

Locally Testing Direct Products in the Low Error Range

Irit Dinur

Weizmann Institute

Dept. of Applied Math and Computer Science

Rehovot, 76100 ISRAEL

irit.dinur@weizmann.ac.il

Elazar Goldenberg

Hebrew University

School of Computer Science and Engineering

Jerusalem, 91904 Israel

elazargo@cs.huji.ac.il

Abstract

Given a function $f : X \rightarrow \Sigma$, its ℓ -wise direct product is the function $F = f^\ell : X^\ell \rightarrow \Sigma^\ell$ defined by: $F(x_1, \dots, x_\ell) = (f(x_1), \dots, f(x_\ell))$. We are interested in the local testability of the direct product encoding (mapping $f \mapsto f^\ell$). Namely, given an arbitrary function $F : X^\ell \rightarrow \Sigma^\ell$, we wish to determine how close it is to f^ℓ for some $f : X \rightarrow \Sigma$, by making two random queries into F . In this work we analyze the case of low acceptance probability of the test. We show that even if the test passes with small probability, $\varepsilon > 0$, already F must have a non-trivial structure and in particular must agree with some f^ℓ on nearly ε of the domain. Moreover, we give a structural characterization of all functions F on which the test passes with probability ε .

Our results can be viewed as a combinatorial analog of the low error ‘low degree test’, that is used in PCP constructions.

1. Introduction

Given a function $f : X \rightarrow \Sigma$, its ℓ -wise direct product is the function $F = f^\ell : X^\ell \rightarrow \Sigma^\ell$ defined by

$$F(x_1, \dots, x_\ell) = (f(x_1), \dots, f(x_\ell)).$$

We think of $|X|$ as being very large compared to Σ , ℓ (for concreteness one may keep in mind $X = [n]$ and $\Sigma = \{0, 1\}$), and view the mapping $f \mapsto f^\ell$ as an encoding of f . In this paper we study the testability of the direct product encoding. Namely, given an arbitrary function $F : X^\ell \rightarrow \Sigma^\ell$, we wish to determine how close it is to f^ℓ for some $f : X \rightarrow \Sigma$, by making the smallest possible number of random queries into F , namely two.

This question has first been studied by Goldreich and Safra [5], who showed that this encoding is testable with a constant number of queries. A very simple two query

test for this encoding was analyzed in [3], where it was shown that if the test succeeds with probability $1 - \delta$, then F agrees with some f^ℓ on at least $1 - O(\delta)$ of the domain. This pretty much pinpoints functions F that pass the test with *high* probability.

In this paper we ask the following question: Which are the functions F that pass the test with an arbitrary probability ε ? We answer this question for all values of $\varepsilon \geq \ell^{-\Omega(1)}$.

One motivation for this question comes from Probabilistically Checkable Proofs (PCPs). It is easy to construct PCPs with small soundness error¹ just by using sequential repetition. However, this increases the number of queries made to the proof. In order to reduce the number of queries the proof f can be replaced by its direct product encoding $F = f^\ell$. However, one must be able to test that the encoded proof F does not cheat, i.e., that F is ‘faithful’ to some underlying f . Moreover, since we are interested in small error, our test must be such that if it passes with probability $\geq \varepsilon$, then we can already conclude that F is sufficiently close to f^ℓ for some f .

Thus our results are analogous to the small-error analysis of the low degree test [10, 1]; and the direct product encoding can be viewed as a combinatorial alternative to the low degree encoding used in small-error PCP constructions. Presumably, one could incorporate this encoding in PCP constructions, but the details are beyond the scope of the current work.

The so-called low-error (or low-acceptance-probability) regime is often more difficult to analyze. One reason is the non-uniqueness of the solution: Clearly F can be a hybrid of $1/\varepsilon$ different legal codewords and still pass the test with probability ε . Thus, this is called the list-decoding regime since one can, at best, guarantee that success of the test

¹The soundness error is the probability that the verifier accepts when it should reject.

implies existence of a list of codewords that have non-trivial agreement with the received word (F in our case).

Let us now formally describe the test T and state our main theorem. Given a function $F : X^\ell \rightarrow \Sigma^\ell$, the test T has a parameter m which we fix to be $m = \ell^c$ for some constant $c = 19/75$, and is as follows:

1. Choose $\mathbf{x} \in X^\ell$ uniformly at random.
2. Choose a random set $I \subset [\ell]$, $|I| = m$, and choose a random $\mathbf{x}' \in X^\ell$ conditioned on $\mathbf{x}'_i = \mathbf{x}_i$ for all $i \in I$.
3. Accept iff $F(\mathbf{x})_I = F(\mathbf{x}')_I$.

This test makes two queries into F , at \mathbf{x} and at \mathbf{x}' . If $F = f^\ell$ for some function f then clearly the test succeeds with probability one. In fact, even if $F = f_1 \times f_2 \times \dots \times f_\ell$ for an ℓ -tuple $\vec{f} = (f_1, \dots, f_\ell)$ of possibly distinct functions $f_i : X \rightarrow \Sigma$ (in the sense that $F(x_1, \dots, x_\ell) = (f_1(x_1), \dots, f_\ell(x_\ell))$) T still accepts with probability one. Our first theorem states that if T passes with probability ε then it is explained by closeness of F to $f_1 \times \dots \times f_\ell$ on some $\varepsilon^{O(1)}$ fraction of the domain.

Theorem 1.1 *Let $F : \mathbf{X} \rightarrow \Sigma^\ell$. If T accepts F with probability ε , then there exists a tuple $\vec{f} = (f_1, \dots, f_\ell)$ of functions $f_i : X \rightarrow \Sigma$ such that for $\Omega(\varepsilon^5)$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$:*

$$\Pr_{i \in [\ell]} [F(\mathbf{x})_i = f_i(\mathbf{x}_i)] \geq 1 - O(\ell^{-\Omega(1)})$$

Let us make a couple of comments about the above theorem.

- The theorem concludes that on many of the tuples \mathbf{x} , $F(\mathbf{x}) \approx \vec{f}(\mathbf{x})$ rather than $F(\mathbf{x}) = \vec{f}(\mathbf{x})$. This weaker conclusion is inherent, as can be seen by taking $F = f^\ell$ and then changing each $F(\mathbf{x})$ arbitrarily in fewer than ℓ/m coordinates. Such a function F will pass the test with high probability, yet is only close to f^ℓ in the above sense.
- A second apparent weakness of this theorem is the fraction $\Omega(\varepsilon^5)$ of tuples that support \vec{f} which fails to fully explain the ε success probability of T . Our second result is a stronger theorem (Theorem 1.3 below) that addresses this issue, and we turn to it shortly.

