with error ε under the Gaussian distribution with seed-length $r = O(\log(n/\varepsilon)(\log\log(n/\varepsilon))^2)$. This nearly matches the recent result of [KM15] upto a $\log\log$ factor. Further, it is known (see e.g [GOWZ10, Lemma 11.1]) that PRGs for halfspaces under the Gaussian distribution imply PRGs for halfspaces over the sphere.

We next prove Corollary 1.5 which derandomizes the Chernoff bound.

Proof of Corollary 1.5. First note that we can assume without loss of generality that each X_i can be sampled with $r_x = O(\log(mn/\varepsilon))$ bits (by ignoring elements which happen with smaller probability). In particular, let each X_i have the same distribution as $h_i(Z)$ for $Z \in_u [m']$ where $m' = 2^{r_x}$ (here we identify $[m']$ with $\{0,1\}^{r_x}$) and some function $h_i : [m'] \to [m]$. Let $\mathcal{G} : \{0,1\}^r \to [m']^n$ be a PRG which $(\varepsilon/2)$-fools (m', n)-generalized halfspaces. Now, let $Y = (h_1(Z_1), h_2(Z_2), \ldots, h_n(Z_n))$, where $(Z_1, \ldots, Z_n) = \mathcal{G}(w)$ for $w \in_u \{0,1\}^r$.

Note that Y can be sampled with $O(\log(mn/\varepsilon) \cdot (\log\log^2(mn/\varepsilon)))$ random bits. We claim that Y satisfies the required guarantees. To see this, define the generalized halfspaces

$$g_+(z) = \mathbb{1}^+\left(\sum_{i=1}^n g_i(h_i(z_i)) - \theta\right), \quad g_-(z) = \mathbb{1}^+\left(\sum_{i=1}^n -g_i(h_i(z_i)) + \theta\right),$$

where

$$\theta = t + \sum_i \mathbb{E}[g_i(X_i)] = t + \sum_{i=1}^{n} \mathbb{E}[g_i(Y_i)].$$

From Corollary 1.3 it follows that

$$|\mathbb{E}[g_+(X)] - \mathbb{E}[g_+(Y)]| \leq \varepsilon/2, \quad |\mathbb{E}[g_-(X)] - \mathbb{E}[g_-(Y)]| \leq \varepsilon/2.$$

From the Chernoff-Hoeffding bound [Hoe63], we have

$$\mathbb{E}[g_+(X) + g_-(X)] = \Pr\left[\left|\sum_{i=1}^{n} g_i(X_i) - \sum_{i=1}^{n} \mathbb{E}[g_i(X_i)]\right| \geq t\right] \leq 2e^{-t^2/2n}.$$

Hence by the triangle inequality,

$$\Pr\left[\left|\sum_{i=1}^{n} g_i(Y_i) - \sum_{i=1}^{n} \mathbb{E}[g_i(Y_i)]\right| \geq t\right] = \mathbb{E}[g_+(Y) + g_-(Y)] \leq 2e^{-t^2/2n} + \varepsilon.$$

□

We next prove Corollary 1.6 about fooling modular tests.

Proof of Corollary 1.6. Let $\mathcal{G} : \{0,1\}^r \to \{0,1\}^n$ be a PRG which fools $(2,n)$-Fourier shapes with error $\varepsilon/\sqrt{Mn}$. We claim that $\mathcal{G}$ fools modular tests with error at most ε.

Let $g(x) = \mathbb{1}(\sum_i a_i x_i \mod M \in S)$ be a modular test, let $X \in_u \{0,1\}^n$ and $Y = \mathcal{G}(y)$ for $y \in_u \{0,1\}^r$. In order to fools modular tests, it suffices that

$$d_{TV}(\sum_i a_i X_i, \sum_i a_i Y_i) \leq \varepsilon.$$

On the other hand, since both these random variables are bounded in the range $\{0, Mn\}$, by Lemma 9.1

$$d_{TV}\left(\sum_i a_i X_i, \sum_i a_i Y_i\right) \leq \sqrt{Mn} \cdot d_{FT}\left(\sum_i a_i X_i, \sum_i a_i Y_i\right) \leq \varepsilon$$

where the last inequality uses the fact that the Fourier transforms of both random variables are $(2,n)$-Fourier shapes by Equation (20). □

Next we prove Corollary 1.7 giving PRGs from combinatorial shapes.

