Direct Sum Testing: The General Case

Irit Dinur
The Weizmann Institute of Science, Rehovot, Israel
http://www.wisdom.weizmann.ac.il/~dinuri/
irit.dinur@weizmann.ac.il

Konstantin Golubev
D-MATH, ETH Zurich, Switzerland
https://people.math.ethz.ch/~golubevk/
golubevk@ethz.ch

Abstract

A function $f : [n_1] \times \cdots \times [n_d] \to \mathbb{F}_2$ is a direct sum if it is of the form $f(a_1, \ldots, a_d) = f_1(a_1) \oplus \ldots \oplus f_d(a_d)$, for some d functions $f_i : [n_i] \to \mathbb{F}_2$ for all $i = 1, \ldots, d$, and where $n_1, \ldots, n_d \in \mathbb{N}$. We present a 4-query test which distinguishes between direct sums and functions that are far from them. The test relies on the BLR linearity test (Blum, Luby, Rubinfeld, 1993) and on the direct product test constructed by Dinur & Steurer (2014).

We also present a different test, which queries the function $(d + 1)$ times, but is easier to analyze.

In multiplicative ± 1 notation, this reads as follows. A d-dimensional tensor with ± 1 entries is called a tensor product if it is a tensor product of d vectors with ± 1 entries, or equivalently, if it is of rank 1. The presented tests can be read as tests for distinguishing between tensor products and tensors that are far from being tensor products.

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms

Keywords and phrases property testing, direct sum, tensor product

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.40

Category RANDOM

Funding *Irit Dinur*: ERC-CoG grant number 772839
Konstantin Golubev: ERC grant number 336283 while at the Weizmann Institute and Bar-Ilan University. Currently, the SNF grant number 200020_169106.

1 Introduction

Let us first fix some notations and definitions. By $[n]$ we mean the set $\{0, 1, 2, \ldots, n\}$. For d positive integers $n_1, \ldots, n_d$, we denote $[\overline{n}; d] = [n_1] \times \cdots \times [n_d]$. For two functions $F, G : X \to Y$, we denote by $\mathrm{dist}(F, G)$ the relative Hamming distance between them, namely $\mathrm{dist}(F, G) = \Pr_{x \in X}[F(x) \neq G(x)]$. We say that $F : X \to Y$ is ε-close to have some Property, if there exists a function $G : X \to Y$ such that g has the Property and $\mathrm{dist}(F, G) \leq \varepsilon$.

Given d functions $f_i : [n_i] \to \mathbb{F}_2$, $i = 1, \ldots, d$, where $n_1, \ldots, n_d \in \mathbb{N}$, their direct sum is the function $f : [\overline{n}; d] \to \mathbb{F}_2$ given by $f(a_1, \ldots, a_d) = f_1(a_1) \oplus f_2(a_2) \oplus \ldots \oplus f_d(a_d)$, where $\oplus$ stands for addition is in the field $\mathbb{F}_2$. We denote $f = f_1 \oplus \cdots \oplus f_d$. We study the testability question: given a function $f : [\overline{n}; d] \to \mathbb{F}_2$ test if it is a direct sum, namely if it belongs to the set

$$DirectSum_{[\overline{n};d]} = \{f_1 \oplus \cdots \oplus f_d \mid f_i : [n_i] \to \mathbb{F}_2,\ i = 1, \ldots, d\}.$$

Direct sum is a natural construction that is often used in complexity for hardness amplification [15, 8, 9, 13, 14]. It is related to the direct product construction: a function $f : [\overline{n}; d] \to \mathbb{F}_2^d$ is the direct product of $f_1, \ldots, f_d$ as above if $f(a_1, \ldots, a_d) = (f_1(a_1), \ldots, f_d(a_d))$ for all $(a_1, \ldots, a_d) \in [\overline{n}; d]$. The testability of direct products has received attention

[7, 5, 4, 10, 6] as abstraction of certain PCP tests. It was not surprising to find [3] that there is a connection between testing direct products to testing direct sum. However, somewhat unsatisfyingly this connection was confined to testing a certain type of *symmetric* direct sum. A symmetric direct sum is a function $f : [n]^d \to \mathbb{F}_2$ that is a direct product with all components equal; namely such that there is a single $g : [n] \to \mathbb{F}_2$ such that

$$f(a_1, \ldots, a_d) = g(a_1) \oplus g(a_2) \oplus \cdots \oplus g(a_d).$$

In [3], a 3-query test was presented for testing if a given f is a symmetric direct sum, and the analysis carried out relying on the direct product test. It was left as an open question to devise and analyze a test for the property of being a (not necessarily symmetric) direct sum.