First, however, let us return to the question of testing whether F is close to the ℓ -th power f^ℓ of a single function $f : X \rightarrow \Sigma$ (rather than to $f_1 \times f_2 \times \dots \times f_\ell$). For this we must consider the modified test T' , which is the same as T except that the last step is now:

- 3'. Choose $s : [\ell] \rightarrow [\ell]$ to be a random permutation on $[\ell]$. Denote by $s(\mathbf{x}') \in X^\ell$ the vector defined by $s(\mathbf{x}')_i = \mathbf{x}'_{s(i)}$. Read $F(\mathbf{x})$ and $F(s(\mathbf{x}'))$ and accept iff for every $i \in I$ $F(\mathbf{x})_i = s^{-1}(F(s(\mathbf{x}')))_i$.

Clearly if $F = f^\ell$ then the test accepts always. We prove via reduction from the main theorem that,

Theorem 1.2 *Let $F : \mathbf{X} \rightarrow \Sigma^\ell$. If T' accepts F with probability ε , then there exists a function $f : X \rightarrow \Sigma$ such that for $\Omega(\varepsilon^6)$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$ it holds:*

$$\Pr_{i \in [\ell]} [F(\mathbf{x})_i = f(\mathbf{x}_i)] \geq 1 - O(\ell^{-\Omega(1)})$$

The proof of this theorem encounters unexpected complications (see Section 4), and it is unclear whether these can be avoided. As previously mentioned, our stronger “structural characterization” below improves this theorem in that the agreement of F with f^ℓ goes from $\varepsilon^{O(1)}$ to $\varepsilon(1 - o(1))$. We remark that both T and T' were essentially considered in [3] modulo a slight technical difference, where their high-acceptance-probability behavior was analyzed.

1.1 The Structural Characterization

Our next result is stronger in that it characterizes (up to lower order terms) functions F on which T' accepts with probability ε . Consider the following “generic” construction of a function F on which T' accepts with probability ε . Choose functions $f_1, \dots, f_t : X \rightarrow \Sigma$. For each function, fix a set $S_i \subseteq \mathbf{X}$ of tuples and set $F(\mathbf{x})$ approximately equal to $f_i(\mathbf{x})$ for all $\mathbf{x} \in S_i$. Outside $\cup S_i$ fix F randomly. Assuming first (for simplicity) that the f_i 's are far from each other (hence the S_i 's are roughly disjoint), it is easy to check that

$$\sum_i \Pr[\mathbf{x}, \mathbf{x}' \in S_i] \geq \varepsilon \implies \Pr[T' \text{ accepts } F] \geq \varepsilon.$$

(neglecting an additive $\ell^{-\Omega(1)}$ term).

Our structural characterization can be viewed as an “inverse theorem” in that for any given F it finds functions f_i and supports $S_i \subseteq \mathbf{X}$ such that essentially the only way T' will accept on a pair \mathbf{x}, \mathbf{x}' , is if they both belong to S_i for some i ,

$$\sum_i \Pr[\mathbf{x}, \mathbf{x}' \in S_i] \geq \varepsilon \iff \Pr[T' \text{ accepts } F] \geq \varepsilon.$$

(again, neglecting an additive $\ell^{-\Omega(1)}$ term).

We also show that at least one f_i must agree with F on at least $\varepsilon(1 - o(1))$ of the domain. This is proven using the eigenvalues of the transition matrix of T' . The precise statement of our theorem is subtle, essentially

since the functions f_i need not be far apart and this, in turn, causes the sets S_i to possibly intersect. An informal version is as follows (the precise statement appears as Theorem 5.1):

Theorem 1.3 (Structural Theorem - Informal) *Let $F : X^\ell \rightarrow \Sigma^\ell$ be a function on which T' accepts with probability $\varepsilon > 0$. There is a list of functions $f_1, \dots, f_t : X \rightarrow \Sigma$ and an i such that F agrees with $(f_i)^\ell$ on at least $\varepsilon(1 - o(1))$ of \mathbf{X} . Moreover, if T' accepts on \mathbf{x}, \mathbf{x}' then with probability $1 - \ell^{-\Omega(1)}$, (i) $F(\mathbf{x})$ and $F(\mathbf{x}')$ agree with f_i for some $1 \leq i \leq t$; (ii) For each i such that $F(\mathbf{x})$ agrees with f_i also $F(\mathbf{x}')$ agrees with f_i .*

In the above, agreement is taken to be agreement on $1 - \ell^{-\Omega(1)}$ fraction of the coordinates.

One consequence of this result is that if T' accepts on \mathbf{x}, \mathbf{x}' then we can *approximately locally decode* F back to f_i . The theorem guarantees that conditioned on T' accepting on \mathbf{x}, \mathbf{x}' , then almost surely there is some i such that for almost all j : the j -th coordinate of $F(\mathbf{x})$ equals $f_i(x_j)$. Let us make a couple of remarks:

- This is related to the issue of locally list decoding the direct product encoding, which was studied in two relatively recent works [6, 7]. In that setting, F is already guaranteed to agree with f^ℓ on an ε fraction of the domain, and the goal is to generate (uniformly) a list of circuits that have oracle access to F , one of which computes f on almost all inputs. Our theorem complements this result in that it removes the need for the assumption about F being ε -close to f^ℓ (rather, we can test whether this holds). Moreover, both testing and decoding can be performed “in one shot” while making the smallest possible number of queries (i.e. two). In addition, it seems that Theorem 1.3 can also be used to give similar local decoding results, but we did not work out the details. We add that [7] were able to extend their results to derandomized direct products, and it would be extremely interesting to similarly derandomize our testing results.
- By reading $d \ll \ell$ coordinates of $F(\mathbf{x})$ we can obtain several values of f_i and ensure that (whp over the possible d -tuples X^d) nearly all of the d values are consistent with a single f_i while still making only two queries into F . Such “consistent reading” behavior is known for low degree tests (see [10, 1, 2]) and is the key to composing PCPs while maintaining small error. We are not aware of any other encoding that shares this property.

1.2 Parallel Repetition

One of the most celebrated results pertaining to the direct product encoding is the parallel repetition theorem of Raz [9]. Without getting into the details let us mention that our test can be viewed as an ℓ -fold repetition of a single-coordinate ‘equality’ test. Parallel repetition theorems could possibly bound the success of a repeated test but would not provide any *structural information* about functions that pass with “non-negligible” probability.

Even so, one may hope to benefit from the proof techniques. Raz’s techniques do not seem to help our setting, in particular since they are “too strong”: they guarantee an upper bound that is exponentially small in ℓ , in contrast to the fact that the success probability of T is meaningful only if it is much larger, at least $1/\ell$ (see Section 6 for an appropriate example).