Proof of Corollary 1.7. Recall that a combinatorial shape $f : [m]^n \to \{0,1\}$ is a function

$$f(x) = h\left(\sum_{i=1}^{n} g_i(x_i)\right)$$

where $g_i : [m] \to \{0,1\}$ and $h : \{0,\ldots,n\} \to \{0,1\}$. Since $\sum_i g_i(x_i) \in \{0,\ldots,n\}$, it suffices to fool the generalized halfspaces

$$f(x) = \sum_i g_i(x_i) - \theta$$

for $\theta \in \{0,\ldots,n\}$ each with error ε/n. Hence the claim follows from Corollary 1.3 about fooling generalized halfspaces. □

References

[ASWZ96] Roy Armoni, Michael E. Saks, Avi Wigderson, and Shiyu Zhou. Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 412–421, 1996.

[BRRY14] Mark Braverman, Anup Rao, Ran Raz, and Amir Yehudayoff. Pseudorandom generators for regular branching programs. *SIAM J. Comput.*, 43(3):973–986, 2014.

[BV10] Joshua Brody and Elad Verbin. The coin problem and pseudorandomness for branching programs. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 30–39, 2010.

[CRSW13] L. Elisa Celis, Omer Reingold, Gil Segev, and Udi Wieder. Balls and bins: Smaller hash families and faster evaluation. *SIAM J. Comput.*, 42(3):1030–1050, 2013.

[De11] Anindya De. Pseudorandomness for permutation and regular branching programs. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity, CCC 2011, San Jose, California, June 8-10, 2011*, pages 221–231, 2011.

[De14] Anindya De. Beyond the central limit theorem: asymptotic expansions and pseudorandomness for combinatorial sums, 2014. ECCC, TR14-125.

[DGJ+09] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS '09)*, 2009.

[DKN10] Ilias Diakonikolas, Daniel Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS '10)*, 2010.

[EGL+98] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Efficient approximation of product distributions. *Random Struct. Algorithms*, 13(1):1–16, 1998.

[GKM14] Parikshit Gopalan, Daniel Kane, and Raghu Meka, 2014. Arxiv: http://arxiv.org/abs/1411.4584.

[GMR+12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 120–129, 2012.

[GMRZ13] Parikshit Gopalan, Raghu Meka, Omer Reingold, and David Zuckerman. Pseudorandom generators for combinatorial shapes. *SIAM J. Comput.*, 42(3):1051–1076, 2013.

[GOWZ10] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *25th Annual IEEE Conference on Computational Complexity*, pages 223–234, 2010.

[GY14] Parikshit Gopalan and Amir Yehudayoff. Inequalities and tail bounds for elementary symmetric polynomials, 2014. no. MSR-TR-2014-131.

[HKM12] Prahladh Harsha, Adam Klivans, and Raghu Meka. An invariance principle for polytopes. *J. ACM*, 59(6):29, 2012.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301), 1963.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 356–364, 1994.

[Kan11a] Daniel M. Kane. k-independent Gaussians fool polynomial threshold functions. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity, CCC 2011, San Jose, California, June 8-10, 2011*, pages 252–261, 2011.

[Kan11b] Daniel M. Kane. A small PRG for polynomial threshold functions of Gaussians. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 257–266, 2011.

[Kan14] Daniel M. Kane. A pseudorandom generator for polynomial threshold functions of Gaussians with subpolynomial seed length. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 217–228, 2014.

[KM15] Pravesh Kothari and Raghu Meka. Almost-optimal pseudorandom generators for spherical caps. In *STOC 2015*, 2015. To appear in STOC 2015.

[KMN11] Daniel M. Kane, Raghu Meka, and Jelani Nelson. Almost optimal explicit johnson-lindenstrauss families. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 14th International Workshop, APPROX 2011, and 15th International Workshop, RANDOM 2011, Princeton, NJ, USA, August 17-19, 2011. Proceedings*, pages 628–639, 2011.

[KNP11] Michal Koucký, Prajakta Nimbhorkar, and Pavel Pudlák. Pseudorandom generators for group products: extended abstract. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 263–272, 2011.

