



# On Approximability of Satisfiable $k$ -CSPs: VI

Amey Bhangale\*      Subhash Khot†      Yang P. Liu‡      Dor Minzer§

## Abstract

We prove local and global inverse theorems for general 3-wise correlations over pairwise-connected distributions. Let  $\mu$  be a distribution over  $\Sigma \times \Gamma \times \Phi$  such that the supports of  $\mu_{xy}$ ,  $\mu_{xz}$ , and  $\mu_{yz}$  are all connected, and let  $f : \Sigma^n \rightarrow \mathbb{C}$ ,  $g : \Gamma^n \rightarrow \mathbb{C}$ ,  $h : \Phi^n \rightarrow \mathbb{C}$  be 1-bounded functions satisfying

$$\left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geq \varepsilon.$$

In this setting, our local inverse theorem asserts that there is  $\delta := \exp(-\varepsilon^{-O_\mu(1)})$  such that with probability at least  $\delta$ , a random restriction of  $f$  down to  $\delta n$  coordinates  $\delta$ -correlates to a product function. To get a global inverse theorem, we prove a restriction inverse theorem for general product functions, stating that if a random restriction of  $f$  down to  $\delta n$  coordinates is  $\delta$ -correlated with a product function with probability at least  $\delta$ , then  $f$  is  $2^{-\text{poly}(\log(1/\delta))}$ -correlated with a function of the form  $L \cdot P$ , where  $L$  is a function of degree  $\text{poly}(1/\delta)$ ,  $\|L\|_2 \leq 1$ , and  $P$  is a product function.

We show applications to property testing and to additive combinatorics. In particular, we show the following result via a density increment argument. Let  $\Sigma$  be a finite set and  $S \subseteq \Sigma \times \Sigma \times \Sigma$  such that: (1)  $(x, x, x) \in S$  for all  $x \in \Sigma$ , and (2) the supports of  $S_{xy}$ ,  $S_{xz}$ , and  $S_{yz}$  are all connected. Then, any set  $A \subseteq \Sigma^n$  with  $|\Sigma|^{-n}|A| \geq \Omega((\log \log \log n)^{-c})$  contains  $x, y, z \in A$ , not all equal, such that  $(x_i, y_i, z_i) \in S$  for all  $i$ . This gives the first reasonable bounds for the restricted 3-AP problem over finite fields.

---

\*Department of Computer Science and Engineering, University of California, Riverside. Supported by the Hellman Fellowship award.

†Department of Computer Science, Courant Institute of Mathematical Sciences, New York University. Supported by the NSF Award CCF-1422159, NSF CCF award 2130816, and the Simons Investigator Award.

‡School of Mathematics, Institute for Advanced Study, Princeton, NJ. This material is based upon work supported by the National Science Foundation under Grant No. DMS-1926686

§Department of Mathematics, Massachusetts Institute of Technology. Supported by NSF CCF award 2227876 and NSF CAREER award 2239160.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Main Results . . . . .	2
1.2	Applications . . . . .	3
1.3	The Swap Norm . . . . .	4
1.4	Subsequent and Future Works . . . . .	6
1.5	Preliminaries . . . . .	6
<b>2</b>	<b>Proof Overview</b>	<b>7</b>
2.1	From 3-Wise Correlation to a Large Swap Norm . . . . .	7
2.2	The 99% Regime of Swap Norm Inverse Theorem . . . . .	8
2.3	Inverse Theorem for the Swap Norm . . . . .	8
2.4	Getting Bounded Functions . . . . .	9
2.5	The Global Inverse Theorem . . . . .	9
<b>3</b>	<b>The Swap Norm and Its Properties</b>	<b>10</b>
3.1	The Box Norm . . . . .	10
3.2	The Relation between the Box Norm and the Swap Norm . . . . .	11
3.3	3-Wise Correlation Implies Large Swap Norm . . . . .	13
<b>4</b>	<b>Proof of Theorem 8: the 99% Regime</b>	<b>17</b>
<b>5</b>	<b>Proof of Theorem 5: the Inverse Theorem for the Swap Norm</b>	<b>18</b>
5.1	An Iterative Scheme . . . . .	18
5.2	Main Induction: Proving Theorem 5 . . . . .	24
<b>6</b>	<b>Correlation to Bounded Product Functions</b>	<b>26</b>
<b>7</b>	<b>Restriction Inverse Theorem</b>	<b>28</b>
7.1	Preliminaries on Product Functions . . . . .	28
7.2	From Random Restrictions to a Direct Product Test . . . . .	30
7.3	Small-Set Expansion . . . . .	34
7.4	Analysis of Direct Product Test . . . . .	35
<b>8</b>	<b>Applications</b>	<b>40</b>
8.1	Restricted 3-APs over $\mathbb{F}_p^n$ and Generalizations . . . . .	40
8.2	Direct Sum Testing . . . . .	44
8.3	Proof of Theorem 1 . . . . .	44

# 1 Introduction

A recent line of research [BKM22,BKM23b,BKM23c,BKM24a,BKM24b] about the approximability of satisfiable constraint satisfaction problems (CSPs in short) identified the following analytical problem as a central component. Suppose that  $\Sigma, \Gamma$  and  $\Phi$  are finite alphabets (thought of as being of constant size), and let  $\mu$  be a distribution over  $\Sigma \times \Gamma \times \Phi$  in which the probability of each atom is at least  $\Omega(1)$ . What triplets of 1-bounded functions  $f: \Sigma^n \rightarrow \mathbb{C}, g: \Gamma^n \rightarrow \mathbb{C}, h: \Phi^n \rightarrow \mathbb{C}$  can achieve a significant 3-wise correlation with respect to  $\mu$ , i.e., satisfy that

$$\left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geq \varepsilon? \tag{1}$$

To get a meaningful answer, one must make some assumptions about the distribution  $\mu$ ; otherwise,  $\mu$  could be supported only on inputs of the form  $(x, x, x)$ , in which case one could take  $f$  to be any function of large 2-norm,  $g(y) = \overline{f(y)}$  and  $h = 1$ . Here and throughout, we will restrict our discussion to the class of pairwise-connected distributions, defined as follows.

**Definition 1.1.** For finite alphabets  $\Sigma, \Gamma, \Phi$ , a distribution  $\mu$  over  $\Sigma \times \Gamma \times \Phi$  is called pairwise-connected if the bipartite graphs  $(\Sigma \cup \Gamma, \text{supp}(\mu_{x,y}))$ ,  $(\Sigma \cup \Phi, \text{supp}(\mu_{x,z}))$ ,  $(\Gamma \cup \Phi, \text{supp}(\mu_{y,z}))$  are all connected.

Earlier works have considered smaller classes of distributions. In [Mos10], Mossel shows that if  $\mu$  is *connected*, then any triplets of functions satisfying (1) must be correlated with low-degree functions. Here, a distribution is called connected if the graph whose vertices are  $\text{supp}(\mu)$ , and whose edges are between points in  $\text{supp}(\mu)$  that differ in exactly one coordinate, is connected. In [BKM22], the authors consider the notion of *Abelian embeddings* and speculate that solutions to (1) are related to Abelian embeddings of  $\mu$ .

**Definition 1.2.** We say  $S \subseteq \Sigma \times \Gamma \times \Phi$  admits an Abelian embedding if there are Abelian group  $(G, +)$  and maps  $\sigma: \Sigma \rightarrow G, \gamma: \Gamma \rightarrow G, \phi: \Phi \rightarrow G$  not all constant, such that  $\sigma(x) + \gamma(y) + \phi(z) = 0$  for all  $(x, y, z) \in S$ . A distribution  $\mu$  over  $\Sigma \times \Gamma \times \Phi$  is said to admit an Abelian embedding if  $\text{supp}(\mu)$  admits an Abelian embedding.

Prior works [BKM23b, BKM24a] considered the class of distributions that do not admit any Abelian embedding (which can easily be seen to be larger than the class of connected distributions), and the more general class of distributions that do not admit embeddings to  $(\mathbb{Z}, +)$ . In the former case, it was proved that triplets of functions satisfying (1) must be correlated with low-degree functions. In the latter case, it was proved that triplets of functions satisfying (1) must be correlated with a functions of the form  $L \cdot \chi \circ \sigma^{\otimes n}, L' \cdot \chi' \circ \gamma^{\otimes n}, L'' \cdot \chi'' \circ \phi^{\otimes n}$ , where  $L, L', L''$  are low-degree functions with bounded 2-norm,  $\chi, \chi', \chi'' \in \widehat{G}^n$  are characters, and  $\sigma, \gamma, \phi$  is an Abelian embedding of  $\mu$  into  $G$ , where  $(G, +)$  is some finite Abelian group.

**The issue of  $(\mathbb{Z}, +)$ -embeddings:** the arguments in [BKM23b, BKM24a] rely on the fact that if  $\mu$  does not admit any embedding into  $(\mathbb{Z}, +)$ , then it can only admit embeddings into Abelian groups of size  $O_\mu(1)$ , of which there are  $O_\mu(1)$  many (up to isomorphism). In particular, one can take a product of all such groups and be sure that the resulting Abelian group will “contain” within it all embeddings of  $\mu$ . This fact completely breaks once  $\mu$  does have embeddings into  $(\mathbb{Z}, +)$ . Indeed, such embedding automatically gives rise to embeddings into  $(\mathbb{Z}_p, +)$  for any prime  $p$ . Despite of that, even in the presence of a  $(\mathbb{Z}, +)$  embedding, it is not clear if there are solutions to (1) that do not fall into the “low-degree times embedding function” mold.

## 1.1 Main Results

### 1.1.1 The Global Inverse Theorem

Our main result is an inverse theorem even in the presence of  $(\mathbb{Z}, +)$  embeddings. In this case, embedding functions are quite arbitrary, and their main notable feature is that they are *product functions*, in the sense that they are product of functions each depending only on one of the input's coordinates. With this in mind, our main result is the following statement.

**Theorem 1.** *Let  $\alpha > 0$ , let  $\Sigma, \Gamma, \Phi$  be finite alphabets, and let  $\mu$  be a pairwise-connected distribution over  $\Sigma \times \Gamma \times \Phi$  in which the probability of each atom is at least  $\alpha$ . Then for every  $\varepsilon > 0$  there exist  $d \in \mathbb{N}$  and  $\delta > 0$  such that the following holds. If  $f: \Sigma^n \rightarrow \mathbb{C}$ ,  $g: \Gamma^n \rightarrow \mathbb{C}$ ,  $h: \Phi^n \rightarrow \mathbb{C}$  are 1-bounded functions such that*

$$\left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geq \varepsilon,$$

*then there exist  $L: \Sigma^n \rightarrow \mathbb{C}$  of degree at most  $d$  and  $\|L\|_2 \leq 1$ , and a product function  $P: \Sigma^n \rightarrow \mathbb{C}$  of the form  $P(x) = \prod_{i=1}^n P_i(x_i)$  where  $|P_i(x_i)| = 1$  for all  $i$  and  $x$ , such that  $|\langle f, L \cdot P \rangle| \geq \delta$ . Quantitatively, we have that  $d = \exp((1/\varepsilon)^{O_\alpha(1)})$ , and  $\delta = \exp(-\exp((1/\varepsilon)^{O_\alpha(1)}))$ .*

### 1.1.2 The Local Inverse Theorem

The main new component in the proof of [Theorem 1](#) is a similar looking local inverse theorem. By that, we mean that given functions  $f, g$  and  $h$  satisfying [\(1\)](#), we prove a similar structural result in a local (and random) part of the domain, given by *random restrictions*. For a function  $f: \Sigma^n \rightarrow \mathbb{C}$ , a subset  $I \subseteq [n]$  of the coordinates, and a setting  $z \in \Sigma^I$ , the *restricted function*  $f_{I \rightarrow z}$  is a function  $f_{I \rightarrow z}: \Sigma^{[n] \setminus I} \rightarrow \mathbb{C}$  defined as  $f_{I \rightarrow z}(y) = f(x_I = y, x_{\bar{I}} = z)$ .

In a random restriction, either the set  $I$  is sampled randomly, the fixing  $z \in \Sigma^I$  is sampled randomly, or both. Below, we use the notation  $I \sim_{1-\alpha} [n]$  to mean that  $I$  is a random subset of  $[n]$  in which each  $i \in [n]$  is included in  $I$  with probability  $1 - \alpha$ , and  $z \sim \nu^I$  means that  $z \in \Sigma^I$  is such that each  $z_i$  is independently distributed according to  $\nu$ .

**Definition 1.3** (Random restriction). Let  $\mu$  be a distribution over  $\Sigma^n$  and write  $\mu = (1 - \alpha)\nu + \alpha\mu'$  for distributions  $\nu, \mu'$ . Then for a function  $f: \Sigma^n \rightarrow \mathbb{C}$ ,  $I \sim_{1-\alpha} [n]$ , and  $z \sim \nu^I$ , we define the random restriction  $f_{I \rightarrow z}: (\Sigma^{[n] \setminus I}, \mu'^{[n] \setminus I}) \rightarrow \mathbb{C}$  as  $f_{I \rightarrow z}(x) := f(x, z)$ .

Our local inverse theorem asserts that if 1-bounded functions  $f, g, h$  satisfy [\(1\)](#), then with noticeable probability, a random restriction of  $f$  (and similarly of  $g, h$ ) correlates to a 1-bounded product function. More precisely:

**Theorem 2.** *Let  $\alpha > 0$ , let  $\Sigma, \Gamma, \Phi$  be finite alphabets, and let  $\mu$  be a pairwise-connected distribution over  $\Sigma \times \Gamma \times \Phi$  in which the probability of each atom is at least  $\alpha$ . Then for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that if  $f: \Sigma^n \rightarrow \mathbb{C}$ ,  $g: \Gamma^n \rightarrow \mathbb{C}$ ,  $h: \Phi^n \rightarrow \mathbb{C}$  are 1-bounded functions satisfying that*

$$\left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geq \varepsilon.,$$

*then, writing  $\mu = (1 - \delta)\nu + \delta U$  where  $U$  is uniform over  $\Sigma$ , we have that:*

$$\Pr_{I \sim_{1-\delta} [n], z \sim \nu^I} \left[ \exists \{P_i: \Sigma \rightarrow \mathbb{C}, \|P_i\|_\infty \leq 1\}_{i \in \bar{I}} \text{ with } \left| \mathbb{E}_{x \sim \Sigma^{\bar{I}}} \left[ f_{I \rightarrow z}(x) \prod_{i \in \bar{I}} P_i(x_i) \right] \right| \geq \delta \right] \geq \delta.$$

Quantitatively,  $\delta(\varepsilon) \geq \exp(-\varepsilon^{-O_\alpha(1)})$ .

**Remark.** A few remarks are in order. First, as far as global inverse theorems are concerned, the low-degree part  $L$  is necessary. Second, the correlation parameter  $\delta$  in Theorem 2 is exponentially better than the one in Theorem 1, and thus it gives better quantitative bounds in some applications. Third, we do not know whether the exponential dependency in Theorem 2 or the double exponential dependency in Theorem 1 is necessary. As far as we know, both results may be true with  $\delta = \varepsilon^{O_\alpha(1)}$ . For instance, in the case no  $(\mathbb{Z}, +)$  embeddings exist, most of the arguments in [BKM24a] give quasi-polynomial type dependency between the two parameters.

## 1.2 Applications

In this section we discuss a few applications of our main results.

### 1.2.1 Restricted 3-APs

Let  $p \geq 3$  be a prime thought of as constant. A restricted 3-AP in  $\mathbb{F}_p^n$  is an arithmetic progression  $x, x + a, x + 2a$  where  $x \in \mathbb{F}_p^n$  and  $a \in \{0, 1\}^n \setminus \{\vec{0}\}$ . The restricted 3-AP problem asks what is the maximum density of a subset  $A \subseteq \mathbb{F}_p^n$  which is free of all restricted 3-AP's, i.e., that does not contain any triplet of the form  $x, x + a, x + 2a \in A$  for some  $x \in \mathbb{F}_p^n$  and  $a \in \{0, 1\}^n \setminus \{\vec{0}\}$ . The restricted 3-AP problem can be seen as a variant of Roth's well known result in the finite field setting [Rot53, Mes95], which was highlighted by Green [Gre].

It follows from the density Hales-Jewett theorem that the density of a set  $A$  that doesn't contain any restricted 3-APs is vanishing with  $n$ , and using the quantitative bounds from [Pol12] one has  $\frac{|A|}{p^n} \leq O\left(\frac{1}{\sqrt{\log^* n}}\right)$ . Below, we state a result that (among other things) gives the first reasonable bounds (i.e., finite number of iterated logarithms) for a set  $A \subseteq \mathbb{F}_p^n$  that doesn't contain any restricted 3-APs.

**Theorem 3.** *Let  $\Sigma$  be a finite set and let  $S \subseteq \Sigma \times \Sigma \times \Sigma$  be a non-empty set satisfying that:*

1.  $(x, x, x) \in S$  for all  $x \in \Sigma$ , and
2.  $\text{supp}(S_{xy}), \text{supp}(S_{xz}), \text{supp}(S_{yz})$  are all connected.

*Then there are constants  $c_\Sigma, C_\Sigma > 0$  such that for any  $A \subseteq \Sigma^n$  with  $\frac{|A|}{|\Sigma|^n} \geq \frac{C_\Sigma}{(\log \log \log n)^{c_\Sigma}}$ , there are  $x, y, z \in A$ , not all equal, such that  $(x_i, y_i, z_i) \in S$  for all  $i \in [n]$ .*

Thus, taking  $\Sigma = \mathbb{F}_p$  and  $S = \{(x, x + a, x + 2a) \mid x \in \mathbb{F}_p, a \in \{0, 1\}\}$ , one sees that  $S$  satisfies the conditions of Theorem 3, and the conclusion regarding the density of 3-AP free sets follows. In fact, we can take an even sparser  $S$  and conclude a similar result. For example, for  $p = 3$  we can take

$$S = \{(0, 0, 0), (1, 1, 1), (2, 2, 2), (0, 1, 2), (1, 2, 0)\},$$

and ensure that a set  $A$  with density exceeding  $\frac{C_\Sigma}{(\log \log \log n)^{c_\Sigma}}$  must contain a triplet  $x, y, z$  not all equal such that  $(x_i, y_i, z_i) \in S$  for all  $i$ . In comparison, the standard restricted 3-AP problem also allows for  $(x_i, y_i, z_i) = (2, 0, 1)$ .

Theorem 3 and the consequence for restricted 3-APs free sets improve upon a recent result of [BKM23a], who proved an analogous statement in the case the common difference  $a$  is allowed to be in  $\{0, 1, 2\}^n \setminus \{\vec{0}\}$ . Our proof of Theorem 3 proceeds by combining Theorem 2 with a density increment argument. This derivation is somewhat similar to the argument in [BKM23a], however the approach we take here is simpler and applies in more generality.

### 1.2.2 Direct Sum Testing in the Low Soundness Regime

Our next application is to the direct sum testing problem [DDG<sup>+</sup>15, DG19, WYZ24]. More specifically, consider the so-called diamond test, introduced in [DG19] and recently analyzed in [WYZ24], which proceeds as follows. Given a function  $f: \Sigma^n \rightarrow \mathbb{F}_p$ , sample  $a, b \in \Sigma^n$ , and  $x \in \{0, 1\}^n$  uniformly and independently. Denoting by  $\phi_x(a, b)$  the point in  $\Sigma^n$  where  $\phi_x(a, b)_i = a_i$  for  $i$ 's such that  $x_i = 1$  and  $\phi_x(a, b)_i = b_i$  for  $i$ 's such that  $x_i = 0$ , the tester reads the values of  $f$  at the points  $a$ ,  $\phi_x(a, b)$ ,  $\phi_x(b, a)$  and  $b$ , and accepts if and only if

$$f(a) - f(\phi_x(a, b)) - f(\phi_x(b, a)) + f(b) = 0.$$

It is clear that any function  $f$  which is a direct sum, i.e. any function of the form  $f(a) = \sum_{i=1}^n f_i(a_i)$ , passes the test with probability 1. For  $p = 2$ , the paper [DG19] proposed the diamond tester as a natural 4-query test of direct sum-ness, and conjectured that it works in the so-called 99% regime. This conjecture was recently confirmed in [WYZ24], wherein the authors showed that if  $f$  passes the test with probability  $1 - \varepsilon$ , then there is a direct sum function  $g$  such that

$$\Pr_{x \in \Sigma^n} [f(x) \neq g(x)] \leq O_{|\Sigma|}(\varepsilon).$$

Using the techniques underlying [Theorem 1](#), we are able to analyze this test in the small soundness regime. More precisely, we show the following result:

**Theorem 4.** *For all finite alphabets  $\Sigma$  and  $\varepsilon > 0$ , there are  $d \in \mathbb{N}$  and  $\varepsilon' > 0$  such that the following holds. Suppose that  $f: \Sigma^n \rightarrow \mathbb{F}_p$  passes the diamond test with probability at least  $\frac{1}{p} + \varepsilon$ . Then there exists  $\alpha \in \mathbb{F}_p \setminus \{0\}$ , a function  $L: \Sigma^n \rightarrow \mathbb{C}$  with  $\|L\|_2 \leq 1$  and degree at most  $d$  and  $P: \Sigma^n \rightarrow \mathbb{C}$  a product function  $P(x) = \prod_{i=1}^n P_i(x_i)$  where  $|P_i(x_i)| = 1$  for all  $i$  and  $x$ , such that*

$$\left| \mathbb{E}_x \left[ \omega_p^{\alpha f(x)} L(x) P(x) \right] \right| \geq \varepsilon',$$

where  $\omega_p$  is a primitive root of unity of order  $p$ . Quantitatively, the bounds on  $d$  and  $\varepsilon'$  are the same as in [Theorem 1](#).

We remark that the complicated form of correlation as in the statement of [Theorem 4](#) is necessary. In the 99% regime, i.e. in the case that the test passes with probability  $1 - \varepsilon$ , our argument allows us to recover a similar to [Theorem 4](#) without the low-degree part  $L$  and with  $\varepsilon' = 1 - O_p(\sqrt{\varepsilon})$ .

### 1.3 The Swap Norm

Underlying the proof of our inverse theorems is the *swap norm*, which is a new norm that can be viewed as an analogue of the  $U^2$ -Gowers' uniformity norm. To define the swap norm, we introduce the following convenient notation.

**Definition 1.4.** Given  $x, y \in \Sigma^n$  we write  $(x', y') \sim (x \leftrightarrow y)$  to denote the following distribution over  $\Sigma^n \times \Sigma^n$ . We sample  $(x', y')$  to satisfy  $(x'_i, y'_i) = (x_i, y_i)$  or  $(y_i, x_i)$  uniformly, independently for each coordinate  $i$ .

In words, given  $x, y \in \Sigma^n$ , the values of coordinates in the distribution  $(x', y') \sim (x \leftrightarrow y)$  is either the same as in  $(x, y)$ , or else it is swapped, and the choice in each coordinate is made independently. With this notation, we define the swap inner product and the swap norm.

**Definition 1.5** (Swap inner product and swap norm). For functions  $f_1, f_2, f_3, f_4 : \Sigma^n \rightarrow \mathbb{C}$ , define

$$\text{swap}(f_1, f_2, f_3, f_4) := \mathbb{E}_{\substack{x \sim \Sigma^n, y \sim \Sigma^n \\ (x', y') \sim (x \leftrightarrow y)}} \left[ f_1(x) f_2(y) \overline{f_3(x') f_4(y')} \right].$$

When  $f_1 = f_2 = f_3 = f_4 = f$ , we abbreviate this notation and write  $\text{swap}(f) := \text{swap}(f, f, f, f)$ .

The form  $\text{swap}(f_1, f_2, f_3, f_4)$  is easily seen to be a multi-linear form, and we show that it satisfies several properties that are reminiscent of Gowers' norms [Gow01]; see Section 3 for details. Among other things, we prove that  $\text{swap}(f)^{1/4}$  is a norm over functions  $f$  (see Corollary 3.6).

To get some intuition as to the relevance of the swap norm to product functions, we note that if  $f(x) = \prod_{i=1}^n P_i(x_i)$  is a product function with  $\|f\|_2 = 1$ , then  $\text{swap}(f) = \|f\|_2^2 = 1$ . More generally, a quick Cauchy-Schwarz argument shows that if  $f$  is  $\varepsilon$ -correlated with a product function  $P$  with  $\|P\|_2 = 1$ , then  $\text{swap}(f) \geq \varepsilon^4$ .<sup>1</sup>

Our local inverse theorem for the swap norm asserts that this is essentially the only possible way to have a large swap norm, at least under random restrictions. More precisely, we show:

**Theorem 5.** *If  $f : \Sigma^n \rightarrow \mathbb{C}$  is a  $B$ -bounded function satisfying that  $\|f\|_2 \leq 1$ ,  $\text{swap}(f)^{1/4} \geq \varepsilon$ , then there is a constant  $\delta(\varepsilon, B) > 0$  such that:*

$$\Pr_{I \sim_{1-\delta(\varepsilon, B)} [n], z \sim \Sigma^I} \left[ \exists \{P_i : \Sigma \rightarrow \mathbb{C}, \|P_i\|_2 \leq 1\}_{i \in \bar{I}} \quad \text{with} \quad \left| \left\langle f_{I \rightarrow z}, \prod_{i \in \bar{I}} P_i \right\rangle \right| \geq \delta(\varepsilon, B) \right] \geq \delta(\varepsilon, B).$$

Quantitatively,  $\delta(\varepsilon, B) \geq (\varepsilon/B)^{O(\varepsilon^{-O(1)})}$ .

**Deducing Theorem 2 from Theorem 5:** to prove Theorem 2, we show that for any triplets of functions  $f, g$  and  $h$  satisfying (1) with respect to some distribution  $\mu$  which is pairwise-connected, it must be the case that the swap norm of  $f$  is noticeable, at least in expectation over appropriately chosen random restrictions. This deduction is shown via a sequence of applications of the Cauchy-Schwarz inequality (which we call the *path trick*, see Definition 3.8). This gets us that a form similar to the one in the definition of the swap inner product, except that the distribution over  $(x, y)$  is not necessarily uniform, is noticeable. We then apply random restrictions to modify the distribution of  $(x, y)$  to be uniform. This step closely mirrors the situation with Szemerédi's theorem: if  $f$  has nontrivial correlation over the distribution supported on  $k$ -APs, then by performing a sequence of Cauchy-Schwarz manipulations, we conclude that  $f$  has a large  $U^{k-1}$  norm.

**The proof of Theorem 5:** the bulk of our effort goes into establishing Theorem 5, and our argument proceeds by induction on  $\varepsilon$ . One challenging aspect of the proof is that for the conclusion to be true, one must consider random restrictions. In other words, there are functions  $f$  that have a noticeable swap norm that have  $o(1)$  correlation with all product functions.<sup>2</sup> This means that our inductive proof strategy must incorporate within it random restrictions all the way, and we defer further discussion to Section 2.