Nevertheless, an earlier proof of a (weaker) parallel repetition theorem due to Feige and Kilian [4] turns out to provide the key to our proof. We elaborate on this shortly.

1.3 Our Proof

The analysis of [3] in the high-acceptance-probability setting proceeds by defining a *majority* function f based on F , by taking for each x the most popular value among all tuples \mathbf{x} containing x . It is shown that if T accepts F with high enough (say 99%) probability then $F \approx f^\ell$. In our low-acceptance-probability setting such an approach cannot succeed, as can be seen by the following example: For each \mathbf{x} let $F(\mathbf{x})$ be $(0, \dots, 0)$ with probability $\frac{1}{2}$ and $(1, \dots, 1)$ with probability $\frac{1}{2}$. Then T accepts with probability $\frac{1}{2}$ while the majority function f is a random function, and surely F is far from f^ℓ . Observe however, that this does not contradict our theorem as F is indeed close to two direct product functions: $\mathbf{0}^\ell$ and $\mathbf{1}^\ell$.

Locally testing a code in the list-decoding regime has been studied in the literature, for example in the low-error low degree test of [1, 10]. The low degree polynomial codes have a high relative distance which is crucial for the low-degree test analyses. Indeed, we were set back by the observation that a combinatorial analysis a la Raz-Safra will not work here².

The key to our proof comes from the work of Feige and Kilian [4] in their analysis of parallel repetition games. They study parallel repetition of so-called miss/match games, and prove a structural dichotomy lemma which easily adapts to our setting. Essentially the

²There are functions F that pass our test whp, but exhibit many “non-transitive triangles”: triples $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{X}$ such that T accepts \mathbf{x}, \mathbf{y} and \mathbf{x}, \mathbf{z} but rejects \mathbf{y}, \mathbf{z} .

lemma says the following: every function $F : X^\ell \rightarrow \Sigma^\ell$ either causes our test T to reject whp, or there are *many* exponentially-small sets $E \subset X^\ell$ on which $F \approx f_{1,E} \times \cdots \times f_{\ell,E}$ (however possibly each E has a distinct $\vec{f}_E = (f_{1,E}, \dots, f_{\ell,E})$).

This lemma leverages noticeable success of the test to deduce a certain structure for F . However, deducing structure on tiny (exponentially small) subsets E is not very useful unless, and this is the key point, these subsets can be glued together in a meaningful way. The main technical work in the proof of Theorem 1.1 goes to showing how to go from local to global agreement and to stitch the tiny E 's together into one big set that agrees with a *single* direct product. We first show that some noticeable fraction of pairs E, E' intersect non-trivially. Then we deduce that such an intersection implies that $\vec{f}_E \approx \vec{f}_{E'}$. Finally we find a ‘‘popular’’ set E that agrees with sufficiently many of the E' 's and proceed to prove that the function f_E agrees with F non-trivially on at least an $\varepsilon^{O(1)}$ fraction of X^ℓ .

We then extend this theorem in two ways to Theorems 1.2 and 1.3. The proofs of both of these theorems encounter unexpected complications on which we elaborate in the corresponding Sections 4 and 5.

Organization of the Paper Section 2 contains basic definitions and lemmas. Sections 3, 4, and 5 contain the proofs of Theorems 1.1, 1.2, and 1.3. We conclude with a discussion of the tightness of the parameters in Section 6.

2 Preliminaries

There is a certain amount of freedom in choosing the parameters, and we have not made an attempt to optimize them. We require that $|X| > \ell^3$. Also throughout the paper always $2\ell^{-1/75} < \varepsilon < 1/48$ (we require ℓ to be large enough for this to be non-void). We fix the remaining parameters as follows: $\rho = \ell^{-1/75}$, $\eta = \ell^{-7/75}$, $m = \ell^{19/75}$.

2.1 General Definitions, Notations, and Tools

Denote $\mathbf{X} = X^\ell$. An element in \mathbf{X} is called an ℓ -tuple or a tuple and is usually denoted by $\mathbf{x}, \mathbf{x}', \mathbf{y}$ etc. . The i -th coordinate of a tuple \mathbf{x} is denoted by $x_i \in X$. For a $k \in [\ell]$, the notion of a k -block is important in our proof:

Definition 2.1 A k -block b is a pair (y, \vec{v}) where $\vec{v} = (i_1, \dots, i_k)$ is a list of indices in increasing order $i_1 < i_2 < \cdots < i_k$ and $y \in X^k$.

We use the letters b, b' to denote k -blocks, and unless stated otherwise we assume that $b = (y, \vec{v}), b' = (y', \vec{v}')$. For a pair of blocks b, b' we use the notation $b \cap b' = \emptyset$ if $\vec{v} \cap \vec{v}' = \emptyset$. The union of disjoint blocks is defined in the natural way: The set of indices is the concatenation of the indices in the original blocks in correct order, and the values of y, y' are concatenated appropriately.

For a tuple $\mathbf{x} \in \mathbf{X}$ and a set of indices $\vec{v} = (i_1, \dots, i_k)$, we denote $\mathbf{x}_b = \mathbf{x}_{\vec{v}} = (x_{i_1}, \dots, x_{i_k})$. We say that a tuple \mathbf{x} contains a block b and denote $b \subset \mathbf{x}$ if $\mathbf{x}_{\vec{v}} = y$. For a k -block b , denote by $\mathbf{X}_b = \{\mathbf{x} \mid \mathbf{x}_{\vec{v}} = y\}$ and similarly $\mathbf{X}_{b,b'} = \{\mathbf{x} \mid \mathbf{x}_{\vec{v}} = y, \mathbf{x}_{\vec{v}'} = y'\}$.

In particular for a 1-block (x, i) , $\mathbf{X}_{(x,i)} = \{\mathbf{x} \mid x_i = x\}$. For $\mathbf{x} \in \mathbf{X}_b$ and $F : \mathbf{X} \rightarrow \Sigma^\ell$ we define $F(\mathbf{x})_b \stackrel{def}{=} F(\mathbf{x})_{\vec{v}}$.