We design and analyze a four-query test which we call the "square in a cube" test, and show that it is a strong absolute local test for being a direct sum. That is, the number of queries is an absolute constant (namely, 4), and the distance from a function to the subspace of direct sums is bounded by some absolute constant (independent of n and d) times the probability of the failure of the test on this function. We also describe a simpler $(d+1)$-query test, whose easy analysis we defer to Section 3.

In order to define the test, we need to introduce the following notation. Given two strings $a, b \in [\overline{n}; d]$ and a set $S \subseteq [d]$, denote by $a_S b$ the string in $[\overline{n}; d]$ whose i-th coordinate equals a_i if $i \in S$ and b_i otherwise.

Test 1 Square in a Cube test.

Given a query access to a function $f : [\overline{n}; d] \to \mathbb{F}_2$:

1. Choose $a, b \in [\overline{n}; d]$ uniformly at random.
2. Choose two subsets $S, T \subset [d]$ uniformly at random, and let $U = S \triangle T$ be their symmetric difference.
3. Accept iff

$$f(a) \oplus f(a_S b) \oplus f(a_T b) \oplus f(a_U b) = 0.$$

We prove the following theorem for Test 1.

▶ **Theorem 1.1** (Main). *There exists an absolute constant $c > 0$ s.t. for all $d \in \mathbb{N}$ and $n_1, \ldots, n_d \in \mathbb{N}$, given $f : [\overline{n}; d] \to \mathbb{F}_2$,*

$$\mathrm{dist}(f, \mathit{DirectSum}_{[\overline{n};d]}) \le c \cdot \Pr_{a,b,S,T}[f(a) \oplus f(a_S b) \oplus f(a_T b) \oplus f(a_{S\triangle T} b) \neq 0]$$

where a, b are chosen independently and uniformly from the domain of f, and S, T are random subsets of $[d]$.

Our proof, similarly to [3], relies on a combination of the BLR linearity testing theorem [2] and the direct product test of [6]. The trick is to find the right combination. We first observe that once we fix a, b, the test is confined to a set of at most 2^d points in the domain, and can be viewed as performing a BLR (affinity rather than linearity) test on this piece of the domain. From the BLR theorem, we deduce an affine linear function on this piece. The next step is to combine the different affine linear functions, one from each piece, into one global direct sum, and this is done by reducing to direct product.

Testing if a tensor has rank 1

An equivalent way to formulate our question is as a test for whether a d-dimensional tensor with ± 1 entries has rank 1. Indeed moving to multiplicative notation and writing $h_i = (-1)^{f_i}$ and $h = (-1)^f$, we are asking whether there are $h_1, \ldots, h_d$ such that

$$h = h_1 \otimes \cdots \otimes h_d.$$

Denoting

$$TensorProduct_{[\overline{n};d]} = \{h_1 \otimes \cdots \otimes h_d \mid h_i : [n_i] \to \{-1, 1\}, , i = 1, \ldots, d\}$$

we have

▶ **Corollary 1.2.** *There exists an absolute constant $c > 0$ s.t. for all $d \in \mathbb{N}$ and $n_1, \ldots, n_d \in \mathbb{N}$, for every $h : [\overline{n}; d] \to \{-1, 1\}$,*

$$\operatorname{dist}(h, TensorProduct_{[\overline{n};d]}) \leq c \cdot \Pr_{a,b,S,T}[h(a) \cdot h(a_S b) \cdot h(a_T b) \cdot h(a_{S \triangle T} b) \neq 1].$$

Structure of the Paper

In Sections 2 and 3 we present two different approaches for testing whether a d-dimensional binary tensor is a tensor product. In Section 4 we discuss possible directions for future research. In Appendix A, we give a proof of the proposition which expands the range of parameters in the direct product test of [6]. This is used in the course of the proof in Section 2.

2 Square in a Cube Test

In this section we present the Square in a Cube Test. Then we introduce the required background: the BLR test for a function being Affine in Subsection 2.1, the direct product test of Dinur & Steurer in Subsection 2.2. Finally, in Subsection 2.3 we prove the main result on the test.

We start by introducing some notation.

Given two vectors $a = (a_1, \ldots, a_d)$, $b = (b_1, \ldots, b_d) \in [\overline{n}; d]$, define

- $\Delta(a, b) = \{i : a_i \neq b_i\} \subseteq [d]$;
- the induced subcube $C_{a,b}$ is the binary cube $\mathbb{F}_2^{\Delta(a,b)}$;
- the projection map $\rho_{a,b} : C_{a,b} \to [\overline{n}; d]$ defined for $x \in C_{a,b}$ as

$$\rho_{a,b}(x)_i = \begin{cases} a_i = b_i, & i \notin \Delta(a, b); \\ b_i, & i \in \Delta(a, b) \text{ and } x_i = 1; \\ a_i, & i \in \Delta(a, b) \text{ and } x_i = 0; \end{cases}$$

The following test is the same as Test 1 in Introduction.

