Ark. Mat., **51** (2013), 29–52
DOI: 10.1007/s11512-011-0145-5

Noise correlation bounds for uniform low degree functions

Per Austrin and Elchanan Mossel

Abstract. We study correlation bounds under pairwise independent distributions for functions with no large Fourier coefficients. Functions in which all Fourier coefficients are bounded by δ are called δ-*uniform*. The search for such bounds is motivated by their potential applicability to hardness of approximation, derandomization, and additive combinatorics.

In our main result we show that $\mathbb{E}[f_1(X_1^1, ..., X_1^n) ... f_k(X_k^1, ..., X_k^n)]$ is close to 0 under the following assumptions:

– the vectors $\{(X_1^j, ..., X_k^j): 1 \le j \le n\}$ are independent identically distributed, and for each j the vector $(X_1^j, ..., X_k^j)$ has a pairwise independent distribution;

– the functions f_i are uniform;

– the functions f_i are of low degree.

We compare our result with recent results by the second author for low influence functions and to recent results in additive combinatorics using the Gowers norm. Our proofs extend some techniques from the theory of hypercontractivity to a multilinear setup.

1. Introduction

1.1. Functionals of pairwise independent distributions

In recent years there has been an extensive study of conditions satisfied by functions $f_1, ..., f_k$ which guarantee that

$$\mathbb{E}[f_1(X_1) ... f_k(X_k)] \approx \prod_{i=1}^{k} \mathbb{E}[f_i(X_i)], \tag{1}$$

Work done while the first author was at the Royal Institute of Technology, funded by ERC Advanced investigator grant 226203 and a grant from the Mittag-Leffler Institute. Second author supported by BSF grant 2004105, NSF CAREER award DMS 0548249, DOD ONR grant N0014-07-1-05-06 and ISF grant 1300/08.

$$\begin{pmatrix} X_1^1 & \dots & X_1^j & \dots & X_1^n \\ \vdots & & \vdots & & \vdots \\ X_i^1 & \dots & X_i^j & \dots & X_i^n \\ \vdots & & \vdots & & \vdots \\ X_k^1 & \dots & X_k^j & \dots & X_k^n \end{pmatrix}$$

Figure 1. The random matrix X. The columns $X^1, ..., X^n$ are independent identically distributed random vectors, and the distribution of the column $X^j = (X_1^j, ..., X_k^j)^T$ is pairwise independent, for each $j \in [n]$.

for certain probability distributions over $(X_1, ..., X_k)$ which are pairwise independent. Recall that the random vector $(X_1, ..., X_k)$ is *pairwise independent* if for all $1 \le i < j \le k$ the random variables X_i and X_j are independent. In the current paper we will consider this problem under the additional assumption that for all $1 \le i \le k$ the random variable X_i is an n-dimensional vector $X_i = (X_i^1, ..., X_i^n) \in \Omega^n$ and that $(X_1^j, ..., X_k^j)$ follow the same (pairwise independent) distribution μ over Ω^k, independently for each $1 \le j \le n$ (see Figure 1). We further assume that Ω is a finite probability space.

In most of the paper we focus on the related problem of finding conditions which guarantee that

$$\mathbb{E}\left[\prod_{i=1}^{k} f_i(X_i)\right] \approx 0, \tag{2}$$

which can be thought of as the special case of (1) when $\prod_{i=1}^k \mathbb{E}[f_i] \approx 0$. In many cases a general bound of type (1) is straightforward to obtain from (2).

The basic example of a condition implying (2) is one of the constituents of the proof of Roth's theorem [16].([1]) Indeed, it is not too hard to show that

$$\left|\mathbb{E}\left[\prod_{i=1}^{3} f_i(X_i)\right]\right| \le \min_{1 \le i \le 3} \|\hat{f}_i\|_\infty, \tag{3}$$

where X_1, X_2 and X_3 are pairwise independent, f_1, f_2 and f_3 are any functions with $\max_{1 \le i \le 3} \|f_i\|_2 \le 1$, and $\hat{f}_1$, $\hat{f}_2$ and $\hat{f}_3$ are their Fourier transforms.

([1]) Roth's original argument considers (X_1, X_2, X_3) which is a uniformly chosen 3-term arithmetic progression in $\mathbb{Z}_p$ but the argument extends immediately to the setup considered here.

Gowers ([5], Theorem 3.2) generalized (3) and showed that:

$$\left|\mathbb{E}\left[\prod_{i=1}^{k} f_i(X_i)\right]\right| \le \min_{1\le i\le k} \|f_i\|_{U^{k-1}}, \tag{4}$$

where

– $X_1, ..., X_k$ is a uniformly chosen k-term arithmetic progression in $\mathbb{Z}_p^n$;

– the functions f_i are all bounded by 1;

– $\|f\|_{U^d}$ is the dth Gowers norm of f (see Definition 2.7).

Note that the uniform distribution over arithmetic progressions $X_1, ..., X_k$ of length $3\le k\le p$ defines a pairwise independent distribution in $(\mathbb{Z}_p^n)^k$. See also [6] and [7] where more general results are obtained for other pairwise independent distributions which are defined by linear equations.

Apart for the additive context, expressions of the form $\prod_{i=1}^{k} f_i(X_i)$ often appear in the study of hardness of approximation in computer science. In this context, a natural condition is that the functions $f_1, ..., f_k$ all have *low influences*. For example, recent results of Samorodnitsky and Trevisan ([17], Lemma 8) show how to utilize the Gowers norms in order to show that (here, $\mathrm{Inf}_j(f_i)$ is the influence of X_i^j on f_i, see e.g. [17] for the exact definition):

$$\left|\mathbb{E}\left[\prod_{i=1}^{k} f_i(X_i)\right]\right| \le O\left(\sqrt{\max_{1\le i\le k}\max_{1\le j\le n} \mathrm{Inf}_j(f_i)}\right) \tag{5}$$

provided that

– $k=2^d$ and $X_1, ..., X_k$ are the elements of a uniformly chosen d-dimensional subspace of $\mathbb{Z}_2^n$;

– the functions f_i are all bounded by 1, and at least one of them has $\mathbb{E}[f_i]=0$.

As a special case this result gives a so-called "inverse theorem" for the dth Gowers norm showing that any function with large dth Gowers norm must have an influential variable. The result also allowed the authors to obtain computational inapproximability results for certain constraint satisfaction problems, assuming the so-called unique games conjecture [8]. The results of [17] include a more general statement which applies in any product group.

A more recent result of the second author ([11], Theorem 1.14) derives a bound similar to (5) by showing that

$$\left|\mathbb{E}\left[\prod_{i=1}^{k} f_i(X_i)\right]\right| \le \Psi^{-1}\left(\max_{1\le i\le k}\max_{1\le j\le n} \mathrm{Inf}_j(f_i)\right), \tag{6}$$

where $\Psi(\varepsilon)=\varepsilon^{O(\log(1/\varepsilon)/\varepsilon)}$, provided that the following holds:

– The distribution μ of $(X_1^j, ..., X_k^j)$ is any *connected* pairwise independent distribution. This means that for all x and y in the support of the distribution, there exists a path from x to y in the support that is obtained by flipping one coordinate at a time;

– The functions f_i are all bounded by 1 and at least one of them has $\mathbb{E}[f_i]=0$.

The proof of (6) is based on showing that if all functions f_i are of degree at most d then ([11], Theorem 4.1)

$$\left|\mathbb{E}\left[\prod_{i=1}^{k} f_i(X_i)\right]\right| \leq C^d \sqrt{\max_{1\leq i\leq k} \max_{1\leq j\leq n} \mathrm{Inf}_j(f_i)} \tag{7}$$

for some absolute constant C provided that

– the distribution μ of $(X_1^j, ..., X_k^j)$ is *any* pairwise independent distribution;

– the functions f_i satisfy $\|f_i\|_2 \leq 1$ for all i and at least one of them has $\mathbb{E}[f_i]=0$.