The following lemma shows that for a Boolean function on \mathbf{X} the expectation remains roughly the same when restricting to a random \mathbf{X}_b . It is used several times in the course of our proof, and is similar to a lemma in [4]. The proof for the precise statement can be found in [8]:

Lemma 2.2 Let X be a set and $n > 1$ an integer, and denote $\mathbf{X} = X^n$. Let $f : \mathbf{X} \rightarrow \{0, 1\}$ with expectation $\mu = \mathbf{E}_{\mathbf{x} \in \mathbf{X}}[f(\mathbf{x})]$. For $(x, i) \in X \times [n]$, denote $\tilde{\mu}_{x,i} = \mathbf{E}_{\mathbf{x} \in \mathbf{X}_{(x,i)}}[f(\mathbf{x})]$. Then,

1. $\Pr_{(x,i)}[|\mu - \tilde{\mu}_{x,i}| \geq 1/\sqrt[3]{n}] \leq 1/\sqrt[3]{n}$
2. $\mathbf{E}_{(x,i)}[(\tilde{\mu}_{x,i})^2] - \mu^2 \leq \frac{\mu}{n}$
3. For $1 \leq r < n$ and an r -block $b = (y, \vec{v})$ denote $\tilde{\mu}_b = \mathbf{E}_{\mathbf{x} \in \mathbf{X}_b}[f(\mathbf{x})]$. Then, $\Pr_b[|\mu - \tilde{\mu}_b| \geq r/\sqrt[3]{n-r}] \leq r/\sqrt[3]{n-r}$.

We conclude with the following standard bounds.

Lemma 2.3 (Chernoff Bound) Let x_1, \dots, x_n i.i.d Bernoulli random variables having $\Pr[x_i = 1] = p$, then: $\Pr[|\sum x_i - pn| > \varepsilon n] < \exp(-\varepsilon^2 n/2)$.

Lemma 2.4 (Chebyshev Bound) Let X be a random variable with expectation μ and variance σ^2 , then, for any $c > 0$: $\Pr[|X - \mu| > c] < \frac{\sigma^2}{c^2}$.

2.2 The Feige-Kilian Dichotomy Lemma

We now turn to describe the Dichotomy Lemma of Feige and Kilian which is the basis for our approach. Without getting into details, in their setting there is a game of questions and answers that is repeated ℓ times. Here a question would be an element of X , and an answer for it would be an element of Σ . More generally, given a k -block (which is essentially a tuple of questions), an answer for it is a tuple $a \in \Sigma^k$.

Fix $F : \mathbf{X} \rightarrow \Sigma^\ell$ for the rest of this section. Let $1 \leq k < \ell$ and let $b = (y, \vec{t})$ be a k -block. The following definitions are quoted from [4].

Definition 2.5 (Live Block) A k -block b is alive if there exists an answer $a \in \Sigma^k$ such that $\Pr_{\mathbf{x} \in \mathbf{X}_b} [F(\mathbf{x})_b = a] \geq \varepsilon$. Such an answer a is called a live answer for b .

Clearly, each block b can have at most $1/\varepsilon$ live answers.

Definition 2.6 Let b be a block and $a \in \Sigma^k$, and let $0 \leq \eta < 1/2$. The pair $(x, i) \in X \times ([\ell] \setminus \vec{t})$ is $1 - \eta$ determined by (b, a) if there exists $\sigma \in \Sigma$ such that $\Pr_{\mathbf{x} \in \mathbf{X}_{b, (x, i)}} [F(\mathbf{x})_i = \sigma \mid F(\mathbf{x})_b = a] \geq 1 - \eta$.

Recall that our goal is to find a direct product $g_1 \times \dots \times g_\ell$ that agrees with F on a noticeable fraction of \mathbf{X} , for some $g_i : X \rightarrow \Sigma$. In the sequel we follow [4] who use notation $g : X \times [\ell] \rightarrow \Sigma$ to group together ℓ functions $g(\cdot, i) : X \rightarrow \Sigma$. Given such a g , we denote by $\vec{g} : \mathbf{X} \rightarrow \Sigma^\ell$ the function defined by $\forall \mathbf{x} = (x_1, \dots, x_\ell) \in \mathbf{X}$, $\vec{g}(\mathbf{x}) \stackrel{\text{def}}{=} (g(x_1, 1), g(x_2, 2), \dots, g(x_\ell, \ell))$

Definition 2.7 (Good Block) A block b is good if b is alive and for every live answer a for it,

$$\Pr_{(x, i) \in X \times ([\ell] \setminus \vec{t})} [(x, i) \text{ is } 1 - \eta \text{ determined by } (b, a)] > 1 - \eta.$$

In that case a is called a good answer for b , and we denote by $g_{b, a} : X \times ([\ell] \setminus \vec{t}) \rightarrow \Sigma$ the function assigning each (x, i) a value σ that maximizes the probability in Definition 2.6. $g_{b, a}$ is called the function that is $1 - \eta$ determined by (b, a) .

For a good block $b = (y, \vec{t})$ and good answer a the function $g_{b, a}$ is only defined on the domain $X \times ([\ell] \setminus \vec{t})$. Therefore, we arbitrarily extend each $g_{b, a}$ to the domain $X \times [\ell]$ demanding only $g_{b, a}(y_i, i) = a_i$ for each $(y_i, i) \in (y, \vec{t})$.

For two vectors $\mathbf{v}, \mathbf{w} \in \Sigma^\ell$ we write $\mathbf{v} \stackrel{1-\eta}{\approx} \mathbf{w}$ to denote $\Pr_{i \in [\ell]} [\mathbf{v}_i = \mathbf{w}_i] \geq 1 - \eta$.

Claim 2.8 Let b be a good block with good answer a . Then for any $\rho > 0$,

$$\Pr_{\mathbf{x} \in \mathbf{X}_b} [F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}_{b, a}(\mathbf{x}) \mid F(\mathbf{x})_b = a] \geq 1 - \frac{2\eta}{\rho}$$

We can now state the Dichotomy Lemma:

Lemma 2.9 (Dichotomy Lemma of [4]) Let $F : \mathbf{X} \rightarrow \Sigma^\ell$, and let $\varepsilon \geq 2\ell^{-1/75}$, then exactly one of following cases holds:

1. (Case 1) The probability that a random k -block is *alive* is at most ε .

2. (Case 2) The probability that a random live k -block is *good* is at least $1 - \varepsilon$.

Proofs of the two above claims can be found in the full version of this paper.

2.3 Agreement

The following definitions will be useful.

Definition 2.10 (Agreement) Fix a k -block b . Let the agreement on b be defined as

$$\mathcal{A}_k(b) = \Pr_{\mathbf{x}, \mathbf{x}' \in \mathbf{X}_b} [F(\mathbf{x})_b = F(\mathbf{x}')_b]$$

and let

$$\mathcal{A}_k = \mathbf{E}_{b: |b|=k} [\mathcal{A}_k(b)]$$

Let k_1, k_2 be integers such that $k_1 + k_2 \leq \ell$. Let b_1 be a k_1 -block and b_2 be a k_2 -block such that $b_1 \cap b_2 = \emptyset$. Define

$$\mathcal{A}_{k_1, k_2}(b_1, b_2) = \Pr_{\mathbf{x}, \mathbf{x}' \in \mathbf{X}_{b_1, b_2}} [F(\mathbf{x})_{b_1} = F(\mathbf{x}')_{b_1}]$$

Observation 2.11 1. $\Pr[T \text{ accepts}] = \mathcal{A}_m$

2. $\mathcal{A}_{k_1, k_2}(b_1, b_2) \geq \mathcal{A}_{k_1 + k_2}(b_1 \cup b_2)$

Lemma 2.12 Let $s, r > 0$ be integers, $s + r \leq \ell$. Fix an s -block b_1 . Then, choosing b_2 disjoint from b_1 ,

$$\mathbf{E}_{b_2} [\mathcal{A}_{s, r}(b_1, b_2)] - \mathcal{A}_s(b_1) \leq \frac{r}{\ell - (r + s)}.$$

Due to space limiting we omit the proof. We remark that the proof is based on Lemma 2.2, and it can be found in the full version of this paper.