Test 2 Square in a Cube test.

Given a query access to a function $f : [\overline{n}; d] \to \mathbb{F}_2$:
1. Choose $a, b \in [\overline{n}; d]$ uniformly at random.
2. Choose $x, y \in C_{a,b}$ uniformly at random.
3. Query f at $\rho_{a,b}(0), \rho_{a,b}(x), \rho_{a,b}(y)$ and $\rho_{a,b}(x \oplus y)$.
4. Accept iff $f(\rho_{a,b}(0)) \oplus f(\rho_{a,b}(x)) \oplus f(\rho_{a,b}(y)) \oplus f(\rho_{a,b}(x \oplus y)) = 0$.

▶ **Theorem 2.1.** *Suppose a function $f : [\overline{n}; d]^d \to \mathbb{F}_2$ passes Test 2 with probability $1 - \varepsilon$ for some $\varepsilon > 0$, then f is $O(\varepsilon)$-close to a tensor product.*

2.1 The BLR affinity test

The Blum-Luby-Rubinfeld linearity test was introduced in [2], where its remarkable properties were proven. Later a simpler proof via Fourier analysis was presented, e.g. see [1]. Below we give a variation of this test for affine functions, see [12, Chapter 1].

▶ **Definition 2.2.** *A function $g : \mathbb{F}_2^d \to \mathbb{F}_2$ is called affine, if there exists a set $S \subseteq [d]$ and a constant $c \in \mathbb{F}_2$ such that for every vector $x \in \mathbb{F}_2^d$*

$$g(x) = c \oplus \bigoplus_{i \in S} x_i.$$

Note that (see [12, Exercise 1.26]) a function g is affine iff for any two vectors $x, y \in \mathbb{F}_2^d$ it satisfies

$$g(0) \oplus g(x) \oplus g(y) \oplus g(x \oplus y) = 0. \tag{1}$$

The BLR test implies that if a function $g : \mathbb{F}_2^d \to \mathbb{F}_2$ satisfies (1) with high probability, then it is close to an affine function.

Test 3 The BLR affinity test.

Given a query access to a function $f : \mathbb{F}_2^d \to \mathbb{F}_2$:
1. Choose $x \sim \mathbb{F}_2^d$ and $y \sim \mathbb{F}_2^d$ independently and uniformly at random.
2. Query g at $0, x, y$ and $x \oplus y$.
3. Accept if $g(0) \oplus g(x) \oplus g(y) \oplus g(x \oplus y) = 0$.

▶ **Theorem 2.3** ([2]). *Suppose $g : \mathbb{F}_2^d \to \mathbb{F}_2$ passes the affinity test with probability $1 - \varepsilon$ for some $\varepsilon > 0$. Then g is ε-close to being affine.*

2.2 Direct Product Test

▶ **Definition 2.4.** *For $k, M, N \in \mathbb{N}$, and k functions $g_1, \ldots, g_k : [N] \to [M]$, their direct product is the function $g : [N]^k \to [M]^k$ denoted $g = g_1 \times \cdots \times g_k$ and defined as $g((x_1, \ldots, x_k)) = (g_1(x_1), \ldots, g_k(x_k))$. A function $g : [N]^k \to [M]^k$, is called a direct product if there exist k functions $g_1, \ldots, g_k : [N] \to [M]$ such that $g = g_1 \times \cdots \times g_k$ for all $(x_1, \ldots, x_k) \in [N]^k$.*

Dinur & Steurer [6] presented a 2-query test, Test 4, that, with constant probability, distinguishes between direct products and functions that are far from direct product.

Test 4 Two-query test $\mathcal{T}(t)$.

Given a query access to a function $g : [N]^k \to [M]^k$:
1. Choose $x \sim \mathbb{F}_2^d$ and $y \sim \mathbb{F}_2^d$ independently and uniformly at random.
2. Query g at $0, x, y$ and $x \oplus y$.
3. Accept if $g(0) \oplus g(x) \oplus g(y) \oplus g(x \oplus y) = 0$.

▶ **Theorem 2.5** ([6, Theorem 1.1]). *Let k, M, N be positive integers, let $t \leq \alpha k$, where $0 < \alpha < 1$, and let $\varepsilon > 0$. Let $g : [N]^k \to [M]^k$ be given such that*

$$\Pr_{A,x,y}(g(x)_A = g(y)_A) \geq 1 - \varepsilon,$$

where A, x, y are chosen w.r.t. the test distribution $\mathcal{T}(t)$. Then there exists a direct product function g' such that $\mathbb{E}_x[\text{dist}(g(x), g'(x))] = O(\varepsilon k/t)$.