<sup>1</sup>This is similar to the way that if  $f$  is correlated with a character then it has noticeable  $U^2$  uniformity norm, and if it is correlated with a "degree  $s$ " polynomial then it has noticeable  $U^{s+1}$  uniformity norm.

<sup>2</sup>For example, one can take  $\Sigma = \{-1, 1\}$  and a random signed quadratic function, namely taking a set of random independent signs  $\{\alpha_{i,j}\}_{1 \leq i < j \leq n}$  and defining  $f(x) = \text{sign}(\sum_{i < j} \alpha_{i,j} x_i x_j)$ .

## 1.4 Subsequent and Future Works

**Subsequent works:** in [BKLM24a], [BKLM24b] we use the results proved herein to settle a conjecture from [BKM22], as well as give reasonable bounds for the density Hales-Jewett theorem over  $[3]^n$ .

More specifically, in [BKLM24a] we prove that [BKM22, Hypothesis 1.6] is true. To state this result, we first recall the natural extension of the definition of Abelian embeddings to  $k$ -ary distributions. We say that a distribution  $\mu$  over  $\Sigma_1 \times \dots \times \Sigma_k$  admits an Abelian embedding if there are Abelian group  $(G, +)$  and maps  $\sigma_i: \Sigma_i \rightarrow G$  not all constant, such that  $\sum_{i=1}^k \sigma_k(x_i) = 0$  for all  $(x_1, \dots, x_k) \in \text{supp}(\mu)$ . The main result of [BKLM24a] is the following assertion:

**Theorem 6.** *Let  $k \geq 3$  and let  $\Sigma_1, \dots, \Sigma_k$  be finite alphabet. Suppose that  $\mu$  is a distribution over  $\Sigma_1 \times \dots \times \Sigma_k$  in which the probability of each atom is at least  $\alpha$ , and further suppose that  $\mu$  does not admit any Abelian embedding. If  $f_i: \Sigma_i^n \rightarrow \mathbb{C}$  are 1-bounded functions for  $i = 1, \dots, k$ , such that*

$$\left| \mathbb{E}_{(x_1, \dots, x_k) \sim \mu^{\otimes n}} \left[ \prod_{i=1}^k f_i(x_i) \right] \right| \geq \varepsilon,$$

*then there exists a function  $L: \Sigma^n \rightarrow \mathbb{C}$  of degree  $O_{\alpha, \varepsilon}(1)$  and  $\|L\|_2 \leq 1$  such that  $\langle f_1, L \rangle \geq \Omega_{\alpha, \varepsilon}(1)$ .*

The paper [BKLM24a] also establishes extensions of this result which are necessary for the application to the density Hales-Jewett problem, and we refer the reader there for further discussion.

In [BKLM24b], we establish reasonable bounds for the density Hales-Jewett theorem [HJ63, FK91, Pol12], which can be thought of as an even more restricted form of the restricted 3-APs problem. In this context, a combinatorial line is a triplet  $x, y, z \in \{0, 1, 2\}^n$ , not all equal, where for each  $i$ ,

$$(x_i, y_i, z_i) \in \{(0, 0, 0), (1, 1, 1), (2, 2, 2), (0, 1, 2)\}.$$

The main result of [BKLM24b] asserts that if  $A \subseteq \{0, 1, 2\}^n$  does not contain any combinatorial line, then  $\frac{|A|}{3^n} \leq \frac{1}{(\log \log \log \log n)^c}$ , where  $c > 0$  is an absolute constant.

**Future works:** in future works, we plan to investigate the algorithmic applications of [Theorem 2](#) for the class of satisfiable constraint satisfaction problems of arity 3, extending the result of [BKM24b] to a wider class of predicates. An example for a predicate of interest which belongs to this larger class (and for which the result of [BKM24b] does not apply) is the rainbow coloring of 3-uniform hypergraph. In this problem, the input is a 3-uniform hypergraph  $\mathcal{H}$  promised to be tripartite, and the goal is to partition the vertices into 3-parts and maximize the number of hyperedges touching all parts.

We also plan to investigate whether one can extend results as [Theorem 3](#) and establish the stronger counting versions. By that, we mean a statement along the lines of: if  $A \subseteq \mathbb{F}_p^n$  is a set of density at least  $\varepsilon$ , then it contains at least  $\delta$  fraction of the restricted 3-APs, where  $\delta = \delta(\varepsilon) > 0$ .

## 1.5 Preliminaries

**General notation.** Let  $[n] = \{1, \dots, n\}$ . For a subset  $I \subseteq [n]$  we write  $\bar{I} = [n] \setminus I$ . For a real number  $x$  we write  $\|x\|_{\mathbb{R}/\mathbb{Z}} := \min_{z \in \mathbb{Z}} |x - z|$ . We let  $\mathbb{C}$  denote complex numbers, and let  $\mathbb{D}$  denote the unit disk. We let  $\text{supp}(\mu)$  denote the support of a distribution  $\mu$ . For a finite set  $\Sigma$  we write  $x \sim \Sigma^n$  to denote sampling  $x$  uniformly randomly from  $\Sigma^n$ .

Our arguments use singular-value-decompositions heavily (abbreviated SVD henceforth), as per the following definition.



**Definition 1.6** (*J-SVD*). For a function  $f : \Sigma^n \rightarrow \mathbb{C}$  we say that its *J-SVD* is a representation  $f = \sum_i \lambda_i g_i h_i$  for  $g_i : \Sigma^J \rightarrow \mathbb{C}$  and  $h_i : \Sigma^I \rightarrow \mathbb{C}$ , where  $\lambda_i$  are singular values and  $g_i, h_i$  are singular vectors when  $f$  is expressed as a matrix  $M \in \mathbb{C}^{\Sigma^J \times \Sigma^I}$  for  $I = [n] \setminus J$  defined as  $M_{x,y} := f(x, y)$ .

## 2 Proof Overview

In this section we give a proof overview for our inverse theorems. We begin by discussing our local inverse theorem, [Theorem 2](#).

### 2.1 From 3-Wise Correlation to a Large Swap Norm

The first step towards proving [Theorem 2](#) is the following statement, which reduces the proof of [Theorem 2](#) to [Theorem 5](#).

**Theorem 7.** *Let  $\mu$  be a pairwise-connected distribution over  $\Sigma \times \Gamma \times \Phi$  for finite sets  $\Sigma, \Gamma, \Phi$ . If 1-bounded functions  $f : \Sigma^n \rightarrow \mathbb{C}$ ,  $g : \Gamma^n \rightarrow \mathbb{C}$ ,  $h : \Phi^n \rightarrow \mathbb{C}$  satisfy that*

$$\left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geq \varepsilon,$$

*then there is  $\alpha \geq \Omega_{|\Sigma|, |\Gamma|, |\Phi|}(1)$ , and distribution  $\nu$  over  $\Sigma$  satisfying  $\mu_x = \alpha U + (1 - \alpha)\nu$  for uniform  $U$  over  $\Sigma$ , such that*

$$\mathbb{E}_{\substack{I \sim_{1-\alpha} [n] \\ z \sim \nu^I}} [\text{swap}(f_{I \rightarrow z}, f_{I \rightarrow z}, f_{I \rightarrow z}, f_{I \rightarrow z})] \geq \varepsilon^{O_\mu(1)}.$$

The idea towards proving [Theorem 7](#) is to enlarge the alphabets  $\Sigma, \Gamma, \Phi$  and distribution  $\mu$  carefully through a sequence of *path tricks*. Formally, the path trick takes an integer  $r$ , sets  $\Sigma^+ \subseteq \Sigma^{2^{r+1}-1}$  and for all  $y_1, z_1, y_2, z_2, \dots, y_{2^r}, z_{2^r}$  such that  $(y_i, z_i), (z_i, y_{i+1}) \in \text{supp}(\mu_{yz})$ , adds the following to  $\text{supp}(\mu^+)$ . Let  $(x_i, y_i, z_i) \in \text{supp}(\mu)$  for  $i = 1, \dots, 2^r$ , and  $(x'_i, y_{i+1}, z_i) \in \text{supp}(\mu)$  for  $i = 1, 2, \dots, 2^r - 1$ . Then,

$$(\{x_1, x'_1, x_2, \dots, x_{2^r}\}, y_1, z_{2^r}) \in \Sigma^+ \times \Gamma \times \Phi$$

is added to  $\text{supp}(\mu^+)$ . We can define similar path tricks that enlarge  $\Gamma$  and  $\Phi$  respectively.

The guarantees of the path trick are given in [Lemma 3.9](#). Informally, if we apply an  $x$ -path trick, enlarging  $\Sigma$ , then  $f^+, g, \tilde{h}$  have correlation at least  $\varepsilon^{O(1)}$ , where  $\tilde{h} : \Phi^n \rightarrow \mathbb{C}$  is still 1-bounded but otherwise arbitrary, and  $f^+ : \Sigma^+ \rightarrow \mathbb{C}$  is defined for  $x = (x_1, x'_1, \dots, x_{2^r-1}) \in (\Sigma^+)^n$  as  $f^+(x) = \prod_{i=1}^{2^r-1} f(x_i) \prod_{i=1}^{2^r-1-1} \overline{f(x'_i)}$ . In other words, the values multiply.

Now, perform the following sequence of three path tricks.

1. Perform a  $y$ -path trick that makes  $xz$  have full support.
2. Perform a  $z$ -path trick that makes  $xy$  have full support.
3. Perform a  $x$ -path trick for  $r = 2$ .

This only affects the value of  $\varepsilon$  polynomially. Let the resulting distribution be  $\mu^+$  over  $\Sigma^+ \times \Gamma^+ \times \Phi^+$ . By inspection, one can check that for all  $a, b \in \Sigma$  that there is some  $y \in \Gamma^+$  and  $z \in \Phi^+$  such that  $((a, a, b), y, z)$  and  $((b, a, a), y, z)$  are both in the support of  $\mu^+$ . Intuitively, this means that the first and third coordinates of  $\Sigma^+ \subseteq \Sigma^3$  can be swapped without affecting the correlation. Because  $f^+(x) = f(x_1)\overline{f(x'_1)}f(x_2)$ , this means that swapping each coordinate of  $x_1, x_2$  should not affect the value of  $f^+$  significantly, and thus  $f$  has large swap norm.

## 2.2 The 99% Regime of Swap Norm Inverse Theorem

The rest of this overview section is devoted to discussing the proof of [Theorem 5](#), which consists of the bulk of the effort in this paper. As mentioned earlier, our argument proceeds by induction on  $\varepsilon$ , and we first address the base case that  $\varepsilon$  is sufficiently close to 1. In this case random restrictions are not necessary, and we establish the following stronger result:

**Theorem 8.** *If  $f : \Sigma^n \rightarrow \mathbb{C}$  satisfies  $\|f\|_2 \leq 1$  and  $\text{swap}(f) \geq 1 - c$  for  $c \leq 0.05$ , then there are  $P_1, \dots, P_n : \Sigma \rightarrow \mathbb{C}$  with  $\|P_i\|_2 = 1$  such that  $|\langle f, P_1 \dots P_n \rangle| \geq 1 - O(c)$ .*

To prove [Theorem 8](#), we build  $P_1, \dots, P_n$  one at a time. Let the  $\{1\}$ -SVD on  $f$  be  $f = \sum \lambda_i g_i h_i$ , where the largest singular value is  $\lambda_1$ . We will take  $P_1 = g_1$ , and note that for any function  $P : \Sigma^{[n] \setminus \{1\}} \rightarrow \mathbb{C}$  we have that  $\langle f, g_1 P \rangle = \lambda_1 \langle h_1, P \rangle$ . We repeat the argument on  $h_1$ : perform an SVD, take the largest singular value and so on, until we end up with a product function. To argue that this process works, it suffices to argue that  $\text{swap}(h_1) \geq (1 + \Omega(\delta))\text{swap}(f)$ , where  $\delta = 1 - \lambda_1^2$ . Towards this end, we expand  $\text{swap}(f)$  (one can check that the orthogonality of  $g$  causes other terms to cancel) and get that

$$1 - c \leq \text{swap}(f) \leq \sum_i \lambda_i^4 \text{swap}(h_i, h_i, h_i, h_i) + 2 \sum_{i < j} \lambda_i^2 \lambda_j^2 \text{swap}(h_i, h_j, h_i, h_j)$$

We require a non-trivial bound on the off-diagonal entries, and we show that  $\text{swap}(h_i, h_j, h_i, h_j) \leq 2/3$  for orthogonal  $h_i, h_j$  ([Lemma 4.2](#)), which holds for  $i \neq j$ . Using this estimate, we can show that  $\lambda_1^2$  is close to 1 and hence  $\delta$  is close to 0. It follows that  $\text{swap}(f) \leq \lambda_1^4 \text{swap}(h_1) + \frac{4}{3}\delta + O(\delta^2)$ , giving  $\text{swap}(h_1) \geq \frac{\text{swap}(f) - \frac{4}{3}\delta + O(\delta^2)}{1 - 2\delta + O(\delta^2)} \geq (1 + \Omega(\delta))\text{swap}(f)$ .

## 2.3 Inverse Theorem for the Swap Norm

We now discuss the proof of [Theorem 5](#). At a high level, our goal will be to perform a sequence of random restrictions and SVDs to eventually get to a function  $g$  such that:

1. **Correlation for  $g$  implies correlation for  $f$ :** if a product function correlates with  $g$ , then a product function correlates with  $f$  up to a multiplicative factor  $\varepsilon^{100}$  loss.
2. **Increase in swap norm:**  $\text{swap}(g)^{1/4} \geq \varepsilon + \varepsilon^{100}$ .
3. **Boundedness:** if  $f$  is bounded, then  $g$  is also bounded.

Once we find such a procedure, we can iterate it until the swap norm of the resulting function becomes close to 1, which happens in  $O(1/\varepsilon^{100})$  iterations by the second item. At that point, we use the base case, namely [Theorem 8](#), and conclude that the resulting function is correlated with a product function. This result then translates to a correlation of a random restriction of  $f$  with a product function, where the correlation parameter is at least  $(\varepsilon^{100})^{1/\varepsilon^{100}} = \exp(-\varepsilon^{-O(1)})$ .

The starting point of our argument is that if  $f$  has large swap norm, then for a random subset  $J \subseteq [n]$ , the largest singular value of the  $J$ -SVD of  $f$  is at least  $\varepsilon^{O(1)}$  in expectation. This follows by [Lemma 3.3](#), which states that the swap norm is the expected *box norm* of  $f$  with respect to a random partition  $(J, \bar{J})$ , and the fact that having a large box norm with respect to some partition implies a large singular value in the corresponding SVD decomposition.

This suggests the following approach: consider the  $J$ -SVD, say  $f = \sum_i \lambda_i g_i h_i$ , and try to come up with a candidate function  $g$  above based in the functions  $g_i$ . A natural attempt, for instance

would be to take  $g$  to be a linear combination of the the singular vectors  $g_1, g_2, \dots$ , possibly only those corresponding to large singular values. However, we do not know how to argue that this works, even in the case where there is only one large singular value. The issue is that there is no clear way to directly relate  $\text{swap}(g_1)\text{swap}(h_1)$  and  $\text{swap}(f)$ .

Instead, we take the following iterative approach. Let  $\lambda gh$  be the first term of the  $J$ -SVD of  $f$ , and write  $\Delta = f - \lambda gh$  – note that  $\|\Delta\|_2^2 = \|f\|_2^2 - \lambda^2 \leq 1 - \varepsilon^{O(1)}$ . If  $\text{swap}(\Delta) \leq \varepsilon^{1000}$ , i.e., is very small, then we know by the fact that  $\text{swap}^{1/4}$  is a norm that  $\text{swap}(\lambda gh) \approx \text{swap}(f)$ . Then  $\text{swap}(g)\text{swap}(h) \geq (1 - o(1))\text{swap}(f)$ , and hence one of  $\text{swap}(g)$  or  $\text{swap}(h)$  is much larger, as desired. Otherwise,  $\text{swap}(\Delta)$  is still quite large and hence the  $J'$ -SVD of  $\Delta$  for a random  $J' \subseteq [n]$  has a large singular value. Let the first term in the  $J'$ -SVD of  $\Delta$  be  $\lambda' g' h'$ , and write

$$\Delta' = \Delta - \lambda' g' h' = f - \lambda gh - \lambda' g' h'.$$

Again, note that  $\|\Delta'\|_2^2 \leq \|\Delta\|_2^2 - \varepsilon^{O(1)}$ . However, now  $gh$  and  $g'h'$  live on different partitions which makes it difficult to relate them to some SVD of  $f$  (we want to use SVDs to increase the  $\text{swap}$  like we did in [Section 2.2](#)). To handle this, we random restrict  $f$  on a common refinement of  $J$  and  $J'$ , i.e.,  $\bar{I} := (J \cap J') \cup (\bar{J} \cap \bar{J}')$ . After random restriction,

$$\Delta'_{I \rightarrow z} = f_{I \rightarrow z} - \lambda g_{I \rightarrow z} h_{I \rightarrow z} - \lambda' g'_{I \rightarrow z} h'_{I \rightarrow z}.$$

Critically,  $g_{I \rightarrow z}, h_{I \rightarrow z}, g'_{I \rightarrow z}, h'_{I \rightarrow z}$  all live on the same partition of  $\bar{I}$ . On average, the random restriction does not increase  $\|\Delta'\|_2$  or decrease  $\text{swap}(f)$  (see [Lemma 3.7](#)). Thus we may continue repeating this process until it terminates – this should happen within  $\varepsilon^{-O(1)}$  steps because  $\|\Delta'\|_2^2$  decreases during each step of the iteration. In the end, after several rounds of random restriction, we are left with a function  $\Delta := f - \sum_{i=1}^T \lambda_i g_i h_i$ , where  $T \leq \varepsilon^{-O(1)}$ ,  $g_i, h_i$  are functions over some partition of  $f$ , and  $\text{swap}(\Delta)$  is very small. At this point, we are able to use a more complicated version of the argument in [Section 2.2](#) to build the desired function  $g$  as a linear combination of the  $g_i$ 's.

There is one point to discuss: the  $g_i, h_i$  may live on the same partition of  $f$ , but may not necessarily be the true SVD of  $f$  on that partition. This is not difficult to fix: after each random restriction, replace the  $g_i, h_i$  with the true SVD up to that many terms. This can only decrease  $\|\Delta\|_2^2$  further, because

$$\min_{\lambda_i, g_i, h_i: 1 \leq i \leq t} \left\| f - \sum_{i=1}^t \lambda_i g_i h_i \right\|_2^2$$

is attained exactly when  $\lambda_i, g_i, h_i$  are the first  $t$  terms of the SVD of  $f$ .

## 2.4 Getting Bounded Functions

The above argument naturally gives rise to product functions  $P$  that are only guaranteed to be  $\ell_2$ -bounded, whereas in [Theorem 2](#) we require the product function to have absolute value 1 always. In [Section 6](#) we prove that if a 1-bounded function  $f$  correlates to an  $\ell_2$ -bounded product function, then  $f$  in fact correlates to a 1-bounded product function, with polynomial loss in the correlation; see [Theorem 10](#) for a formal statement.

## 2.5 The Global Inverse Theorem

Finally, to get [Theorem 1](#) we use the following restriction inverse theorem, extending the ones from [[BKM23c](#), [BKM24a](#)]:

**Theorem 9.** Let  $f : \Sigma^n \rightarrow \mathbb{C}$  be 1-bounded and  $\mu = (1 - \rho)\nu + \rho U$  where  $U$  is uniform over  $\Sigma$ . Suppose that for  $\varepsilon > 0$  it holds that

$$\Pr_{I \sim_{1-\rho}[n], z \sim \nu^I} \left[ \exists \{P_i : \Sigma \rightarrow \mathbb{C}, \|P_i\|_\infty \leq 1\}_{i \in \bar{I}} \text{ with } \left| \mathbb{E}_{x \sim \Sigma^{\bar{I}}} \left[ f_{I \rightarrow z}(x) \prod_{i \in \bar{I}} P_i(x_i) \right] \right| \geq \varepsilon \right] \geq \varepsilon.$$

Then there exist 1-bounded  $P_i : \Sigma \rightarrow \mathbb{C}$  for  $i = 1, \dots, n$ , and a function  $L : \Sigma^n \rightarrow \mathbb{C}$  of degree at most  $D := D(\rho, \varepsilon)$  with  $\|L\|_2 \leq 1$  such that:

$$\left| \mathbb{E}_{x \sim \mu^{\otimes n}} \left[ f(x) L(x) \prod_{i=1}^n P_i(x_i) \right] \right| \geq \varepsilon',$$

where quantitatively  $\varepsilon' \geq \exp(-C_\Sigma(\rho^{-1} \log(1/\varepsilon))^{O(1)})$  and  $D \leq C_\Sigma(\varepsilon\rho)^{-O(1)}$  for some constant  $C_\Sigma$ .

In words, [Theorem 9](#) asserts that if a random restriction of  $f$  is correlated with a product function with noticeable probability, then  $f$  itself must be correlated with a product of a low-degree function with a product function.

The proof of [Theorem 9](#) proceeds along similar lines as in [[BKM23c](#), [BKM24a](#)], with some differences. In previous works, a key component of the proof was a small-set expansion statement, asserting that if  $h : \Sigma^n \rightarrow \Sigma^N$  is a function such that

$$\Pr_{\substack{x \sim \Sigma^n \\ y \sim_\rho x}} [h(x) = h(y)] \geq \varepsilon$$

(where  $y \sim_\rho x$  is the distribution where for each coordinate  $i$  independently,  $y_i = x_i$  with probability  $\rho$ , and else  $y_i \sim \Sigma$ ), then there exists  $a \in \Sigma^N$  such that  $h(x) = a$  for at least  $\varepsilon^{O_\rho(1)}$  of the  $x$ 's. Here, we have to prove a similar looking statement in the case that  $h : \Sigma^n \rightarrow \mathbb{D}^N$  (where  $\mathbb{D}$  is the unit disk), and we know that  $h(x)$  and  $h(y)$  are close in  $\ell_2$  (as opposed to exactly equal) with probability at least  $\varepsilon$ . In this case, we show that there exists  $a \in (\mathbb{R}/\mathbb{Z})^N$  such that  $h(x)$  is close in  $\ell_2$  to  $a$  for at least  $\varepsilon^{O_\rho(1)}$  of the  $x$ 's.

### 3 The Swap Norm and Its Properties

#### 3.1 The Box Norm

We begin by discussing the *box norm*, which is intimately related to the swap norm.

**Definition 3.1** (Box inner product and box norm). For functions  $f_1, f_2, f_3, f_4 : \Sigma^n \rightarrow \mathbb{C}$  and  $S \subseteq [n]$ , define the box inner product as

$$\mathbf{box}_S(f_1, f_2, f_3, f_4) := \mathbb{E}_{\substack{x_S, y_S \sim \Sigma^S \\ x_{\bar{S}}, y_{\bar{S}} \sim \Sigma^{\bar{S}}}} \left[ f_1(x_S, x_{\bar{S}}) f_2(y_S, y_{\bar{S}}) \overline{f_3(y_S, x_{\bar{S}})} f_4(x_S, y_{\bar{S}}) \right].$$

In the case that  $f_1 = f_2 = f_3 = f_4 = f$ , we simplify the notation and write  $\mathbf{box}_S(f) = \mathbf{box}_S(f, f, f, f)$ .

It is easily seen that  $\mathbf{box}_S(f_1, f_2, f_3, f_4)$  is a multi-linear form, and it is known that  $\mathbf{box}_S(f)^{1/4}$  is a norm. Indeed, the following lemma implies that:

**Lemma 3.2.** *Let  $f_1, f_2, f_3, f_4 : \Sigma^n \rightarrow \mathbb{C}$  and  $S \subseteq [n]$ . Then*

$$|\mathbf{box}_S(f_1, f_2, f_3, f_4)|^2 \leq \mathbf{box}_S(f_1, f_3, f_1, f_3) \mathbf{box}_S(f_2, f_4, f_2, f_4),$$

$$|\mathbf{box}_S(f_1, f_2, f_3, f_4)|^4 \leq \mathbf{box}_S(f_1) \mathbf{box}_S(f_2) \mathbf{box}_S(f_3) \mathbf{box}_S(f_4),$$

and  $\mathbf{box}_S(f) \leq \|f\|_2^4$  for any  $f : \Sigma^n \rightarrow \mathbb{C}$ .

*Proof.* For the first assertion,

$$\begin{aligned} |\mathbf{box}_S(f_1, f_2, f_3, f_4)|^2 &= \left| \mathbb{E}_{x_S, y_S \sim \Sigma^S} \left( \mathbb{E}_{x \sim \Sigma^{\bar{S}}} f_1(x_S, x) \overline{f_3(y_S, x)} \right) \left( \mathbb{E}_{x \sim \Sigma^{\bar{S}}} \overline{f_4(x_S, x)} f_2(y_S, x) \right) \right|^2 \\ &\leq \mathbb{E}_{x_S, y_S \sim \Sigma^S} \left| \mathbb{E}_{x \sim \Sigma^{\bar{S}}} f_1(x_S, x) \overline{f_3(y_S, x)} \right|^2 \mathbb{E}_{x_S, y_S \sim \Sigma^S} \left| \mathbb{E}_{x \sim \Sigma^{\bar{S}}} \overline{f_4(x_S, x)} f_2(y_S, x) \right|^2 \\ &= \mathbf{box}_S(f_1, f_3, f_1, f_3) \mathbf{box}_S(f_2, f_4, f_2, f_4), \end{aligned}$$

where we have applied the Cauchy-Schwarz inequality. Repeating this argument (but with the coupling  $f_1, f_4$  and  $f_2, f_3$ ) gives that  $\mathbf{box}_S(f_1, f_2, f_3, f_4)^2 \leq \mathbf{box}_S(f_1, f_4, f_1, f_4) \mathbf{box}_S(f_2, f_3, f_2, f_3)$ , which gives  $\mathbf{box}_S(f_1, f_3, f_1, f_3)^2 \leq \mathbf{box}_S(f_1) \mathbf{box}_S(f_3)$ , hence the second assertion.