3 Local to Global

In this subsection we prove our first main result.

Theorem 1.1 Let $F : \mathbf{X} \rightarrow \Sigma^\ell$ be a function that the test T accepts with probability 3ε , ($\varepsilon \geq 2\ell^{-1/75}$), then there exists a function $g : X \times [\ell] \rightarrow \Sigma$ such that, for $\varepsilon^5/16$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$ it holds:

$$F(\mathbf{x})_i \stackrel{1-9\rho}{\approx} \vec{g}(\mathbf{x}).$$

We begin with the following lemma, showing that if the test T accepts with some probability then most live blocks are good.

Lemma 3.1 If the test T accepts a function F with probability 3ε , ($\varepsilon \geq 2\ell^{-1/75}$), then there exists $m/2 \leq k \leq m$, such that at least ε of the k -blocks are alive, and at least $1 - \varepsilon$ of the live k -blocks are good.

Proof: By Lemma 2.9 either Case 1 or Case 2 apply to F . If Case 1 doesn't apply, then there must be at least ε live k -blocks, of which at least $1 - \varepsilon$ are good and the lemma is proven.

It remains to prove that if Case 1 applies to F , then the test T accepts with probability at most 3ε , thereby contradicting the hypothesis of the lemma.

So assume there are at most ε live k -blocks. Let us rewrite in which the way the test T selects the tuples \mathbf{x}, \mathbf{x}' :

1. Choose a random k -block b' .
2. Pick a random $m-k$ -block b'' , such that $b'' \cap b' = \emptyset$, and let $b = (y, \vec{t})$ be the m -block that is obtained from the union of b' and b'' .
3. Pick a random $\mathbf{x} \in \mathbf{X}_b$.
4. Pick a random $\mathbf{x}' \in \mathbf{X}_b$.

Clearly b is a random m -block, and the distribution over \mathbf{x}, \mathbf{x}' is identical to the distribution induced by T . Now we examine the probability of T in terms of the agreement.

$$\begin{aligned} \Pr[T \text{ accepts}] &= \mathcal{A}_m = \mathbf{E}_b \mathcal{A}_m(b) \\ &= \mathbf{E}_{b', b''} \mathcal{A}_m(b' \cup b'') \quad (1) \\ &\leq \mathbf{E}_{b', b''} \mathcal{A}_{k, m-k}(b', b'') \end{aligned}$$

Where the first equality and last inequality are obtained from Observation 2.11. We separate the expectation in (1) into two parts: the blocks b' that are alive, and those who are not.

The live blocks can contribute up to ε , since they appear with probability $\leq \varepsilon$.

It is left to bound the contribution of each of the non alive blocks: Let b' be a non-alive k block, then according to Lemma 2.12:

$$\mathbf{E}_{b''} [\mathcal{A}_{k, m-k}(b', b'')] - \mathcal{A}_k[b'] \leq \frac{m-k}{\ell - (m-k)} \leq \frac{m}{\ell/2} < \varepsilon$$

Since b' is non alive block, then $\mathcal{A}_k[b'] < \varepsilon$, and therefore: $\mathbf{E}_{b''} [\mathcal{A}_{k, m-k}(b', b'')] < 2\varepsilon$.

Altogether the expectation in (1) is bounded by 3ε . ■

From now until the end of the proof, unless stated otherwise, a block is assumed to be a k -block. Let us define an indicator variable $I(\mathbf{x}, b, b', a, a')$ to be equal 1 iff $\mathbf{x} \in \mathbf{X}_{b, b'}$ and $F(\mathbf{x})_b = a$ and $F(\mathbf{x})_{b'} = a'$ and a, a' are good for b, b' respectively. Now set $I(\mathbf{x}, b, b') \stackrel{\text{def}}{=} \sum_{a, a'} I(\mathbf{x}, b, b', a, a')$. Clearly $I(\mathbf{x}, b, b')$ is either zero or one and it is one exactly if both $F(\mathbf{x})_b$ is a good answer for b and $F(\mathbf{x})_{b'}$ is a good answer for b' . Let \mathcal{D}_1 be a distribution on triples (b, b', \mathbf{x}) defined by choosing

two random k -blocks b, b' such that $b \cap b' = \emptyset$ and a tuple \mathbf{x} containing b, b' . (We recall that for blocks $b = (y, \vec{t})$ and $b' = (y', \vec{t}')$ $b \cap b' = \emptyset$ iff $\vec{t} \cap \vec{t}' = \emptyset$). We first prove that

Lemma 3.2 $\mathbf{E}_{(\mathbf{x}, b, b') \sim \mathcal{D}_1} [I(\mathbf{x}, b, b')] \geq \varepsilon^2$.

We will then consider a graph whose vertices are the blocks and whose edges are roughly between pairs (b, b') for which $\Pr_{\mathbf{x}} [I(\mathbf{x}, b, b')]$ is large. We will then choose a block b^* that has maximal degree in this graph and prove that g_{b^*, a^*} is the global function we are seeking for an appropriate good answer a^* (recall that the function $g_{b, a}$ was defined in Definition 2.7).

Proof: Let us examine another distribution \mathcal{D}_2 defined by first choosing a uniform tuple $\mathbf{x} \in \mathbf{X}$ and then choosing two blocks $b, b' \subset \mathbf{x}$ independently at random.

We prove Lemma 3.2 in two steps. First we prove that

$$\mathbf{E}_{(\mathbf{x}, b, b') \sim \mathcal{D}_2} [I(\mathbf{x}, b, b')] \geq 3\varepsilon^2 \quad (2)$$

and then we argue that $\mathcal{D}_1, \mathcal{D}_2$ are close enough for our needs.