▶ Remark 2.6. The above formulation of Theorem 2.5 is slightly more general than the original statement in [6], as there it is proved for $0 < \alpha < 1/2$. In order to show that the Theorem holds for $0 < \alpha < 1$, we prove the following reduction statement:

> If a function g passes Test $\mathcal{T}(t)$ with probability at least $1-\varepsilon$ for $t = \alpha k$ with $1/2 \le \alpha < 1$, then g passes Test $\mathcal{T}(t')$ with probability at least $1-\varepsilon'$ for $t' = \alpha' k$, where $0 < \alpha' < 1/2$, $\varepsilon' = r\varepsilon/\alpha$ and r is a positive integer.

This reduction shows that Theorem 2.5 is true as it is stated for $t = \alpha k$ for all $0 < \alpha < 1$, as the reduction affects only the constant in the $O(\cdot)$ notation.

For a more detailed explanation, see Appendix A.

2.3 Proof of Theorem 2.1

For a positive integer D, we denote by $\mu_{2/3}(\mathbb{F}_2^D)$ the distribution on $\mathbb{F}_2^D$, where each coordinate, independently, is equal to 0 with probability 1/3 and to 1 with probability 2/3.

We use the following proposition in the course of the proof.

▶ **Proposition 2.7.** *Let $S \subseteq [D]$ be a set and $\chi_S : \mathbb{F}_2^D \to \mathbb{F}_2$ be the corresponding linear function, i.e., $\chi_S(x) = \bigoplus_{i\in S} x_i$. Suppose*

$$\Pr_{x\sim\mu_{2/3}(\mathbb{F}_2^D)} (\chi_S(x) = 0) > \frac{2}{3},$$

then $S = \emptyset$.

Proof. Consider $(-1)^{\chi_S}$. Then

$$\Pr_{x\sim\mu_{2/3}(\mathbb{F}_2^D)} (\chi_S(x) = 0) = \Pr_{x\sim\mu_{2/3}(\mathbb{F}_2^D)} \Big((-1)^{\chi_S(x)} = 1\Big).$$

Also the following holds

$$\frac{1}{3} < \left|2 \Pr_{x\sim\mu_{2/3}(\mathbb{F}_2^D)} \Big((-1)^{\chi_S(x)} = 1\Big) - 1\right| = \left|\mathbb{E}_{x\sim\mu_{2/3}(\mathbb{F}_2^D)}(-1)^{\chi_S(x)}\right| =$$

$$\left|\prod_{i\in[D]} \mathbb{E}_{x_i\sim\mu_{2/3}(\mathbb{F}_2)}(-1)^{x_i}\right| = \left|\left(-\frac{1}{3}\right)^{|S|}\right| = \left(\frac{1}{3}\right)^{|S|},$$

and the statement follows. ◀

Proof of Theorem 2.1. Assume Test 2 rejects a function $f : [\overline{n}; d] \to \mathbb{F}_2$ with probability less than ε, i.e.,

$$\Pr_{\substack{a,b\sim[\overline{n};d]\\ x,y\sim C_{a,b}}} (f_{a,b}(0) \oplus f_{a,b}(x) \oplus f_{a,b}(y) \oplus f_{a,b}(x\oplus y) = 0) > 1-\varepsilon,$$

where all distributions are uniform, and $f_{a,b}$ is a shorthand for $f \circ \rho_{a,b}$. Then there exists $a \in [\overline{n}; d]$ such that

$$\Pr_{\substack{b\sim[\overline{n};d]\\ x,y\sim C_{a,b}}} (f_{a,b}(0) \oplus f_{a,b}(x) \oplus f_{a,b}(y) \oplus f_{a,b}(x\oplus y) = 0) > 1-\varepsilon.$$

Note that the operations re-indexing the domain $[\overline{n}; d]$[1], as well as *flipping* a function, i.e., adding the constant one function to it element-wise, preserve the distance between functions. Hence, w.l.o.g. we can assume for convenience that $a = (0, \ldots, 0)$ and that $f(a) = 0$.

We write C_b for $C_{a,b}$ and f_b for $f_{a,b}$. Then for every $b \in [\overline{n}; d]$,

$$\Pr_{x,y \sim C_b} \left(f_b(0) \oplus f_b(x) \oplus f_b(y) \oplus f_b(x \oplus y) = 0\right) = 1 - \varepsilon_b.$$

The BLR theorem (Theorem 2.3) implies that for each $b \in [\overline{n}; d]$ there exists a subset $S(b) \subseteq \Delta(a, b)$, such that

$$\Pr_{x \sim C_b} \big(f_b(x) = \chi_{S(b)}(x)\big) = 1 - \varepsilon_b.$$

▶ Remark 2.8. By the BLR theorem, there should be the 'greater or equal to' sign instead of the equality. We assume equality for convenience.