The bound (6) is then derived from (7) by applying certain truncation arguments. These results of [11] do not use any algebraic symmetries nor the Gowers norm. Rather, they were based on extending Lindeberg's proof of the central limit theorem [10] using invariance and generalizing recent works [12] and [15].

We note that the results of [11] later implied results by the authors of this paper [1] which gave stronger and more general inapproximability results than those obtained in [17]. It was further noted in [11] that many of the additive applications involve pairwise independent distributions.

1.2. Our results

Motivated by these lines of work in additive number theory and hardness of approximation we wish to obtain weaker conditions that guarantee (2). Indeed our main result, Theorem 3.2, shows that

$$\left|\mathbb{E}\left[\prod_{i=1}^{k} f_i(X_i)\right]\right| \leq C^d \|\hat{f}_1\|_\infty \prod_{i=2}^{k} \|f_i\|_2 \tag{8}$$

for some constant C which only depends on the pairwise independent distribution μ, where

– $\|\hat{f}_1\|_\infty = \max |\hat{f}_1(\sigma)|$ denotes the size of the largest Fourier coefficient of f_1;

– $X_1, ..., X_k$ are pairwise independent as in Figure 1;

– the functions f_i are of Fourier degree at most d; in other words, all of their Fourier coefficients at levels above d are 0.

We also give some basic extensions of this. As a first simple corollary we give, in Corollary 3.6, a result of type (1) with similar error bounds as our main theorem. Elaborating on this extension we show in Corollary 3.8 that in the case when (1) does not hold, one can find three Fourier coefficients $\hat{f}_{i_1}(\sigma_1)$, $\hat{f}_{i_2}(\sigma_2)$ and $\hat{f}_{i_3}(\sigma_3)$ which are all of non-negligible magnitude, and which "intersect" in the sense that σ_1, σ_2 and σ_3 share some variable $j \in [n]$. Results of this type are often useful in applications to hardness of approximation.

We note that the conditions on the underlying distribution and uniformity are very weak while the condition on the Fourier degree of the function is very strong. By a simple application of Hölder's inequality, we will see in Proposition 3.9 that the results extend to functions which are "almost low-degree" in the sense that the high-degree parts have small ℓ_k norm.

As mentioned above, the proofs of [11] work by first establishing the result (7) for arbitrary low-degree polynomials and then performing a truncation argument, giving (6) where the degree requirements have been traded for an additional requirement on the pairwise independent distribution (and a requirement that the functions are bounded). Hence, the work presented in this paper may be viewed as an important step in establishing similar results for a wider family of functions. Note that our result (8) is strictly stronger than (7) as the bound is stated in terms of the largest Fourier coefficient instead of the largest influence (and that it suffices that only one of the functions has small coefficients, as opposed to (7) where all the functions are required to have small influences).

A very natural question to ask is to what extent the (rather severe) degree restriction can be relaxed. Unfortunately, this restriction cannot be removed completely, since for the pairwise independent relation corresponding to the Gowers norm, it is known, using examples due to Gowers [4] and Furstenberg and Weiss [3], that there are functions with large U^3 norm but no large Fourier coefficients. However, it is quite possible that the degree restriction can be removed provided one is willing to require a bit more of the pairwise independent distribution. In particular, if one as in (6) requires that μ is *connected*, the counterexample given by the Gowers norm is excluded. Such a restriction, while generally too strong in the additive combinatorics settings, is often quite natural in applications to hardness of approximation and social choice.

1.3. Applications

The applications we present mostly concern functions of low Fourier degree with no large Fourier coefficients. We show that such functions cannot "distinguish" between truly independent distributions and pairwise independent product distributions. In particular we show that such functions defined over $\mathbb{Z}_p^n$ always have small Gowers norm. This implies that for functions of low Fourier degree all of the U^k norms are equivalent for $k \geq 2$. Moreover, such functions cannot distinguish the uniform distribution over arithmetic progressions from the uniform distributions over the product space.

1.4. Proof idea

The proof of (8) is based on induction on the degree and the number of variables. In a way it is similar to inductive proofs for deriving hypercontractive estimates for polynomials of random variables, see, e.g., [12]. Naturally the setup is different as each polynomial is applied on different random variables. The pairwise independence property is crucial in the proof as it shows that certain second order terms vanish.

1.5. Paper structure

In Section 2 we recall some background in Fourier analysis and noise correlation. In Section 3 we derive the main result and some corollaries. In Section 4 we derive some applications of the main result. In Section 5 we discuss potential extensions of the main result.

2. Preliminaries

2.1. Notation

Let Ω be a finite set and let μ be a probability distribution on Ω. The following notation will be used throughout the paper.

– $(\Omega^n, \mu^{\otimes n})$ denotes the product space $\Omega \times \ldots \times \Omega$, endowed with the product distribution.

– $\alpha(\mu) := \min\{\mu(x) : x \in \Omega \text{ and } \mu(x) > 0\}$ denotes the minimum non-zero probability of any atom in Ω under the distribution μ.

– $L^2(\Omega, \mu)$ denotes the space of functions from Ω to $\mathbb{C}$. We define the inner product on $L^2(\Omega, \mu)$ by $\langle f, g\rangle := \mathbb{E}_{x \in (\Omega,\mu)}[f(x)\overline{g(x)}]$, and the ℓ_p norm by $\|f\|_p := \mathbb{E}_{x \in (\Omega,\mu)}[|f|^p]^{1/p}$.

For a probability distribution μ on $\Omega_1 \times \ldots \times \Omega_k$ (not necessarily a product distribution) and $i \in [k]$, we use μ_i to denote the marginal distribution on Ω_i. Such a distribution μ is said to be pairwise independent if for every $1 \leq i < j \leq k$, every $a \in \Omega_i$

and every $b \in \Omega_j$ it is true that

$$\Pr_{x \in (\Omega_1 \times \ldots \times \Omega_k, \mu)}[x_i = a \text{ and } x_j = b] = \mu_i(a)\mu_j(b).$$

2.2. Fourier decomposition

In this subsection we recall some background in Fourier analysis that will be used in the paper.

Let q be a positive integer (not necessarily a prime power), and let (Ω, μ) be a finite probability space with $|\Omega| = q$, which is non-degenerate in the sense that $\mu(x) > 0$ for every $x \in \Omega$. Let $\chi_0, ..., \chi_{q-1} \colon \Omega \to \mathbb{C}$ be an orthonormal basis for the space $L^2(\Omega, \mu)$ with respect to the scalar product $\langle \cdot, \cdot \rangle$. Furthermore, we require that this basis has the property that $\chi_0 = \mathbf{1}$, where $\mathbf{1}$ is the function which is identically 1 on Ω.

We remark that since the choice of basis is essentially arbitrary, one can take $\chi_0, ..., \chi_{q-1}$ to be an $\mathbb{R}$-valued basis rather than a $\mathbb{C}$-valued one (which can be desirable in the case when one works exclusively with $\mathbb{R}$-valued functions). The only place in the paper where this distinction makes a difference is the final part of Theorem 3.2, where this is stated explicitly.

In the complex-valued case when μ is the uniform distribution we can take the standard Fourier basis $\chi_y(x) = \exp(2\pi i x y / q)$, where we identify Ω with $\mathbb{Z}_q$ in some canonical way and i here denotes the imaginary unit.