For the third assertion we note that by Cauchy-Schwarz again

$$\begin{aligned} \mathbf{box}_S(f, f, f, f) &= \mathbb{E}_{\substack{x_S, y_S \sim \Sigma^S \\ x_{\bar{S}}, y_{\bar{S}} \sim \Sigma^{\bar{S}}}} \left[ f(x_S, x_{\bar{S}}) f(y_S, y_{\bar{S}}) \overline{f(y_S, x_{\bar{S}})} \overline{f(x_S, y_{\bar{S}})} \right] \\ &\leq \left( \mathbb{E}_{\substack{x_S, y_S \sim \Sigma^S \\ x_{\bar{S}}, y_{\bar{S}} \sim \Sigma^{\bar{S}}}} |f(x_S, x_{\bar{S}})|^2 |f(y_S, y_{\bar{S}})|^2 \mathbb{E}_{\substack{x_S, y_S \sim \Sigma^S \\ x_{\bar{S}}, y_{\bar{S}} \sim \Sigma^{\bar{S}}}} |f(y_S, x_{\bar{S}})|^2 |f(x_S, y_{\bar{S}})|^2 \right)^{1/2} \\ &= \|f\|_2^4. \quad \square \end{aligned}$$

### 3.2 The Relation between the Box Norm and the Swap Norm

Next, we observe that the swap inner product of 4 functions is the expected box inner product of the same 4 functions over a randomly chosen partition.

**Lemma 3.3.** *We have that*

$$\mathbf{swap}(f_1, f_2, f_3, f_4) = \mathbb{E}_{S \subseteq [n]} [\mathbf{box}_S(f_1, f_2, f_3, f_4)].$$

*Proof.* Looking at the definition of  $\mathbf{swap}(f_1, f_2, f_3, f_4)$ , we denote by  $S$  the random set of coordinates where  $(x'_i, y'_i) = (y_i, x_i)$ , i.e. where there was swap. Conditioning on  $S$ , it can be seen that the expectation is equal to exactly  $\mathbf{box}_S(f_1, f_2, f_3, f_4)$ , and as the distribution of  $S$  is uniform over all subsets of  $[n]$  the claim follows.  $\square$

Using this connection, [Lemma 3.2](#) implies analogous properties for the swap norm.

**Lemma 3.4.** *We have that  $\mathbf{swap}(f_1, f_2, f_3, f_4) \leq \|f_1\|_2 \|f_2\|_2 \|f_3\|_2 \|f_4\|_2$ .*

*Proof.* By Lemma 3.3

$$\begin{aligned}
\text{swap}(f_1, f_2, f_3, f_4) &= \mathbb{E}_{S \subseteq [n]} [\text{box}_S(f_1, f_2, f_3, f_4)] \\
&\leq \mathbb{E}_{S \subseteq [n]} \left[ \text{box}_S(f_1)^{1/4} \text{box}_S(f_2)^{1/4} \text{box}_S(f_3)^{1/4} \text{box}_S(f_4)^{1/4} \right] \\
&\leq \|f_1\|_2 \|f_2\|_2 \|f_3\|_2 \|f_4\|_2. \quad \square
\end{aligned}$$

We also get the following Cauchy-Schwarz-Gowers type inequalities for the swap norms.

**Lemma 3.5.** *For any  $f_1, f_2, f_3, f_4$  we have that:*

$$|\text{swap}(f_1, f_2, f_3, f_4)|^2 \leq \text{swap}(f_1, f_2, f_1, f_2) \text{swap}(f_3, f_4, f_3, f_4) \quad (2)$$

and

$$\text{swap}(f_1, f_2, f_1, f_2)^2 \leq \text{swap}(f_1, f_1, f_1, f_1) \text{swap}(f_2, f_2, f_2, f_2). \quad (3)$$

*Proof.* The first follows from the identity

$$\begin{aligned}
\text{swap}(f_1, f_2, f_3, f_4) &= \mathbb{E}_{x, y \sim \Sigma^n} \left[ f_1(x) f_2(y) \cdot \mathbb{E}_{(x', y') \sim (x \leftrightarrow y)} \overline{f_3(x') f_4(y')} \right] \\
&= \mathbb{E}_{x, y \sim \Sigma^n} \left[ \mathbb{E}_{(x', y') \sim (x \leftrightarrow y)} f_1(x') f_2(y') \cdot \mathbb{E}_{(x', y') \sim (x \leftrightarrow y)} \overline{f_3(x') f_4(y')} \right]. \quad (4)
\end{aligned}$$

The second equality follows because  $\mathbb{E}_{(x', y') \sim (x \leftrightarrow y)} \overline{f_3(x') f_4(y')}$  is the same for all  $(x'', y'')$  in the support of  $(x \leftrightarrow y)$ . Now, (2) follows from the Cauchy-Schwarz inequality. For (3), we note:

$$\begin{aligned}
\text{swap}(f_1, f_2, f_1, f_2)^2 &= \left( \mathbb{E}_S \text{box}_S(f_1, f_2, f_1, f_2) \right)^2 \\
&\leq \left( \mathbb{E}_S \text{box}_S(f_1, f_1, f_1, f_1)^{1/2} \text{box}_S(f_2, f_2, f_2, f_2)^{1/2} \right)^2 \\
&\leq \mathbb{E}_S \text{box}_S(f_1, f_1, f_1, f_1) \mathbb{E}_S \text{box}_S(f_2, f_2, f_2, f_2) \\
&= \text{swap}(f_1, f_1, f_1, f_1) \text{swap}(f_2, f_2, f_2, f_2).
\end{aligned}$$

Here, the first equality is Lemma 3.3, and the first inequality is Lemma 3.2. □

Applying Lemma 3.5 implies that  $\text{swap}(f)^{1/4}$  is a norm.

**Corollary 3.6.**  *$\text{swap}(f)^{1/4}$  is a norm on functions  $f$ .*

*Proof.* Clearly it obeys scaling and nonnegativity by (4). By Lemma 3.5 we get that:

$$\begin{aligned}
\text{swap}(f_1 + f_2) &= \sum_{(a, b, c, d) \in \{1, 2\}^4} \text{swap}(f_a, f_b, f_c, f_d) \\
&\leq \sum_{(a, b, c, d) \in \{1, 2\}^4} \text{swap}(f_a)^{1/4} \text{swap}(f_b)^{1/4} \text{swap}(f_c)^{1/4} \text{swap}(f_d)^{1/4} \\
&= (\text{swap}(f_1)^{1/4} + \text{swap}(f_2)^{1/4})^4. \quad \square
\end{aligned}$$

The next lemma proves that  $\text{swap}(f)^{1/2}$  is non-decreasing in expectation under random restrictions.

**Lemma 3.7.** *For a subset  $I \subseteq [n]$  and  $f : \Sigma^n \rightarrow \mathbb{C}$  we have that*

$$\text{swap}(f, f, f, f)^{1/2} \leq \mathbb{E}_{z \sim \Sigma^I} \left[ \text{swap}(f_{I \rightarrow z}, f_{I \rightarrow z}, f_{I \rightarrow z}, f_{I \rightarrow z})^{1/2} \right].$$

*Proof.* Note that

$$\begin{aligned} \text{swap}(f, f, f, f) &= \mathbb{E}_{\substack{x \sim \Sigma^I, y \sim \Sigma^I \\ (x', y') \sim (x \leftrightarrow y)}} \text{swap}(f_{I \rightarrow x}, f_{I \rightarrow y}, f_{I \rightarrow x'}, f_{I \rightarrow y'}) \\ &\leq \mathbb{E}_{\substack{x \sim \Sigma^I, y \sim \Sigma^I \\ (x', y') \sim (x \leftrightarrow y)}} \text{swap}(f_{I \rightarrow x})^{1/4} \text{swap}(f_{I \rightarrow y})^{1/4} \text{swap}(f_{I \rightarrow x'})^{1/4} \text{swap}(f_{I \rightarrow y'})^{1/4} \\ &\leq \left( \mathbb{E}_{\substack{x \sim \Sigma^I, y \sim \Sigma^I \\ (x', y') \sim (x \leftrightarrow y)}} \text{swap}(f_{I \rightarrow x})^{1/2} \text{swap}(f_{I \rightarrow y})^{1/2} \mathbb{E}_{\substack{x \sim \Sigma^I, y \sim \Sigma^I \\ (x', y') \sim (x \leftrightarrow y)}} \text{swap}(f_{I \rightarrow x'})^{1/2} \text{swap}(f_{I \rightarrow y'})^{1/2} \right)^{1/2} \\ &= \mathbb{E}_{\substack{x \sim \Sigma^I, y \sim \Sigma^I \\ (x', y') \sim (x \leftrightarrow y)}} \text{swap}(f_{I \rightarrow x})^{1/2} \text{swap}(f_{I \rightarrow y})^{1/2} = \left( \mathbb{E}_{x \sim \Sigma^I} \text{swap}(f_{I \rightarrow x})^{1/2} \right)^2. \end{aligned}$$

Here the first inequality uses [Lemma 3.5](#) and the second uses Cauchy-Schwarz.  $\square$

### 3.3 3-Wise Correlation Implies Large Swap Norm

The goal of this section is to establish [Theorem 7](#).

#### 3.3.1 Path Tricks

In this section we discuss the *path trick*, which is a tool that allows us to modify the underlying distribution  $\mu$  to get additional desirable properties from it. We start by defining the distribution resulting from a path trick of  $r$  steps on  $\mu$ .

**Definition 3.8.** Given a distribution  $\mu$  on  $\Sigma \times \Gamma \times \Phi$  and integer  $r \geq 1$ , define the  $x$  path trick distribution  $\mu^+$  on  $\Sigma^+ \times \Gamma \times \Phi$  (where  $\Sigma^+ \subseteq \Sigma^{2^r-1}$ ) as follows:

1. Sample  $y_1 \sim \mu_y$ .
2. Sample  $(x_1, z_1)$  from  $\mu$ , conditioned on  $y = y_1$ .
3. Sample  $(x'_1, y_2)$  from  $\mu$ , conditioned on  $z = z_1$ .
4. Repeat steps 2 and 3, and in the final step sample  $(x_{2^r}, y_{2^{r-1}+1})$  from  $\mu$  conditioned on  $z = z_{2^r-1}$ .

$\mu^+$  is over the sequences  $((x_1, x'_1, \dots, x'_{2^{r-1}-1}, x_{2^r}), y_1, z_{2^r-1})$  generated by the process above.

The  $y$  and  $z$  path tricks can be defined symmetrically. One way to think of the  $x$ -path trick is as considering the bipartite graph between  $\Gamma$  and  $\Phi$  that has edges between  $(y, z)$  if there is  $x$  such that  $(x, y, z)$  is in the support of  $\mu$ , and in that case the edge is labeled by  $x$ . Path tricks correspond to walks in this bipartite graph, and the resulting symbol in the first coordinate corresponds to the

edge labels collected throughout the walk. The following result from [BKM22] asserts that if  $f, g, h$  have non-trivial correlation with respect to  $\mu$ , then  $f^+, g$  and  $h'$  have non-trivial correlation with respect to the path trick distribution  $\mu^+$ , where  $f^+$  is an alternating conjugated product of  $f$  along the labels and  $h'$  may be an arbitrary 1-bounded function. We reproduce its proof below for the reader's convenience.

**Lemma 3.9** (Path trick). *Let  $\mu$  be a distribution on  $\Sigma \times \Gamma \times \Phi$ . If 1-bounded functions  $f : \Sigma^n \rightarrow \mathbb{C}$ ,  $g : \Gamma^n \rightarrow \mathbb{C}$  and  $h : \Phi^n \rightarrow \mathbb{C}$  satisfy  $|\mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)]| \geq \varepsilon$ , then for any  $r \geq 1$  and  $\mu^+$  defined as above for an  $r$ -step path trick, there are 1-bounded functions  $f^+ : (\Sigma^+)^n \rightarrow \mathbb{C}$ ,  $g : \Gamma^n \rightarrow \mathbb{C}$ ,  $\tilde{h} : \Phi^n \rightarrow \mathbb{C}$  satisfying:*

$$\left| \mathbb{E}_{(x,y,z) \sim (\mu^+)^{\otimes n}} [f^+(x)g(y)\tilde{h}(z)] \right| \geq \varepsilon^{2^r}.$$

Additionally for  $x = (x_1, x'_1, \dots, x_{2^{r-1}}) \in (\Sigma^+)^n$  it holds that  $f^+(x) = \prod_{i=1}^{2^{r-1}} f(x_i) \prod_{i=1}^{2^{r-1}-1} \overline{f(x'_i)}$ .

*Proof.* We start by performing a simple reduction on  $h$ . Define

$$\tilde{h}(z) := \mathbb{E}_{(x,y,z') \sim \mu^{\otimes n}} [f(x)g(y) | z' = z].$$

Then note that

$$\left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)] \right| = \left| \mathbb{E}_{z \sim \mu_z} [h(z)\overline{\tilde{h}(z)}] \right| \leq \|h\|_2 \|\tilde{h}\|_2 \leq \|\tilde{h}\|_2 = \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)\tilde{h}(z)]^{1/2}.$$

Hence  $|\mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)\tilde{h}(z)]| \geq \varepsilon^2$ . The remainder of the proof involves repeated application of Cauchy-Schwarz. We show the following by induction on  $r$ :

$$\left| \mathbb{E}_{\substack{x_1, x'_1, \dots, x_{2^{r-1}}, x'_{2^{r-1}} \\ y_1, \dots, y_{2^{r-1}+1} \\ z_1, \dots, z_{2^{r-1}}} \left[ \prod_{i=1}^{2^{r-1}} f(x_i) \overline{f(x'_i)} \cdot g(y_1) \overline{g(y_{2^{r-1}+1})} \right] \right| \geq \varepsilon^{2^r}. \quad (5)$$

Let us check the base case  $r = 1$ .

$$\varepsilon^2 \leq \left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)\tilde{h}(z)] \right| = \left| \mathbb{E}_{\substack{z_1 \\ x_1, y_1, x'_1, y_2}} [f(x_1) \overline{f(x'_1)} g(y_1) \overline{g(y_2)}] \right|.$$

The inductive step will follow by squaring this, and using Cauchy-Schwarz on the variables other than  $y_{2^{r-1}+1}$ . Precisely,

$$\begin{aligned} \varepsilon^{2^{r+1}} &\leq \left| \mathbb{E}_{\substack{x_1, x'_1, \dots, x_{2^{r-1}}, x'_{2^{r-1}} \\ y_1, \dots, y_{2^{r-1}+1} \\ z_1, \dots, z_{2^{r-1}}} \left[ \prod_{i=1}^{2^{r-1}} f(x_i) \overline{f(x'_i)} \cdot g(y_1) \overline{g(y_{2^{r-1}+1})} \right] \right|^2 \\ &\leq \mathbb{E}_{y_{2^{r-1}+1}} \left| \mathbb{E}_{\substack{x_1, x'_1, \dots, x_{2^{r-1}}, x'_{2^{r-1}} \\ y_1, \dots, y_{2^{r-1}} \\ z_1, \dots, z_{2^{r-1}}} \left[ \prod_{i=1}^{2^{r-1}} f(x_i) \overline{f(x'_i)} \cdot g(y_1) \right] \right|^2. \end{aligned}$$



Note that this last expression exactly equals (5) for  $r + 1$ , completing the induction. Finally, the definition of  $\tilde{h}$  gives that the LHS expression in (5) can be expressed as:

$$\left| \mathbb{E}_{\substack{x_1, x'_1, \dots, x_{2^{r-1}} \\ y_1, \dots, y_{2^{r-1}} \\ z_1, \dots, z_{2^{r-1}}}} \left[ \prod_{i=1}^{2^{r-1}} f(x_i) \prod_{i=1}^{2^{r-1}-1} \overline{f(x'_i)} \cdot g(y_1) \tilde{h}(z_{2^{r-1}}) \right] \right|.$$

So we can set  $f^+(x) = \prod_{i=1}^{2^{r-1}} f(x_i) \prod_{i=1}^{2^{r-1}-1} \overline{f(x'_i)}$ . This is exactly what we wanted to prove.  $\square$

**Pairwise-connectedness:** We remark that it is easy to observe that if  $\mu$  is pairwise-connected, then  $\mu^+$  is also pairwise-connected. In fact, it improves connectivity between two of the coordinates: if  $r$  is taken so that  $2^r$  is larger than  $|\Gamma|, |\Phi|$ , then  $\text{supp}(\mu_{y,z}^+) = \Gamma \times \Phi$ ; see [BKM24a, Lemma 3.12] for a proof. This fact will be useful for us later on.

### 3.3.2 A Symmetry Argument

Now we prove [Theorem 7](#) by applying a sequence of path tricks, and then applying a symmetry argument. We start by applying the following sequence of three path tricks.

1. Apply a  $y$  path trick with parameter  $r$  satisfying  $2^r \geq \max\{|\Sigma|, |\Phi|\}$ .
2. Apply a  $z$  path trick with parameter  $r$  satisfying  $2^r \geq \max\{|\Sigma|, |\Gamma^+|\}$ .
3. Apply a  $x$  path trick with parameter  $r = 2$ .

After this sequence of operations, we know that there is a distribution  $\mu^+$  on  $\Sigma^+ \times \Gamma^+ \times \Phi^+$  such that  $\Sigma^+ \subseteq \Sigma^3$ , and 1-bounded functions  $g : (\Gamma^+)^n \rightarrow \mathbb{C}$ ,  $h : (\Phi^+)^n \rightarrow \mathbb{C}$  such that

$$\left| \mathbb{E}_{\substack{(x,y,z) \sim (\mu^+)^n \\ x=(x_1, x_2, x_3)}} f(x_1) \overline{f(x_2)} f(x_3) g(y) h(z) \right| \geq \varepsilon^{O(1)}. \quad (6)$$

By doing the path tricks in this specific order, we are able to guarantee the following nice property of  $\text{supp}(\mu^+)$  that allows us to then exploit symmetry.

**Lemma 3.10.** *Let distribution  $\mu^+$  over  $\Sigma^+ \times \Gamma^+ \times \Phi^+$  be as defined above. For any  $a, b \in \Sigma$  there are  $y \in \Gamma^+, z \in \Phi^+$  such that both  $((a, a, b), y, z) \in \text{supp}(\mu^+)$  and  $((b, a, a), y, z) \in \text{supp}(\mu^+)$ .*

*Proof.* After the first  $y$  path trick, by pairwise connectivity the distribution  $\mu'$  over  $\Sigma \times \Gamma^+ \times \Phi$  is pairwise-connected and satisfies that  $\mu'_{xz}$  has full support. Let  $z \in \Phi$  be chosen arbitrarily, and let  $y_a, y_b$  be such that  $(a, y_a, z) \in \text{supp}(\mu')$  and  $(b, y_b, z) \in \text{supp}(\mu')$ . Let  $\mu''$  be the distribution over  $\Sigma \times \Gamma^+ \times \Phi^+$  after applying the  $z$  path trick to  $\mu'$ . Again by pairwise connectivity,  $\mu''$  is pairwise-connected and the support of  $\mu''_{xy}$  is full, so there is some  $z' \in \Phi^+$  such that  $(a, y_b, z') \in \text{supp}(\mu'')$ . Also, note that for  $\vec{z} := (z, z, \dots, z)$  we have that  $(a, y_a, \vec{z}), (b, y_b, \vec{z}) \in \text{supp}(\mu'')$ . To conclude, set  $y := y_b$  and  $z := \vec{z}$  and note that  $(a, a, b)$  corresponds to the path  $y_b \rightarrow z' \rightarrow y_b \rightarrow \vec{z}$ , and  $(b, a, a)$  corresponds to the path  $y_b \rightarrow \vec{z} \rightarrow y_a \rightarrow \vec{z}$ .  $\square$

We are now ready to prove [Theorem 7](#), restated below.

**Theorem 7.** Let  $\mu$  be a pairwise-connected distribution over  $\Sigma \times \Gamma \times \Phi$  for finite sets  $\Sigma, \Gamma, \Phi$ . If 1-bounded functions  $f : \Sigma^n \rightarrow \mathbb{C}$ ,  $g : \Gamma^n \rightarrow \mathbb{C}$ ,  $h : \Phi^n \rightarrow \mathbb{C}$  satisfy that

$$\left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geq \varepsilon,$$

then there is  $\alpha \geq \Omega_{|\Sigma|, |\Gamma|, |\Phi|}(1)$ , and distribution  $\nu$  over  $\Sigma$  satisfying  $\mu_x = \alpha U + (1 - \alpha)\nu$  for uniform  $U$  over  $\Sigma$ , such that

$$\mathbb{E}_{\substack{I \sim_{1-\alpha}[n] \\ z \sim \nu^I}} \left[ \text{swap}(f_{I \rightarrow z}, f_{I \rightarrow z}, f_{I \rightarrow z}, f_{I \rightarrow z}) \right] \geq \varepsilon^{O_\mu(1)}.$$

*Proof.* Consider  $\mu^+$  as above, and define the distribution  $\mu^-$  as follows. Sample  $a, b \in \Sigma$  uniformly at random and sample  $c \in \{a, b\}$  uniformly at random. Now let  $y, z$  be as guaranteed by [Lemma 3.10](#), i.e.,  $((a, c, b), y, z) \in \text{supp}(\mu^+)$  and  $((b, c, a), y, z) \in \text{supp}(\mu^+)$ . Include  $((a, c, b), y, z)$  in  $\mu^-$  with probability mass  $\frac{1}{2|\Sigma|^2}$ . Let  $\alpha > 0$  and  $\nu$  be such that  $\mu^+ = \alpha\mu^- + (1 - \alpha)\nu$ . Applying a random restriction and using [\(6\)](#) tells us that

$$\mathbb{E}_{\substack{I \sim_{1-\alpha}[n] \\ ((u_1, u_2, u_3), v, w) \sim \nu^I}} \left| \mathbb{E}_{\substack{(x,y,z) \sim (\mu^-)^{\bar{I}} \\ x=(a,c,b)}} f_{I \rightarrow u_1}(a) \overline{f_{I \rightarrow u_2}(c)} f_{I \rightarrow u_3}(b) g_{I \rightarrow v}(y) h_{I \rightarrow w}(z) \right| \geq \varepsilon^{O(1)}. \quad (7)$$

Define the function  $\tilde{f}_{I,u} : \text{supp}(\mu_x^-)^{\bar{I}} \rightarrow \mathbb{C}$  as:

$$\tilde{f}_{I,u}(a, c, b) := \mathbb{E}_{a', b' \sim (a \leftrightarrow b)} f_{I \rightarrow u_1}(a') \overline{f_{I \rightarrow u_2}(c)} f_{I \rightarrow u_3}(b').$$

Note that by choice of  $y, z$  (via [Lemma 3.10](#)) we have that that

$$\mathbb{E}_{\substack{(x,y,z) \sim (\mu^-)^{\bar{I}} \\ x=(a,c,b)}} f_{I \rightarrow u_1}(a) \overline{f_{I \rightarrow u_2}(c)} f_{I \rightarrow u_3}(b) g_{I \rightarrow v}(y) h_{I \rightarrow w}(z) = \mathbb{E}_{\substack{(x,y,z) \sim (\mu^-)^{\bar{I}} \\ x=(a,c,b)}} \tilde{f}_{I,u}(a, c, b) g_{I \rightarrow v}(y) h_{I \rightarrow w}(z).$$

Because  $g, h$  are 1-bounded this, together with [\(7\)](#), implies that

$$\begin{aligned} \varepsilon^{O(1)} &\leq \mathbb{E}_{\substack{I \sim_{1-\alpha}[n] \\ (u_1, u_2, u_3) \sim \nu_x^I}} \left\| \tilde{f}_{I,u} \right\|_{L^2((\mu_x^-)^{\bar{I}})} \\ &= \mathbb{E}_{\substack{I \sim_{1-\alpha}[n] \\ (u_1, u_2, u_3) \sim \nu_x^I}} \left( \mathbb{E}_{(a,c,b) \sim (\mu_x^-)^{\bar{I}}} \left[ |f_{I \rightarrow u_2}(c)|^2 \left| \mathbb{E}_{(a',b') \sim (a \leftrightarrow b)} f_{I \rightarrow u_1}(a') f_{I \rightarrow u_3}(b') \right|^2 \right] \right)^{1/2} \\ &\leq \mathbb{E}_{\substack{I \sim_{1-\alpha}[n] \\ (u_1, u_2, u_3) \sim \nu_x^I}} \left( \mathbb{E}_{(a,c,b) \sim (\mu_x^-)^{\bar{I}}} \left[ \left| \mathbb{E}_{(a',b') \sim (a \leftrightarrow b)} f_{I \rightarrow u_1}(a') f_{I \rightarrow u_3}(b') \right|^2 \right] \right)^{1/2} \\ &= \mathbb{E}_{\substack{I \sim_{1-\alpha}[n] \\ (u_1, u_2, u_3) \sim \nu_x^I}} \text{swap}(f_{I \rightarrow u_1}, f_{I \rightarrow u_3}, f_{I \rightarrow u_1}, f_{I \rightarrow u_3})^{1/2} \\ &\leq \left( \mathbb{E}_{\substack{I \sim_{1-\alpha}[n] \\ (u_1, u_2, u_3) \sim \nu_x^I}} \text{swap}(f_{I \rightarrow u_1}) \right)^{1/4} \left( \mathbb{E}_{\substack{I \sim_{1-\alpha}[n] \\ (u_1, u_2, u_3) \sim \nu_x^I}} \text{swap}(f_{I \rightarrow u_3}) \right)^{1/4}, \end{aligned}$$

where we have applied Cauchy-Schwarz and [Lemma 3.5 \(3\)](#). To conclude, note that for  $(x_1, x_2, x_3) \sim \mu_x^+$  that both  $x_1$  and  $x_3$  are distributed as  $\mu_x$ , and thus if  $\nu'$  is the distribution of  $u_1$  (or  $u_3$ ) for  $(u_1, u_2, u_3) \sim \nu$ , then  $(1 - \alpha)\nu' + \alpha U = \mu_x$ . This exactly implies the conclusion of [Theorem 7](#).  $\square$

## 4 Proof of [Theorem 8](#): the 99% Regime

The goal of the section is to establish [Theorem 8](#). The key component in the proof is the following lemma, which then implies [Theorem 8](#) by induction.

**Lemma 4.1.** *Let  $f: \Sigma^n \rightarrow \mathbb{C}$  be a function such that  $\|f\|_2 \leq 1$  and  $\text{swap}(f) = 1 - \varepsilon$  for  $\varepsilon \leq 0.05$ . Let  $\lambda_1$  be the maximum singular value under the  $\{1\}$ -SVD of  $f$ . Then:*

- $\lambda_1^2 \geq 1 - 3\varepsilon$ .
- If  $\delta = 1 - \lambda_1^2$  and  $f = \sum_{i=1}^m \lambda_i g_i h_i$  is the  $\{1\}$ -SVD then  $\text{swap}(h_1) \geq (1 - \varepsilon)(1 + \Omega(\delta))$ .

We require a weak bound on  $\text{swap}$  applied to orthogonal functions.