Let $F : [\overline{n}; d] \to \mathbb{F}_2^d$ be a function defined as follows. For each $b \in [\overline{n}; d]$, the set $S(b) \subseteq \Delta(a, b)$ can be viewed as a subset of $[d]$, since $\Delta(a, b) \subseteq [d]$. Then $F(b)$ is defined as the element of $\mathbb{F}_2^d$ corresponding to the set $S(b)$.

We now show that F passes Test 4 with high probability and hence is close to a direct product.

Let $b \in [\overline{n}; d]$ be chosen uniformly at random, and let $b' \in [\overline{n}; d]$ be chosen with respect to the following distribution $D(b)$. For each $i \in [d]$,

$$b'_i = \begin{cases} b_i, & \text{w.p. } 3/4; \\ \text{chosen uniformly at random from } [n] \setminus \{b_i\}, & \text{w.p. } 1/4. \end{cases}$$

Note that the distribution on pairs (b, b'), where b is chosen uniformly from $[\overline{n}; d]$ and b' w.r.t. $D(b)$, is equivalent to the following: for each $i \in [d]$,

$$\begin{cases} b_i = b'_i \text{ chosen uniformly from } [n], & \text{w.p. } 3/4; \\ b_i \neq b'_i \text{ both chosen uniformly from } [n] & \text{w.p. } 1/4. \end{cases} \tag{2}$$

In particular, it is symmetric in the sense that choosing $b' \sim [\overline{n}; d]$ uniformly at random first, and then $b \sim D(b')$, leads to the same distribution on pairs (b, b') as the one described above.

For such a pair (b, b') define distribution $\mathcal{D}_{b,b'}$ on $[\overline{n}; d]$ as follows. For a vector $x \sim \mathcal{D}_{b,b'}$,

$$x_i = \begin{cases} 0, & & \text{if } i \in \Delta(b, b'); \\ 0, & \text{w.p. } 1/3; & \multirow{2}{*}{\text{if } i \notin \Delta(b, b').} \\ b_i = b'_i & \text{w.p. } 2/3. & \end{cases}$$

Note that the distribution $\mathcal{D}_{b,b'}$ is supported on a binary cube of dimension $d - |\Delta(b, b')|$ inside $[\overline{n}; d]$. Denote

$$\varepsilon_{b,b'} = \Pr_{x \sim \mathcal{D}_{b,b'}} \big(f(x) \neq \chi_{F(b)(x)}\big).$$

We claim that the following holds

$$\varepsilon_b = \Pr_{x \sim C_b} \big(f(x) \neq \chi_{F(b)(x)}\big) = \mathop{\mathbb{E}}_{b' \sim D(b)} \varepsilon_{b,b'}. \tag{3}$$

---

[1] By this we mean selecting permutations π_i on $[n_i]$ for $i = 1, \ldots, d$, and setting $f^{\pi_1, \ldots, \pi_d}(x_1, \ldots, x_d) = f(\pi_1(x_1), \ldots, \pi_d(x_d))$

To see (3) note that since b is chosen uniformly, b' is chosen w.r.t. $D(b)$, and $x \sim \mathcal{D}_{b,b'}$, the resulting distribution for x is

$$x_i = \begin{cases} 0, & \text{w.p. } 1/2; \\ b_i & \text{w.p. } 1/2, \end{cases}$$

which is exactly the uniform distribution on C_b.

We now show that

$$\Pr_{\substack{b\sim[\overline{n};d] \\ b'\sim D(b)}} \left(\varepsilon_{b,b'} + \varepsilon_{b',b} > \frac{1}{3} \right) < 6\varepsilon \tag{4}$$

First note that it follows from the definitions that

$$\mathop{\mathbb{E}}_{b\sim[\overline{n};d]} \mathop{\mathbb{E}}_{b'\sim D(b)} \varepsilon_{b,b'} = \mathop{\mathbb{E}}_{b\sim[\overline{n};d]} \varepsilon_b = \varepsilon.$$

And by the symmetry of the distribution on pairs (b, b'),

$$\mathop{\mathbb{E}}_{b\sim[\overline{n};d]} \mathop{\mathbb{E}}_{b'\sim D(b)} \varepsilon_{b',b} = \mathop{\mathbb{E}}_{b'\sim D(b)} \mathop{\mathbb{E}}_{b\sim[\overline{n};d]} \varepsilon_{b',b} = \varepsilon.$$