For $\sigma \in \mathbb{Z}_q^n$, define $\chi_\sigma \colon \Omega^n \to \mathbb{C}$ as $\bigotimes_{i \in [n]} \chi_{\sigma_i}$, i.e.,

$$\chi_\sigma(x_1, ..., x_n) = \prod_{i \in [n]} \chi_{\sigma_i}(x_i).$$

It is well-known and easy to check that the functions $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^n}$ form an orthonormal basis for the product space $L^2(\Omega^n, \mu^{\otimes n})$. Thus, every function $f \in L^2(\Omega^n, \mu^{\otimes n})$ can be written as

$$f(x) = \sum_{\sigma \in \mathbb{Z}_q^n} \hat{f}(\sigma) \chi_\sigma(x),$$

where $\hat{f} \colon \mathbb{Z}_q^n \to \mathbb{C}$ is defined by $\hat{f}(\sigma) = \langle f, \chi_\sigma \rangle$. The most basic properties of $\hat{f}$ are summarized by Fact 2.1, which is an immediate consequence of the orthonormality of $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^n}$.

Fact 2.1. *We have*

$$\mathbb{E}[fg] = \sum_{\sigma \in \mathbb{Z}_q^n} \hat{f}(\sigma)\hat{g}(\sigma), \quad \mathbb{E}[f] = \hat{f}(\mathbf{0}) \quad \textit{and} \quad \mathrm{Var}[f] = \sum_{\sigma \neq \mathbf{0}} \hat{f}(\sigma)^2.$$

We refer to the transform $f \mapsto \hat{f}$ as the *Fourier transform*, and $\hat{f}$ as the *Fourier coefficients* of f. We remark that the article "the" is somewhat inappropriate, since the transform and coefficients in general depend on the choice of basis $\{\chi_i\}_{i \in \mathbb{Z}_q}$. However, we will always be working with some fixed (albeit arbitrary) basis, and hence there should be no ambiguity in referring to the Fourier transform as if it were unique. Furthermore, most of the important properties of $\hat{f}$ are actually basis-independent. In particular Definitions 2.2–2.4 and Fact 2.5 do not depend on the choice of Fourier basis.

Before proceeding, let us introduce some useful notation in relation to the Fourier transform.

Definition 2.2. A *multi-index* is a vector $\sigma \in \mathbb{Z}_q^n$, for some q and n. The *support* of a multi-index σ is $S(\sigma) = \{i : \sigma_i > 0\} \subseteq [n]$. We extend notation defined for $S(\sigma)$ to σ in the natural way, and write e.g. $|\sigma|$ instead of $|S(\sigma)|$, $i \in \sigma$ instead of $i \in S(\sigma)$, and so on.

Definition 2.3. The (*Fourier*) *degree* $\deg(f)$ of $f \in L^2(\Omega^n, \mu^{\otimes n})$ is the infimum of all $d \in \mathbb{Z}$ such that $\hat{f}(\sigma) = 0$ for all σ with $|\sigma| > d$.

The degree of f is one of its most important properties. In general, the smaller $\deg(f)$ is, the more "nicely behaved" f is. When $\deg(f) \le d$, we will refer to f as a *degree-d polynomial* in $L^2(\Omega^n, \mu^{\otimes n})$.

Definition 2.4. For $f\colon \Omega^n \to \mathbb{C}$ and $d \in \mathbb{Z}$, the function $f^{\le d}\colon \Omega^n \to \mathbb{C}$ is defined by

$$f^{\le d} = \sum_{|\sigma| \le d} \hat{f}(\sigma) \chi_\sigma.$$

We define $f^{<d}$, $f^{=d}$, $f^{>d}$ and $f^{\ge d}$ analogously.

Another fact which is sometimes useful is the following trivial bound on the ℓ_∞ norm of χ_σ (recall that $\alpha(\mu)$ is the minimum non-zero probability of any atom in μ).

Fact 2.5. *Let $(\Omega^n, \mu^{\otimes n})$ be a product space with Fourier basis $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^n}$. Then for any $\sigma \in \mathbb{Z}_q^n$,*

$$\|\chi_\sigma\|_\infty \le \alpha(\mu)^{-|\sigma|/2}.$$

2.3. Noise correlation

In this section we introduce the notion of a noisy inner product and noise correlation.

Various special cases of noise correlation have been the focus of much work, as we discuss below. Informally, the noise correlation between two functions f and g measure how much $f(x)$ and $g(y)$ correlate on random inputs x and y which are correlated. We remark that the name “noise correlation” is a slight misnomer and that “correlation under noise” would be a more descriptive name—we are not looking at how well a random variable correlates with noise, but rather how well a collection of random variables correlate with each other in the presence of noise.

Definition 2.6. Let (Ω, μ) be a product space with $\Omega = \Omega_1 \times \ldots \times \Omega_k$, and let $f_1, \ldots, f_k$ be functions such that $f_i \in L^2(\Omega_i^n, \mu_i^{\otimes n})$. The *noisy inner product* of $f_1, \ldots, f_k$ with respect to μ is

$$\langle f_1, f_2, \ldots, f_k \rangle_\mu = \mathbb{E}\left[\prod_{i=1}^k f_i\right].$$

The *noise correlation* of $f_1, \ldots, f_k$ with respect to μ is

$$\langle f_1, f_2, \ldots, f_k \rangle_\mu - \prod_{i=1}^k \mathbb{E}[f_i]$$

As it can take some time to get used to Definition 2.6, let us write out the product $\langle f_1, \ldots, f_k \rangle_\mu$ more explicitly. Let $f_i \colon \Omega_i^n \to \mathbb{C}$ be functions on the product space Ω_i^n, and let μ be some probability distribution on $\Omega = \Omega_1 \times \ldots \times \Omega_k$. Then,

$$\langle f_1, \ldots, f_k \rangle_\mu = \mathbb{E}_X\left[\prod_{i=1}^k f_i(X_i)\right],$$

where X is a $k \times n$ random matrix such that each column of X is a sample from (Ω, μ), independently of the other columns, and X_i refers to the ith row of X.

The notation $\langle f_1, \ldots, f_k \rangle_\mu$ is a new notation for quantities studied before in e.g. [11], its applications [1] and [14] and additive number theory. The focus of the current paper is where $X_1, \ldots, X_k$ are pairwise independent, though noisy inner products are of much interest also in cases for non-pairwise independent distributions including in percolation, theoretical computer science and social choice, see e.g. [2], [9], [12] and [13].

2.3.1. The Gowers norm

An instance of the noisy inner product which has been the focus of much attention in recent years is the Gowers norm, which we will now define. Let p be a prime. For a function $f\colon \mathbb{Z}_p^n \to \mathbb{C}$ and a "direction" $Y \in \mathbb{Z}_p^n$, the "derivative" of f in direction Y, $f_Y\colon \mathbb{Z}_p^n \to \mathbb{C}$ is defined by $f_Y(X) = f(X+Y)\overline{f(X)}$. Repeating, we define $f_{Y_1,\ldots,Y_d}(X) = (f_{Y_1,\ldots,Y_{d-1}})_{Y_d}(X) = \prod_{S\subseteq[d]} \mathcal{C}^{|S|+1} f\big(X + \sum_{i\notin S} Y_i\big)$, where $\mathcal{C}$ denotes the complex conjugation operator.

Definition 2.7. Let $f\colon \mathbb{Z}_p^n \to \mathbb{C}$. The *d-th Gowers norm* of f, denoted $\|f\|_{U^d}$, is defined by

$$\|f\|_{U^d}^{2^d} = \mathbb{E}[f_{Y_1,\ldots,Y_d}(X)],$$

where the expected value is over a random $X \in \mathbb{Z}_p^n$ and d random directions $Y_1, \ldots, Y_d$.