Let $a_{\mathbf{x}} = \Pr_{b \subset \mathbf{x}} [F(\mathbf{x})_b \text{ is alive for } b]$, and $g_{\mathbf{x}} = \Pr_{b \subset \mathbf{x}} [F(\mathbf{x})_b \text{ is good for } b]$. Observe that

$$\mathbf{E}_{\mathbf{x}} [(g_{\mathbf{x}})^2] = \Pr_{(b, b', \mathbf{x}) \sim \mathcal{D}_2} [I(\mathbf{x}, b, b')].$$

We would like to connect the probability that T accepts with the expectation of $a_{\mathbf{x}}$. However, the test T checks consistency on blocks of size m , while $a_{\mathbf{x}}$ refers to blocks of size k . Therefore, we consider a new test T_k which acts as follows:

- Choose a random k block b' .
- Pick a random $m-k$ block b'' , such that $b' \cap b'' = \emptyset$, and let $b = b' \cup b''$.
- Pick $\mathbf{x}, \mathbf{x}' \in \mathbf{X}_b$ uniformly at random.
- Accept iff $F(\mathbf{x})_{b'} = F(\mathbf{x}')_{b'}$.

Claim 3.3 $\Pr[T_k \text{ accepts}] \geq \Pr[T \text{ accepts}]$.

Let $s_{\mathbf{x}}$ denote the probability of T_k succeeding conditioned on choosing \mathbf{x} as the first tuple. $s_{\mathbf{x}} = a_{\mathbf{x}} \cdot \Pr[T_k \text{ succeeds on } b, \mathbf{x} \mid \mathbf{x}, F(\mathbf{x})_b \text{ is alive for } b] + (1 - a_{\mathbf{x}}) \cdot \Pr[T_k \text{ succeeds on } b, \mathbf{x} \mid \mathbf{x}, F(\mathbf{x})_b \text{ is not alive for } b] \leq a_{\mathbf{x}} \cdot 1 + (1 - a_{\mathbf{x}}) \cdot \varepsilon \leq a_{\mathbf{x}} + \varepsilon$ So $a_{\mathbf{x}} \geq s_{\mathbf{x}} - \varepsilon$. Now $\mathbf{E}[a_{\mathbf{x}}] \geq \mathbf{E}[s_{\mathbf{x}}] - \varepsilon = \Pr[T_k \text{ succeeds}] - \varepsilon \geq 2\varepsilon$, and therefore $\mathbf{E}[(a_{\mathbf{x}})^2] \geq \mathbf{E}[a_{\mathbf{x}}]^2 \geq 4\varepsilon^2$. Note that from Lemma 3.1 we get that: $\mathbf{E}[g_{\mathbf{x}}] \geq (1 - \varepsilon)\mathbf{E}[a_{\mathbf{x}}]$, yielding to $\mathbf{E}[(g_{\mathbf{x}})^2] \geq (1 - \varepsilon)^2 \mathbf{E}[a_{\mathbf{x}}]^2 \geq 3\varepsilon^2$ So (2) is established.

Now we have to connect between the distributions \mathcal{D}_1 and \mathcal{D}_2 : Define A as the event of selecting b, b' such that $b \cap b' = \emptyset$.

Claim 3.4 Fix any values of b_1, b_2 and $\mathbf{x}_0 \supset b_1, b_2$ then:
 $\Pr_{\mathcal{D}_1}[\mathbf{x} = \mathbf{x}_0 \text{ and } b = b_1 \text{ and } b' = b_2] = \Pr_{\mathcal{D}_2}[\mathbf{x} = \mathbf{x}_0 \text{ and } b = b_1 \text{ and } b' = b_2 | A]$

Let us calculate

$$\Pr_{\mathcal{D}_2}[A] = \frac{\binom{\ell}{k} \cdot \binom{\ell-k}{k}}{\binom{\ell}{k}^2} \geq (1 - 2k/\ell)^k \geq 1 - 2k^2/\ell. \quad (3)$$

Now let us calculate $\Pr_{\mathcal{D}_2}[I(\mathbf{x}, b, b') | A]$ using Bayes' rule:

$$\begin{aligned} \mathbf{E}_{(\mathbf{x}, b, b') \sim \mathcal{D}_1}[I(\mathbf{x}, b, b')] &= \Pr_{\mathcal{D}_2}[I(\mathbf{x}, b, b') | A] \\ &= \frac{\Pr_{\mathcal{D}_2}[I(\mathbf{x}, b, b') \text{ and } A]}{\Pr_{\mathcal{D}_2}[A]} \geq \Pr_{\mathcal{D}_2}[I(\mathbf{x}, b, b')] - \Pr_{\mathcal{D}_2}[\bar{A}] \end{aligned}$$

Plugging in (2) and (3) we get a lower bound of ε^2 , and we are done. \blacksquare

Our next step is to define a graph on the blocks. Recall that the number of good answers for any block b is at most $1/\varepsilon$, since each good answer is also alive. Let us choose randomly for each good block b a good answer a_b (and an arbitrary answer for the non-good blocks). In expectation over these random choices, (and by Lemma 3.2)

$$\mathbf{E}_{(\mathbf{x}, b, b') \sim \mathcal{D}_1}[I(\mathbf{x}, b, b', a_b, a_{b'})] \geq \frac{\varepsilon^2}{1/\varepsilon^2} = \varepsilon^4. \quad (4)$$

Therefore let us fix some deterministic choice of a_b per b that attains this expectation.

For blocks b_1, b_2 , such that $b_1 \cap b_2 = \emptyset$ let $a_1 = a_{b_1}, a_2 = a_{b_2}$ and if b_1, b_2 are good then let $g_1 = g_{b_1, a_1}, g_2 = g_{b_2, a_2}$. Define $b_1 \sim b_2$ iff $\mathbf{E}_{\mathbf{x}}[I(\mathbf{x}, b_1, b_2, a_1, a_2)] \geq \varepsilon^4/2$. So by (4) and Markov's inequality $\Pr(b_1 \sim b_2) \geq \varepsilon^4/2$ where b_1, b_2 are random blocks such that $b_1 \cap b_2 = \emptyset$ (as implied by the distribution \mathcal{D}_1).

We will prove next that almost always if $b_1 \sim b_2$ then $g_1 \approx g_2$. Let us recall that since b_1 is a good block with good answer a_1 then by Claim 2.8,

$$\Pr_{\mathbf{x} \in \mathbf{X}_{b_1}} \left[F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}_1(\mathbf{x}) \mid F(\mathbf{x})_{b_1} = a_1 \right] \geq 1 - \frac{2\eta}{\rho} \quad (5)$$

Suppose we could replace X_{b_1} by X_{b_1, b_2} , namely, prove that

$$\Pr_{\mathbf{x} \in \mathbf{X}_{b_1, b_2}} \left[F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}_1(\mathbf{x}) \mid F(\mathbf{x})_{b_1} = a_1 \right] \geq 1 - \frac{2\eta}{\rho}. \quad (6)$$

The only difference between (5) and (6) is the domain from which \mathbf{x} is chosen. Similarly suppose this could be done for b_2 and g_2 . In that case we would be on our way

to showing that in fact $g_1 \approx g_2$ essentially since $b_1 \sim b_2$ implies that on a non-negligible fraction of $\mathbf{x} \in X_{b_1, b_2}$, $F(\mathbf{x})$ agrees both with g_1 and with g_2 .