Combined together, the previous two equations imply that

$$\mathop{\mathbb{E}}_{b\sim[\overline{n};d]} \mathop{\mathbb{E}}_{b'\sim D(b)} (\varepsilon_{b,b'} + \varepsilon_{b',b}) = 2\varepsilon,$$

and by the Markov inequality, Inequality 4 follows. By the definition of $\varepsilon_{b,b'}$,

$$\Pr_{x\sim\mathcal{D}_{b,b'}} \left(\chi_{F(b)}(x) = \chi_{F(b')}(x) \right) > 1 - (\varepsilon_{b,b'} + \varepsilon_{b',b}).$$

which is equivalent to

$$\Pr_{x\sim\mathcal{D}_{b,b'}} \left(\chi_{F(b)\Delta F(b')}(x) = 1 \right) > 1 - (\varepsilon_{b,b'} + \varepsilon_{b',b}).$$

Proposition 2.7 implies that if $1 - (\varepsilon_{b,b'} + \varepsilon_{b',b}) > \frac{2}{3}$, then

$$F(b)_{C_b \cap C_{b'}} = F(b')_{C_b \cap C_{b'}}.$$

By Theorem 2.5, the function $F : [\overline{n}; d] \to \mathbb{F}_2^d$ is close to a direct product, i.e., there exist d functions $F_1, \ldots, F_d : [n] \to \mathbb{F}_2$ such that

$$\Pr_{b\sim[\overline{n};d]} (F(b) = (F_1(b_1), \ldots, F_d(b_d))) \geq 1 - O(\varepsilon).$$

Therefore,

$$\Pr_{b\sim[\overline{n};d]} \left(f(b) = \bigoplus_{i=1}^{d} F_i(b_i) \right) \geq 1 - O(\varepsilon). \qquad \blacktriangleleft$$

3 The Shapka Test

In this section we present a different test for whether a tensor is a tensor product. It queries the tensor at $(d+2)$ places at most, but the proof is simpler than for the previous test.

In [11], Kaufman and Lubotzky showed an interesting connection between the theory of high-dimensional expanders and property testing. Namely, they showed that $\mathbb{F}_2$-coboundary expansion of a 2-dimensional complete simplicial complex implies testability of whether a symmetric $\mathbb{F}_2$-matrix is a tensor square of a vector. The following test is inspired by their work and in a way generalizes it. However, since the description below does not employ neither terminology nor machinery of high-dimensional expanders, we refer to [11] for the connection between this theory and property testing.

Given two strings $a, b \in [\overline{n}; d]$, for $i \in [d]$ denote by $a^i_b \in [\overline{n}; d]$ the vector which coincides with a in every coordinate except for the i-th one, where it coincides with b, i.e.,

$$(a^i_b)_j = \begin{cases} a_j, & \text{if } j \neq i; \\ b_i, & \text{if } j = i. \end{cases}$$

For a string $a \in [\overline{n}; d]$, and a number $x \in [n_i]$, we write a^i_x for the string which is equal to a in every coordinate except for the i-th one, where it is equal to x, i.e.,

$$a^i_x = (a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_d).$$

Test 5 The Shapka Test.

Given a query access to a function $f : [\overline{n}; d] \to \mathbb{F}_2$:
1. Choose $a, b \in [\overline{n}; d]$ uniformly at random.
2. Define the query set $Q_{a,b} \subseteq [\overline{n}; d]$ to consist of a, a^j_b for all $j \in [d]$, and also b if d is even.
3. Query f at the elements of $Q_{a,b}$.
4. Accept iff $\bigoplus_{q \in Q_{a,b}} f(q) = 0$.

▶ Remark 3.1. Shapka is the Russian word for a winter hat (derived from Old French *chape* for a *cap*). The name *the Shapka test* comes from the fact that the set $Q_{a,b}$ consists of the two top layers of the induced binary cube $C_{a,b}$ (and also the bottom layer if d is even).

▶ **Theorem 3.2.** *Suppose a function $f : [\overline{n}; d] \to \mathbb{F}_2$ passes Test 5 with probability $1 - \varepsilon$ for some $\varepsilon > 0$, then f is ε-close to a tensor product.*

Proof. Let δ be the relative Hamming distance from f to the subspace of direct sums, i.e., for every direct sum $g : [\overline{n}; d] \to \mathbb{F}_2$ it holds that

$$Pr_{x \sim [\overline{n}; d]} (f(x) \neq g(x)) \geq \delta.$$

For a vector $a \in [\overline{n}; d]$, let us define the local view of f from a, that is d functions $f^a_1, \ldots, f^a_d$, where $f^d_i : [n_i] \to \mathbb{F}_2$, $i = 1, \ldots, d$, that are defined as follows. For $1 \leq i \leq d-1$, and $x \in [n_i]$,

$$f^a_i(x) = f(a^i_x).$$

For $i = d$, the definition of $f^a_d : [n_d] \to \mathbb{F}_2$ depends on the parity of d and goes as follows

$$\begin{cases} f^a_d(x) = f(a^d_x), & \text{if } d \text{ is odd}, \\ f^a_d(x) = f(a^d_x) \oplus f(a), & \text{if } d \text{ is even}. \end{cases}$$