This norm was introduced by Gowers [5] in a Fourier-analytic proof of Szemerédi's theorem [18] and has since been used extensively in additive number theory. The Gowers norm can be written as a noisy inner product. Indeed, we can write

$$\|f\|_{U^d}^{2^d} = \mathbb{E}\Bigg[\prod_{S\subseteq[d]} g_S(X_S)\Bigg] = \langle g_\varnothing, \ldots, g_{[d]}\rangle_\mu,$$

where we define $g_S\colon\mathbb{Z}_p^n \to \mathbb{C}$ by $g_S(X) = \mathcal{C}^{|S|+1} f(X)$, and the collection $\{X_S\}_{S\subseteq[d]}$ of random variables is defined by $X_S = X + \sum_{i\notin S} Y_i$ for a uniformly random $X \in \mathbb{Z}_p^n$ and independent uniformly random directions $Y_1, \ldots, Y_d \in \mathbb{Z}_p^n$.

2.3.2. Noisy inner products under pairwise independence

This paper focuses on noisy inner products under pairwise independent distributions. The interest in this special case comes from applications in computer science and additive number theory. We briefly mention a few of these applications.

– In computer science there is interest in pairwise independent distributions in hardness of approximation, in particular those of small support. See [1] where the results of [11] were used to derive hardness results based on pairwise independence.

– As mentioned above, the Gowers norm and the Gowers inner-product are both noisy inner products. Note that the collections of vectors $X + \sum_{i\in S} X_i : S \subseteq [d]$ is pairwise (in fact 3-wise as long as $d \geq 2$) independent.

– Another noisy inner product that is closely related to additive applications is obtained by considering arithmetic progressions. For concreteness consider again the case, where all the functions are of $\mathbb{Z}_p^n \to \{0,1\}$ and let $k<p$. Given k such functions $f_1, ..., f_k$ we let

$$\langle f_1, ..., f_k \rangle_\mu = \mathbb{E}\left[\prod_{i=1}^{k} f_i(iX+Y)\right],$$

where X and Y are independent and uniformly chosen in $\mathbb{Z}_p^n$ (note that $iX+Y$ and $jX+Y$ are independent for $1 \le i < j \le k$). If A is an indicator of a set then the number of k-term progressions in A is in fact

$$p^{2n} \langle A, A, ..., A \rangle_\mu.$$

3. Main theorem

In this section, we state and prove our main theorem. First we define the parameter which controls how good bounds we get.

Definition 3.1. Let $f_1, ..., f_k$ be a collection of functions. By $\deg_{-2}(f_1, ..., f_k)$ we denote the sum of the $k-2$ smallest degrees of $f_1, ..., f_k$.

We can now state the main theorem.

Theorem 3.2. *Let (Ω, μ) be a pairwise independent product space with $\Omega = \Omega_1 \times ... \times \Omega_k$. Then there is a constant C depending only on μ such that the following holds*:

Let $f_1, ..., f_k$ be functions $f_i \in L^2(\Omega_i^n, \mu_i^{\otimes n})$. Denote by $\delta := \max_{\sigma \in \mathbb{Z}_q^n} |\hat{f}_1(\sigma)|$ the size of the largest Fourier coefficient of f_1, and let $D := \deg_{-2}(f_1, ..., f_k)$ denote the sum of the $k-2$ smallest degrees of $f_1, ..., f_k$. Then,

$$|\langle f_1, ..., f_k \rangle_\mu| \le C^D \delta \prod_{i=2}^{k} \|f_i\|_2.$$

Furthermore, one can always take $C = \left(k\sqrt{(q-1)/\alpha}\right)^3$, where $\alpha = \min_i \alpha(\mu_i)$. If μ is balanced, i.e., if all marginals μ_i are uniform, then there is a choice of complex Fourier basis such that one can take $C = \left(k\sqrt{q-1}\right)^3$.

We remark that, while Theorem 3.2 is very limited because of its requirement on the degrees of the f_i's, the lack of any other assumptions is nice. In particular, we do not need to assume that the f_i's are bounded, nor do we need any assumptions on μ beyond the pairwise independence condition.

Proof. We prove this by induction over n. If $n=0$, the statement is easily verified (either $D=-\infty$, or $D=0$, depending on whether one of the functions is 0 or not).([2])

Write $f_i=g_i+h_i$, where

$$g_i=\sum_{1\notin\sigma}\hat{f}(\sigma)\chi_\sigma \quad \text{and} \quad h_i=\sum_{1\in\sigma}\hat{f}(\sigma)\chi_\sigma,$$

i.e., h_i is the part of f_i which depends on X^1 (the first column of X), and g_i is the part which does not depend on X^1. Then

$$\langle f_1,...,f_k\rangle_\mu=\mathbb{E}_X\left[\prod_{i=1}^k f_i(X_i)\right]=\sum_{T\subseteq[k]}\mathbb{E}_X\left[\prod_{i\notin T}g_i(X_i)\prod_{i\in T}h_i(X_i)\right].$$

For $T\subseteq[k]$, define

$$E(T)=\mathbb{E}_X\left[\prod_{i\notin T}g_i(X_i)\prod_{i\in T}h_i(X_i)\right].$$

The key ingredient will be the following lemma, bounding $|E(T)|$.

Lemma 3.3. *Let* $\varnothing\subseteq T\subseteq[k]$. *Then*

– *if* $T=\varnothing$, *we have*

$$|E(T)|\le C^D\delta\prod_{i=2}^k\|g_i\|_2;$$

– *if* $1\le|T|\le2$, *we have*

$$E(T)=0;$$

– *if* $|T|\ge3$, *we have*

$$|E(T)|\le C^{D+2}\left(\frac{\sqrt{(q-1)/\alpha}}{C}\right)^{|T|}\delta\prod_{i\notin T\{1\}\neq1}\|g_i\|_2\prod_{i\in T\{1\}\neq1}\|h_i\|_2.$$

Before proving the lemma, let us see how to use it to finish the proof of Theorem 3.2.

([2]) We point out that $f_i\in L^2(\Omega_i^0,\mu_i^{\otimes 0})$ does not formally make sense. However in this case, the appropriate way to view f_i is as an element of $L^2(\Omega_i^N,\mu_i^{\otimes N})$ which only depends on the n first coordinates, for some large value of N. In particular, for the case $n=0$ we have that f_i is a constant.

Write $\|h_i\|_2 = \tau_i \|f_i\|_2$ for some $\tau_i \in [0,1]$, so that $\|g_i\|_2 = \sqrt{1-\tau_i^2}\|f_i\|_2$ (by orthogonality of the Fourier decomposition). By plugging in the different cases of Lemma 3.3, we can then bound $\langle f_1, ..., f_k\rangle_\mu$ by

$$
\begin{aligned}
(9)\quad |\langle f_1, ..., f_k\rangle_\mu| &\le \sum_T |E(T)| \\
&\le C^D \delta \prod_{i=2}^{k} \|g_i\|_2 \\
&\quad + \sum_{|T|\ge 3} C^{D+2} \left(\frac{\sqrt{(q-1)/\alpha}}{C} \right)^{|T|} \delta \prod_{i \notin T\{1\} \neq 1} \|g_i\|_2 \prod_{i \in T\{1\} \neq 1} \|h_i\|_2 \\
&= C^D \delta \prod_{i=2}^{k} \|f_i\|_2 \Bigg(\prod_{i=2}^{k} \sqrt{1-\tau_i^2} + \sum_{|T|\ge 3} C^2 \left(\frac{\sqrt{(q-1)/\alpha}}{C} \right)^{|T|} \\
&\quad \times \prod_{i \notin T\{1\} \neq 1} \sqrt{1-\tau_i^2} \prod_{i \in T\{1\} \neq 1} \tau_i \Bigg).
\end{aligned}
$$

Hence, it suffices to bound the "factor" inside the large parenthesis in (9) by 1 in order to complete the proof of Theorem 3.2.