[KRS12] Zohar Shay Karnin, Yuval Rabani, and Amir Shpilka. Explicit dimension reduction and its applications. *SIAM J. Comput.*, 41(1):219–249, 2012.

[LC67] P. M. Lewis and C. L. Coates. *Threshold Logic*. John Wiley, New York, 1967.

[LLSZ97] Nathan Linial, Michael Luby, Michael E. Saks, and David Zuckerman. Efficient construction of a small hitting set for combinatorial rectangles in high dimension. *Combinatorica*, 17(2):215–234, 1997.

[LRTV09] Shachar Lovett, Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Pseudorandom bit generators that fool modular sums. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 12th International Workshop, APPROX 2009, and 13th International Workshop, RANDOM 2009, Berkeley, CA, USA, August 21-23, 2009. Proceedings*, pages 615–630, 2009.

[Lu02] Chi-Jen Lu. Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22(3):417–434, 2002.

[MZ09] Raghu Meka and David Zuckerman. Small-bias spaces for group products. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 12th International Workshop, APPROX 2009, and 13th International Workshop, RANDOM 2009, Berkeley, CA, USA, August 21-23, 2009. Proceedings*, pages 658–672, 2009.

[MZ13] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM J. Comput.*, 42(3):1275–1301, 2013.

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *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 J. on Comput.*, 22(4):838–856, 1993.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. Comput. System Sci.*, 52(1):43–52, 1996.

[Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4), 2008.

[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, May 1-4, 1999, Atlanta, Georgia, USA*, pages 159–168, 1999.

[RS10] Yuval Rabani and Amir Shpilka. Explicit construction of a small epsilon-net for linear threshold functions. *SIAM J. Comput.*, 39(8):3501–3520, 2010.

[RTV06] Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Pseudorandom walks on regular digraphs and the RL vs. L problem. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 457–466, 2006.

[SSS95] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff-hoeffding bounds for applications with limited independence. *SIAM J. Discrete Math.*, 8(2):223–250, 1995.

[SZ99] Michael E. Saks and Shiyu Zhou. $BPSPACE(S) \subseteq DSPACE(S^{3/2})$. *J. Comput. Syst. Sci.*, 58(2):376–403, 1999.

[Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. *Random Struct. Algorithms*, 11(4):345–367, 1997.

A Proofs from Section 3

Proof of Lemma 3.5. Let $I_{i,k}$ be the indicator function of the event that $h(i) = k$. Note that $h(v) = \sum_{i,j,k} I_{i,k} I_{j,k} v_i^2 v_j^2$. Therefore,

$$h(v)^p = \sum_{i_1,\ldots,i_p,j_1,\ldots,j_p} \sum_{k_1,\ldots,k_p} \prod_{t=1}^{p} I_{i_t,k_t} I_{j_t,k_t} \prod_{t=1}^{p} v_{i_t}^2 v_{j_t}^2.$$

Let $R(i_t, j_t, k_t)$ be 0 if for some t, t' $k_t \neq k_t'$ but one of i_t or j_t equals $i_{t'}$ or $j_{t'}$ and otherwise be equal to m^{-T} where T is the number of distinct values taken by i_t or j_t. Notice that by the δ-biasedness of h that

$$\mathbb{E}\left[\prod_{t=1}^{p} I_{i_t,k_t} I_{j_t,k_t}\right] \leq R(i_t, j_t, k_t) + \delta.$$

Combining with the above we find that

$$\begin{aligned}
\mathbb{E}[h(v)^p] &\leq \sum_{i_1,\ldots,i_p,j_1,\ldots,j_p} \sum_{k_1,\ldots,k_p} (R(i_t, j_t, k_t) + \delta) \prod_{t=1}^{p} v_{i_t}^2 v_{j_t}^2 \\
&\leq \sum_{i_1,\ldots,i_p,j_1,\ldots,j_p} \sum_{k_1,\ldots,k_p} R(i_t, j_t, k_t) \prod_{t=1}^{p} v_{i_t}^2 v_{j_t}^2 + \delta m^p \sum_{i_1,\ldots,i_p,j_1,\ldots,j_p} \prod_{t=1}^{p} v_{i_t}^2 v_{j_t}^2 \\
&\leq \sum_{i_1,\ldots,i_p,j_1,\ldots,j_p} \sum_{k_1,\ldots,k_p} R(i_t, j_t, k_t) \prod_{t=1}^{p} v_{i_t}^2 v_{j_t}^2 + \delta m^p \|v\|_2^{4p}.
\end{aligned}$$