Given a collection of d functions, $g_i : [n_i] \to \mathbb{F}_2$, $i = 1, \ldots, d$, recall that their direct sum is the function $g_1 \oplus \cdots \oplus g_d$ such that for a vector $x \in [\overline{n}; d]$ the following holds

$$g_1 \oplus \cdots \oplus g_d = \bigoplus_{i \in [d]} g_i(x_i).$$

The following holds for any $[\overline{n}; d]$,

$$\left(f - f_1^a \oplus \cdots \oplus f_d^a)\right)(b_1, \ldots, b_d) = \bigoplus_{q \in Q_{a,b}} f(q). \tag{5}$$

As $f_1^a \oplus \cdots \oplus f_d^a$ is a direct sum, it is at least δ-far from f, and hence for any $a \in [\overline{n}; d]$,

$$\Pr_{b \sim [\overline{n}; d]} \left(\left(f - f_1^a \oplus \cdots \oplus f_d^a\right)(b) = 1 \right) \geq \delta. \tag{6}$$

Assume now that f fails Test 5 with probability ε, i.e.,

$$\varepsilon = \Pr_{a,b \sim [\overline{n}; d]} \left(\bigoplus_{q \in Q_{a,b}} f(q) = 1 \right).$$

Combining this equality with (5) and (6), we get the following

$$\varepsilon = \mathop{\mathbb{E}}_{a \sim [\overline{n}; d]} \Pr_{b \sim [\overline{n}; d]} \left(\left(f - f_a^1 \oplus \cdots \oplus f_a^d\right)(b_1, \ldots, b_d) = 1 \right) \geq \left(\mathop{\mathbb{E}}_{a \sim [\overline{n}; d]} \delta \right) = \delta,$$

which completes the proof. ◀

4 Further Directions

Below we present possible directions for future research.

1. Can the original function $f : [\overline{n}; d] \to \mathbb{F}_2$ be reconstructed by a voting scheme using the Shapka Test 5?
2. It is plausible that the Square in the Cube test 2 can be analyzed by the Fourier transform approach similarly to the analysis of the BLR test.
3. Another test in the spirit of the Shapka Test is the following.

Test 6 The Shapka Test.

Given a query access to a function $f : [\overline{n}; d] \to \mathbb{F}_2$:

a. Choose $a, b \in [\overline{n}; d]$ uniformly at random.
b. Choose $x \in C_{a,b}$ uniformly at random.
c. Query f at $\rho_{a,b}(0), \rho_{a,b}(x), \rho_{a,b}(1)$ and $\rho_{a,b}(x \oplus 1)$.
d. Accept iff $f(\rho_{a,b}(0)) \oplus f(\rho_{a,b}(x)) \oplus f(\rho_{a,b}(1)) \oplus f(\rho_{a,b}(x \oplus 1)) = 0$.

We conjecture that this test is also good, i.e., if a function passes the test with high probability then it is close to a tensor product.

References

1 Mihir Bellare, Don Coppersmith, JOHAN Hastad, Marcos Kiwi, and Madhu Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.

2 Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of computer and system sciences*, 47(3):549–595, 1993.

3 Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. *SIAM Journal on Computing*, 46(4):1336–1369, 2017.

4 Irit Dinur and Elazar Goldenberg. Locally testing direct product in the low error range. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 613–622. IEEE, 2008.

5 Irit Dinur and Omer Reingold. Assignment testers: Towards a combinatorial proof of the PCP theorem. *SIAM Journal on Computing*, 36(4):975–1024, 2006.

6 Irit Dinur and David Steurer. Direct product testing. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 188–196. IEEE, 2014.

7 Oded Goldreich and Shmuel Safra. A combinatorial consistency lemma with application to proving the PCP theorem. *SIAM Journal on Computing*, 29(4):1132–1154, 2000.

8 Russell Impagliazzo, Ragesh Jaiswal, and Valentine Kabanets. Approximate list-decoding of direct product codes and uniform hardness amplification. *SIAM Journal on Computing*, 39(2):564–605, 2009.

9 Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. Uniform direct product theorems: simplified, optimized, and derandomized. *SIAM Journal on Computing*, 39(4):1637–1665, 2010.