Let $\tau = \max_{i \ge 2} \tau_i$. Then the factor in (9) can be bound by

$$
(10) \qquad \sqrt{1-\tau^2} + \tau^2 \sum_{i=3}^{k} \binom{k}{i} \left(\frac{\sqrt{(q-1)/\alpha}}{C^{1/3}} \right)^i,
$$

where in the sum the value of i corresponds to the size of the set T and we assumed that $C > 1$ and then used that $C^{2-i} \le C^{-i/3}$ for $i \ge 3$. To bound (10), we use the following simple lemma.

Lemma 3.4. *For every $k \ge 3$,*

$$
\sum_{i=3}^{k} \binom{k}{i} \frac{1}{k^i} \le \frac{1}{2}.
$$

Proof. Since $\binom{k}{i} \le k^i / i!$ we have

$$
\sum_{i=3}^{k} \binom{k}{i} \frac{1}{k^i} \le \sum_{i=3}^{k} \frac{1}{i!} \le e - \frac{5}{2} \le \frac{1}{2},
$$

where the second inequality is by the Taylor expansion $e = \sum_{i=0}^{\infty} 1/i! \ge \sum_{i=0}^{k} 1/i!$. $\square$

Hence, if $C\geq(k\sqrt{(q-1)/\alpha})^3$, the factor in (9) is bounded by

$$\sqrt{1-\tau^2}+\tfrac{1}{2}\tau^2\leq 1.$$

This concludes the proof of Theorem 3.2. We have, however, not yet addressed the claim that if the marginals μ_i are uniform, there is a Fourier basis such that C can be chosen as $\left(k\sqrt{q-1}\right)^3$. See the comment after the proof of Lemma 3.3 for this claim. □

We now prove the lemma used in the previous proof.

Proof of Lemma 3.3. The case $T=\varnothing$ is a direct application of the induction hypothesis, since the functions g_i depend on at most $n-1$ variables (and have $\deg_{-2}(g_1,...,g_k)\leq D$).

For $i\in[k]$, write

$$h_i(x)=\sum_{j=1}^{q-1}\chi_{i,j}(x_1)h_{i,j}(x_2,...,x_n)$$

for a Fourier basis $\chi_{i,0}=1,\chi_{i,1},...,\chi_{i,q-1}$ of $L^2(\Omega_i,\mu_i)$. Denoting the jth column of X by X^j, and writing $\mathbb{E}_{X^2,...,X^n}$ for the average over $X^2,...,X^n$ we can write $E(T)$ as

$$\begin{aligned} E(T) &= \mathbb{E}_{X^2,...,X^n}\left[\prod_{i\notin T}g_i(X_i)\mathbb{E}_{X^1}\left[\prod_{i\in T}h_i(X_i)\right]\right] \\ &= \mathbb{E}_{X^2,...,X^n}\left[H_T(X)\prod_{i\notin T}g_i(X_i)\right], \end{aligned}$$

where

$$H_T(X)=\mathbb{E}_{X^1}\left[\prod_{i\in T}h_i(X_i)\right]=\sum_{\sigma\in[q-1]^T}\mathbb{E}_{X^1}\left[\prod_{i\in T}\chi_{i,\sigma_i}(X_i^1)\right]\prod_{i\in T}h_{i,\sigma_i}(X_i).$$

Now for $1\leq|T|\leq 2$, the pairwise independence of μ gives that for any $\sigma\in[q-1]^T$,

$$\mathbb{E}_{X^1}\left[\prod_{i\in T}\chi_{i,\sigma_i}(X_i^1)\right]=\prod_{i\in T}\mathbb{E}[\chi_{i,\sigma_i}]=0.$$

Hence in this case $H_T(X)=0$ and by extension $E(T)=0$.

Thus, only the case $|T|\geq 3$ remains. By Hölder's inequality, we can bound

$$\mathbb{E}_{X^1}\left[\prod_{i\in T}\chi_{i,\sigma_i}(X_i^1)\right]\leq\prod_{i\in T}\|\chi_{i,\sigma_i}\|_{|T|}. \tag{11}$$

By Fact 2.5, $\|\chi_{i,\sigma_i}\|_\infty$ can be bound by

$$\sqrt{\frac{1}{\alpha(\mu_i)}} \le \sqrt{\frac{1}{\min_i \alpha(\mu_i)}} = \sqrt{\frac{1}{\alpha}}.$$

Hence we can bound the above by $(1/\alpha)^{|T|/2}$.

Plugging this into $E(T)$ gives

$$E(T) \le \left(\frac{1}{\alpha}\right)^{|T|/2} \mathbb{E}_{X^2,\dots,X^n}\Bigg[\sum_{\sigma\in[q-1]^T} \prod_{i\in T} h_{i,\sigma_i}(X_i) \prod_{i\notin T} g_i(X_i)\Bigg].$$

For $\sigma\in[q-1]^T$, let D_σ be the sum of the $k-2$ smallest degrees of the polynomials $\{g_i : i\notin T\}\cup\{h_{i,\sigma_i} : i\in T\}$. Since g_i and h_{i,σ_i} are functions of $n-1$ variables, we can use the induction hypothesis to get a bound of

$$E(T) \le \left(\frac{1}{\alpha}\right)^{|T|/2} \sum_{\sigma\in[q-1]^T} C^{D_\sigma}\delta \prod_{i\in T\{1\}\ne 1} \|h_{i,\sigma_i}\|_2 \prod_{i\notin T\{1\}\ne 1} \|g_i\|_2.$$

But since the h_{i,σ_i}'s have strictly smaller degrees than the corresponding f_i's, D_σ is bounded by $D-|T|+2$, and hence we have that

$$\begin{aligned} E(T) &\le \alpha^{-|T|/2} C^{D-|T|+2} \sum_{\sigma\in[q-1]^T} \delta \prod_{i\in T\{1\}\ne 1} \|h_{i,\sigma_i}\|_2 \prod_{i\notin T\{1\}\ne 1} \|g_i\|_2 \\ &\le C^{D+2}\left(\frac{\sqrt{(q-1)/\alpha}}{C}\right)^{|T|} \delta \prod_{i\in T\{1\}\ne 1} \|h_i\|_2 \prod_{i\notin T\{1\}\ne 1} \|g_i\|_2, \end{aligned}$$

where we used the fact that $\sum_{j\in[q-1]} \|h_{i,j}\|_2 \le \sqrt{q-1}\|h_i\|_2$ (by the Cauchy–Schwarz inequality and orthogonality of the functions $h_{i,j}$).

To obtain the bound for $|E(T)|$, we can simply negate one of the functions g_i for $i\notin T$ or h_i for $i\in T$, so that $E(T)$ is negated and the calculations above produce an upper bound on $-E(T)$. This concludes the proof of Lemma 3.3. □

Remark 3.5. In the case when the marginal distributions μ_i are uniform, one can take as basis of (Ω,μ) the standard Fourier basis $\chi_y(x)=e^{2\pi i yx/q}$ (where we identify the elements x of Ω with $\mathbb{Z}_q$ and i denotes the imaginary unit). For this basis, $\|\chi_j\|_\infty=1$ and hence (11) can be bound by 1 rather than by $1/\sqrt{\alpha}$, which implies that for this basis, we can choose $C=\left(k\sqrt{q-1}\right)^3$.

3.1. Corollaries

We proceed with some corollaries of Theorem 3.2. The first says that if all non-empty Fourier coefficients of f_1 are small, then the noise correlation is small.

Corollary 3.6. *Assume the setting of Theorem 3.2, but with $\|f_i\|_2 \le 1$ for each i and*

$$\delta := \max_{1 \le i \le k-2} \max_{\sigma \neq \mathbf{0}} |\hat{f}_i(\sigma)|.$$

Then,

$$\left| \langle f_1, ..., f_k \rangle_\mu - \prod_{i=1}^{k} \mathbb{E}[f_i] \right| \le \delta(k-2)C^D, \tag{12}$$

where C and D are as in Theorem 3.2.