Next we consider

$$\sum_{k_1,\ldots,k_p} R(i_t, j_t, k_t)$$

for fixed values of $i_1, \ldots, i_p, j_1, \ldots, j_p$. We claim that it is at most $m^{-S/2}$ where S is again the number of distinct elements of the form i_t or j_t that appear in this way an odd number of times. Letting T be the number of distinct elements of the form i_t or j_t, the expression in question is m^{-T} times the number of choices of k_t so that each value of i_t or j_t appears with only one value of k_t. In other words this is m^{-T} times the number of functions $f : \{i_t, j_t\} \to [m]$ so that $f(i_t) = f(j_t)$ for all t. This last relation splits $\{i_t, j_t\}$ into equivalence classes given by the transitive closure of the operation that $x \sim y$ if $x = i_t$ and $y = j_t$ for some t. We note that any x that appears an odd number of times as an i_t or j_t must be in an equivalence class of size at least 2 because it must appear at least once with some other element. Therefore, the number of equivalence classes, E is at least $T - S/2$. Thus, the sum in question is at most $m^{-T} m^E \leq m^{-S/2}$. Therefore, we have that

$$\mathbb{E}[h(v)^p] \leq (2p)! \sum_{\text{Multisets } M \subset [n], |M| = 2p} m^{-\{\text{Odd}(M)\}/2} \prod_{i \in M} v_i^2 + \delta m^p \|v\|_2^{4p}.$$

Where $\mathrm{Odd}(M)$ is the number of elements occurring in M an odd number of times. This equals

$$\begin{aligned}
\mathbb{E}[h(v)^p] &\leq (2p)! \sum_{k=0}^{p} \sum_{\text{Multisets } M \subset [n], |M|=2p, \mathrm{Odd}(M)=2k} m^{-k} \prod_{i \in M} v_i^2 + \delta m^p \|v\|_2^{4p} \\
&\leq (2p)! \sum_{k=0}^{p} m^{-k} \sum_{i_1,\ldots,i_{2k}} \sum_{j_1,\ldots,j_{p-k}} \prod v_{i_t}^2 \prod v_{j_t}^4 + \delta m^p \|v\|_2^{4p} \\
&= (2p)! \sum_{k=0}^{p} \left(\frac{\|v\|_2^4}{m} \right)^k \|v\|_4^{4(p-k)} + \delta m^p \|v\|_2^{4p} \\
&\leq O(p)^{2p} \left(\frac{\|v\|_2^4}{m} \right)^p + O(p)^{2p} \|v\|_4^{4p} + \delta m^p \|v\|_2^{4p}.
\end{aligned}$$

Note that the second line above comes from taking M to be the multiset

$$\{i_1, i_2, \ldots, i_{2k}, j_1, j_1, j_2, j_2, \ldots, j_{p-k}, j_{p-k}\}.$$

This completes our proof. □

Proof of Lemma 3.6. Let X_i denote the indicator random variable which is 1 if $h(i) = j$ and 0 otherwise. Let $Z = \sum_i v_i X_i$. Now, if h were a truly random hash function, then, by Hoeffding's inequality,

$$\Pr\left[|Z - \|v\|_1 / m| \geq t\right] \leq 2 \exp\left(-t^2 / 2 \sum_i v_i^2 \right).$$

Therefore, for a truly random hash function and even integer $p \geq 2$, $\|Z\|_p = O(\|v\|_2)\sqrt{p}$. Therefore, for a δ-biased hash family, we get $\|Z\|_p^p \leq O(p)^{p/2} \|v\|_2^p + \|v\|_1^p \delta$. Hence, by Markov's inequality, for any $t > 0$,

$$\Pr\left[|Z - \|v\|_1 / m| \geq t\right] \leq \frac{O(p)^{p/2} \|v\|_2^p + \|v\|_1^p \delta}{t^p}.$$

□

B Proofs from Section 4

Proof of Lemma 4.2. First we note that since for any complex random variable, Z, that

$$\mathbb{E}\left[|Z|^k\right] = 2^{O(k)} \mathbb{E}[|\Re(Z)|^k + |\Im(Z)|^k]$$

and $\mathrm{Var}(Z) = \mathrm{Var}(\Re(Z)) + \mathrm{Var}(\Im(Z))$, it suffices to prove our lemma when Z is a real-valued random variable.