**Lemma 4.2.** *If  $\langle h_1, h_2 \rangle = 0$  and  $\|h_1\|_2 = \|h_2\|_2 = 1$ , then  $\text{swap}(h_1, h_2, h_1, h_2) \leq 2/3$ .*

*Proof.* Let  $\alpha, \beta$  be uniform complex numbers with  $|\alpha| = |\beta| = 1/\sqrt{2}$ . Note that

$$\begin{aligned} 1 &\geq \mathbb{E}_{\alpha, \beta} [\text{swap}(\alpha h_1 + \beta h_2, \alpha h_1 + \beta h_2, \alpha h_1 + \beta h_2, \alpha h_1 + \beta h_2)] \\ &= \frac{1}{4} \text{swap}(h_1) + \frac{1}{4} \text{swap}(h_2) + \text{swap}(h_1, h_2, h_1, h_2) \\ &\geq \frac{1}{2} \sqrt{\text{swap}(h_1) \text{swap}(h_2)} + \text{swap}(h_1, h_2, h_1, h_2) \\ &\geq \frac{3}{2} \text{swap}(h_1, h_2, h_1, h_2), \end{aligned}$$

where the first inequality uses [Lemma 3.4](#) and the last uses [Lemma 3.5](#).  $\square$

Now we can proceed to the proof of [Lemma 4.1](#).

*Proof of [Lemma 4.1](#).* Writing  $f = \sum_i \lambda_i g_i h_i$  in the  $\{1\}$ -SVD, one can check that

$$1 - \varepsilon = \text{swap}(f, f, f, f) = \sum_i \lambda_i^4 \text{swap}(h_i) + 2 \sum_{i < j} \lambda_i^2 \lambda_j^2 \text{swap}(h_i, h_j, h_i, h_j).$$

Applying [Lemma 4.2](#) gives

$$1 - \varepsilon \leq \sum_i \lambda_i^4 \text{swap}(h_i) + 2 \sum_{i < j} \lambda_i^2 \lambda_j^2 \cdot \frac{2}{3} = \sum_i \lambda_i^4 \text{swap}(h_i) + \frac{2}{3} \left( \left( \sum_i \lambda_i^2 \right)^2 - \sum_i \lambda_i^4 \right), \quad (8)$$

which is at most  $\frac{1}{3} \sum_i \lambda_i^4 + \frac{2}{3}$  as  $\text{swap}(h_i) \leq 1$ . Thus  $\sum_i \lambda_i^4 \geq 1 - 3\varepsilon$ , and as  $\sum_i \lambda_i^2 \leq 1$  we conclude that  $\lambda_1^2 \geq 1 - 3\varepsilon$ . Denoting  $\delta = 1 - \lambda_1^2$ , we get that  $\delta \leq 3\varepsilon$  and also that  $\sum_{j \neq 1} \lambda_j^2 \leq \delta$ . Using these notations we conclude from (8) that

$$1 - \varepsilon \leq \sum_i \lambda_i^4 \text{swap}(h_i) + \frac{2}{3} \left( 1 - \sum_i \lambda_i^4 \right) \leq \lambda_1^4 \text{swap}(h_1) + \frac{4}{3} \delta + O(\delta^2),$$

where the final inequality uses  $\lambda_1^4 \geq 1 - 2\delta$ . Thus,

$$\text{swap}(h_1) \geq \frac{1 - \varepsilon - \frac{4}{3}\delta - O(\delta^2)}{(1 - \delta)^2} \geq (1 - \varepsilon)(1 + \Omega(\delta)),$$

for sufficiently small  $\varepsilon$  (recall that  $\delta \leq 3\varepsilon$ ). This completes the proof.  $\square$

We now prove [Theorem 8](#), restated below.

**Theorem 8.** *If  $f : \Sigma^n \rightarrow \mathbb{C}$  satisfies  $\|f\|_2 \leq 1$  and  $\text{swap}(f) \geq 1 - c$  for  $c \leq 0.05$ , then there are  $P_1, \dots, P_n : \Sigma \rightarrow \mathbb{C}$  with  $\|P_i\|_2 = 1$  such that  $|\langle f, P_1 \dots P_n \rangle| \geq 1 - O(c)$ .*

*Proof.* We prove the following claim by induction on  $n$ : there is a constant  $C$  such that for all sufficiently small  $\varepsilon$ , if  $\text{swap}(f) \geq 1 - \varepsilon$  then there is a product function  $P = P_1 \dots P_n$  with  $\|P\|_2 = 1$  and  $\langle f, P \rangle \geq 1 - C\varepsilon$ . It is evident for  $n = 1$ .

Now let  $n > 1$ . Perform the SVD as in [Lemma 4.1](#) and let  $\lambda_1^2 = 1 - \delta$ . We will set  $P = g_1 P'$  where  $\langle h_1, P' \rangle \geq 1 - C(\varepsilon - \Omega(\delta))$  by induction, because  $\text{swap}(h_1) \geq (1 - \varepsilon)(1 + \Omega(\delta))$ . Note that

$$\langle f, P \rangle = \lambda_1 \langle h_1, P' \rangle \geq (1 - \delta)^{1/2} (1 - C(\varepsilon - \Omega(\delta))) \geq 1 - C\varepsilon$$

where the last inequality holds provided that  $C$  is a sufficiently large constant.  $\square$

## 5 Proof of [Theorem 5](#): the Inverse Theorem for the Swap Norm

The goal of this section is to establish [Theorem 5](#). At a high level, the proof of the result proceeds by induction on  $\varepsilon$ . We show that given a function  $f$ , we can do a process consisting of taking random restrictions and SVD decompositions to get a function  $f'$  whose (normalized) value is significantly larger than that of  $f$ . Additionally, the function  $f'$  will have the property that if it is correlated with a product function, then  $f$  is also correlated with a product function (perhaps a different one). Here and throughout, the normalized value of  $f$  is  $\text{swap}(f)^{1/2} / \|f\|_2^2$ ; it will be easier to work with as we are considering restrictions of  $f$  that may have different 2-norms.

### 5.1 An Iterative Scheme

Fix a function  $f : \Sigma^n \rightarrow \mathbb{C}$  with  $\sqrt{\text{swap}(f)} \geq \varepsilon^2$  as in the statement of [Theorem 5](#). In this section we provide an iterative argument that alternately peels off large singular vectors for random partitions and performs random restrictions, with the goal of increasing  $\text{swap}(f)$  so that we can apply induction. At each time step in the iteration we maintain a subset  $R_t \subseteq [n]$  (a random subset of some density), a partition  $R_t = S_t \cup T_t$  (which is uniformly random conditioned on  $R_t$ ), and a function  $f_t : \Sigma^{R_t} \rightarrow \mathbb{C}$  which is a random restriction of  $f$  onto  $R_t$ .

Let us now formally define the iterative scheme. Set  $f_0 = f$ ,  $t = 0$ ,  $R_0 = [n]$ , and take  $S_0, T_0$  as a uniformly random partition of  $R_0$ . We choose  $\gamma = \Theta(\varepsilon^8)$ , to be determined more precisely later.

1. Define  $\Delta_t := f_t - \sum_{i=1}^t \lambda_i g_i h_i$ , where  $\lambda_1, \dots, \lambda_t$  are the  $t$  largest singular values of the  $S_t$ -SVD of  $f_t$  with corresponding singular vectors  $g_i : \Sigma^{S_t} \rightarrow \mathbb{C}$ ,  $h_i : \Sigma^{T_t} \rightarrow \mathbb{C}$ .
2. If  $\text{swap}(\Delta_t)^{1/2} < \gamma \|f_t\|_2^2$  or  $t \geq 8\gamma^{-2}$ , terminate.
3. If  $\text{swap}(\Delta_t)^{1/2} \geq \gamma \|f_t\|_2^2$  and  $t < 8\gamma^{-2}$ , let  $R_t = S' \cup T'$  be a uniformly random partition of  $R_t$ . Set  $S_{t+1} := S_t \cap S'$  and  $T_{t+1} := T_t \cap T'$ . Set  $R_{t+1} := S_{t+1} \cup T_{t+1}$ .
4. Let  $f_{t+1} := (f_t)_{(R_t \setminus R_{t+1}) \rightarrow z}$  for random  $z \sim \Sigma^{R_t \setminus R_{t+1}}$ . Increase  $t$  by 1 and iterate.

We use the following potential function to analyze the progress of the iterative scheme:

$$\Phi_t := \frac{\varepsilon^{-5} \text{swap}(f_t)^{1/2} - \|\Delta_t\|_2^2}{\|f_t\|_2^2}.$$

The idea behind the potential function is that during each iteration, if it does not terminate,  $\Delta_t$  is subtracting off one extra term in the SVD, so  $\|\Delta_t\|_2^2$  should be decreasing and thus increases  $\Phi_t$ . Thus, if  $\text{swap}(f_t)$  stays the same throughout the process, we will eventually have that  $\text{swap}(\Delta_t)$  must be small.

The first step in the analysis is to argue that with non-negligible probability,  $\Phi_t$  does not decrease significantly under a random restriction.

**Lemma 5.1.** *Let  $f, g : \Sigma^n \rightarrow \mathbb{C}$  be functions such that  $f$  is  $B$ -bounded. Let  $I \subseteq [n]$ ,  $C > 1$ , and  $\delta > 0$ . Then with probability at least  $\frac{\delta \|f\|_2^2}{CB^2}$  over  $z \sim \Sigma^I$ :*

$$\frac{C \text{swap}(f_{I \rightarrow z})^{1/2} - \|g_{I \rightarrow z}\|_2^2}{\|f_{I \rightarrow z}\|_2^2} \geq \frac{C \text{swap}(f)^{1/2} - \|g\|_2^2}{\|f\|_2^2} - 2\delta \quad \text{and} \quad \|f_{I \rightarrow z}\|_2^2 \geq \frac{\delta}{C + 2\delta + \frac{\|g\|_2^2}{\|f\|_2^2}} \|f\|_2^2. \quad (9)$$

*Proof.* Let  $\Phi := \frac{C \text{swap}(f)^{1/2} - \|g\|_2^2}{\|f\|_2^2}$ . By Lemma 3.7 we know that:

$$\mathbb{E}_{z \sim \Sigma^I} [C \text{swap}(f_{I \rightarrow z})^{1/2} - \|g_{I \rightarrow z}\|_2^2 - (\Phi - 2\delta) \|f_{I \rightarrow z}\|_2^2] \geq C \text{swap}(f)^{1/2} - \|g\|_2^2 - (\Phi - 2\delta) \|f\|_2^2 = 2\delta \|f\|_2^2.$$

As the random variable under the expectation on the left hand side is at most  $CB^2$  always (since  $f$  is  $B$ -bounded), it follows from an averaging argument that

$$\Pr_{z \sim \Sigma^I} \left[ C \text{swap}(f_{I \rightarrow z})^{1/2} - \|g_{I \rightarrow z}\|_2^2 - (\Phi - 2\delta) \|f_{I \rightarrow z}\|_2^2 \geq \delta \|f\|_2^2 \right] \geq \frac{\delta \|f\|_2^2}{CB^2}.$$

Any  $z$  for which this event holds satisfies the conclusion of the lemma. Indeed, the first inequality in (9) follows by bounding  $\delta \|f\|_2^2 \geq 0$  and re-arranging, and the second inequality follows by bounding  $\text{swap}(f_{I \rightarrow z})^{1/2} \leq \|f_{I \rightarrow z}\|_2^2$  (which holds by Lemma 3.4),  $\|g_{I \rightarrow z}\|_2^2 \geq 0$ ,  $\Phi \geq -\|g\|_2^2 / \|f\|_2^2$  and rearranging.  $\square$

### 5.1.1 Analyzing a Single Iteration

We now state and prove the main iteration lemma, stating that each step is successful with noticeable probability, wherein a successful step is one in which the potential  $\Phi_t$  increases significantly and the 2-norm of  $f_t$  does not decrease too much.

**Lemma 5.2.** *Let  $f_t, R_t = S_t \cup T_t, \Delta_t$  be defined as above. If  $\text{swap}(\Delta_t)^{1/2} \geq \gamma \|f_t\|_2^2$ , then with probability at least  $\Omega(\gamma^4 \varepsilon^5 \|f_t\|_2^6 / B^6)$ , it holds that  $\Phi_{t+1} \geq \Phi_t + \gamma^2/4$  and  $\|f_{t+1}\|_2^2 \geq \Omega(\varepsilon^5 \gamma^2) \|f_t\|_2^2$ .*

*Proof.* Let  $R_t = S' \cup T'$  be a uniformly random partition, as described in the iterative scheme. Define  $\tilde{\Delta} := \Delta_t / \|\Delta_t\|_2$ , so that  $\|\tilde{\Delta}\|_2 = 1$  and  $\text{swap}(\tilde{\Delta}) \geq \gamma^2 \|f_t\|_2^4 \|\Delta_t\|_2^{-4}$  by the assumption that the process has not terminated. By Lemma 3.3 we have that  $\mathbb{E}_S[\text{box}_S(\tilde{\Delta})] = \text{swap}(\tilde{\Delta}) \geq \gamma^2 \|f_t\|_2^4 \|\Delta_t\|_2^{-4}$ , and as  $\text{box}_S(\tilde{\Delta}) \leq 1$  for all  $S$  by Lemma 3.2, we get by an averaging argument that

$$\Pr_{S'} \left[ \text{box}_{S'}(\tilde{\Delta}) \geq \frac{\gamma^2 \|f_t\|_2^4}{2 \|\Delta_t\|_2^4} \right] \geq \frac{\gamma^2 \|f_t\|_2^4}{2 \|\Delta_t\|_2^4} \geq \frac{\gamma^2 \|f_t\|_2^4}{2B^4}.$$

In the last inequality we used the fact that  $\|\Delta_t\|_2^4 \leq \|f_t\|_2^4 \leq B^4$ . Let  $\kappa_1 \geq \kappa_2 \geq \dots$  be the singular values of  $\tilde{\Delta} \in \mathbb{C}^{\Sigma^{S'} \times \Sigma^{T'}}$ . With these notations, we get that with probability at least  $\gamma^2 \|f_t\|_2^4 B^{-4}/2$  over  $S'$

$$\kappa_1^2 \geq \sum_i \kappa_i^4 = \text{box}_{S'}(\tilde{\Delta}) \geq \frac{\gamma^2 \|f_t\|_2^4}{2 \|\Delta_t\|_2^4}. \quad (10)$$

Let  $g : \Sigma^{S'} \rightarrow \mathbb{C}$  and  $h : \Sigma^{T'} \rightarrow \mathbb{C}$  be the unit singular vectors corresponding to  $\kappa_1$  and define  $\Delta'_t := \Delta_t - \|\Delta_t\|_2 \kappa_1 g h$ . Using (10) it follows that

$$\|\Delta'_t\|_2^2 = \|\Delta_t\|_2^2 - \|\Delta_t\|_2^2 \kappa_1^2 \leq \|\Delta_t\|_2^2 - \frac{\gamma^2 \|f_t\|_2^4}{2 \|\Delta_t\|_2^2} \leq \|\Delta_t\|_2^2 - \frac{\gamma^2}{2} \|f_t\|_2^2, \quad (11)$$

where the last transition uses the fact that  $\|\Delta_t\|_2 \leq \|f_t\|_2$ .

Now we will take the random restriction and invoke [Lemma 5.1](#). Let  $I := R_t \setminus R_{t+1}$ ,  $\delta = \gamma^2/8$ ,  $C = \varepsilon^{-5}$ , and use [Lemma 5.1](#) for  $f := f_t, g := \Delta'_t$  to get that with probability at least  $\frac{\gamma^2 \varepsilon^5 \|f_t\|_2^2}{8B^2}$  over  $z \sim \Sigma^I$ ,

$$\frac{\varepsilon^{-5} \text{swap}((f_t)_{I \rightarrow z})^{1/2} - \|(\Delta'_t)_{I \rightarrow z}\|_2^2}{\|(f_t)_{I \rightarrow z}\|_2^2} \geq \frac{\varepsilon^{-5} \text{swap}(f_t)^{1/2} - \|\Delta'_t\|_2^2}{\|f_t\|_2^2} - \frac{\gamma^2}{4},$$

and

$$\|f_{t+1}\|_2^2 = \|(f_t)_{I \rightarrow z}\|_2^2 \geq \Omega(\gamma^2 \varepsilon^5) \|f_t\|_2^2$$

(we used  $\|\Delta'_t\|_2^2 \leq \|f_t\|_2^2$  to replace the  $\|g\|_2^2 / \|f\|_2$  term in [Lemma 5.1](#) by 1). By (11) we know that  $\frac{\varepsilon^{-5} \text{swap}(f_t)^{1/2} - \|\Delta'_t\|_2^2}{\|f_t\|_2^2} \geq \Phi_t + \frac{\gamma^2}{4}$ . Thus, the conclusion follows from the fact that  $f_{t+1} = (f_t)_{I \rightarrow z}$  and  $\|(\Delta'_t)_{I \rightarrow z}\|_2^2 \geq \|\Delta_{t+1}\|_2^2$ . Indeed, the latter inequality follows because

$$(\Delta'_t)_{I \rightarrow z} = f_{t+1} - \sum_{i=1}^t \lambda_i (g_i)_{I \rightarrow z} (h_i)_{I \rightarrow z} - \|\Delta_t\|_2 \kappa_1 g_{I \rightarrow z} h_{I \rightarrow z},$$

and  $(g_1)_{I \rightarrow z}, \dots, (g_t)_{I \rightarrow z}$  and  $g_{I \rightarrow z}$  are all functions from  $\Sigma^{S_{t+1}} \rightarrow \mathbb{C}$ , and  $(h_1)_{I \rightarrow z}, \dots, (h_t)_{I \rightarrow z}$  and  $h_{I \rightarrow z}$  are all functions from  $\Sigma^{T_{t+1}} \rightarrow \mathbb{C}$ , and that the SVD is the optimal way to reduce the  $\ell_2$  norm of a matrix by subtracting rank one matrices.  $\square$

### 5.1.2 Analyzing the Termination State

We next analyze the functions  $f_t$  and  $\Delta_t$  upon the termination of the process. In the following lemma, we show that with noticeable probability, either the process terminated with  $\Delta_t$  with small value while  $\text{swap}(f_t)^{1/2} / \|f_t\|_2^2$  has not decrease by much, or else  $\text{swap}(f_t)^{1/2} / \|f_t\|_2^2$  is noticeably larger than  $\text{swap}(f)^{1/2} / \|f\|_2^2$ . Denote by  $\tau$  the time step on which the iteration scheme ends, so either  $\tau = 8\gamma^{-2}$  or  $\text{swap}(\Delta_\tau)^{1/2} < \gamma \|f_\tau\|_2^2$ .

**Lemma 5.3.** *Let  $T = 8\gamma^{-2}$ . Over the randomized process described above, with probability at least  $(\varepsilon \|f\|_2 / B)^{O(T^2)}$ , we either have (**Case 1**) that:*

$$\frac{\text{swap}(f_T)^{1/2}}{\|f_T\|_2^2} \geq \frac{\text{swap}(f)^{1/2}}{\|f\|_2^2} + \varepsilon^5 \quad \text{and} \quad \|f_T\|_2^2 \geq (\varepsilon \gamma)^{O(T)} \|f\|_2^2,$$

or (**Case 2**) there is some  $\tau \in \{1, 2, \dots, T\}$  such that  $\text{swap}(\Delta_\tau)^{1/2} < \gamma \|f_\tau\|_2^2$ ,  $\|f_t\|_2^2 \geq (\varepsilon \gamma)^{O(t)}$  for all  $t \leq \tau$ , and

$$\frac{\text{swap}(f_\tau)^{1/2}}{\|f_\tau\|_2^2} \geq \frac{\text{swap}(f)^{1/2}}{\|f\|_2^2} - \varepsilon^5.$$

*Proof.* Consider the event that the conclusion of [Lemma 5.2](#) is true at each iteration  $i \leq \tau$ . Under this we get that  $\|f_i\|_2^2 \geq (\varepsilon\gamma)^{O(i)}\|f\|_2^2$  for all  $i = 0, 1, \dots, T$ . The probability of this is at least  $(\varepsilon\gamma/B)^{O(T)} \cdot \prod_{i=0}^t \|f_i\|_2^{O(1)} \geq (\varepsilon\gamma\|f\|_2/B)^{O(T^2)}$ .

In this event, we get that  $\Phi_t \geq \Phi_0 + t\gamma^2/4$ , for  $t \leq \tau$ . Thus

$$\varepsilon^{-5} \left( \frac{\text{swap}(f_t)^{1/2}}{\|f_t\|_2^2} - \frac{\text{swap}(f)^{1/2}}{\|f\|_2^2} \right) \geq t \frac{\gamma^2}{4} + \frac{\|\Delta_t\|_2^2}{\|f_t\|_2^2} - \frac{\|\Delta_0\|_2^2}{\|f\|_2^2} \geq t \frac{\gamma^2}{4} - 1,$$

because  $\Delta_0 = f$ . If  $t = 8\gamma^{-2}$ , we are in Case 1. Otherwise, Case 2 holds.  $\square$

Intuitively, if the first case of [Lemma 5.3](#) holds then we have successfully made progress by increasing the value of  $f_t$ . At that point we could appeal to the inductive hypothesis and conclude that  $f_t$  is correlated with a product function, and thus be done. Otherwise, the second case holds, and we know that we have not lost too much in value when passing from  $f$  to  $f_t$  but at the same time we know that  $\text{swap}(\Delta_t)$  is small, meaning that, in a sense, we have exhausted all of the meaningful parts in the SVD decompositions of  $f$ . In the next subsection, we show how to use this information to construct a function  $f'$  related to  $f$  that has significantly larger value, such that correlations of  $f'$  with product functions are related to correlation of  $f$  with product functions.

### 5.1.3 Analyzing Case 2 in [Lemma 5.3](#)

In this section we provide a way to increment  $\text{swap}(f)$  in the second case in [Lemma 5.3](#).

**Lemma 5.4.** *If  $\text{swap}(\Delta_t)^{1/2} \leq c^2\varepsilon^8\|f_t\|_2^2$ , for sufficiently small  $c$ , and  $\frac{\text{swap}(f_t)^{1/2}}{\|f_t\|_2^2} \in [\varepsilon^2/2, 0.99]$  for some  $t = O(\varepsilon^{-16})$ , then there are  $B\varepsilon^{-O(1)}\|f_t\|_2^{-1}$ -bounded functions  $g : \Sigma^{S_t} \rightarrow \mathbb{C}, h : \Sigma^{T_t} \rightarrow \mathbb{C}$  with  $\|g\|_2 = \|h\|_2 = 1$  and constant  $\eta_{g,h}$  with  $|\eta_{g,h}| \geq \Omega(\varepsilon^8)$  such that at least one of the following holds:*

1.  $\text{swap}(g) \geq (1 + \Omega(\varepsilon^2)) \frac{\text{swap}(f_t)}{\|f_t\|_2^2}$  and for any  $I \subseteq S_t, z \in \Sigma^I$ , and  $P : \Sigma^{S_t \setminus I} \rightarrow \mathbb{C}$  it holds that

$$\langle (f_t)_{I \rightarrow z}, Ph \rangle = \eta_{g,h} \|f_t\|_2 \langle g_{I \rightarrow z}, P \rangle.$$

2.  $\text{swap}(h) \geq (1 + \Omega(\varepsilon^2)) \frac{\text{swap}(f_t)}{\|f_t\|_2^4}$  and for any  $I \subseteq T_t, z \in \Sigma^I$ , and  $P : \Sigma^{T_t \setminus I} \rightarrow \mathbb{C}$  it holds that

$$\langle (f_t)_{I \rightarrow z}, gP \rangle = \eta_{g,h} \|f_t\|_2 \langle h_{I \rightarrow z}, P \rangle.$$

Before proving this, we need a technical lemma, asserting that the singular vectors of large singular values are bounded in  $\ell_\infty$ . This helps to prove that  $g, h$  in [Lemma 5.4](#) are bounded.

**Lemma 5.5.** *Let  $f \in \mathbb{C}^{S \times T}$  and let  $g \in \mathbb{C}^S$  and  $h \in \mathbb{C}^T$  with  $\|g\|_2 = \|h\|_2 = 1$  be singular vectors of  $f$  with singular value  $\lambda$ . Then  $g, h$  are  $B/\lambda$ -bounded.*

*Proof.* Thinking of  $f$  as a matrix and  $g$  and  $h$  as vectors, we have that  $g = \frac{1}{\lambda} f^* h$ . Thus

$$\|g\|_\infty \leq \frac{B\|h\|_1}{\lambda} \leq \frac{B\|h\|_2}{\lambda} = \frac{B}{\lambda}. \quad \square$$

*Proof of [Lemma 5.4](#).* By scaling, we may assume that  $\|f_t\|_2 = 1$ . For simplicity of notation, let  $f := f_t$ , consider the  $S_T$ -SVD decomposition of  $f$ , say  $f = \sum_i \lambda_i g_i h_i$  where  $\lambda_i \geq \lambda_{i+1}$  for all  $i$ .

Define

$$\Delta := f - \sum_{i \in [t], \lambda_i \geq \frac{c\varepsilon^4}{t}} \lambda_i g_i h_i$$

and recall that  $\Delta_t = f - \sum_{i \in [t]} \lambda_i g_i h_i$ . By [Corollary 3.6](#), we know that

$$\text{swap}(\Delta)^{1/4} \leq \text{swap}(\Delta_t)^{1/4} + \text{swap}(\Delta - \Delta_t)^{1/4} \leq c\varepsilon^4 + \sum_{i \in [t], \lambda_i < \frac{c\varepsilon^4}{t}} \lambda_i \leq 2c\varepsilon^4.$$

Thus,  $\Delta$  has a small value and we henceforth work with it. For simplicity, relabel  $t$  as the number of  $\lambda_i$  subtracted in the definition of  $\Delta$ , so that  $\Delta = f - \sum_{i=1}^t \lambda_i g_i h_i$ . We remark that (by the upper bound on  $t$  in the statement) we get that  $\lambda_i \geq \Omega(c\varepsilon^{20})$  for each  $i = 1, \dots, t$ .

Let  $F = f - \Delta = \sum_{i=1}^t \lambda_i g_i h_i$ . By [Corollary 3.6](#) we know that

$$\text{swap}(F)^{1/4} \geq \text{swap}(f)^{1/4} - \text{swap}(\Delta)^{1/4} \geq \text{swap}(f)^{1/4}(1 - 4c\varepsilon^2). \quad (12)$$

The remaining analysis is split into two cases depending on the size of  $\lambda_1$ .