Proof. We prove the claim by induction on k. The case $k=2$ is trivial. For the induction hypothesis let $g_1(x)=f_1(x)-\mathbb{E}[f_1]$. Then by Theorem 3.2,

$$|\langle f_1, ..., f_k \rangle_\mu - \mathbb{E}[f_1]\langle f_2, ..., f_k \rangle_\mu| = |\langle g_1, f_2, ..., f_k \rangle_\mu| \le \delta C^D$$

and by the induction hypothesis

$$\left| \mathbb{E}[f_1]\langle f_2, ..., f_k \rangle_\mu - \prod_{i=1}^{k} \mathbb{E}[f_i] \right| = |\mathbb{E}[f_1]| \left| \langle f_2, ..., f_k \rangle_\mu - \prod_{i=2}^{k} \mathbb{E}[f_i] \right| \le (k-3)\delta C^D.$$

The proof follows. □

A more careful examination of the proof above reveals that in the case where the noise correlation is large there should be a basis element with large weight in one of the functions that is correlated with some other functions. Specifically we have the following result.

Corollary 3.7. *Assume the setting of Theorem 3.2 but with $D=\sum_{i=1}^{k} \deg(f_i)$, the sum of the degrees of all the functions, and $\|f_i\|_2 \le 1$ for each f_i.*

Then for all $\delta > 0$, if

$$\left| \langle f_1, ..., f_k \rangle_\mu - \prod_{i=1}^{k} \mathbb{E}[f_i] \right| > 2\delta(k-2)C^D, \tag{13}$$

then there exists $1 \le i \le k-2$ and a non-empty multi-index σ such that

$$|\hat{f}_i(\sigma)| > \delta \quad \text{and} \quad |\mathbb{E}[\chi_\sigma^i f_{i+1} \cdots f_k]| > \delta^2 C^D,$$

where C is the constant from Theorem 3.2.

Proof. From the proof of Corollary 3.6 it follows that if (13) holds then there exists $1 \le i \le k-2$ such that

$$|\langle g_i, f_{i+1}, ..., f_k \rangle_\mu| > 2\delta C^D,$$

where $g_i = f_i - \mathbb{E}[f_i]$. Write $g_i = \sum_{\sigma\in A} \hat{g}_i(\sigma)\chi^i_\sigma + h_i$, where A is the set of all σ for which $|\hat{g}_i(\sigma)| > \delta$. Then by Theorem 3.2 it follows that

$$|\mathbb{E}[h_i f_{i+1} \ldots f_k]| < \delta C^D,$$

which implies that

$$\left|\mathbb{E}\left[\left(\sum_{\sigma\in A} \hat{g}_i(\sigma)\chi^i_\sigma\right) f_{i+1} \ldots f_k\right]\right| > \delta C^D.$$

Writing

$$t(\sigma) = \mathbb{E}[\chi^i_\sigma f_{i+1} \ldots f_k]$$

for $\sigma\in A$, we see that $\sum_{\sigma\in A} |\hat{g}_i(\sigma)t(\sigma)| > \delta C^D$. As $\sum_{\sigma\in A} |\hat{g}_i(\sigma)|^2 \le 1$ it follows that

$$\sum_{\sigma\in A} |\hat{g}_i(\sigma)t(\sigma)| > \delta C^D \sum_{\sigma\in A} |\hat{g}_i(\sigma)|^2,$$

which implies that there exists a σ with

$$|\mathbb{E}[\chi^i_\sigma f_{i+1} \ldots f_k]| = |t(\sigma)| > \delta C^D |\hat{g}_i(\sigma)| \ge \delta^2 C^D. \tag{14}$$

The proof follows. □

Next we apply the previous corollary to (14) and the functions $f_{i+1}, \ldots, f_k, \chi^i_\sigma$ to obtain that $|\mathbb{E}[f_{j+1} \ldots f_k \chi^i_\sigma \chi^j_{\sigma'}]|$ is large for some $j > i$ and σ'. Continuing in this manner we obtain the following corollary.

Corollary 3.8. *Assume the setting of Theorem 3.2 but with $D = \sum_{i=1}^k \deg(f_i)$, the sum of the degrees of all the functions, and $\|f_i\|_2 \le 1$ for each f_i.*

Then for all $\delta > 0$, if

$$\left|\langle f_1, \ldots, f_k\rangle_\mu - \prod_{i=1}^k \mathbb{E}[f_i]\right| > C^D \delta, \tag{15}$$

then there exists a set $I \subseteq [k]$ with $|I| \ge 3$ and for every $i\in I$ a non-zero multi-index $\sigma(i)$ such that

$$|\hat{f}_i(\sigma)| > \left(\frac{\delta}{2k}\right)^{2^k} \quad \text{for all } i\in I;$$

$$|\{i : a \in S(\sigma(i))\}| \ge 3 \quad \text{for all } a \in \bigcup_{i\in I} S(\sigma(i))$$

(the 3 above may be replaced by $r+1$ if the distributions involved are r-wise independent).

Proof. Define $\delta_0 = \delta^{1/2}$, and $\delta_i = \delta_{i-1}^2/2k$. We show by induction on r that it is possible to find disjoint $I, J \subseteq [k]$, where I is of size at least r and for all $i \in I$ there exists a non-zero multi-index $\sigma(i)$ such that

$$|\hat{f}_i(\sigma(i))| > \delta_r = \frac{\delta^{2^{r-1}}}{(2k)^{2^r-1}} > \left(\frac{\delta}{2k}\right)^{2^r} \tag{16}$$

and further

$$\mathbb{E}\left[\prod_{i\in I} \chi^i_{\sigma(i)} \prod_{j\in J} f_j\right] > C^D \delta_{r+1}. \tag{17}$$

The base case $r=1$ is established by the previous claim. The induction step is proved by noting that if J is non-empty and $j \in J$, then we may apply the previous claim to the sequence of functions $\{f_j\}_{j\in J}$ followed by the functions $\chi^i(\sigma(i))$. We then obtain (16) and (17) with δ_{r+1} and sets I' and J', where J' is of size one smaller than J. When we stop with $J = \varnothing$ and $r \le k$ we obtain that J is empty and therefore

$$\mathbb{E}\left[\prod_{i\in I} \chi^i_{\sigma(i)}\right] > C^D \delta_{k+1} > 0.$$

This together with the pairwise independence implies that for all $a \in \bigcup_{i\in I} S(\sigma(i))$,

$$|\{i : a \in S(\sigma(i))\}| \ge 3$$

as needed. □

We finally note while all of the results above are stated for low-degree polynomials, they also apply for polynomials that are almost low-degree. Indeed Hölder's inequality implies the following result.

Proposition 3.9. *Assume the setting of Theorem 3.2 with k functions satisfying $\|f_i\|_k \le 1$ and $\|f_i^{>d}\|_k \le \varepsilon$ for all i. Then*

$$|\langle f_1, \ldots, f_k\rangle_\mu - \langle f_1^{\le d}, \ldots, f_k^{\le d}\rangle_\mu| \le k\varepsilon(1+\varepsilon)^{k-1}.$$

Proof. The proof follows by using Hölder's inequality k times, each time replacing f_i by $f_i^{\le d}$. Note that $\|f_i^{\le d}\|_k \le \|f_i\|_k + \|f_i^{>d}\|_k \le 1+\varepsilon$, so that when making the ith replacement, the error incurred is bound by

$$\left(\prod_{j=1}^{i-1} \|f_j^{\le d}\|_k\right) \|f_i^{>d}\|_k \left(\prod_{j=i+1}^{k} \|f_j\|_k\right) \le (1+\varepsilon)^{i-1}\varepsilon. \qquad \square$$

4. Applications

The first application is a "weak inverse theorem" for the Gowers norm. From Theorem 3.2 and the fact that

$$\|f\|_{U^2} = \left(\sum_{\sigma} |\hat{f}^4(\sigma)|\right)^{1/4}$$

we immediately obtain the following result.