10 Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query PCPs. *SIAM Journal on Computing*, 41(6):1722–1768, 2012.

11 Tali Kaufman and Alexander Lubotzky. High dimensional expanders and property testing. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 501–506. ACM, 2014.

12 Ryan O'Donnell. *Analysis of boolean functions.* Cambridge University Press, 2014.

13 Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR lemma. *Journal of Computer and System Sciences*, 62(2):236–266, 2001.

14 Luca Trevisan. List-decoding using the XOR lemma. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 126–135. IEEE, 2003.

15 Andrew C Yao. Theory and application of trapdoor functions. In *23rd Annual Symposium on Foundations of Computer Science (sfcs 1982)*, pages 80–91. IEEE, 1982.

A Appendix: Proof of Remark 2.6

In this section we show that Theorem 2.5 holds for a wider range of parameters than in its original formulation in [6]. This was used in the course of the proof of 2.1.

In [6], Dinur and Steurer proved Theorem 2.5 for $0 < \alpha < 1/2$. The following reduction shows that the theorem is true for all $0 < \alpha < 1$ by a reduction from $1/2 \leq \alpha < 1$ to some $0 < \alpha' < 1/2$. Recall that Test 4 makes two queries according to the distribution $\mathcal{T}(t)$, which is the following distribution: (1) Choose a set $A \subset [k]$ of size t uniformly at random. (2) Choose $x, y \in [N]^k$ uniformly at random, conditioned $x_A = y_A$.

▶ **Proposition A.1.** *Let $agr(g, \alpha)$ denote the probability that a function g passes Test 4 with respect to distribution $\mathcal{T}(\alpha k)$. If $agr(g, \alpha) \geq 1 - \varepsilon$ for some $1/2 \leq \alpha \leq 1$, then $agr(g, \alpha') \geq 1 - r\varepsilon$ for $0 < \alpha' \leq 1/2$, where $r = \left\lceil \frac{1}{2(1-\alpha)} \right\rceil$ and $\alpha' = 1 - (1 - \alpha)r$.*

In addition, is $agr(g, 1/2) \geq 1 - \varepsilon$, then also $agr(g, \alpha - 1/k) \geq 1 - 2\varepsilon$.

Proof. Fix a function $g : [N]^k \to [M]^k$, and suppose $agr(g, \alpha) \geq 1 - \varepsilon$ for some $1/2 \leq \alpha < 1$, i.e.,

$$\Pr_{A,x,y \sim \mathcal{T}(\alpha k)} (g(x)_A = g(y)_A) \geq 1 - \varepsilon.$$

We will show that $agr(g, \alpha') > 1 - r\varepsilon$ where $r = \left\lceil \frac{1}{2(1-\alpha)} \right\rceil$ and $\alpha' = 1 - (1 - \alpha)r$. Note that α' satisfies $0 < \alpha' \leq 1/2$.

Given a pair of random vectors x_0, x_r and a set A distributed according to $\mathcal{T}(\alpha' k)$, we construct a sequence of vectors $x_1, \ldots, x_{r-1}$ such that for all $1 \leq i \leq r$, the pair x_{i-1}, x_i is distributed according to $\mathcal{T}(\alpha k)$.

The complement of A has size $(1 - \alpha)rk$. Partition it randomly into r parts of equal size $(1 - \alpha)k$, $[k] \setminus A = B_1 \cup \cdots \cup B_r$. Denote $C_i = [k] \setminus B_i$ for all $1 \leq i \leq r$.

For each $1 \leq i \leq r - 1$, construct x_i such that it agrees with x_0 on the coordinates in $[k] \setminus \bigcup_{j=1}^{i} B_j$ and with x_r on the rest of the coordinates $\bigcup_{j=i}^{i} B_j$. Then for each $1 \leq i \leq r$, x_i agrees with x_{i-1} on the set C_i of the size αk. Therefore,

$$\Pr \left(g(x_{i-1})_{A_i} = g(x_i)_{A_i} \right) \geq 1 - \varepsilon.$$

Hence,

$$1 - r \cdot \varepsilon \leq \Pr \left(\forall\, 1 \leq i \leq r : \; g(x_{i-1})_{A_i} = g(x_i)_{A_i} \right) \leq \Pr_{A_r,x,y \sim \mathcal{T}(\alpha' k)} \left(g(x_0)_{A_r} = g(x_r)_{A_r} \right).$$

The case of $\alpha' = 1/2$ has to be treated separately. In this case there is a reduction to $\alpha'' = 1/2 - 1/k$ as follows. Given two vectors x_0, x_2 distributed w.r.t. $\mathcal{T}(k/2 - 1)$ construct an intermediate random vector x_1 which agrees on exactly half of the coordinates with both x_0 and x_2. ◀