We can now compute the expectation of $(\sum_i Z_i)^k$ by expanding out the polynomial in question and computing the expectation of each term individually. In particular, we have that

$$\mathbb{E}\left[\left|\sum_i Z_i\right|^k\right] = \sum_{i_1,\ldots,i_k} \mathbb{E}\left[\prod_{j=1}^k Z_{i_j}\right].$$

Next we group the terms above by the set S of indices that occur as i_j for some j. Thus, we get

$$\sum_{m=1}^{k} \sum_{|S|=m} \sum_{\substack{i_1,\ldots,i_k \in S \\ \{i_j\}=S}} \mathbb{E}\left[\prod_{j=1}^k Z_{i_j}\right].$$

We note that the expectation in question is 0 unless for each $j \in S$, Z_j occurs at least twice in the product. Therefore, the expectation is 0 unless $m \le k/2$ and overall is at most $B^{k-2m}\prod_{j\in S}\mathrm{Var}(Z_j)$. Thus, the expectation in question is at most

$$\sum_{m=1}^{k/2} \sum_{|S|=m} m^k B^{k-2m} \prod_{j\in S} \mathrm{Var}(Z_j).$$

Next, note that by expanding out $(\sum_i \mathrm{Var}(Z_i))^m$ we find that $\sigma^{2m} \ge m! \sum_{|S|=m}\prod_{j\in S}\mathrm{Var}(Z_j)$. Therefore, the expectation in question is at most

$$\begin{aligned}
\sum_{m=1}^{k/2} 2^{O(k)} m^{k-m} B^{k-2m}\sigma^{2m} &\le 2^{O(k)} \sum_{m=0}^{k/2} k^{k-m} B^{k-2m}\sigma^{2m} \\
&\le 2^{O(k)}\left(k^{k/2}\sigma^k + k^k B^k\right) \\
&\le 2^{O(k)}(\sigma\sqrt{k} + Bk)^k,
\end{aligned}$$

as desired. □

C Proofs from Section 5

Proof of Lemma 5.4. First note that $t \in \{0,\ldots,T\}$ satisfying the hypothesis exists since $\|v\|_2^2 \in [1,n]$. For $\ell \in [n]$, let $I(\ell)$ be the indicator random variable which is 1 if $\ell \in B_t$. Since $|B_t| = 2^t$, $\Pr[I(\ell)=1] = 2^t/n$. If we set $V = \|v_t\|^2$,

$$V = \sum_\ell v_\ell^2 I(\ell)$$

$$\mathbb{E}[V] = \|v\|^2 \frac{2^t}{n} \in [1/2, 1].$$

By the pairwise independence of σ,

$$\begin{aligned}\mathbb{E}[V^2] &= \sum_{\ell=1}^{n} v_\ell^4 I(\ell) + \sum_{\ell\neq\ell'=1}^{n} v_\ell^2 v_{\ell'}^2 I(\ell)I(\ell') \\ &\leq \sum_{\ell=1}^{n} v_\ell^4 \frac{2^t}{n} + \sum_{\ell\neq\ell'=1}^{n} v_\ell^2 v_{\ell'}^2 \frac{2^{2t}}{n} \\ &\leq \frac{2^t}{n}\|v\|_4^4 + \mathbb{E}[V]^2.\end{aligned}$$

Therefore,

$$\mathrm{Var}(V) = \mathbb{E}[V^2] - \mathbb{E}[V]^2 \leq \frac{2^t}{n}\|v\|^4 \leq \frac{2^t\|v\|_2^2}{n}\|v\|_\infty^2 \leq \frac{1}{16}$$

Thus, by Chebyshev's inequality,

$$\Pr[|V - \mathbb{E}[V]| > 1/3] \leq 9/16$$

In particular, with probability at least $7/16$, $V = \left\|v^t\right\|_2^2 \in [1/6, 4/3]$. □

Proof of Lemma 5.6. Let $\ell = 2C_1\log(1/\delta)$ and $T = \Theta(\log^5(1/\delta))$ to be chosen later. Let $\mathcal{H} = \{h : [n] \to [T]\}$ to be a δ'-biased family for $\delta' = \exp(-C(\log(1/\delta)))$ for C a sufficiently large constant.