**Case 1:**  $\lambda_1 \geq 1 - \varepsilon^2/10^{10}$ . By [Lemma 3.4](#) and [Corollary 3.6](#) we know that

$$\begin{aligned} \text{swap}(\lambda_1 g_1 h_1)^{1/4} &\geq \text{swap}(F)^{1/4} - \text{swap}\left(\sum_{i=2}^t \lambda_i g_i h_i\right)^{1/4} \\ &\geq \text{swap}(f)^{1/4}(1 - 4c\varepsilon^2) - \left\| \sum_{i=2}^t \lambda_i g_i h_i \right\|_2 \\ &\geq 0.999 \text{swap}(f)^{1/4}, \end{aligned}$$

where we have used that  $\left\| \sum_{i=2}^t \lambda_i g_i h_i \right\|_2 \leq \sqrt{1 - \lambda_1^2} \leq \varepsilon/10^5$ . Because  $\text{swap}(g_1 h_1) = \text{swap}(g_1)\text{swap}(h_1)$  we conclude that because  $\text{swap}(f) \leq 0.99$ ,

$$\text{swap}(g_1) \geq (1 + \Omega(1))\text{swap}(f) \quad \text{or} \quad \text{swap}(h_1) \geq (1 + \Omega(1))\text{swap}(f).$$

Assume the latter by symmetry. In this case, let  $g = g_1, h = h_1$ . Then item 2 follows because

$$\langle f_{I \rightarrow z}, gP \rangle = \left\langle \sum_i \lambda_i g_i (h_i)_{I \rightarrow z}, g_1 P \right\rangle = \lambda_1 \langle h_{I \rightarrow z}, P \rangle.$$

and  $g_1, h_1$  are  $O(B)$ -bounded by [Lemma 5.5](#), and  $\eta_{g,h} := \lambda_1$ .

**Case 2:**  $\lambda_1 \leq 1 - \varepsilon^2/10^{10}$ . Let  $\delta = 1 - \lambda_1^2$ . Define

$$\text{swap}_{T_t}(F, F, F, F) := \mathbb{E}_{\substack{x, y \sim \Sigma^{R_t} \\ (x', y') \sim (x \leftrightarrow_{T_t} y)}} F(x)F(y)\overline{F(x')} \overline{F(y')},$$

where  $(x', y') \sim (x \leftrightarrow_{T_t} y)$  denotes that  $(x'_i, y'_i) = (x_i, y_i)$  for  $i \in R_t \setminus T_t$ , and otherwise  $(x'_i, y'_i) = (x_i, y_i)$  or  $(y_i, x_i)$  with probability  $1/2$  for  $i \in T_t$ . Then we can see that

$$\text{swap}(f)(1 - 16c\varepsilon^2) \leq \text{swap}(F) \leq \text{swap}_{T_t}(F) = \sum_{i=1}^t \lambda_i^4 \text{swap}(h_i, h_i, h_i, h_i) + 2 \sum_{1 \leq i < j \leq t} \lambda_i^2 \lambda_j^2 \text{swap}(h_i, h_j, h_i, h_j). \quad (13)$$



Here, the first inequality is by (12), and the second follows because (using (4))

$$\begin{aligned}
\text{swap}(f, f, f, f) &= \mathbb{E}_{x, y \sim \Sigma^n} \left| \mathbb{E}_{(x', y') \sim (x \leftrightarrow y)} f_1(x') f_2(y') \right|^2 \\
&= \mathbb{E}_{x, y \sim \Sigma^n} \left| \mathbb{E}_{(x', y') \sim (x \leftrightarrow y)} \mathbb{E}_{(x'', y'') \sim (x' \leftrightarrow_{T_t} y')} f_1(x'') f_2(y'') \right|^2 \\
&\leq \mathbb{E}_{x, y \sim \Sigma^n} \left| \mathbb{E}_{(x', y') \sim (x \leftrightarrow y)} \mathbb{E}_{(x'', y'') \sim (x' \leftrightarrow_{T_t} y')} f_1(x'') f_2(y'') \right|^2 \\
&= \mathbb{E}_{x, y \sim \Sigma^n} \left| \mathbb{E}_{(x', y') \sim (x \leftrightarrow_{T_t} y)} f_1(x') f_2(y') \right|^2 \\
&= \text{swap}_{T_t}(f, f, f, f).
\end{aligned}$$

We further split into two cases depending on the size of the second term in (13).

**Case 2(a): The second term is large.** Consider the case that

$$2 \sum_{1 \leq i < j \leq t} \lambda_i^2 \lambda_j^2 \text{swap}(h_i, h_j, h_i, h_j) \geq \frac{\delta}{4} \text{swap}(f).$$

In this case let  $\beta_1, \dots, \beta_t \in \mathbb{C}$  be i.i.d. unit complex numbers, let  $\alpha_i = t^{-1/2} \beta_i$ , and  $g := \sum_{i=1}^t \overline{\alpha_i} g_i$ . Then  $\|g\|_2 = 1$  and

$$\langle f_{I \rightarrow z}, gP \rangle = \left\langle \sum_i \lambda_i g_i (h_i)_{I \rightarrow z}, \left( \sum_{i=1}^t \overline{\alpha_i} g_i \right) P \right\rangle = \left\langle \sum_{i=1}^t \lambda_i \alpha_i (h_i)_{I \rightarrow z}, P \right\rangle = t^{-1/2} \langle h_{I \rightarrow z}, P \rangle$$

for  $h := \sum_{i=1}^t \beta_i \lambda_i h_i$ , so  $\|h\|_2 \leq 1$ , so the second part of item 2 holds with  $\eta_{g, h} = t^{-1/2} \geq \Omega(\varepsilon^8)$ . A calculation yields that

$$\begin{aligned}
\mathbb{E}_{\beta_1, \dots, \beta_t} \text{swap}(h) &= \sum_{i, j, k, \ell \in [t]} \lambda_i \lambda_j \overline{\lambda_k \lambda_\ell} \text{swap}(h_i, h_j, h_k, h_\ell) \mathbb{E}[\beta_i \beta_j \overline{\beta_k \beta_\ell}] \\
&= \sum_{i=1}^t \lambda_i^4 \text{swap}(h_i, h_i, h_i, h_i) + 4 \sum_{1 \leq i < j \leq t} \lambda_i^2 \lambda_j^2 \text{swap}(h_i, h_j, h_i, h_j) \\
&\geq \text{swap}(f)(1 - 16c\varepsilon^2) + 2 \sum_{1 \leq i < j \leq t} \lambda_i^2 \lambda_j^2 \text{swap}(h_i, h_j, h_i, h_j) \\
&\geq \text{swap}(f)(1 - 16c\varepsilon^2) + \frac{\delta}{4} \text{swap}(f) \geq \text{swap}(f)(1 + \Omega(\varepsilon^2)),
\end{aligned}$$

for sufficiently small  $c$ . Here, the second equality follows because  $\mathbb{E}[\beta_i \beta_j \overline{\beta_k \beta_\ell}] = 1$  only if  $\{i, j\} = \{k, \ell\}$ , and otherwise is 0. Also, the first inequality uses (13). Thus, item 2 holds for some choice of  $\beta_1, \dots, \beta_t$ . The functions  $g$  and  $h$  are  $B\varepsilon^{-O(1)}$  bounded by Lemma 5.5 because  $\lambda_i \geq \varepsilon^{O(1)}$  for all  $i$ .

**Case 2(b): The second term is small.** Precisely, that

$$2 \sum_{1 \leq i < j \leq t} \lambda_i^2 \lambda_j^2 \text{swap}(h_i, h_j, h_i, h_j) \leq \frac{\delta}{4} \text{swap}(f).$$

In this case, we know that

$$\sum_{i=1}^t \lambda_i^4 \text{swap}(h_i, h_i, h_i, h_i) \geq \text{swap}(f)(1 - 16c\varepsilon^2) - \frac{\delta}{4} \text{swap}(f) \geq (1 - \delta/3) \text{swap}(f).$$

Now we remove all small  $\lambda_i$  to get that

$$\sum_{1 \leq i \leq t, \lambda_i^2 > \frac{\delta \text{swap}(f)}{10}} \lambda_i^4 \text{swap}(h_i, h_i, h_i, h_i) \geq (1 - \delta/3) \text{swap}(f) - \delta \text{swap}(f)/10 \geq (1 - \delta/2) \text{swap}(f).$$

To conclude, note that

$$\max_{i: \lambda_i^2 > \frac{\delta \text{swap}(f)}{10}} \text{swap}(h_i) \geq \frac{(1 - \delta/2) \text{swap}(f)}{\sum_i \lambda_i^4} \geq \frac{(1 - \delta/2) \text{swap}(f)}{\lambda_1^2} \geq (1 + \Omega(\delta)) \text{swap}(f).$$

Thus, picking  $i$  achieving the maximum on the left hand side, we may set  $g = g_i, h = h_i$  and recall from above that  $\langle f_{I \rightarrow z}, gP \rangle = \lambda_i \langle (h_i)_{I \rightarrow z}, P \rangle$ , where  $\lambda_i^2 \geq \Omega(\varepsilon^6)$ . Thus the second part of item 2 holds for  $\eta_{g,h} := \lambda_i \geq \Omega(\varepsilon^3)$ . Finally,  $g_i, h_i$  are  $B\varepsilon^{-O(1)}$  bounded by [Lemma 5.5](#).  $\square$

## 5.2 Main Induction: Proving [Theorem 5](#)

Now we apply the iteration scheme of [Section 5.1](#) to prove [Theorem 5](#), restated below, by an inductive argument.

**Theorem 5.** *If  $f : \Sigma^n \rightarrow \mathbb{C}$  is a  $B$ -bounded function satisfying that  $\|f\|_2 \leq 1$ ,  $\text{swap}(f)^{1/4} \geq \varepsilon$ , then there is a constant  $\delta(\varepsilon, B) > 0$  such that:*

$$\Pr_{I \sim_{1-\delta(\varepsilon, B)} [n], z \sim \Sigma^I} \left[ \exists \{P_i : \Sigma \rightarrow \mathbb{C}, \|P_i\|_2 \leq 1\}_{i \in \bar{I}} \text{ with } \left| \langle f_{I \rightarrow z}, \prod_{i \in \bar{I}} P_i \rangle \right| \geq \delta(\varepsilon, B) \right] \geq \delta(\varepsilon, B).$$

Quantitatively,  $\delta(\varepsilon, B) \geq (\varepsilon/B)^{O(\varepsilon^{-O(1)})}$ .

*Proof.* We prove the following statement by induction on  $\varepsilon$  and  $B$ : under the hypotheses of [Theorem 5](#), there is a constant  $\delta(\varepsilon, B)$  such that

$$\mathbb{E}_{\substack{I \sim_{1-\delta(\varepsilon, B)} [n] \\ z \sim \Sigma^I}} \left[ \sup \left\{ \left| \langle f_{I \rightarrow z}, \prod_{i \in \bar{I}} P_i \rangle \right| : P_i : \Sigma \rightarrow \mathbb{C}, \|P_i\|_2 \leq 1 \text{ for } i \in \bar{I} \right\} \right] \geq \delta(\varepsilon, B). \quad (14)$$

[Theorem 5](#) follows from (14), because  $f$  is  $B$ -bounded. If  $\varepsilon \geq 0.99$  then (14) follows from [Theorem 8](#). For the inductive step, note that for each  $i = 0, 1, \dots, t$ , the distribution of set  $R_i$  is  $R_i \sim_{2^{-i}} [n]$ . Also, conditioned on  $R_i, S_i \sim_{1/2} R_i$ . Let  $\gamma = c\varepsilon^8$  for a sufficiently small constant  $c$ . The conclusion of [Lemma 5.3](#) holds with probability at least  $q(\varepsilon, B) := (\varepsilon/B)^{O(T^2)}$ . In Case 1 of [Lemma 5.3](#), let  $\tau := 8\gamma^{-2}$ , and in Case 2 let  $\tau$  be as in the statement.

We first consider Case 1, when  $\tau = 8\gamma^{-2}$ . If so then

$$\frac{\text{swap}(f_\tau)^{1/2}}{\|f_\tau\|_2^2} \geq \frac{\text{swap}(f)^{1/2}}{\|f\|_2^2} + \varepsilon^5 \geq \varepsilon^2 + \varepsilon^5$$

by [Lemma 5.3](#). Let  $\varepsilon' := (\varepsilon^2 + \varepsilon^5)^{1/2} = \varepsilon + \Omega(\varepsilon^4)$  and  $B' := B\|f\|_2/\|f_\tau\|_2 \geq B\varepsilon^{-O(\tau)}$  by [Lemma 5.3](#). By induction, a random restriction of  $f_\tau$  correlates to a product function, i.e.,

$$\mathbb{E}_{\substack{I \sim_{1-\delta(\varepsilon', B')} R_t \\ z \sim \Sigma^I}} \left[ \sup \left\{ \left| \left\langle (f_\tau)_{I \rightarrow z}, \prod_{i \in R_t \setminus I} P_i \right\rangle \right| : P_i : \Sigma \rightarrow \mathbb{C}, \|P_i\|_2 \leq 1 \text{ for } i \in R_t \setminus I \right\} \right] \geq \delta(\varepsilon', B') \|f_t\|_2 \\ \geq \delta(\varepsilon', B') \varepsilon^{O(\tau)}.$$

Because restricting onto  $(R_t \setminus I) \sim_{\delta(\varepsilon', B')} R_\tau$  is equivalent in distribution to restricting onto a set sampled  $\sim_{2^{-\tau}\delta(\varepsilon', B')} [n]$ , we conclude that:

$$\mathbb{E}_{\substack{I \sim_{1-2^{-\tau}\delta(\varepsilon', B')} [n] \\ z \sim \Sigma^I}} \left[ \sup \left\{ \left| \left\langle f_{I \rightarrow z}, \prod_{i \in \bar{I}} P_i \right\rangle \right| : P_i : \Sigma \rightarrow \mathbb{C}, \|P_i\|_2 \leq 1 \text{ for } i \in \bar{I} \right\} \right] \geq q(\varepsilon, B) \delta(\varepsilon', B') \varepsilon^{O(\tau)}.$$

Thus this case follows by induction for the choice (where  $\tau \leq \varepsilon^{-O(1)}$ ) for

$$\delta(\varepsilon, B) := q(\varepsilon, B) \varepsilon^{O(\tau)} \delta(\varepsilon + \Omega(\varepsilon^4), B\varepsilon^{-O(\tau)}).$$

In Case 2, by symmetry we may assume that item 2 of [Lemma 5.4](#) holds. Then with probability at least  $q(\varepsilon, B)/2$  over  $R_t \sim_{2^{-\tau}} [n]$ ,  $z \sim \Sigma^{[n] \setminus R_t}$ , and  $S_t \sim_{1/2} R_t$ , that there are functions  $g, h$  satisfying item 2 of [Lemma 5.4](#). By item 2 of [Lemma 5.4](#) we know that

$$\text{swap}(h) \geq (1 + \Omega(\varepsilon^2)) \frac{\text{swap}(f_\tau)}{\|f_\tau\|_2^4} \geq (1 + \Omega(\varepsilon^2)) \left( \frac{\text{swap}(f)^{1/2}}{\|f\|_2^2} - \varepsilon^5 \right)^2 \geq (1 + \Omega(\varepsilon^2)) \text{swap}(f),$$

as  $\text{swap}(f)^{1/2} \geq \varepsilon^2$  and  $\|f\|_2 = 1$ . Let  $\varepsilon' = \varepsilon + \Omega(\varepsilon^3)$  and  $B' = B\varepsilon^{-O(\tau)}$  so that  $h$  is  $B'$ -bounded (recall that  $\|f_\tau\|_2^2 \geq \varepsilon^{O(\tau)}$ ). By induction we know that a random restriction of  $h$  correlates to a product function, i.e.,

$$\mathbb{E}_{\substack{I \sim_{1-\delta(\varepsilon', B')} T_t \\ z \sim \Sigma^I}} \left[ \sup \left\{ \left| \left\langle h_{I \rightarrow z}, \prod_{i \in T_t \setminus I} P_i \right\rangle \right| : P_i : \Sigma \rightarrow \mathbb{C}, \|P_i\|_2 \leq 1 \text{ for } i \in T_t \setminus I \right\} \right] \geq \delta(\varepsilon', B'). \quad (15)$$

In particular, note that the random restriction in the above equation is independent of  $S_t$ . For a fixed restriction  $I \sim_{1-\delta(\varepsilon', B')} T_t$ ,  $z \sim \Sigma^I$  let  $P_{I,z} := \prod_{i \in T_t \setminus I} P_i$  denote the product function guaranteed in (15). Then, we know that:

$$\mathbb{E}_{\substack{I \sim_{1-\delta(\varepsilon', B')} T_t \\ z \sim \Sigma^I}} \mathbb{E}_{z' \sim \Sigma^{S_t}} \left| \langle (f_\tau)_{(S_t, I) \rightarrow (z', z)}, g_{S_t \rightarrow z'} P_{I,z} \rangle \right| \geq \mathbb{E}_{\substack{I \sim_{1-\delta(\varepsilon', B')} T_t \\ z \sim \Sigma^I}} \left| \langle (f_\tau)_{I \rightarrow z}, g P_{I,z} \rangle \right| \\ = \eta_{g,h} \|f_\tau\|_2 \mathbb{E}_{\substack{I \sim_{1-\delta(\varepsilon', B')} T_t \\ z \sim \Sigma^I}} \left| \langle h_{I \rightarrow z}, P_{I,z} \rangle \right| \\ \geq \varepsilon^{O(\tau+1)} \delta(\varepsilon', B'),$$

where the first transition is by the triangle inequality, the second transition is by [Lemma 5.4](#), and the third one is by the choice of  $P_{I,z}$ . The result follows because:

1. The hypotheses required to apply [Lemma 5.4](#) occur with probability at least  $q(\varepsilon, B)$ .
2.  $g_{S_t \rightarrow z'}$  is a constant bounded by  $B\varepsilon^{-O(\tau)}$  (by [Lemma 5.4](#)), and
3.  $(f_\tau)_{(S_t, I) \rightarrow (z', z)}$  has identical distribution as  $f_{J \rightarrow x}$  for  $J \sim_{1-2^{-(\tau+1)}\delta(\varepsilon', B')}$  and  $x \sim \Sigma^J$ .

This completes the induction, by setting  $\delta(\varepsilon, B) = B^{-1}\varepsilon^{O(\tau+1)}q(\varepsilon, B)\delta(\varepsilon', B')$ . Solving the recursion for  $\delta$  gives the quantitative bounds claimed in the theorem.  $\square$

## 6 Correlation to Bounded Product Functions

Note that [Theorem 5](#) says that (under random restriction) there is an  $\ell_2$ -bounded product function that correlates with the 1-bounded function  $f$ . In this section, we prove that this random restriction must in fact correlate to a 1-bounded product function.

**Theorem 10.** *Let  $f : \Sigma^n \rightarrow \mathbb{C}$  be a function with  $\|f\|_4 \leq 1$  and  $P_i : \Sigma \rightarrow \mathbb{C}$  be such that  $\|P_i\|_2 \leq 1$  and*

$$\left| \mathbb{E}_{x \sim \Sigma^n} [f(x) \prod_{i=1}^n P_i(x_i)] \right| \geq \delta.$$

*Then there are functions  $P'_i : \Sigma \rightarrow \mathbb{D}$  such that*

$$\left| \mathbb{E}_{x \sim \Sigma^n} [f(x) \prod_{i=1}^n P'_i(x_i)] \right| \geq \delta^{O(1)}.$$

Fix  $f$  and  $P$  as in the setting of the theorem. By making small perturbations in  $P$  we may assume that  $P(x) \neq 0$  for all  $x$ . Our proof relies on the following standard approximation of intervals by exponential sums.

**Lemma 6.1.** *For any real numbers  $R \geq L$  and  $\eta$  there is  $c : \mathbb{R} \rightarrow \mathbb{C}$  and function  $b(x) := \int_{-\infty}^{\infty} c(\theta) e^{2\pi i \theta x}$  satisfies:*

- $\int_{-\infty}^{\infty} |c(\theta)| d\theta \leq \left( \frac{R-L}{\eta} \right)^{1/2}$ , and
- For  $x \in [L + \eta/2, R - \eta/2]$  it holds that  $b(x) = 1$
- For  $x \in (-\infty, L - \eta/2] \cup [R + \eta/2, \infty)$  it holds that  $b(x) = 0$ ,
- For all  $x \in \mathbb{R}$  it holds that  $0 \leq b(x) \leq 1$ .

*Proof.* Let  $I_{[A, B]}$  be the indicator function of the interval  $[A, B]$ . Let  $b = \eta^{-1} I_{[L, R]} * I_{[-\eta/2, \eta/2]}$ . All items except the first are evident. For the first item, for a function  $g : \mathbb{R} \rightarrow \mathbb{C}$  let  $\widehat{g}(\theta) := \int_{-\infty}^{\infty} g(x) e^{-2\pi i \theta x} dx$  be the Fourier transform. By Fourier inversion,

$$\begin{aligned} \int_{-\infty}^{\infty} |c(\theta)| d\theta &= \frac{1}{\eta} \int_{-\infty}^{\infty} |\widehat{I_{[L, R]}}(\theta)| |\widehat{I_{[-\eta/2, \eta/2]}}(\theta)| d\theta \\ &\leq \frac{1}{\eta} \left( \int_{-\infty}^{\infty} |\widehat{I_{[L, R]}}(\theta)|^2 d\theta \right)^{1/2} \left( \int_{-\infty}^{\infty} |\widehat{I_{[-\eta/2, \eta/2]}}(\theta)|^2 d\theta \right)^{1/2} = \left( \frac{R-L}{\eta} \right)^{1/2}, \end{aligned}$$

by Parseval's identity.  $\square$

Now we move towards the proof of [Theorem 10](#). Write  $P_i = Q_i R_i$  where  $R_i : \Sigma \rightarrow \mathbb{R}_{>0}$ , and  $|Q_i(x)| = 1$  for all  $x \in \Sigma$ . By replacing  $f \rightarrow f \prod_{i=1}^n Q_i$  we may assume that  $P_i = R_i$  in fact, i.e.,  $P_i$  take nonnegative real values.

Define the set  $S_{L,R} := \{x \in \Sigma^n : e^L \leq P(x) \leq e^R\}$  and let  $\mu(S_{L,R}) := \Pr_{x \in \Sigma^n}[x \in S_{L,R}]$  and  $A_{L,R} := \mathbb{E}_{x \in \Sigma^n}[f(x)P(x)1_{x \in S_{L,R}}]$ . Let  $D = \log(16/\delta^2)$ .

**Lemma 6.2.** *It holds that  $A_{-D,D} \geq \delta/2$ .*

*Proof.* First we bound

$$\left| \mathbb{E}_{x \sim \Sigma^n} [f(x)P(x)1_{P(x) \leq \delta^2/16}] \right| \leq \frac{\delta}{4},$$

because  $\|f\|_1 \leq 1$ . Also, by Hölder's inequality and  $\|f\|_4 \leq 1$

$$\left| \mathbb{E}_{x \sim \Sigma^n} [f(x)P(x)1_{P(x) \leq 16/\delta^2}] \right| \leq \left| \mathbb{E}_{x \sim \Sigma^n} [|P(x)|^{4/3} 1_{P(x) \leq 16/\delta^2}] \right|^{3/4} \leq \frac{\delta}{4} \mathbb{E}_{x \sim \Sigma^n} |P(x)|^2 \leq \frac{\delta}{4},$$

because  $f$  is 1-bounded and  $\|P\|_2 \leq 1$ . □

*Proof of Theorem 10.* Let  $\alpha, \eta < 1$  be parameters to be chosen later. Let  $-D - 1 \leq L \leq D$  and  $R = L + \alpha$ . Let  $c_{L,R}(\theta)$  be the coefficients guaranteed by [Lemma 6.1](#). Then

$$\int_{-\infty}^{\infty} c_{L,R}(\theta) \left( \mathbb{E}_{x \sim \Sigma^n} [f(x)e^{2\pi i \theta \ln P(x)}] \right) d\theta = \mathbb{E}_{x \sim \Sigma^n} [f(x)g_{L,R}(x)],$$

where  $g_{L,R}(x) = b(\ln P(x))$  and  $b$  is the function in [Lemma 6.1](#). By the properties of  $g_{L,R}$ ,

$$\left| \mathbb{E}_{x \in \Sigma^n} [f(x)g_{L,R}(x)] \right| \geq \left| \mathbb{E}_{x \sim \Sigma^n} f(x)1_{x \in S_{L,R}} \right| - \mu(S_{L-\eta/2, L+\eta/2}) - \mu(S_{R-\eta/2, R+\eta/2}).$$

Furthermore, we can bound (using  $-D - 1 \leq L \leq D$ )

$$\begin{aligned} \left| \mathbb{E}_{x \sim \Sigma^n} f(x)1_{x \in S_{L,R}} \right| &\geq e^{-L} \left| \mathbb{E}_{x \sim \Sigma^n} f(x)P(x)1_{x \in S_{L,R}} \right| - \mathbb{E}_{x \sim \Sigma^n} [|e^{-L}P(x) - 1| \cdot |f(x)|1_{x \in S_{L,R}}] \\ &\geq e^{-L} A_{L,R} - 2\alpha\mu(S_{L,R}), \end{aligned}$$

because  $|e^\alpha - 1| \leq 2\alpha$  (for  $\alpha \leq 1$ ) and  $\|f\|_1 \leq 1$ . Overall, we have concluded that

$$\begin{aligned} \int_{-\infty}^{\infty} c_{L,R}(\theta) \left( \mathbb{E}_{x \sim \Sigma^n} [f(x)e^{2\pi i \theta \ln P(x)}] \right) d\theta \\ \geq e^{-L} A_{L,R} - 2\alpha\mu(S_{L,R}) - \mu(S_{L-\eta/2, L+\eta/2}) - \mu(S_{R-\eta/2, R+\eta/2}). \end{aligned}$$

The expectation of the RHS over  $L \sim [-D - 1, D]$  chosen uniformly at random is at least

$$\frac{\alpha A_{-D,D}}{e^{-D}(2D+1)} - \frac{4\alpha^2}{2D+1} - \frac{4\eta}{2D+1} \geq \frac{\alpha\delta^3}{32(2D+1)} - \frac{4\alpha^2}{2D+1} - \frac{4\eta}{2D+1}, \quad (16)$$

where the last inequality is by [Lemma 6.2](#). Here, we used the fact that

$$\mathbb{E}_L [e^{-L} A_{L,R}] \geq e^{-D} \mathbb{E}_L [A_{L,R}] \geq e^{-D} A_{-D,D} \frac{\alpha}{2D+1},$$

and that similar calculations give

$$\mathbb{E}_L[\mu(S_{L,R})] \leq \frac{2\alpha}{2D+1}, \quad \mathbb{E}_L[\mu(S_{L-\eta/2,L+\eta/2})] \leq \frac{2\eta}{2D+1}, \quad \mathbb{E}_L[\mu(S_{R-\eta/2,R+\eta/2})] \leq \frac{2\eta}{2D+1}.$$

Choosing  $\alpha = \delta^3/1000$  and  $\eta = \delta^6/400000$  the expression in (16) is at least  $\frac{\delta^6}{100000(2D+1)}$ . Hence

$$\int_{-\infty}^{\infty} c_{L,R}(\theta) \left( \mathbb{E}_{x \sim \Sigma^n} [f(x)e^{2\pi i \theta \ln P(x)}] \right) d\theta \geq \frac{\delta^6}{100000(2D+1)}$$

for some choice of  $L$ . Because  $\int_{-\infty}^{\infty} |c_{L,R}(\theta)| d\theta \leq (\alpha/\eta)^{1/2} = 20/\delta$ , we conclude that

$$\left| \mathbb{E}_{x \sim \Sigma^n} [f(x)e^{2\pi i \theta \ln P(x)}] \right| \geq \frac{\delta^7}{2000000(2D+1)}$$

for some  $\theta$ . This concludes the proof, because  $e^{2\pi i \theta \ln P(x)}$  is a product function in which each component is into the unit disc.  $\square$

**Proof of Theorem 2:** Theorem 2 immediately follows by combining Theorems 5, 7 and 10.

## 7 Restriction Inverse Theorem

The goal of this section is to establish Theorem 9.