Proposition 4.1. *Let $f\colon \mathbb{Z}_p^n \to \mathbb{C}$ have Fourier degree d, have $\|f\|_2 = 1$ and let $k \geq 2$. If the kth Gowers norm of f satisfies $\|f\|_{U^k} > \varepsilon$, then there exists a multi-index $\sigma \in \mathbb{Z}_p^n$ such that*

$$|\hat{f}(\sigma)| \geq \left(\frac{\varepsilon}{(2^k\sqrt{q-1})^{3d}}\right)^{2^k},$$

where the Fourier coefficient is with respect to the standard Fourier basis. In particular,

$$\|f\|_{U^2} \geq \left(\frac{\varepsilon}{(2^k\sqrt{q-1})^{3d}}\right)^{2^k}.$$

This implies that for functions of low Fourier degree, all U^k norms for constant $k \geq 2$ are equivalent. We next obtain a similar result for arithmetic progressions using Theorem 3.2 and Corollary 3.8.

Proposition 4.2. *Let $(X_1, \dots, X_k)$ have the uniform distribution over arithmetic progressions of length k in $\mathbb{Z}_p^n$, where $3 \leq k \leq p$. Let $Y_1, \dots, Y_k$ be independent identically and uniformly distributed in $\mathbb{Z}_p^n$. Let $f_1, \dots, f_k\colon \mathbb{Z}_p^n \to \mathbb{C}$ have Fourier degree d and $\|f_i\|_2 \leq 1$ for all i. Then, if*

$$|\mathbb{E}[f_1(X_1) \dots f_k(X_k)] - \mathbb{E}[f_1(Y_1) \dots f_k(Y_k)]| > \varepsilon,$$

then the following is true with respect to the standard Fourier basis:

(1) *None of the functions f_i are δ-uniform with*

$$\delta = \frac{\varepsilon}{(k\sqrt{q-1})^{3dk}}.$$

(2) *There exist indices $1 \leq i(1) < i(2) < i(3) \leq k$ and multi-indices*

$$\sigma(1), \sigma(2), \sigma(3) \in \mathbb{Z}_p^n, \quad \textit{with } \sigma(1) \cap \sigma(2) \cap \sigma(3) \neq \varnothing,$$

such that

$$|\hat{f}_{i(j)}(\sigma(j))| \geq \left(\frac{\varepsilon}{k(k\sqrt{q-1})^{3dk}}\right)^{2^k}$$

for $1 \leq j \leq 3$.

We note that the two results above may be interpreted as certain types of derandomization results which can be defined in further generality. The basic setup is that there are $2k$ vectors $X_1, ..., X_k$ and $Y_1, ..., Y_k$. All of the vectors have the same distribution which is uniform in some product space Ω^n. However, the Y_i's are independent while the X_i's are only pairwise independent. How can the two distributions be distinguished? One way to distinguish is to consider functions f_i of X_i (resp. Y_i) and to show that $\prod_{i=1}^k f_i(X_i)$ is far in expectation from $\prod_{i=1}^k f_i(Y_i)$. Our results show that if the functions f_i are uniform and of low degree then it is impossible to have such a distinguisher.

We finally note that for all the applications considered here, the results hold assuming the functions are close in the kth norm to functions of low degree by Proposition 3.9.

5. Possible extensions

We briefly discuss some comments regarding possible extensions of the main result.

5.1. Invariance

The result of [11] shows under stronger conditions the *invariance* of the functions $f_1, ..., f_k$. In other words: they show that the distribution of $(f_1, ..., f_k)$ under the pairwise distribution is close to the distribution under the product distribution with the same marginals as μ.

One would not expect that such a strong conclusion will hold here. Consider for instance the following example. Let $f\colon \{-1,1\}^n \to \mathbb{R}$ be defined by

$$f(x) = \frac{(x_1-1)(x_2+...+x_n)}{n^{1/2}}.$$

Then f has Fourier degree 2, variance $\Theta(1)$, and coefficients of order $n^{-1/2}$. Define a distribution μ on triples of strings $(x,y,z) \in (\{-1,1\}^n)^3$, by letting, for each $i \in [n]$, the distribution on the ith coordinate be the uniform distribution over (x_i, y_i, z_i) satisfying $x_i y_i z_i = 1$. Then μ is balanced pairwise independent. Now consider the distribution of $(f(x), f(y), f(z))$ and compare it with the distribution of $(f(\tilde{x}), f(\tilde{y}), f(\tilde{z}))$ for $\tilde{x}$, $\tilde{y}$ and $\tilde{z}$ independent uniformly random strings of $\{-1,1\}^n$. The distribution of $(f(x), f(y), f(z))$ is supported only on points where at least one of the coordinates is 0 (since one of x_1, y_1 and z_1 is always 1). On the other hand, the distribution of $(f(\tilde{x}), f(\tilde{y}), f(\tilde{z}))$ has an $\Omega(1)$ fraction of its support on points such that all three of $|f(\tilde{x})|$, $|f(\tilde{y})|$ and $|f(\tilde{z})|$ are lower bounded by $\Omega(1)$. Hence the two distributions are not close, even though the Fourier coefficients of f can be made arbitrarily small by increasing n.

The same reasoning shows that we cannot hope for invariance even if all moments on up to $k-1$ variables match. For instance, even if $X_1, ..., X_k$ are $(k-1)$-wise independent it is not necessarily the case that the distribution of $(f(X_1), ..., f(X_k))$ is close to a product distribution.

5.2. Relaxed degree conditions

As mentioned before, the previous works [11] and [12] established results of the type discussed here by first deriving the results for low degree polynomials and then applying "truncation arguments" to obtain results for general bounded functions. It seems that in the context of the current paper these truncation arguments are more challenging.

Indeed, it is well-known that in general, large Gowers norm does not imply large Fourier coefficients (consider e.g. the function $f(X)=(-1)^{\sum_{i=1}^{n-1} x_i x_{i+1}}$ over $\mathbb{Z}_2^n$), and hence one cannot hope to drop the requirement of small Fourier degree and generalize our theorem to general bounded functions.

However, improvements are still possible. First, it is possible that under additional conditions on the pairwise independent marginal distributions, the requirement on low Fourier degree can be dropped completely. We discuss this below.

A second, closely related possible improvement, is to slightly relax the strong Fourier degree requirements. In particular, one can hope that a similar bound can be derived for functions with exponentially small Fourier tails, i.e., functions f such that the total Fourier mass on the high-degree part decays exponentially, $\|f^{>d}\|_2^2 \le (1-\gamma)^d$ for some $\gamma>0$. Such functions arise naturally in many applications, e.g., when functions are evaluated on slightly noisy inputs. Hence, it is natural to ask whether the following extension of our result can be true.

Question 5.1. Let (Ω, μ) be a pairwise independent product space with $\Omega=\Omega_1 \times ... \times \Omega_k$. Is it true that for every $\gamma>0$ and $\varepsilon>0$, there exists a constant $\delta:=\delta(\gamma, \varepsilon)>0$ such that the following holds? If $f_1, ..., f_k \in L^2(\Omega_i^n, \mu_i^{\otimes n})$ satisfy

$$\begin{aligned} \|f_i\|_\infty &\le 1 && \text{for } i\in[k];\\ \|f_i^{\ge d}\|_2^2 &\le (1-\gamma)^d && \text{for } i\in[k] \text{ and } d\in[n];\\ |\hat{f}_1(\sigma)| &\le \delta && \text{for } \sigma\in\mathbb{Z}_q^n, \end{aligned}$$

then

$$\langle f_1, ..., f_k\rangle_\mu \le \varepsilon.$$

An affirmative answer to Question 5.1 would also have consequences for completely dropping the degree requirement under additional conditions on the marginal distributions.