So how do we convert (5) to (6)? The idea is that for a random b_2 , \mathbf{X}_{b_1, b_2} is a random restriction of X_{b_1} which cannot change probabilities too much:

Claim 3.5 Fix a good block b and let $g = g_{b, a_b}$. Then for at least $1 - \frac{2k}{\sqrt[3]{\ell-k}}$ of the blocks b' ,

$$\Pr_{\mathbf{x} \in \mathbf{X}_{b, b'}} \left[F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}(\mathbf{x}) \mid F(\mathbf{x})_b = a_b \right] \geq 1 - \frac{2}{\varepsilon} \left(\frac{2\eta}{\rho} + \frac{k}{\sqrt[3]{\ell-k}} \right) \geq 1 - 6\varepsilon^5.$$

We skip the proof due to space limiting.

Constructing a graph on the blocks. We now construct a graph whose vertices are all the k -blocks in two steps. First, place an edge between b_1 and b_2 iff $b_1 \sim b_2$. Using (3) we know that $\Pr_{b_1, b_2} [b_1 \sim b_2] \geq \Pr_{b_1, b_2: b_1 \cap b_2 = \emptyset} [b_1 \sim b_2] - \Pr[b_1 \cap b_2 \neq \emptyset] \geq \varepsilon^4/2 - 2k^2/\ell$. Hence the graph is pretty dense. Next, for each block b remove (if exist) edges to all blocks b' which violate Claim 3.5, namely, blocks b' for which

$$\Pr_{\mathbf{x} \in \mathbf{X}_{b, b'}} \left[F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}(\mathbf{x}) \mid F(\mathbf{x})_b = a_b \right] < 1 - 6\varepsilon^5.$$

These are blocks on which the transition from \mathbf{X}_b to $\mathbf{X}_{b, b'}$ causes a big change. Claim 3.5 implies that the fraction of edges removed is at most $\frac{4k}{\sqrt[3]{\ell-k}}$. The final graph has edge density at least $\varepsilon^4/2 - 2k^2/\ell - \frac{4k}{\sqrt[3]{\ell-k}} \geq \varepsilon^4/4$.

Concluding the Proof of Theorem 1.1 Let us fix b^* to be a vertex with maximal degree in this graph, and $g = g_{b^*, a_{b^*}}$ will be our global function. The last step in our proof is to show that

$$\Pr_{\mathbf{x}} [\vec{g}(\mathbf{x}) \stackrel{1-9\rho}{\approx} F(\mathbf{x})] \geq \varepsilon^5/16.$$

Let b be a neighbor of b^* in the graph. We first show that

$$\Pr_{(x, i)} [g_{b, a_b}(x, i) = g(x, i)] \geq 1 - 4\rho. \quad (7)$$

Indeed, by Claim 3.5 we know that

$$\Pr_{\mathbf{x} \in \mathbf{X}_{b, b^*}} \left[F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}(\mathbf{x}) \mid F(\mathbf{x})_{b^*} = a^* \right] \geq 1 - 6\varepsilon^5$$

and

$$\Pr_{\mathbf{x} \in \mathbf{X}_{b, b^*}} \left[F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}_{b, a_b}(\mathbf{x}) \mid F(\mathbf{x})_b = a_b \right] \geq 1 - 6\varepsilon^5.$$

On the other hand, since $b \sim b^*$ we know that $\Pr_{\mathbf{x} \in \mathbf{X}_{b, b^*}} [F(\mathbf{x})_b = a_b \text{ and } F(\mathbf{x})_{b^*} = a_{b^*}] \geq \varepsilon^4/2$.

Putting these three equations together we deduce that on at least a fraction $\varepsilon^4/2 - 12\varepsilon^5 \geq \varepsilon^4/4$ of \mathbf{X}_{b,b^*} we have $\vec{g}(\mathbf{x}) \stackrel{1-\rho}{\approx} F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x})$, so $\vec{g}(\mathbf{x}) \stackrel{1-2\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x})$. We now need the following claim.

Claim 3.6 *Let $g_1, g_2 : X \times [t] \rightarrow \Sigma$ be two functions, and let $\beta = \Pr_{(x,i)}[g_1(x,i) \neq g_2(x,i)] > 0$. The fraction of tuples $\mathbf{x} \in X^t$ for which $|\Pr_i[g_1(\mathbf{x}_i, i) \neq g_2(\mathbf{x}_i, i)] - \beta| \geq \beta/2$ is at most $\frac{4}{\beta t}$.*

We apply Claim 3.6 on the space \mathbf{X}_{b,b^*} (which for our purpose is the same as X^t with $t = \ell - 2k$). We deduce that if $\beta = \Pr[g(x, i) \neq g_{b,a_b}(x, i)] > 4\rho$ then the fraction of tuples $\mathbf{x} \in \mathbf{X}_{b,b^*}$ for which $\vec{g}(\mathbf{x}) \stackrel{1-2\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x})$ is at most $\frac{4}{\beta(\ell-2k)}$, and cannot be as large as $\varepsilon^4/4$. So (7) is established.

Choose now a random block b and a random $\mathbf{x} \in \mathbf{X}_b$ (so clearly \mathbf{x} is uniform in \mathbf{X}). b is a neighbor of b^* with probability at least $\varepsilon^4/4$. Conditioned on that and based on Claim 2.8, with probability at least $\varepsilon(1 - \frac{2\rho}{\rho}) > \varepsilon/2$ \mathbf{x} is such that $F(\mathbf{x})_b = a_b$ and also $F(\mathbf{x}) \stackrel{1-\rho}{\approx} \vec{g}_{b,a_b}(\mathbf{x})$. Again by (7), using Claim 3.6 we know that the fraction of tuples on which $\vec{g}_{b,a_b}(\mathbf{x}) \stackrel{1-6\rho}{\approx} \vec{g}(\mathbf{x})$ is small ($< \frac{4}{\beta(\ell-2k)} < \varepsilon^{74}$). On all other tuples we must have (by the triangle inequality) that $F(\mathbf{x}) \stackrel{1-7\rho}{\approx} \vec{g}(\mathbf{x})$. Altogether, this holds for at least $\varepsilon^5/16$ of the tuples $\mathbf{x} \in \mathbf{X}$, and Theorem 1.1 is established. ■

4 From T to T'

In this section we discuss Theorem 1.2 that shows that if F passes the test T' then it noticeably agrees not just with $g : X \times [\ell] \rightarrow \Sigma$ but rather with f^ℓ for $f : X \rightarrow \Sigma$.