### 7.1 Preliminaries on Product Functions

In this section we collect preliminary facts about product functions and their *distance*. For  $I \subseteq [n]$  define  $\mathcal{F}(I)$  to be the set of product functions  $P : \Sigma^I \rightarrow \mathbb{C}$  such that  $|P_i(x)| = 1$  for all  $x \in \Sigma^n$  and  $i \in [n]$ . We start by discussing how to convert such a product function  $P : \Sigma^n \rightarrow \mathbb{C}$  into a vector.

Throughout this section we will use the notation  $O_{\rho,\gamma}(\cdot)$  to denote that we are suppressing constants depending on  $\rho, \gamma$ , with polynomial dependence on the parameters being suppressed. For example, if  $\rho, \gamma < 1$ , this means  $O((\rho\gamma)^{-O(1)})$ . It will be convenient for us to fix a special reference input  $\sigma \in \Sigma$ .

**Definition 7.1.** Let  $P(x) = \prod_{i=1}^n P_i(x_i)$  be a product function with  $P_i(\sigma) \in \mathbb{R}_{\geq 0}$  for all  $i \in [n]$ . Define  $\pi(P) \in \mathbb{R}^{2n|A|}$  as  $\pi(P)_{i,\sigma'} = P_i(\sigma') \in \mathbb{C}$ . For general  $\tilde{P} = \prod_{i=1}^n \tilde{P}_i$ , let  $c_1, \dots, c_n \in \mathbb{C}$  with  $|c_i| = 1$  be such that  $c_i \tilde{P}_i(\sigma) \in \mathbb{R}_{\geq 0}$ . Then define  $\pi(\tilde{P}) := \pi(\prod_{i=1}^n c_i \tilde{P}_i)$ .

Informally,  $\pi(P)$  is defined by concatenating  $P_i(\sigma')$  for  $i \in [n], \sigma' \in \Sigma$ , if  $P_i(\sigma)$  are real. Note that  $\pi(P)$  is a  $2n|A|$  dimensional because  $P_i(\sigma')$  are complex numbers. The reason we are taking the reference point  $\sigma$  and normalizing appropriately so that the  $P(\sigma)$  is real valued is to ensure that  $\pi(P)$  is invariant under rotation by a unit complex number. Our next goal is to establish that  $|\langle P, Q \rangle|$  for product functions  $P, Q$  is related to  $\|\pi(P) - \pi(Q)\|_2$ . The statement below asserts that if  $\pi(P)$  and  $\pi(Q)$  are far, then  $P$  and  $Q$  can only have a small correlation. Throughout this section, the implicit constant  $\Omega$  hides factors that depend on the probability of the smallest atom in  $\mu$ .

**Lemma 7.2.** *Let  $P, Q : \Sigma^n \rightarrow \mathbb{C}$  be 1-bounded product functions. Then*

$$|\langle P, Q \rangle| \leq \exp(-\Omega(\|\pi(P) - \pi(Q)\|_2^2)).$$

*Proof.* We may assume that  $P_i(\sigma), Q_i(\sigma) \in \mathbb{R}_{\geq 0}$  for all  $i \in [n]$ . Because  $|\langle P, Q \rangle| = \prod_{i=1}^n |\langle P_i, Q_i \rangle|$ , it suffices to argue that for each  $i$ ,

$$|\langle P_i, Q_i \rangle| \leq 1 - \Omega(\|P_i - Q_i\|_2^2).$$

Fix  $i$  and let  $\theta \in \mathbb{C}$  be of absolute value 1 such that  $|\langle P_i, Q_i \rangle| = \theta \langle P_i, Q_i \rangle$ . Then

$$|\langle P_i, Q_i \rangle| = \operatorname{Re}(\langle \theta P_i, Q_i \rangle) = \frac{1}{2} \mathbb{E}_x [|P_i(x)|^2 + |Q_i(x)|^2 - |\theta P_i(x) - Q_i(x)|^2].$$

Denote  $\tau_i = \|P_i - Q_i\|_2^2$ . If  $|P_i(x)| \leq 1 - \frac{1}{8}\tau_i$  or  $|Q_i(x)| \leq 1 - \frac{1}{8}\tau_i$  for some  $x$ , then (using the 1-boundedness of  $P_i, Q_i$ ) we immediately get from the previous inequality that  $|\langle P_i, Q_i \rangle| \leq 1 - \Omega(\tau_i)$ , and we are done. Hence, we assume henceforth that  $|P_i(x)|, |Q_i(x)| > 1 - \frac{1}{8}\tau_i$  for all  $x$ . We get that

$$|\langle P_i, Q_i \rangle| \leq 1 - \frac{1}{2} \mathbb{E}_x [|\theta P_i(x) - Q_i(x)|^2].$$

In the rest of the argument we show that  $\mathbb{E}_x [|\theta P_i(x) - Q_i(x)|^2] \geq \Omega(\tau_i)$  by case analysis. If  $|\theta - 1| \leq \frac{1}{2}\sqrt{\tau_i}$ , then by the triangle inequality

$$\mathbb{E}_x [|\theta P_i(x) - Q_i(x)|^2]^{1/2} \geq \mathbb{E}_x [|P_i(x) - Q_i(x)|^2]^{1/2} - |1 - \theta| \mathbb{E}_x [|P_i(x)|^2]^{1/2} \geq \frac{1}{2}\sqrt{\tau_i},$$

and so  $\mathbb{E}_x [|\theta P_i(x) - Q_i(x)|^2] \geq \frac{1}{4}\tau_i$  and we are done. Else,  $|\theta - 1| > \frac{1}{2}\sqrt{\tau_i}$ , and then

$$\mathbb{E}_x [|\theta P_i(x) - Q_i(x)|^2] \geq \Omega(|\theta P_i(\sigma) - Q_i(\sigma)|^2) \geq |P_i(\sigma)\operatorname{Re}(\theta) - Q_i(\sigma)|^2 + |P_i(\sigma)\operatorname{Im}(\theta)|^2.$$

If  $|\operatorname{Im}(\theta)| \geq \frac{1}{4}\sqrt{\tau_i}$  then the second term on the right hand side is already  $\frac{1}{16}\tau_i$ , and we are done. Else,  $|\operatorname{Im}(\theta)| \leq \frac{1}{4}\sqrt{\tau_i}$  and  $|\theta - 1| > \frac{1}{2}\sqrt{\tau_i}$ , so it follows that  $\operatorname{Re}(\theta) \leq 1 - \frac{1}{4}\sqrt{\tau_i}$ . If  $\operatorname{Re}(\theta) \leq 0$  then  $|P_i(\sigma)\operatorname{Re}(\theta) - Q_i(\sigma)| \geq |Q_i(\sigma)| \geq 1/2$ , and we are done. Else,

$$|P_i(\sigma)\operatorname{Re}(\theta) - Q_i(\sigma)| \geq |Q_i(\sigma)| - |P_i(\sigma)\operatorname{Re}(\theta)| \geq 1 - \frac{1}{8}\tau_i - \left(1 - \frac{1}{4}\sqrt{\tau_i}\right) \cdot 1 = \frac{1}{8}\sqrt{\tau_i}. \quad \square$$

We will require a partial converse to [Lemma 7.2](#). It essentially says that if  $\|\pi(P) - \pi(Q)\|_2^2 \leq \varepsilon$ , then  $\|P - Q\|_2^2$  is small up to some transformations.

**Lemma 7.3.** *If  $P, Q \in \mathcal{F}([n])$  then there is some  $|c| = 1$  such that  $\|cP - Q\|_2^2 \leq \|\pi(P) - \pi(Q)\|_2^2$ .*

*Proof.* Let  $|c_i| = 1$  be such that  $c_i \langle P_i, Q_i \rangle \in \mathbb{R}_{\geq 0}$ , and let  $c = \prod c_i$ . Note that

$$\|P_i - Q_i\|_2^2 = 2 - \langle P_i, Q_i \rangle - \langle Q_i, P_i \rangle,$$

and hence  $|\langle P_i, Q_i \rangle| \geq 1 - \|P_i - Q_i\|_2^2/2$ .

$$\begin{aligned} \|cP - Q\|_2^2 &= 2 - \mathbb{E}_x \prod_{i=1}^n c_i P_i(x_i) \overline{Q_i(x_i)} - \mathbb{E}_x \prod_{i=1}^n \overline{c_i P_i(x_i)} Q_i(x_i) \\ &= 2 - 2 \prod_{i=1}^n |\langle P_i, Q_i \rangle| \leq 2 - 2 \prod_{i=1}^n (1 - \|P_i - Q_i\|_2^2/2) \\ &\leq \sum_{i=1}^n \|P_i - Q_i\|_2^2 = \|\pi(P) - \pi(Q)\|_2^2, \end{aligned}$$

as desired. □

Now we will describe a procedure which takes a function  $f$  and produces a small list of product functions which essentially capture the set of all product functions with nontrivial correlation to  $f$ . We begin by defining the list of all product functions that are correlated with  $f$ .

**Definition 7.4.** Given a 1-bounded function  $f : \Sigma^n \rightarrow \mathbb{C}$ , define  $\text{List}_\varepsilon[f]$  to be the set of product functions  $P \in \mathcal{F}([n])$  with  $|\langle f, P \rangle| \geq \varepsilon$ .

The next result extracts from  $\text{List}_\varepsilon[f]$  a short list that is essentially a net.

**Lemma 7.5** ([BKM24a, Lemma 12.16]). *Let  $f : \Sigma^n \rightarrow \mathbb{C}$  be a 1-bounded function. There is some  $\text{ShortList}_{\varepsilon, \delta}[f] \subseteq \text{List}_\varepsilon[f]$  satisfying:*

1.  $|\text{ShortList}_{\varepsilon, \delta}[f]| \leq \frac{1}{\varepsilon^2 - \delta}$ , and
2. For all  $P \in \text{List}_\varepsilon[f]$  there is  $P' \in \text{ShortList}_{\varepsilon, \delta}[f]$  with  $|\langle P, P' \rangle| \geq \delta$ .

We will require a lemma which says that if  $f$  correlates to a product function, then it does so with nonnegligible probability under random restriction.

**Lemma 7.6.** *Let  $f : \Sigma^n \rightarrow \mathbb{C}$  be a 1-bounded function, and  $P$  be a 1-bounded product function with  $|\langle f, P \rangle| \geq \varepsilon$ . For any subset  $I \subseteq [n]$ , it holds that*

$$\Pr_{z \sim \Sigma^I} [|\langle f_{I \rightarrow z}, P_{I \rightarrow z} \rangle| \geq \varepsilon/2] \geq \varepsilon/2.$$

*Proof.* Follows by a simple averaging argument, as  $f, P$  are 1-bounded. □

## 7.2 From Random Restrictions to a Direct Product Test

We will require the following theorem which turns local agreement into global agreement. If  $D = 0$ , it would follow by standard small-set-expansion arguments.

**Theorem 11.** *Let  $M, n$  be positive integers and  $\rho, \gamma \in (0, 1)$ . Let  $f : \{0, 1\}^n \rightarrow \mathbb{R}^M$  be such that:*

$$\Pr_{x \sim \rho^{[n]}, y \sim_{1-\gamma} x} [\|f(x) - f(y)\|_2 \leq D] \geq \varepsilon.$$

*Then*

$$\Pr_{x, y \sim \rho^{[n]}} [\|f(x) - f(y)\|_2 \leq O_{\rho, \gamma}(D \log(1/\varepsilon))] \geq \varepsilon^{O_{\rho, \gamma}(1)}.$$

*Proof.* Deferred to [Section 7.3](#). □

The hypothesis of [Theorem 9](#) guarantees that a random restriction of  $f$  correlates to a product function on  $\Sigma^n$ . We can think of the product function that correlates with  $f_{I \rightarrow z}$  as an element of  $(\mathbb{R}^{2^{|\Sigma|}})^I$ . Consider  $F : \{0, 1\}^n \rightarrow (\mathbb{R}^{2^{|\Sigma|}})^{\leq n}$  as the map from  $I$  to the correlating product function (later we argue that the product function correlating to  $f_{I \rightarrow z}$  does not heavily depend on  $z$ ). We will prove that this function  $F[I]$  satisfies the following *direct product test*.

**Definition 7.7** (Direct product test). For  $\rho, \alpha > 0$ , we say that a function  $F : \{0, 1\}^n \rightarrow (\mathbb{R}^K)^{\leq n}$  for integer  $K$  and parameter  $D$  passes the  $\text{DP}(\rho, \alpha)$  with probability:

$$\Pr_{\substack{C \sim_{\alpha \rho} [n] \\ A, B \sim_{\rho} [n]: A, B \supseteq C}} [\|F[A]|_C - F[B]|_C\|_2 \leq D].$$

Here, the probability is over  $C$  sampled from  $\sim_{\rho} [n]$  and  $A, B$  sampled from  $\sim_{\rho} [n]$  conditioned on  $A \supseteq C$  and  $B \supseteq C$ .



Then, we prove the following inverse theorem for the direct product test of [Definition 7.7](#).

**Theorem 12.** *Let  $n, K$  be positive integers such that  $F : \{0, 1\}^n \rightarrow ([0, 1]^K)^{\leq n}$  with parameter  $D > 1$  passes  $\text{DP}(\rho, \alpha)$  with probability at least  $\varepsilon$ . Then there is a function  $g : [n] \rightarrow [0, 1]^K$  with*

$$\Pr_{A \sim \rho^{[n]}} \left[ \|g|_A - F[A]\|_2 \leq O_{\rho, \alpha}(D \log(1/\varepsilon) + \log(1/\varepsilon)^{3/2}) \right] \geq \varepsilon^{O_{\rho, \alpha}(1)}.$$

Also, for each  $x \in [n]$  there is some  $A \subseteq [n]$  such that  $g(x) = F[A]_x$ .

The remainder of the section contains the proof of [Theorem 9](#) given [Theorems 11](#) and [12](#).

### 7.2.1 Local structure

Note that the short lists of  $f_{\bar{I} \rightarrow z}$  (as in [Lemma 7.5](#)) may depend on both  $I$  and  $z$ . Our next goal will be to define short lists that do not depend on  $z$  and use these to define a direct product test.

Let  $K > C_\rho \log(1/\varepsilon)^{O(1)}$  and  $\zeta < \varepsilon^{O_\rho(1)}$  for sufficiently large constant  $C_\rho$ . For  $I \subseteq [n]$  define

$$W_I := \{P \in \mathcal{F}(I) : \Pr_{z \in \Sigma^{\bar{I}}} [\exists Q \in \text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\bar{I} \rightarrow z}], \|\pi(P) - \pi(Q)\|_2^2 \leq K] \geq \zeta\},$$

i.e.,  $W_I$  is the set of product functions that are close to something in  $\text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\bar{I} \rightarrow z}]$  for a noticeable fraction of assignments  $z$  to  $\bar{I}$ . The next lemma produces a short list capturing  $W_I$ .

**Lemma 7.8.** *There is a subset  $SW_I \subseteq W_I$  satisfying:*

1.  $|SW_I| \leq O(\varepsilon^{-2} \zeta^{-1})$ .
2. For all  $P \in W_I$  there is  $Q \in SW_I$  with  $\|\pi(P) - \pi(Q)\|_2^2 \leq 4K$ .

*Proof.* For  $Q \in \mathcal{F}(I)$  define

$$S(Q) := \{(z, P) : P \in \text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\bar{I} \rightarrow z}], \|\pi(P) - \pi(Q)\|_2^2 \leq K\}.$$

Let  $SW_I$  consist of a maximal set of  $Q \in W_I$  such that  $S(Q)$  are disjoint. We first prove that item 2 is satisfied. Consider a  $P \in W_I$  where  $(z, Q') \in S(P) \cap S(Q)$  for some  $Q \in SW_I$ . Then  $\|\pi(P) - \pi(Q)\|_2^2 \leq 2\|\pi(P) - \pi(Q')\|_2^2 + 2\|\pi(Q') - \pi(Q)\|_2^2 \leq 4K$ , by the triangle inequality.

Towards item 1, for a subset  $S \subseteq \Sigma^{\bar{I}} \times \mathcal{F}(I)$  define its *index* as  $\mathcal{I}(S) = \sum_{(z, P) \in S} \mu(z)$ , where  $\mu$  is the measure over  $z \in \Sigma^{\bar{I}}$ . Then

$$\zeta |SW_I| \leq \mathcal{I}(\cup_{Q \in SW_I} S(Q)) \leq O(\varepsilon^{-2}),$$

where the first inequality is by disjointness and the definition of  $S(Q)$ , and the second inequality is by [Lemma 7.5](#). Indeed, note that if  $(z, P) \in S(Q)$  then  $P \in \text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\bar{I} \rightarrow z}]$ , and each short list has size at most  $O(\varepsilon^{-2})$ .  $\square$

Our next goal is to argue that for many  $I' \subseteq_{1/2} I \subseteq_\rho [n]$  it holds that  $SW_{I'}$  is nonempty. This is by a small-set-expansion argument.

**Lemma 7.9.** *It holds that*

$$\Pr_{I' \subseteq_{\rho/2} [n]} [|SW_{I'}| > 0] \geq \Omega(\varepsilon^7).$$

*Proof.* For each  $z \in \Sigma^{\overline{I}}$ , let  $Q_z \in \text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\overline{I} \rightarrow z}]$  be random. We argue that

$$\Pr_{\substack{I \subseteq \rho[n], z \sim \Sigma^{\overline{I}} \\ I' \subseteq_{1/2} I, z', z'' \sim \Sigma^{I \setminus I'}}} \left[ \|\pi(Q_{(z, z')}) - \pi(Q_{(z, z'')})\|_2^2 \leq O(\log(1/\varepsilon)) \right] \geq \Omega(\varepsilon^7). \quad (17)$$

Indeed, by the hypothesis in [Theorem 9](#), with probability  $\varepsilon$  over  $(I, z)$  there is some  $P$  with  $|\langle f_{\overline{I} \rightarrow z}, P \rangle| \geq \varepsilon$ . By [Lemma 7.6](#),  $|\langle f_{\overline{I} \rightarrow (z, z')}, P|_{I'} \rangle| \geq \varepsilon/2$  with probability at least  $\varepsilon/2$  over  $z'$ , and thus there is some  $Q \in \text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\overline{I} \rightarrow (z, z')}]$  with  $|\langle P|_{I'}, Q \rangle| \geq \varepsilon^2/10$ , and hence  $\|\pi(P|_{I'}) - \pi(Q)\|_2^2 \leq O(\log(1/\varepsilon))$  by [Lemma 7.2](#).

Hence with probability at least  $(\varepsilon/2)^2$  conditioned on  $(I, z)$ , there are

$$Q' \in \text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\overline{I} \rightarrow (z, z')}] \quad \text{and} \quad Q'' \in \text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\overline{I} \rightarrow (z, z'')}]$$

with  $\|\pi(P|_{I'}) - \pi(Q')\|_2^2 \leq O(\log(1/\varepsilon))$  and  $\|\pi(P|_{I'}) - \pi(Q'')\|_2^2 \leq O(\log(1/\varepsilon))$ , so  $\|\pi(Q') - \pi(Q'')\|_2^2 \leq O(\log(1/\varepsilon))$  by the triangle inequality. The probability that these specific  $Q', Q''$  are picked from the short lists is  $\Omega(\varepsilon^4)$ , by [Lemma 7.5](#). Thus (17) follows.

Now for  $K = C \log(1/\varepsilon)^{O(1)}$  and small  $\zeta = \varepsilon^{O_\rho(1)}$ , the lemma follows from [Theorem 11](#).  $\square$

Now we define a relaxed version of  $W_I$ . Define

$$\widetilde{W}_I := \{P \in \mathcal{F}(I) : \Pr_{z \in \Sigma^{\overline{I}}} [\exists Q \in \text{ShortList}_{\varepsilon/4, \varepsilon^2/100}[f_{\overline{I} \rightarrow z}], \|\pi(P) - \pi(Q)\|_2^2 \leq O(K + \log(1/\varepsilon))] \geq \zeta \varepsilon/4\}.$$

We can define  $\widetilde{SW}_I \subseteq \widetilde{W}_I$  as in [Lemma 7.8](#). Finally we argue that if  $|SW_{I'}| > 0$  and  $P \in SW_{I'}$  then for any  $I'' \subseteq I'$  it holds that  $P|_{I''} \in \widetilde{SW}_{I''}$ .

**Lemma 7.10.** *Let  $P \in SW_{I'}$  and  $I'' \subseteq I'$ . Then  $P|_{I''} \in \widetilde{SW}_{I''}$ .*

*Proof.* Let  $z \sim \Sigma^{\overline{I}}$  and  $Q \in \text{ShortList}_{\varepsilon/2, \varepsilon^2/10}[f_{\overline{I} \rightarrow z}]$  be such that  $\|\pi(P) - \pi(Q)\|_2^2 \leq K$ . Such  $Q$  exists with probability at least  $\zeta$  over  $z$  because  $P \in SW_{I'}$ . By [Lemma 7.6](#) we know that

$$\Pr_{z' \sim \Sigma^{I' \setminus I''}} \left[ |\langle f_{\overline{I} \rightarrow (z, z')}, Q|_{I''} \rangle| \geq \varepsilon/4 \right] \geq \varepsilon/4.$$

Thus with probability at least  $\zeta \varepsilon/4$  over  $(z, z')$ ,  $Q|_{I''} \in \text{List}_{\varepsilon/4}[f_{\overline{I} \rightarrow (z, z')}]$ . By the definition of  $\text{ShortList}_{\varepsilon/4, \varepsilon^2/100}$  and [Lemma 7.2](#), we know that there is  $P' \in \text{ShortList}_{\varepsilon/4, \varepsilon^2/100}$  with  $\|\pi(P') - \pi(Q|_{I''})\|_2^2 \leq O(\log(1/\varepsilon))$ . Thus

$$\|\pi(P') - \pi(P|_{I''})\|_2^2 \leq 2\|\pi(P') - \pi(Q|_{I''})\|_2^2 + 2\|\pi(P) - \pi(Q)\|_2^2 \leq O(K + \log(1/\varepsilon)).$$

This holds with probability at least  $\zeta \varepsilon/4$  over  $(z, z')$  as desired.  $\square$

## 7.2.2 Applying the Direct Product Test

Define a function  $F : \{0, 1\}^n \rightarrow (\mathbb{R}^{2^{|\Sigma|}})^{\leq n}$  as follows. Let  $I \subseteq [n]$  and define  $F[I]$  to be a uniformly random element of  $\widetilde{SW}_I$  (and apply  $\pi$ ) if it is nonempty, and otherwise a random point. The following lemma bounds the success probability of the direct product test.

**Lemma 7.11.** *It holds that*

$$\Pr_{\substack{I' \subseteq_{\rho/2} [n] \\ I'', I''' \subseteq_{1/2} I'}} \left[ \|F[I'']_{I'' \cap I'''} - F[I''']_{I'' \cap I'''}\|_2^2 \leq O(K + \log(1/\varepsilon)) \right] \geq \Omega(\zeta^2 \varepsilon^{11}).$$

*Proof.* By Lemma 7.9 we know that  $|SW_{I'}| > 0$  with probability at least  $\Omega(\varepsilon^7)$ . Fix  $P \in SW_{I'}$ . By Lemma 7.10 we know that  $P|_{I''} \in \widetilde{SW}_{I''}, \widetilde{SW}_{I''}$ . The result follows from  $|\widetilde{SW}_{I''}|, |\widetilde{SW}_{I''}| \leq O(\zeta^{-1}\varepsilon^{-2})$  by Lemma 7.8 and the triangle inequality.  $\square$

We now have all the pieces necessary to prove Theorem 9.

*Proof of Theorem 9.* By Lemma 7.11 and Theorem 12 there is a 1-bounded global function  $g : [n] \rightarrow \mathbb{C}^{|\Sigma|}$  and constants  $K' \leq O_\rho((K + \log(1/\varepsilon)) \log(1/\varepsilon)^{O(1)})$  and  $\varepsilon' \geq (\varepsilon\zeta)^{O_\rho(1)}$  such that

$$\Pr_{I \subseteq \rho/4[n]} [\|g|_I - F[I]\|_2^2 \leq K'] \geq \varepsilon'. \quad (18)$$

Let  $G(x) = \prod_{i=1}^n g(i)_{x_i}$  be the product function corresponding to  $g$ . Let  $\gamma = \varepsilon^3/(CK')$  for sufficiently large constant  $C$ .