Let $p = c\log(1/\delta)/\log\log(1/\delta))$ for a constant c to be chosen later. Let $v \in [0,1]^n$ with $\|v\|_2^2 \geq C_2\log^5(1/\delta)$ and note that if $\left\|v_{h^{-1}(j)}\right\|_2^2 \geq \|v\|_2^2/\ell$ for some $j \in [T]$, then $h(v) \geq \|v\|_2^4/\ell^2$ (recall the definition of $h(v)$ from Equation (2)). Therefore, by Lemma 3.5 and Markov's inequality, the probability that this happens is at most

$$\begin{aligned}\frac{\mathbb{E}[h(v)^p]\ell^{2p}}{\|v\|_2^{4p}} &\leq \left(\frac{\ell^{2p}}{\|v\|_2^{4p}}\right)\left(O(p)^{2p}\left(\frac{\|v\|_2^4}{T}\right)^p + O(p)^{2p}\|v\|_4^{4p} + T^p\|v\|_2^{4p}\delta'\right) \\ &\leq O\left(\frac{p^2\ell^2}{T}\right)^p + O\left(\frac{p^2\ell^2}{\|v\|_2^2}\right)^p + T^p\ell^{2p}\delta' \\ &\leq O(\log(1/\delta))^{-p} + O(\log(1/\delta))^{7p}\delta' \\ &< \delta,\end{aligned}$$

for a suitable choice of the constant c and $\delta' = \exp(-C\log(1/\delta))$.

Now suppose that $\left\|v_{h^{-1}(j)}\right\|_2^2 < \|v\|_2^2/\ell$ for all $j \in [T]$. Let $I = \{j : \left\|v_{h^{-1}(j)}\right\|^2 \geq \|v\|_2^2/2T\}$. Then,

$$\|v\|_2^2 \leq |I| \cdot (\|v\|_2^2/\ell) + T \cdot \|v\|_2^2/(2T).$$

Therefore, we must have $|I| \geq \ell/2$. This proves the claim. □

D Proofs from Section 7

Proof. Let $\alpha = n^{-1/3}, \beta = n^{-1/36}$.

Note that $|L| \leq \mathsf{Tvar}(f)/\alpha \leq n^{2/9}$. Since $h \in_u \mathcal{H}$ is k-wise independent, for any index $j \in [t]$,

$$\Pr[|L \cap h^{-1}(j)| > k/2] \leq \binom{|L|}{k/2}\left(\frac{1}{t}\right)^{k/2} \leq \left(\frac{\mathsf{Tvar}(f)}{\alpha}\right)^{k/2}\left(\frac{1}{t}\right)^{k/2} \leq O\left(n^{-5/18}\right)^{k/2}. \tag{21}$$

Define $v \in \mathbb{R}^n$ by $v_j = \sigma^2(f_j)$ if $j \in S$ and 0 otherwise. Now,

$$\|v\|_2^2 = \sum_{j\in S}\sigma^4(f_j) \leq \max_{j\in S}\sigma^2(f_j)\sum_{j\in S}\sigma^2(f_j) \leq \mathsf{Tvar}(f)\alpha.$$

By Lemma 3.6 applied to v, we get that for any $j \in [t]$,

$$\Pr_{h\in\mathcal{H}}\left[\sum_{\ell\in S:h(\ell)=j}\sigma^2(f_\ell) \geq \frac{\mathsf{Tvar}(f)}{t} + \frac{(\mathsf{Tvar}(f))^{1/2}\alpha^{1/4}}{2} \leq \beta\right] \leq O(k)^{k/2}\alpha^{k/4} = O(k)^{k/2}n^{-\Omega(k)}. \tag{22}$$

This completes the proof. □