In particular, for marginal distributions whose support is *connected* in the sense described in Section 1.1, by [11] it is known that applying a small amount of noise to each of the functions $f_1, ..., f_k$ does not change $\langle f_1, ..., f_k\rangle_\mu$ by much.

Since applying noise gives exponentially decaying Fourier tails, an affirmative answer to Question 5.1 would imply that for connected marginal distributions, the condition on the Fourier degree of the functions can be dropped completely.

The statement of Question 5.1 allows for much weaker bounds on the error ε than we had in Theorem 3.2, where the error bound was $\lambda(d,\delta)\prod_{i=2}^k \|f_i\|_2$ (with $\lambda(d,\delta)=\delta C^d$). One cannot hope for such a strong error bound in the setting of Question 5.1 (with $\lambda(d,\delta)$ replaced by some function $\lambda(\gamma,\delta)$ depending on the rate of decay of the Fourier tails, rather than the degree), as illustrated by the following example communicated to us by Hamed Hatami, Shachar Lovett, Alex Samorodnitsky and Julia Wolf: consider a pairwise independent distribution μ on $\{0,1\}^k$ in which the first $\approx \log k$ bits are chosen uniformly at random, and the remaining bits are sums of different subsets of the first $\log k$ bits. This distribution is not connected in the sense described above, but that can easily be arranged by adding a small amount of noise to μ, which will not have any significant impact on the calculations which follow. Let $f\colon \{0,1\}^n \to \{0,1\}$ be the function which returns 1 on the all-zeros string, and 0 otherwise. Then, one has that

$$\langle f, ..., f\rangle_\mu = \Pr[X_1 = ... = X_k = 0] \approx 2^{-n\log k},$$

whereas $\|f\|_2 = 2^{-n/2}$ and hence the product $\prod_{i=2}^k \|f\|_2$ equals $2^{-n(k-1)/2}$ so that

$$\lambda(\gamma,\delta)\prod_{i=2}^k \|f\|_2 = \lambda(\gamma,\delta)2^{-n(k-1)/2} \ll \langle f, ..., f\rangle_\mu.$$

One may argue that it is more reasonable to bound $\langle f_1, ..., f_k\rangle_\mu$ in terms of e.g. the ℓ_k norms of the f_i's rather than the ℓ_2 norms. We do not know of any counterexample to such a strengthening of Question 5.1.

5.3. A partial solution to Question 5.1 by Hamed Hatami

We were recently informed by Hamed Hatami (personal communication) that Question 5.1 admits a positive answer in the case when the distribution μ is the support of $(L_1, ..., L_k)$, where the L_i are distinct linear forms over the same additive groups. This follows since given a value of ε we may choose d large enough so that $\|f_1^{\geq d}\|_2 \leq \varepsilon/2$ and thus applying the Cauchy–Schwarz inequality yields that

$$\langle f_1^{\geq d}, f_2, ..., f_k \rangle_\mu \leq \frac{\varepsilon}{2}.$$

On the other hand applying the Gowers–Cauchy–Schwarz inequality using the fact that $f_2, ..., f_k$ are bounded one can obtain that

$$\langle f_1^{<d}, f_2, ..., f_k \rangle_\mu \leq \|f_1^{<d}\|_{U^{k-1}}.$$

Thus choosing $\delta(\varepsilon, d)$ sufficiently small and using the main result of the paper we obtain that the last quantity is at most $\varepsilon/2$.

Acknowledgements. We are grateful to Hamed Hatami, Shachar Lovett, Alex Samorodnitsky and Julia Wolf for communicating us the example in Section 5.2. We would also like to thank Madhur Tulsiani for stimulating and helpful discussions. We thank Hamed Hatami for providing the partial solution to Question 5.1. Finally, we are grateful to Ryan O'Donnell and the two anonymous referees for their many comments which improved the presentation of the paper.

References

1. Austrin, P. and Mossel, E., Approximation resistant predicates from pairwise independence, *Comput. Complexity* **18** (2009), 249–271.
2. Benjamini, I., Kalai, G. and Schramm, O., Noise sensitivity of boolean functions and applications to percolation, *Inst. Hautes Études Sci. Publ. Math.* **90** (1999), 5–43.
3. Furstenberg, H. and Weiss, B., A mean ergodic theorem for $(1/N)\sum_{n=1}^N f(T^n x)\cdot g(T^{n^2} x)$, in *Convergence in Ergodic Theory and Probability* (*Columbus, OH, 1993*), Ohio State Univ. Math. Res. Inst. Publ. **5**, pp. 193–227, de Gruyter, Berlin, 1996.
4. Gowers, W. T., A new proof of Szemerédi's theorem for arithmetic progressions of length four, *Geom. Funct. Anal.* **8** (1998), 529–551.
5. Gowers, W. T., A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** (2001), 465–588.
6. Gowers, W. T. and Wolf, J., The true complexity of a system of linear equations, *Proc. Lond. Math. Soc.* **100** (2010), 155–176.
7. Green, B. and Tao, T., The primes contain arbitrarily long arithmetic progressions, *Ann. of Math.* **167** (2008), 481–547.
8. Khot, S., On the power of unique 2-prover 1-round games, in *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing* (*Montreal, QC, 2002*), pp. 767–775, ACM, New York, 2002.

9. Khot, S., Kindler, G., Mossel, E. and O'Donnell, R., Optimal inapproximability results for max-cut and other 2-variable CSPs?, *SIAM J. Comput.* **37** (2007), 319–357.
10. Lindeberg, J. W., Eine neue Herleitung des exponential-Gesetzes in der Wahrscheinlichkeitsrechnung, *Math. Z.* **15** (1922), 211–235.
11. Mossel, E., Gaussian bounds for noise correlation of functions, *Geom. Funct. Anal.* **19** (2010), 1713–1756.
12. Mossel, E., O'Donnell, R. and Oleszkiewicz, K., Noise stability of functions with low influences: invariance and optimality, *Ann. of Math.* **171** (2010), 295–341.
13. O'Donnell, R., *Computational Applications of Noise Sensitivity*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2003.
14. Raghavendra, P., Optimal algorithms and inapproximability results for every CSP, in *Proceedings of the* 40*th Annual ACM Symposium on Theory of Computing STOC'08* (*Victoria, BC*, *2008*), pp. 245–254, ACM, New York, 2008.
15. Rotar, V. I., Limit theorems for polylinear forms, *J. Multivariate Anal.* **9** (1979), 511–530.
16. Roth, K. F., On certain sets of integers, *J. Lond. Math. Soc.* **28** (1953), 245–252.
17. Samorodnitsky, A. and Trevisan, L., Gowers uniformity, influence of variables, and PCPs, in *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing* (*Seattle, WA*, *2006*), pp. 11–20, ACM, New York, 2006.
18. Szemerédi, E., On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 299–345.

Per Austrin
Department of Computer Science
University of Toronto
10 King's College Road
Toronto, ON M5S 3GK
Canada
austrin@cs.toronto.edu

Elchanan Mossel
University of California, Berkeley
367 Evans Hall
Berkeley, CA 94720-3860
U.S.A.
mossel@stat.berkeley.edu
and
Weizmann Institute of Science
Rehovot 76100
Israel

Received August 19, 2010
published online April 29, 2011