By (18) and the definition of  $\widetilde{SW}_I$ , we know that

$$\Pr_{I \subseteq \rho/4[n], z \sim \Sigma^I} [\exists Q \in \mathcal{F}(I), |\langle f_{I \rightarrow z}, Q \rangle| \geq \varepsilon/2, \|\pi(G|_I) - \pi(Q)\|_2^2 \leq O(K')] \geq \varepsilon'\zeta.$$

For this fixed pair  $(I, z)$  we conclude that

$$\Pr_{I' \subseteq \gamma I, z' \sim \Sigma^{I \setminus I'}} [|\langle f_{I' \rightarrow (z, z')}, Q_{(I \setminus I') \rightarrow z'} \rangle| \geq \varepsilon/4, \|\pi(G|_{I'}) - \pi(Q_{(I \setminus I') \rightarrow z'})\|_2^2 \leq c\varepsilon^2] \geq \varepsilon/8. \quad (19)$$

This follows because  $|\langle f_{I' \rightarrow (z, z')}, Q_{(I \setminus I') \rightarrow z'} \rangle| \geq \varepsilon/4$  with probability at least  $\varepsilon/4$  by Lemma 7.6, and also,  $\mathbb{E}[\|G|_{I'} - Q_{(I \setminus I') \rightarrow z'}\|_2^2] \leq O(\gamma K')$ , and hence

$$\Pr[\|\pi(G|_{I'}) - \pi(Q_{(I \setminus I') \rightarrow z'})\|_2^2 > c\varepsilon^2] \leq O(\gamma K'/(c\varepsilon^2)) \leq \varepsilon/8,$$

by Markov's inequality and the choice of  $\gamma$ . Let  $|\alpha| = 1$  be such that

$$\|\alpha G|_{I'} - Q_{(I \setminus I') \rightarrow z'}\|_2^2 \leq c\varepsilon^2,$$

which exists by Lemma 7.3. Thus

$$\begin{aligned} \text{Stab}_0((f\overline{G})_{I \rightarrow (z, z')}) &= |\langle f_{I \rightarrow (z, z')}, \alpha G|_{I'} \rangle|^2 \\ &\geq \left( |\langle f_{I' \rightarrow (z, z')}, Q_{(I \setminus I') \rightarrow z'} \rangle| - \|\alpha G|_{I'} - Q_{(I \setminus I') \rightarrow z'}\|_2 \right)^2 \\ &\geq \frac{\varepsilon^2}{64}. \end{aligned}$$

Thus

$$\text{Stab}_{1-\rho\gamma/4}(f\overline{G}) = \mathbb{E}_{\substack{I \subseteq \rho/4[n], I' \subseteq \gamma I \\ z \sim \Sigma^I, z' \sim \Sigma^{I \setminus I'}}} [\text{Stab}_0((f\overline{G})_{I \rightarrow (z, z')})] \geq \varepsilon'\zeta \cdot \frac{\varepsilon}{8} \cdot \frac{\varepsilon^2}{64} \geq \frac{\varepsilon'\zeta\varepsilon^3}{512}.$$

Thus Theorem 9 follows for degree  $D := O(\rho\gamma \log(1/\delta'))$  for  $\delta' := \frac{\varepsilon'\zeta\varepsilon^3}{2000}$ .  $\square$

### 7.3 Small-Set Expansion

The goal of this section is to establish [Theorem 11](#). We require the following standard result on small-set expansion in the biased hypercube.

**Theorem 13** (Small-set expansion). *Let  $\rho, \gamma \in (0, 1)$ , and  $A \subseteq \{0, 1\}^n$ . Let  $\mu(A) := \Pr_{x \sim \rho^{[n]}}[x \in A]$ . There is a constant  $c = \Omega_{\rho, \gamma}(1)$  such that for all  $A$ ,  $\Pr_{x \sim \rho^{[n]}, y \sim_{1-\gamma}(x)}[x, y \in A] \leq \mu(A)^{1+c}$ .*

*Proof.* Let  $T_\gamma: L_2(\{0, 1\}^n; \mu_\rho^{\otimes n}) \rightarrow L_2(\{0, 1\}^n; \mu_\rho^{\otimes n})$  be the averaging operator defined as  $T_\gamma f(x) = \mathbb{E}_{y \sim_{1-\gamma}(x)} f(y)$ . Then

$$\Pr_{x \sim \rho^{[n]}, y \sim_{1-\gamma}(x)}[x, y \in A] = \langle 1_A, T_\gamma 1_A \rangle \leq \|1_A\|_2 \|T_\gamma 1_A\|_2.$$

By [[O'D14](#), Chapter 10] we have that  $\|T_\gamma 1_A\|_2 \leq \|1_A\|_{2-c}$  for some  $c = c(\rho, \gamma) > 0$ , giving us that

$$\Pr_{x \sim \rho^{[n]}, y \sim_{1-\gamma}(x)}[x, y \in A] \leq \|1_A\|_2 \|1_A\|_{2-c} = \mu(A)^{1/2} \cdot \mu(A)^{1/(2-c)} \leq \mu(A)^{1+c/4}. \quad \square$$

Our first goal is to apply [Theorem 13](#) to establish an analogue of [Theorem 11](#) in the setting where the dimension  $m$  is small.

**Lemma 7.12.** *Let  $m, n$  be positive integers and  $\rho, \gamma \in (0, 1)$ . Let  $f: \{0, 1\}^n \rightarrow \mathbb{R}^m$  be such that:*

$$\Pr_{x \sim \rho^{[n]}, y \sim_{1-\gamma} x}[\|f(x) - f(y)\|_2 \leq D] \geq \varepsilon.$$

Then

$$\Pr_{x, y \sim \rho^{[n]}}[\|f(x) - f(y)\|_2 \leq 2Dm] \geq (\varepsilon/2)^{O_{\rho, \gamma}(1)}.$$

*Proof.* For all  $v \in \mathbb{R}^m$ , let

$$S_v = \{x \in \{0, 1\}^n : f(x)_i \in [v_i, v_i + 2D\sqrt{m}] \text{ for all } i \in [m]\}.$$

Note that each  $x \in \mathbb{R}^n$ ,  $\int_{v \in \mathbb{R}^m} 1_{x \in S_v} = (2D\sqrt{m})^m$ . Also, for each  $x, y \in \mathbb{R}^m$  with  $\|x - y\|_2 \leq D$  it holds that

$$\int_{v \in \mathbb{R}^m} 1_{x, y \in S_v} = \prod_{i=1}^m (2D\sqrt{m} - |x_i - y_i|) \geq (2D\sqrt{m})^m \left(1 - \sum_{i=1}^m \frac{|x_i - y_i|}{2D\sqrt{m}}\right) \geq (2D\sqrt{m})^m / 2.$$

Combining these along with small-set-expansion ([Theorem 13](#))

$$\begin{aligned} \varepsilon &\leq \Pr_{x \sim \rho^{[n]}, y \sim_{1-\gamma} x}[\|f(x) - f(y)\|_2 \leq D] \\ &\leq \frac{2}{(2D\sqrt{m})^m} \mathbb{E}_{x \sim \rho^{[n]}, y \sim_{1-\gamma} x} \left[ \int_{v \in \mathbb{R}^m} 1_{x, y \in S_v} \right] \\ &\leq \frac{2}{(2D\sqrt{m})^m} \int_{v \in \mathbb{R}^m} \mu(S_v)^{1+c} \\ &\leq 2 \max_v \mu(S_v)^c, \end{aligned}$$

where we have applied [Theorem 13](#). Thus,  $\mu(S_v) \geq (\varepsilon/2)^{O_{\rho, \nu}(1)}$  for some  $v \in \mathbb{Z}^m$ . To conclude, note for any  $x, y \in S_v$  that  $\|f(x) - f(y)\|_2 \leq 2Dm$  and

$$\Pr_{x, y \sim \rho^{[n]}}[\|f(x) - f(y)\|_2 \leq 2Dm] \geq \mu(S_v)^2 \geq (\varepsilon/2)^{O_{\rho, \gamma}(1)}. \quad \square$$

Now to prove [Theorem 11](#) we will essentially randomly project the values of  $f(x)$  down to a constant number of dimensions, and then apply [Lemma 7.12](#). Towards this we first prove that a random projection indeed preserves the  $\ell_2$  norm of pairwise distances in a way that we need.

**Lemma 7.13.** *Let  $u, v \in \mathbb{R}^M$  and let  $g_1, \dots, g_m \sim \mathcal{N}(0, I)$  be uniform Gaussian vectors in  $M$  dimensions. Then for  $u' = m^{-1/2}(\langle u, g_1 \rangle, \dots, \langle u, g_m \rangle) \in \mathbb{R}^m$  and  $v' = m^{-1/2}(\langle v, g_1 \rangle, \dots, \langle v, g_m \rangle)$  it holds that*

$$\Pr[\|u' - v'\|_2^2 > 2\|u - v\|_2^2] \leq e^{-m/10} \quad \text{and} \quad \Pr[\|u' - v'\|_2^2 < \delta\|u - v\|_2^2] \leq (10\delta)^{m/2}.$$

*Proof.* We may assume without loss of generality that  $v = 0$  and that  $\|u\|_2 = 1$ . The first bound then follows from the Hanson-Wright inequality, because each  $\langle u, g_i \rangle$  is Gaussian with variance  $\|u\|_2^2 = 1$ . For the second bound let  $\eta_i = \langle u, g_i \rangle / \|u\|_2$ , and note that

$$\Pr[\|u'\|_2^2 \leq \delta\|u\|_2^2] = \int_{\sum_{i=1}^m \eta_i^2 \leq \delta m} \prod_{i=1}^m \frac{1}{\sqrt{2\pi}} e^{-\eta_i^2/2} \leq (2\pi)^{-m/2} \text{Vol}(\sqrt{\delta m} B_m) \leq (10\delta)^{m/2},$$

where  $B_m$  is the unit ball in  $m$  dimensions. □

At this point, we are ready to proceed to the proof of [Theorem 11](#).

*Proof of Theorem 11.* Let  $m, \delta$  be parameters chosen later. By Markov's inequality and [Lemma 7.13](#) there are  $g_1, \dots, g_m \in \mathbb{R}^M$  such that  $g(u) = m^{-1/2}(\langle u, g_1 \rangle, \dots, \langle u, g_m \rangle)$  satisfies:

$$\Pr_{x, y \sim \rho[n]} [\|g(u) - g(v)\|_2 < \delta^{1/2} \|f(u) - f(v)\|_2] \leq 3(10\delta)^{m/2} \tag{20}$$

and

$$\Pr_{x \sim \rho[n], y \sim 1-\gamma x} [\|g(u) - g(v)\|_2 \leq 2D] \geq \varepsilon - 3e^{-m/10} \geq \varepsilon/2$$

for  $m = 100C_{\rho, \gamma} \log(1/\varepsilon)$  for sufficiently large constant  $C_{\rho, \gamma} = O_{\rho, \gamma}(1)$ . Combining the latter equation with [Lemma 7.12](#) gives

$$\Pr_{x, y \sim \rho[n]} [\|g(x) - g(y)\|_2 \leq 4Dm] \geq (\varepsilon/2)^{O_{\rho, \gamma}(1)} \geq \varepsilon^{O_{\rho, \gamma}(1)}. \tag{21}$$

Let  $\delta$  be sufficiently small so that  $6(10\delta)^{m/2} < \varepsilon^{O_{\rho, \gamma}(1)}$ , here the constant  $C_{\rho, \gamma}$  in the definition of  $m$  is sufficiently large. Combining (21) with (20) gives

$$\Pr_{x, y \sim \rho[n]} [\|f(u) - f(v)\|_2 \leq 4\delta^{-1/2} Dm] \geq \varepsilon^{O_{\rho, \gamma}(1)} - 3(10\delta)^{m/2} \geq \varepsilon^{O_{\rho, \gamma}(1)},$$

as desired. □

## 7.4 Analysis of Direct Product Test

In this section we prove [Theorem 12](#), restated below.

**Theorem 12.** *Let  $n, K$  be positive integers such that  $F : \{0, 1\}^n \rightarrow ([0, 1]^K)^{\leq n}$  with parameter  $D > 1$  passes  $\text{DP}(\rho, \alpha)$  with probability at least  $\varepsilon$ . Then there is a function  $g : [n] \rightarrow [0, 1]^K$  with*

$$\Pr_{A \sim \rho[n]} \left[ \|g|_A - F[A]\|_2 \leq O_{\rho, \alpha}(D \log(1/\varepsilon) + \log(1/\varepsilon)^{3/2}) \right] \geq \varepsilon^{O_{\rho, \alpha}(1)}.$$

Also, for each  $x \in [n]$  there is some  $A \subseteq [n]$  such that  $g(x) = F[A]_x$ .

We first rephrase the distribution in the direct product test. Note that we may find constants  $\gamma, \gamma' \in (0, 1)$  such that the distribution over  $A, B, C$  in [Definition 7.7](#) is equivalent to sampling  $A \sim_\rho [n]$ ,  $B \sim_{1-\gamma} A$ , and  $C \subseteq_{\gamma'} A \cap B$ . Here  $B \sim_{1-\gamma} A$  means that  $A, B$  are  $1 - \gamma$  correlated.

Let us work with this formulation from now on. Our first goal is to establish that  $\|F[A]|_{A \cap B} - F[B]|_{A \cap B}\|_2$  is small.

**Lemma 7.14.** *Under the hypotheses of [Theorem 12](#) it holds that*

$$\Pr_{A \sim_\rho [n], B \sim_{1-\gamma} A} \left[ \|F[A]|_{A \cap B} - F[B]|_{A \cap B}\|_2 \leq O_{\rho, \gamma}(D + \log(1/\varepsilon)^{1/2}) \right] \geq \frac{2\varepsilon}{3}.$$

*Proof.* Follows by the fact that  $C \subseteq_{\gamma'} A \cap B$  for some  $\gamma' > 0$  and the Chernoff bound.  $\square$

For the remainder of the section, let  $D' = O(D + \log(1/\varepsilon)^{1/2})$ .

**Definition 7.15** (Consistent). We say that  $A, B$  are  $D'$ -consistent if  $\|F[A]|_{A \cap B} - F[B]|_{A \cap B}\|_2 \leq D'$ .

If  $A, B$  are  $D'$ -consistent we sometimes write that  $A \in \text{Cons}(B)$  or  $B \in \text{Cons}(A)$ .

**Definition 7.16** (Good). We say that  $A$  is good if  $\Pr_{B \sim_{1-\gamma} A} [B \in \text{Cons}(A)] \geq \varepsilon/3$ .

By an averaging argument, we know that  $\Pr_{A \sim_\rho [n]} [A \text{ is good}] \geq \varepsilon/3$ . Let  $r$  be a parameter to be determined.

**Definition 7.17** (Excellent). For some  $D'' > D'$ , we say that  $A$  is excellent if  $A$  is good and

$$\Pr_{B \sim_{1-\gamma} A} \left[ \mathbb{E}_{B' \sim_{1-\gamma} A} [\|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2^2 \mid B' \in \text{Cons}(A)] > (D'')^2 \text{ and } B \in \text{Cons}(A) \right] \leq r.$$

Now we establish for large enough  $D''$  that the probability that  $A$  is good but not excellent is exponentially small.

**Lemma 7.18.** *If  $D'' > C_{\rho, \gamma} D'$  then*

$$\Pr_{A \sim_\rho [n]} [A \text{ is good but not excellent}] \leq \exp(-\Omega_{\rho, \gamma}(D''))/r.$$

*Proof.* Let  $C := C_{\rho, \gamma}$ . For a real number  $X \geq 0$ , let  $E(A, B, B', X)$  be the event that

$$\|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2^2 \geq C \|F[B]|_{A \cap B \cap B'} - F[B']|_{A \cap B \cap B'}\|_2^2 + X.$$

The main claim is that  $\Pr_{A, B, B'} [E] \leq \exp(-\Omega_{\rho, \gamma}(X))$ . To see this, consider first fixing  $B, B'$ . Note for  $i \in B \cap B'$ , there is some probability  $p > 0$  (for each coordinate iid, independent of  $n$ ), such that  $i \in A$ . Thus, the claim follows from Bernstein's inequality.

Note that when  $E(A, B, B', X)$  does not hold and  $(A, B), (A, B')$  are  $D'$ -consistent:

$$\begin{aligned} \|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2^2 &\leq X + C \|F[B]|_{A \cap B \cap B'} - F[B']|_{A \cap B \cap B'}\|_2^2 \\ &\leq X + 2C \|F[B]|_{A \cap B \cap B'} - F[A]|_{A \cap B \cap B'}\|_2^2 \\ &\quad + 2C \|F[A]|_{A \cap B \cap B'} - F[B']|_{A \cap B \cap B'}\|_2^2 \\ &\leq X + 2CD + 2CD \leq X + O(D). \end{aligned}$$

where we have used the triangle inequality and consistency of  $(A, B)$  and  $(A, B')$ . As  $\Pr_{A,B,B'}[E] \leq \exp(-\Omega_{\rho,\gamma}(X))$ , we get that

$$\mathbb{E}_{A,B,B'} \left[ \mathbb{1}_{B,B' \in \text{Cons}(A), A \text{ good}} \exp \left( \|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2^2 / C' \right) \right] \leq \exp(O_{\gamma,\rho}(D)), \quad (22)$$

for sufficiently large constant  $C'$  depending on  $\rho, \gamma$ . Consider good  $A$  and  $B \in \text{Cons}(A)$  with

$$\mathbb{E}_{B'} \left[ \mathbb{1}_{B' \in \text{Cons}(A)} \exp \left( \|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2^2 / C' \right) \right] \leq \exp(O_{\gamma,\rho}(D)).$$

Then

$$\mathbb{E}_{B'} \left[ \mathbb{1}_{B' \in \text{Cons}(A)} \|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2^2 \right] \leq O_{\gamma,\rho}(D)$$

follows by convexity of  $\exp$ . The lemma follows from Markov's inequality applied to (22).  $\square$

Now we define how to generate a global function  $g_A$  for each set  $A \subseteq [n]$ . Ultimately, our goal is to prove that these  $g_A$  agree for a nonnegligible fraction of  $A$ .

**Definition 7.19** (Voting). Define the function  $g_A : [n] \rightarrow [0, 1]^K$  as

$$g_A(x) = \min_{v = F[B']_x : B' \in \{0,1\}^n} \mathbb{E}_{B \sim_{1-\gamma} A} \left[ \mathbb{1}_{B \in \text{Cons}(A), x \in B} \|v - F[B]_x\|_2^2 \right].$$

In other words,  $g_A(x)$  is the minimizer over all  $F[B]_x$  of the expected distance between it and a random  $B$  that is consistent with  $A$ .

We require a lemma on the conditions of product distributions.

**Lemma 7.20.** *Let  $U = (U_1, \dots, U_n)$  be a product distribution. Let  $\tilde{U}$  be  $U$  conditioned on an event with probability at least  $d$  and  $\tilde{U}_1, \dots, \tilde{U}_n$  be the marginal distributions of  $\tilde{U}$ . Then*

$$\sum_{i=1}^n d_{\text{TV}}(U_i, \tilde{U}_i)^2 \leq \log(1/d).$$

*Proof.* Use the following sequence of inequalities:

$$\log(1/d) \geq d_{\text{KL}}(\tilde{U} \| U) \geq \sum_{i=1}^n d_{\text{KL}}(\tilde{U}_i, U_i) \geq \sum_{i=1}^n d_{\text{TV}}(\tilde{U}_i, U_i)^2.$$

The first is because  $\tilde{U}$  is  $U$  conditioned on an event of probability at most  $d$ , the second is because  $U$  is a product distribution, and the last is by Pinsker's inequality.  $\square$

For the remainder of the analysis it will be useful to define the *degree* of a set  $A$ .

**Definition 7.21** (Degree). Define  $\text{deg}(A) := \Pr_{A' \sim_{1-\gamma} A} [A' \in \text{Cons}(A)]$ .

We first prove that for  $A$  with nonnegligible degree that the votes are consistent with  $F[A]$ .

**Lemma 7.22.** *For all  $A$  it holds that  $\|(g_A)|_A - F[A]\|_2 \leq O_{\rho,\gamma}(D'' + \log(1/\text{deg}(A)))^{1/2}$ .*

*Proof.* By the definition of consistency we know that

$$\begin{aligned}
(D')^2 &\geq \mathbb{E}[\|F[B]_{A \cap B} - F[A]_{A \cap B}\|_2^2 \mid B \in \text{Cons}(A)] \\
&= \sum_{x \in A} \mathbb{E}[1_{x \in B} \|F[B]_x - F[A]_x\|_2^2 \mid B \in \text{Cons}(A)] \\
&\geq \sum_{x \in A} \mathbb{E}[1_{x \in B} \|F[B]_x - (g_A)_x\|_2^2 \mid B \in \text{Cons}(A)]
\end{aligned}$$

where the final inequality is by the optimality of  $g_A$ . By the triangle inequality we conclude that

$$\begin{aligned}
4(D')^2 &\geq \sum_{x \in A} \mathbb{E}[1_{x \in B} \|F[A]_x - (g_A)_x\|_2^2 \mid B \in \text{Cons}(A)] \\
&= \sum_{x \in A} \Pr[x \in B \mid B \in \text{Cons}(A)] \|F[A]_x - (g_A)_x\|_2^2.
\end{aligned}$$

Now,  $\Pr[x \in B \mid B \in \text{Cons}(A)] \geq \Omega_{\rho, \gamma}(1)$  except for  $O_{\rho, \gamma}(\log(1/\deg(A)))$  many  $x \in A$ , by [Lemma 7.20](#). Also,  $\|F[A]_x - (g_A)_x\|_2^2 \leq O(1)$  for all  $x$ . Thus,

$$\|F[A] - (g_A)|_A\|_2^2 = \sum_{x \in A} \|F[A]_x - (g_A)_x\|_2^2 \leq O_{\rho, \gamma}((D')^2 + \log(1/\deg(A)))$$

as desired.  $\square$

Now we prove that if  $A$  is excellent, then  $F[B]$  agrees with  $g_A$  for most  $B$  such that  $(A, B)$  is  $D'$ -consistent.

**Lemma 7.23.** *If  $A$  is excellent then*

$$\Pr_{B \sim_{1-\gamma} A} [\|g_A|_B - F[B]\|_2 > C_{\rho, \gamma} D'' \text{ and } B \in \text{Cons}(A)] \leq O(r/\varepsilon).$$

*Proof.* Fix  $B$  satisfying the condition in [Definition 7.17](#). By definition,

$$(D'')^2 \geq \mathbb{E}_{B' \sim_{1-\gamma} A} [\|F[B]|_{B \cap B'} - F[B']_{B \cap B'}\|_2^2 \mid B' \in \text{Cons}(A)].$$

By the triangle inequality, and optimality of  $g_A$ ,

$$\begin{aligned}
&\mathbb{E}_{B' \sim_{1-\gamma} A} [\|g_A|_{B \cap B'} - F[B]|_{B \cap B'}\|_2^2 \mid B' \in \text{Cons}(A)] \\
&\leq 2 \mathbb{E}_{B' \sim_{1-\gamma} A} [\|F[B]|_{B \cap B'} - F[B']_{B \cap B'}\|_2^2 + \|g_A|_{B \cap B'} - F[B']_{B \cap B'}\|_2^2 \mid B' \in \text{Cons}(A)] \\
&\leq O((D'')^2).
\end{aligned}$$

Rewrite the above equation as

$$O((D'')^2) \geq \sum_{x \in B} \Pr_{B' \sim_{1-\gamma} A} [x \in B' \mid B' \in \text{Cons}(A)] \|(g_A)_x - F[B]_x\|_2^2.$$

To conclude, recall by [Lemma 7.20](#) that because  $\Pr[B' \in \text{Cons}(A)] \geq \varepsilon/3$  because  $A$  is good, that for all but  $O_{\rho, \gamma}(\log(1/\varepsilon))$  many  $x \in B$  that

$$\Pr_{B' \sim_{1-\gamma} A} [x \in B' \mid B' \in \text{Cons}(A)] \geq \Omega_{\rho, \gamma}(1).$$

This completes the proof.  $\square$



Finally, we prove that for almost all pairs of consistent  $A, A'$  that the votes are close in  $\ell_2$ . The proof closely resembles that of [Lemma 7.18](#).

**Lemma 7.24.** *For  $D''' > C_{\rho, \gamma} D''$ , We have that*

$$\Pr_{A \sim_{\rho} [n], A' \sim_{1-\gamma} A} \left[ \|g_A - g_{A'}\|_2 \leq D''', \text{ and } \deg(A), \deg(A') \geq \varepsilon^{O(1)} \right] \geq \varepsilon^{O(1)}.$$

*Proof.* We first argue that there are many pairs  $(A, A')$  that are  $D'$ -consistent and  $\deg(A), \deg(A') \geq \varepsilon^{O(1)}$ . Indeed, let  $S = \{A : \deg(A) \geq \varepsilon/3\}$ , i.e., the set of good  $A$ . Then

$$\begin{aligned} & \Pr[A' \in \text{Cons}(A), A \in S, A' \in S] \\ & \geq \Pr[A' \in \text{Cons}(A)] - \Pr[A' \in \text{Cons}(A), A \notin S] - \Pr[A' \in \text{Cons}(A), A' \notin S] \\ & \geq \varepsilon - \varepsilon/3 - \varepsilon/3 \geq \varepsilon/3. \end{aligned}$$

By [Lemma 7.18](#) at least  $\varepsilon/4$  of pairs  $(A, A')$  also have that both  $A, A'$  are excellent as long as  $\exp(-\Omega_{\rho, \gamma}(D''))/r \leq \varepsilon/100$  (we will choose  $r$  later to ensure this).

We now prove the stronger claim that over consistent pairs  $(A, A')$ , the probability that  $B \sim_{1-\gamma} A$  and  $B' \sim_{1-\gamma} A'$  are inconsistent is very unlikely. In other words,  $B$  and  $B'$  vote similarly towards  $g_A$  and  $g_{A'}$ . Formally, we want to prove that:

$$\begin{aligned} & \Pr_{\substack{A \sim_{\rho} [n], A' \sim_{1-\gamma} A \\ B \sim_{1-\gamma} A, B' \sim_{1-\gamma} A'}} [A', B \in \text{Cons}(A), B' \in \text{Cons}(A') \text{ and } \|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2 > D'''] \\ & \leq \exp(-\Omega_{\rho, \gamma}(D''')). \end{aligned}$$

Towards proving this, define  $E(A, A', B, B')$  to be the event that

$$\|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2 > C_{\rho, \gamma} \|F[B]|_{A \cap A' \cap B \cap B'} - F[B']|_{A \cap A' \cap B \cap B'}\|_2 + D'''/2.$$

We first argue that  $\Pr_{A, A', B, B'}[E(A, A', B, B')] \leq \exp(-\Omega_{\rho, \gamma}(D'''))$ . Indeed, fix  $B, B'$  and note that for  $i \in B, B'$  there is some probability  $p > 0$  (independent of  $n$ ) such that  $i \in A \cap A'$ . Thus this follows by a Chernoff bound. When  $E(A, A', B, B')$  does not hold and  $(A, A'), (A, B)$ , and  $(A', B')$  are all consistent pairs:

$$\begin{aligned} \|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2 & \leq D'''/2 + C \|F[B]|_{A \cap A' \cap B \cap B'} - F[B']|_{A \cap A' \cap B \cap B'}\|_2 \\ & \leq D'''/2 + C \|F[B]|_{A \cap A' \cap B \cap B'} - F[A]|_{A \cap A' \cap B \cap B'}\|_2 \\ & \quad + C \|F[A]|_{A \cap A' \cap B \cap B'} - F[A']|_{A \cap A' \cap B \cap B'}\|_2 \\ & \quad + C \|F[A']|_{A \cap A' \cap B \cap B'} - F[B']|_{A \cap A' \cap B \cap B'}\|_2 \\ & \leq D'''/2 + CD' + CD' + CD' \leq D'''. \end{aligned}$$

Here we have used the triangle inequality and consistency. Restrict to  $(A, A')$  satisfying

$$\Pr_{B \sim_{1-\gamma} A, B' \sim_{1-\gamma} A'} [A', B \in \text{Cons}(A), B' \in \text{Cons}(A') \text{ and } \|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2 > D'''] \leq r,$$

for  $r$  chosen later. At least  $\varepsilon/6 - \exp(-\Omega_{\rho, \gamma}(D'''))/r \geq \varepsilon/12$  fraction of  $(A, A')$  satisfy this. Say that  $(B, B')$  is a *valid pair* if  $\|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2 \leq D'''$ ,

$$\|g_A|_B - F[B]\|_2 \leq O_{\rho, \gamma}(D'') \quad \text{and} \quad \|g_{A'}|_{B'} - F[B']\|_2 \leq O_{\rho, \gamma}(D'').$$

By [Lemma 7.23](#) we know that at most  $O(r)$  fraction of pairs  $(B, B')$  are not valid.

For such  $(A, A')$ :

$$(D''')^2 \geq \mathbb{E}_{B, B'} [\|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2^2 \mid B \in \text{Cons}(A), B' \in \text{Cons}(A'), (B, B') \text{ valid}],$$

and by the triangle inequality:

$$\begin{aligned} & \mathbb{E}_{B, B'} [\|g_A|_{B \cap B'} - g_{A'}|_{B \cap B'}\|_2^2 \mid B \in \text{Cons}(A), B' \in \text{Cons}(A'), (B, B') \text{ valid}] \\ & \leq 3 \mathbb{E}_{B, B'} \left[ \|g_A|_B - F[B]\|_2^2 + \|g_{A'}|_{B'} - F[B']\|_2^2 + \|F[B]|_{B \cap B'} - F[B']|_{B \cap B'}\|_2^2 \right. \\ & \quad \left. \mid B \in \text{Cons}(A), B' \in \text{Cons}(A'), (B, B') \text{ valid} \right] \leq O_{\rho, \gamma}((D''')^2). \end{aligned}$$

Rewriting this gives us:

$$O_{\rho, \gamma}((D''')^2) \geq \sum_{x \in [n]} \Pr_{B, B'} [x \in B \cap B' \mid B \in \text{Cons}(A), B' \in \text{Cons}(A'), (B, B') \text{ valid}] \|(g_A)_x - (g_{A'})_x\|_2^2.$$

Because  $B \cap B'$  is a product distribution, and the event  $B \in \text{Cons}(A), B' \in \text{Cons}(A'), (B, B')$  has probability at least  $\varepsilon^2/9 - O(r) \geq \varepsilon^2/10$ , by [Lemma 7.20](#) we get that

$$\Pr_{B, B'} [x \in B \cap B' \mid B \in \text{Cons}(A), B' \in \text{Cons}(A'), (B, B') \text{ valid}] \geq \Omega_{\rho, \gamma}(1)$$

on all but  $O_{\rho, \gamma}(\log(1/\varepsilon))$  many  $x \in [n]$ . This completes the proof.  $\square$

Now [Theorem 12](#) immediately follows from combining [Lemma 7.24](#) with [Theorem 11](#), when choosing  $r = e^{-C'_{\rho, \gamma} D''}$  for appropriately chosen constant  $C'_{\rho, \gamma} > 0$ .

*Proof of [Theorem 12](#).* By [Lemma 7.24](#) and [Theorem 11](#) there is a function  $g : [n] \rightarrow [0, 1]^K$  such that

$$\Pr_{A \sim \rho[n]} [\|g - g_A\|_2 \leq O_{\rho, \gamma}(D''' \log(1/\varepsilon)), \text{deg}(A) \geq \varepsilon^{O(1)}] \geq \varepsilon^{O_{\rho, \gamma}(1)}.$$

Here, we may assume that  $\text{deg}(A) \geq \varepsilon^{O(1)}$  by restricting  $g_A$  only to  $A$  with  $\text{deg}(A) \geq \varepsilon^{O(1)}$ , and then applying [Theorem 11](#). To conclude, note that [Lemma 7.22](#) gives

$$\|g|_A - F[A]\|_2 \leq \|g - g_A\|_2 + \|F[A] - g_A|_A\|_2 \leq O_{\rho, \gamma}(D''' \log(1/\varepsilon))$$

for  $\text{deg}(A) \geq \varepsilon^{O(1)}$ , and recall that  $D''' \leq O_{\rho, \gamma}(D + \log(1/\varepsilon)^{1/2})$ .  $\square$

## 8 Applications

### 8.1 Restricted 3-APs over $\mathbb{F}_p^n$ and Generalizations

In this section we establish [Theorem 3](#) by a density increment argument.

**Theorem 3.** *Let  $\Sigma$  be a finite set and let  $S \subseteq \Sigma \times \Sigma \times \Sigma$  be a non-empty set satisfying that:*

1.  $(x, x, x) \in S$  for all  $x \in \Sigma$ , and
2.  $\text{supp}(S_{xy}), \text{supp}(S_{xz}), \text{supp}(S_{yz})$  are all connected.

Then there are constants  $c_\Sigma, C_\Sigma > 0$  such that for any  $A \subseteq \Sigma^n$  with  $\frac{|A|}{|\Sigma|^n} \geq \frac{C_\Sigma}{(\log \log \log n)^{c_\Sigma}}$ , there are  $x, y, z \in A$ , not all equal, such that  $(x_i, y_i, z_i) \in S$  for all  $i \in [n]$ .

Throughout this section, let  $m = |\Sigma|$ . Let  $\alpha = m^{-n}|A|$  and define  $f_A : \Sigma^n \rightarrow \mathbb{R}$  as  $f_A(x) = A(x) - \alpha$ , where we abuse notation to write  $A(x)$  to be the indicator function of  $A$ . Naïvely, we wish to have a distribution  $\mu$  supported on  $S$  such that  $\mu_x, \mu_y, \mu_z$  are all uniform. Unfortunately this is not always possible. Instead we settle for an approximate version of uniformity which we then random restrict so that  $\mu_x$  becomes uniform.

Throughout this section, assume that  $\alpha \geq C_m(\log \log \log n)^{-c_m}$ . Call a triple  $(x, y, z)$  such that  $x, y, z$  are not all equal and  $(x_i, y_i, z_i) \in S$  for all  $i \in [n]$  a *valid triple*.

**Lemma 8.1.** *Let  $\mu$  be the distribution over  $\Sigma \times \Sigma \times \Sigma$  with mass  $\frac{1}{m} - \frac{\delta}{m\sqrt{n}}$  on  $(a, a, a)$  for  $a \in \Sigma$ , and  $\frac{1}{|S|-m} \cdot \frac{\delta}{\sqrt{n}}$  on each  $\vec{a}$  for  $\vec{a} \in S \setminus \{(a, a, a) : a \in \Sigma\}$ . If  $A$  has no valid triples, then there are 1-bounded functions  $g : \Sigma^n \rightarrow \mathbb{C}, h : \Sigma^n \rightarrow \mathbb{C}$  such that:*

$$\max \left\{ \left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f_A(x)g(y)h(z)] \right|, \left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [g(x)f_A(y)h(z)] \right|, \left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [g(x)h(y)f_A(z)] \right| \right\} \geq \alpha^3/7 - \exp(-\Omega(\delta\sqrt{n})). \quad (23)$$

Also, it holds that  $\mathbb{E}_{x \sim \mu_x^{\otimes n}} [A(x)] \geq \alpha - 2\delta$ .

*Proof.* Note that if  $A$  has no valid triples, then

$$\mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [A(x)A(y)A(z)] = \Pr_{(x,y,z) \sim \mu^{\otimes n}} [x = y = z] \leq \exp(-\Omega(\delta\sqrt{n})).$$

Thus the first part follows from expanding  $A = \alpha + f_A$ . The second part follows because

$$\mathbb{E}_{x \sim \mu_x^{\otimes n}} [A(x)] \geq \mathbb{E}_{x \sim \Sigma^n} [A(x)] - d_{\text{TV}}(\mu_x^{\otimes n}, \Sigma^n) \geq \alpha - 2\delta,$$

by applying Pinsker's inequality. □

For the remainder of the section we let  $\gamma = \exp(-1/\alpha^{C'_m})$  for sufficiently large constant  $C'_m$  (depending on  $m$ ) and let  $\delta = \gamma^3/11000$ . Assume without loss of generality that the maximum in Lemma 8.1 is attained by the first expectation. Now we will apply our inverse theorem to randomly restrict  $f_A$  so that it correlates with a product function over the uniform distribution on  $\Sigma^n$  (or we obtain a density increment by other means). For  $A \subseteq \Sigma^n$  let  $\mu(A) = m^{-n}|A|$  denote the density of  $A$  with respect to the uniform measure on  $\Sigma^n$ .

**Lemma 8.2.** *Assume that (23) holds. Then there is a subset  $I \subseteq [n]$  with  $|I| \leq n - n^{1/3}$  and  $u \in \Sigma^n$  such that either:*

- $\mu(A_{I \rightarrow u}) \geq \alpha + \delta$ , or
- $\mu(A_{I \rightarrow u}) \geq \alpha - 10\delta/\gamma$  and there is a product function  $P := P_1 \dots P_n$ , where  $P_i : \Sigma \rightarrow \mathbb{D}$  for each  $i$ , such that

$$\mathbb{E}_{x \sim \Sigma^I} [(f_A)_{I \rightarrow u}(x)P(x)] \geq \gamma.$$

*Proof.* Applying [Theorem 2](#) (recall that  $\mu$  is pairwise-connected by item 2 of [Theorem 3](#)) we know that there is some  $\beta \geq \Omega_m\left(\frac{\delta\gamma}{\sqrt{n}}\right) \geq 2n^{-2/3}$  such that:

$$\Pr_{\substack{I \sim_{1-\beta}[n] \\ u \sim \nu^I}} \left[ \sup\{|\langle (f_A)_{I \rightarrow u}, P \rangle| : P \text{ is 1-bounded product function}\} \geq \gamma \right] \geq \gamma,$$

for some distribution  $\nu$  satisfying  $\mu = (1 - \beta)\nu + \beta U$ . Indeed, we first pass via a random restriction from the distribution  $\mu$  on  $n$  coordinates to a distribution  $\mu'$  on  $\Theta(\delta\sqrt{n})$  coordinates, where the support of  $\mu'$  is  $S$ , the probability of each atom in  $\mu'$  is  $\Omega_m(1)$ , the marginal of  $\mu'$  on  $x$  is uniform, and get the an expectation analogous to (23) is still at least  $\alpha^3/16$  with probability at least  $\alpha^3/16$ . We then invoke [Theorem 2](#) on this expectation.

Thus item 2 holds unless

$$\Pr_{\substack{I \sim_{1-\beta}[n] \\ u \sim \nu^I}} \left[ \mu(A_{I \rightarrow u}) \geq \alpha - 10\delta/\gamma \right] \leq 1 - 0.99\gamma. \quad (24)$$

In that case, by [Lemma 8.1](#),

$$\mathbb{E}_{\substack{I \sim_{1-\beta}[n] \\ u \sim \nu^I}} [\mu(A_{I \rightarrow u})] = \mathbb{E}_{x \sim \mu^{\otimes n}} [A(x)] \geq \alpha - 2\delta.$$

Combining this with (24) implies item 1.  $\square$

If item 1 of [Lemma 8.2](#) holds, then we are done (we have obtained a density increment). Otherwise, item 2 holds. In this case, for simplicity rename  $\bar{I}$  as  $[n]$  again (recall that  $|\bar{I}| \geq n^{1/3}$  by [Lemma 8.2](#)), and rename  $A_{I \rightarrow u}$  as  $A$ . Thus  $(f_A)_{I \rightarrow u}$  gets renamed as  $f_A = A - \alpha$ .

In this case we require a more specialized kind of random restriction. We will try to find subsets  $T \subseteq [n]$  and restrict to  $x \in \Sigma^n$  such that  $x_t = x_{t'}$  for all  $t, t' \in T$ . We hope for two things to be true for this subset  $T$ . First,  $\prod_{t \in T} P_t$  should be nearly-constant. Second, the density of  $A$  does not drop significantly under this restriction. While this is clearly not true for a fixed  $T$ , we prove that we can choose  $T$  with sufficient randomness to ensure this.

We now formally define the operation which forces coordinates to take the same value.

**Definition 8.3.** For a function  $g : \Sigma^n \rightarrow \mathbb{C}$  and  $T \subseteq [n]$ , define  $g_{=T} : \Sigma^{n-|T|+1} \rightarrow \mathbb{C}$  as follows. For  $x \in \Sigma$  and  $y \in \Sigma^{[n] \setminus T}$  let  $v \in \Sigma^n$  be the vector where  $v_i = y_i$  for  $i \in [n] \setminus T$  and  $v_i = x$  otherwise. Then  $g_{=T}(x, y) := g(v)$ . For brevity, for disjoint sets  $T_1, \dots, T_N$  we write  $g_{=T_1, \dots, T_N} = (\dots (g_{=T_1}) \dots)_{=T_N}$ .

We now prove that if  $T$  is a sufficiently random set, then  $\mathbb{E}[g_{=T}]$  is close to  $\mathbb{E}[g]$ .

**Lemma 8.4.** *Let  $S \subseteq [n]$  and let  $g : \Sigma^n \rightarrow \mathbb{C}$  be 1-bounded. Then*

$$\left| \mathbb{E}_{x \sim \Sigma^n} g(x) - \mathbb{E}_{T \subseteq S, |T|=k} \mathbb{E}_{x \sim \Sigma^{n-k+1}} g_{=T}(x) \right| \leq \frac{2mk}{\sqrt{|S|}}.$$

*Proof.* Let  $\nu$  be the distribution on  $\Sigma^n$  over the  $v$  vector (as defined in [Definition 8.3](#)) obtained by sampling  $T \subseteq S$ ,  $|T| = k$  and  $x \in \Sigma^{n-k+1}$ . It suffices to prove that  $d_{\text{TV}}(\nu, \Sigma^n) \leq \frac{2mk}{\sqrt{|S|}}$  because  $g$  is 1-bounded. To prove this, further let  $\nu'$  to be the distribution where we sample  $T \subseteq S$  and set  $x_t = a$  for all  $t \in T$ , where  $a$  is some fixed element of  $\Sigma$ . We will prove that  $d_{\text{TV}}(\nu', \Sigma^n) \leq \frac{2mk}{\sqrt{|S|}}$ .

Towards this, we may assume that  $S = [n]$ . The claim is evident when  $k = 1$ . For  $k > 1$  note that sampling  $T \subseteq [n]$  can be achieved by sampling  $T' \subseteq [n]$  with  $|T'| = k - 1$  and then sampling  $t \in [n] \setminus T'$  and setting  $T = T' \cup \{t\}$ . Thus the result follows by induction.  $\square$

With these tools in hand we are ready to prove [Theorem 3](#).

*Proof of Theorem 3.* We may assume that item 2 of [Lemma 8.2](#) holds, or else we have already obtained a density increment. Let  $v_1, \dots, v_n : \Sigma \rightarrow \mathbb{R}/\mathbb{Z}$  be such that  $P_j(x) = e^{2\pi i v_j(x)}$  for all  $j \in [n]$ ,  $x \in \Sigma$ .

Let  $\zeta = \frac{1}{8m^2}$ , and let  $N = n^\zeta$ . By the Pigeonhole principle, we can find disjoint sets  $S_1, \dots, S_N$  of size  $|S_i| = \sqrt{n}/2$  such that for every  $i \in [N]$  and  $j, j' \in S_i$ , it holds that  $\|v_j - v_{j'}\|_\infty \leq n^{-\frac{1}{2m}}$ . Let  $v$  be an arbitrary vector in  $S_i$ , and let  $1 \leq k_i \leq n^{\frac{1}{4m}}$  be such that  $\|k_i v\|_\infty \leq n^{-\frac{1}{4m^2}}$  – such a  $k_i$  exists by Dirichlet approximation.

Now let  $T_i$  be a uniformly random subset of  $S_i$  of size  $k_i$  for all  $i \in [N]$  and perform the following sequence of restrictions. For all  $i \in [N]$  force that  $x_j = x_{j'}$  for all  $j, j' \in T_i$ , and then uniformly random restrict all coordinates in  $[n] \setminus (T_1 \cup \dots \cup T_N)$ . By [Lemma 8.4](#) we know that:

$$\mathbb{E}_{T_i \subseteq S_i, |T_i|=k_i, i \in [N]} \mathbb{E}_{u \sim \Sigma^{[n] \setminus \cup_i T_i}} [\mu((A_{=T_1, \dots, T_N})_{I \rightarrow u})] \geq \alpha - 10\delta/\gamma - \frac{4Nmn^{\frac{1}{4m}}}{n^{1/4}} \geq \alpha - 11\delta/\gamma. \quad (25)$$

Similarly, by [Lemma 8.4](#) we know that

$$\mathbb{E}_{T_i \subseteq S_i, |T_i|=k_i, i \in [N]} \mathbb{E}_{u \sim \Sigma^{[n] \setminus \cup_i T_i}} \mathbb{E}_{x \sim \Sigma^N} [((f_A)_{=T_1, \dots, T_N})_{I \rightarrow u}(x)(P_{=T_1, \dots, T_N})_{I \rightarrow u}(x)] \geq \gamma - \frac{4Nmn^{\frac{1}{4}}}{n^{1/4}} \geq \gamma/2. \quad (26)$$

Let  $Q = (P_{=T_1, \dots, T_N})_{I \rightarrow u}$ . We will argue that  $Q$  is nearly constant. For  $x, y \in \Sigma^N$ ,

$$Q(x) - Q(y) = \exp\left(2\pi i \sum_{j=1}^N \sum_{t \in T_j} v_t(y)\right) \left( \exp\left(2\pi i \sum_{j=1}^N \sum_{t \in T_j} (v_t(x) - v_t(y))\right) - 1 \right) \prod_{i \in ([n] \setminus \cup_i T_i)} P_i(u_i).$$

For  $j \in [N]$  and let  $v \in T_j$  be such that  $\|k_j v\|_\infty \leq n^{-\frac{1}{4m^2}}$ . Then

$$\left\| \sum_{t \in T_j} v_t \right\|_\infty \leq \|k_j v\|_\infty + \sum_{t \in T_j} \|v_t - v\|_\infty \leq n^{-\frac{1}{4m^2}} + k_j n^{-\frac{1}{2m}} \leq 2n^{-\frac{1}{4m^2}}.$$

Combining this with the above yields that  $|Q(x) - Q(y)| \leq 100Nn^{-\frac{1}{4m^2}} = 100n^{-\frac{1}{8m^2}}$ . Letting  $A' := (A_{=T_1, \dots, T_N})_{I \rightarrow u}$ , combining this with the above [\(26\)](#) then implies that:

$$\mathbb{E}_{T_i \subseteq S_i, |T_i|=k_i, i \in [N]} \mathbb{E}_{u \sim \Sigma^{[n] \setminus \cup_i T_i}} [|\mu(A') - \alpha|] \geq \gamma/2 - 100n^{-\frac{1}{8m^2}} \geq \gamma/3,$$

and thus

$$\Pr_{T_i \subseteq S_i, |T_i|=k_i, i \in [N]} \mathbb{E}_{u \sim \Sigma^{[n] \setminus \cup_i T_i}} [|\mu(A') - \alpha| \geq \gamma/6] \geq \gamma/6.$$

If

$$\Pr_{T_i \subseteq S_i, |T_i|=k_i, i \in [N]} \mathbb{E}_{u \sim \Sigma^{[n] \setminus \cup_i T_i}} [\mu(A') > \alpha - \gamma/6] > 1 - \gamma/6$$

then there is some event  $A'$  where  $\mu(A') > \alpha - \gamma/6$  while  $|\mu(A') - \alpha| \geq \gamma/6$ , so  $\mu(A') \geq \alpha + \gamma/6$  as desired. Otherwise, we know that

$$\Pr_{T_i \subseteq S_i, |T_i|=k_i, i \in [N]} \mathbb{E}_{u \sim \Sigma^{[n] \setminus \cup_i T_i}} [\mu(A') \leq \alpha - \gamma/6] \geq \gamma/6.$$

Finally, recall that  $\delta = \gamma^3/11000$ . Combining this with (25), which says that

$$\mathbb{E}_{\substack{T_i \subseteq S_i, |T_i|=k_i, i \in [N] \\ u \sim \Sigma^{[n] \setminus \cup_i T_i}}} [\mu(A')] \geq \alpha - 11\delta/\gamma \geq \alpha - \gamma^2/1000$$

gives that there is some event  $A'$  with  $\mu(A') \geq \alpha + \gamma^2/72$ .

The claimed quantitative dependence in [Theorem 3](#) holds because  $\gamma = \exp(-1/\alpha^C)$ , and  $A' \subseteq \Sigma^N$  for  $N \geq n^{\frac{1}{24m^2}}$  (at some point we labelled  $n^{1/3}$  as  $n$ ). Thus, after  $O(1/\gamma)$  iterations the dimension would be at least  $n^{(1/24m^2)^{O(1/\gamma)}}$  and we would end up with a set with density  $1 - o(1)$ , which clearly must contain a valid triplet.  $\square$

## 8.2 Direct Sum Testing

In this section we prove [Theorem 4](#), and we fix  $f$  as in the statement of [Theorem 4](#) throughout. We first note that the direct sum test is equivalent to the following test: sample  $I \subseteq_{1/2} [n]$ ,  $x, x' \sim \mathbb{F}_p^I$ ,  $x, x' \sim \mathbb{F}_p^{\bar{I}}$  and check that  $f(x, y) - f(x, y') - f(x', y) + f(x', y') = 0$ . Thus, it follows that

$$\mathbb{E} \left[ \sum_{j \in \mathbb{F}_p} \omega_p^{j(f(x,y) - f(x,y') - f(x',y) + f(x',y'))} \right] \geq \left( \frac{1}{p} + \varepsilon \right) p = 1 + p\varepsilon,$$

and so  $\mathbb{E}_x \left[ \sum_{j \neq 0} \omega_p^{j(f(x,y) - f(x,y') - f(x',y) + f(x',y))} \right] \geq p\varepsilon$ . It follows there is  $j \neq 0$  such that, letting  $F(z) = \omega_p^{jf(z)}$ , we have that

$$\left| \mathbb{E} \left[ F(x, y) \overline{F(x', y)} \overline{F(x, y')} F(x', y') \right] \right| \geq \varepsilon.$$

We note that the left hand side is exactly  $\text{swap}(F)$ , so applying [Theorems 5, 9](#) and [10](#) we get that there is  $L: \Sigma^n \rightarrow \mathbb{C}$  with  $\|L\|_2 \leq 1$  and degree at most  $D = \exp((1/\varepsilon)^{O(1)})$ , a product function  $P = P_1 \cdots P_n$  where  $P_i: \Sigma \rightarrow \mathbb{C}$  is 1-bounded, such that  $|\langle F, LP \rangle| \geq \delta$ , and  $\delta = \exp(-\exp((1/\varepsilon)^{O(1)}))$ . Letting  $G = \frac{1}{\|L\|_4} F \bar{L}$ , we get that  $\|G\|_4 \leq 1$  and

$$|\langle G, P \rangle| = \frac{|\langle F, LP \rangle|}{\|L\|_4} \geq \frac{\delta}{2^{O(D)}},$$

where we used  $\|L\|_4 \leq 2^{O(D)}$  which follows by hypercontractivity. Applying [Theorem 10](#) again gives that there is a product function  $Q = Q_1 \cdots Q_n$  where  $Q_i: \Sigma \rightarrow \mathbb{D}$  for each  $i$ , such that  $|\langle G, Q \rangle| \geq \frac{\delta^{O(1)}}{2^{O(D)}}$ , and plugging in the definition of  $G$  finishes the proof.

## 8.3 Proof of [Theorem 1](#)

[Theorem 1](#) follows by combining [Theorems 2](#) and [9](#) and then applying [Theorem 10](#) as in [Section 8.2](#).

## References

[BKLM24a] Amey Bhangale, Subhash Khot, Yang P. Liu, and Dor Minzer. On approximability of satisfiable  $k$ -csps: VII. 2024+.

- [BKLM24b] Amey Bhangale, Subhash Khot, Yang P. Liu, and Dor Minzer. Reasonable bounds for combinatorial lines of length three. 2024+.
- [BKM22] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable  $k$ -csps: I. In *STOC 2022*, pages 976–988. ACM, 2022.
- [BKM23a] Amey Bhangale, Subhash Khot, and Dor Minzer. Effective bounds for restricted 3-arithmetic progressions in  $\mathbb{F}_p^n$ . *Electron. Colloquium Comput. Complex.*, TR23-116, 2023. To appear in *Discrete Analysis*.
- [BKM23b] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable  $k$ -csps: II. In *STOC 2023*, pages 632–642. ACM, 2023.
- [BKM23c] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable  $k$ -csps: III. In *STOC 2023*, pages 643–655. ACM, 2023.
- [BKM24a] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable  $k$ -csps: IV. In *STOC 2024*, pages 1423–1434. ACM, 2024.
- [BKM24b] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable  $k$ -csps: V. *CoRR*, abs/2408.15377, 2024.
- [DDG<sup>+</sup>15] Roe David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. In *ITCS 2015*, pages 327–336, 2015.
- [DG19] Irit Dinur and Konstantin Golubev. Direct sum testing: The general case. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2019.
- [FK91] Hillel Furstenberg and Yitzhak Katznelson. A density version of the Hales-Jewett theorem. *Journal d’Analyse Mathématique*, 57(1):64–119, 1991.
- [Gow01] W. T. Gowers. A new proof of Szemerédi’s theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.
- [Gre] Ben Green. 100 open problems. *manuscript*.
- [HJ63] AW Hales and RI Jewett. Regularity and positional games. *Transactions of the American Mathematical Society*, 106(2):222–229, 1963.
- [Mes95] Roy Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *Journal of Combinatorial Theory, Series A*, 71(1):168–172, 1995.
- [Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010.
- [O’D14] Ryan O’Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.
- [Pol12] DHJ Polymath. A new proof of the density Hales-Jewett theorem. *Annals of Mathematics*, pages 1283–1327, 2012.
- [Rot53] Klaus F Roth. On certain sets of integers. *J. London Math. Soc*, 28(104-109):3, 1953.
- [WYZ24] Alek Westover, Edward Yu, and Kai Zheng. New direct sum tests. *arXiv preprint arXiv:2409.10464*, 2024. to appear in ITCS 2025.