Theorem 1.2 *Let $F : \mathbf{X} \rightarrow \Sigma^\ell$ be a function that the test T' accepts with probability $> 5\varepsilon$, ($\varepsilon \geq 2\ell^{-1/93}$), then there exists a function $f : X \rightarrow \Sigma$ such that, for $\Omega(\varepsilon^8)$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$ it holds:*

$$F(\mathbf{x})_i \stackrel{1-29\rho}{\approx} \vec{f}(\mathbf{x}).$$

Let $s : [\ell] \rightarrow [\ell]$ be a permutation. For a vector v we denote $s(v)$ to be the vector defined by $(s(v))_i = v_{s(i)}$. We partition the space \mathbf{X} into equivalence classes such that each class is the set of all permutations of a given \mathbf{x} :

$$C(\mathbf{x}) = \{s(\mathbf{x}) \mid s \text{ is a permutation}\}$$

A function $G : \mathbf{X} \rightarrow \Sigma^\ell$ is called ‘folded’ if it is consistent on every equivalence class, i.e. for all \mathbf{x}, s : $G(s(\mathbf{x})) = s(G(\mathbf{x}))$.

The proof goes by reduction from T to T' . Namely, we randomly reduce F to a ‘folded’ function G . We then apply Theorem 1.1 on G and get a function $g : X \times [\ell] \rightarrow \Sigma$ that agrees on a non-negligible part of the domain of G , and with a little more work we get a function $g : X \rightarrow \Sigma$. For each G we get a (possibly) different g , so it is not immediate to deduce that F too agrees with g on a non-negligible part of the domain. Instead, we first show that the only way F can pass the test with good probability is if it is already somewhat ‘folded’. In other words, on a random equivalence class there are relatively few different values that are supported by at least ε fraction of the class. It is then possible to relate the support of G to the support of F and deduce that F agrees with \vec{g} noticeably.

The proof for this Theorem appears in the full version of the paper.

5 The Structural Theorem

We have already seen (in Theorem 1.2) that if T' accepts the function F with probability ε , then there exists a function $f : X \rightarrow \Sigma$ such that $F(\mathbf{x}) \stackrel{1-O(\rho)}{\approx} \vec{f}(\mathbf{x})$ for $\Omega(\varepsilon^6)$ fraction of the tuples $\mathbf{x} \in \mathbf{X}$. In this section we fully characterize the structure of all functions F on which T' accepts with probability ε . Consider the ‘generic’ example for such a function F as in Section 1.1. Our structural characterization can be viewed as an ‘inverse theorem’ in that for any given F it finds functions f_i and supports $S_i \subseteq \mathbf{X}$ (on which $F(\mathbf{x}) \approx \vec{f}_i(\mathbf{x})$) such that essentially the only way T' will accept on a pair \mathbf{x}, \mathbf{x}' , is if they both belong to S_i for some i . (neglecting an additive $\ell^{-\Omega(1)}$ term). In fact, we prove a stronger statement: whenever T' accepts on \mathbf{x}, \mathbf{x}' then (i) there is an i for which $\mathbf{x}, \mathbf{x}' \in S_i$ and (ii) for all j s.t. $\mathbf{x} \notin S_j$ also $\mathbf{x}' \notin S_j$. This implies the following consistency behavior: if we condition on $\mathbf{x} \in S_i$ for a fixed i , then T' will whp only accept pairs \mathbf{x}, \mathbf{x}' for which also $\mathbf{x}' \in S_i$. Two subtle issues need to be addressed:

1. The number of f_i ’s that agree with F on a non-negligible fraction of \mathbf{X} can be huge, if we allow f_i ’s that are too close to each other. One would like to ‘cluster’ the similar f_i ’s and place one representative from each cluster in the final list. This is tricky but possible, as one needs to ensure that the different clusters are ‘well separated’ so that whenever \mathbf{x} supports the cluster of f_i and \mathbf{x}' does not, T' will reject whp.
2. The next subtlety lies with the possible overlap between the S_i ’s. Even after clustering has been performed, it may happen that S_i and S_j will have a

large intersection. In that case the events $\mathbf{x}, \mathbf{x}' \in S_i$ are not disjoint for different i 's, and possibly even $\sum_i \Pr[\mathbf{x}, \mathbf{x}' \in S_i] \gg 1$.

A finer statement is obtained by considering all possible intersections

$$R_J = (\cap_{j \in J} S_j) \cap (\cap_{j \notin J} \bar{S}_j) \quad J \subseteq [t]$$

noting that the R_J 's are disjoint. We show that whp if T' accepts on a pair \mathbf{x}, \mathbf{x}' then they both belong to exactly the same R_J , and $J \neq \emptyset$.

For $f : X \rightarrow \Sigma$ and $\gamma \in (0, 1)$ we denote $\text{supp}_\gamma(f) \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbf{X} \mid F(\mathbf{x}) \stackrel{1-\gamma}{\approx} \bar{f}(\mathbf{x})\}$ and say that \mathbf{x} γ -supports f if $\mathbf{x} \in \text{supp}_\gamma(f)$. Throughout this section we use the following parameters: $\rho_0 = 23\rho, \delta = \rho_0^8$ and $\varepsilon_0 = 2\ell^{-1/75}$.

Theorem 5.1 (Formal Version of Theorem 1.3) *Let $F : \mathbf{X} \rightarrow \Sigma^\ell$ be a function that the test T' accepts with probability $\alpha > \ell^{-1/150}$. Then there exist functions $f_1, \dots, f_t : X \rightarrow \Sigma$ (with $t < \ell^{O(1)}$) and radii $\rho_1, \dots, \rho_t \in [\rho_0, 2\rho_0]$ such that the following holds. Let $S_j = \text{supp}_{\rho_j}(f_j)$, and for each $J \subseteq [t]$, let $R_J = (\cap_{j \in J} S_j) \cap (\cap_{j \notin J} \bar{S}_j)$. Then*

1. $1 \geq \sum_{J \neq \emptyset} \Pr_{\mathbf{x}, \mathbf{x}'}[\mathbf{x}, \mathbf{x}' \in R_J \mid T' \text{ accepts on } \mathbf{x}, \mathbf{x}'] \geq 1 - O(\rho_0 + \varepsilon_0)/\alpha = 1 - \ell^{-\Omega(1)}$.
2. Set $\varepsilon_J = |R_J|/|\mathbf{X}|$. Then $\Pr_{\mathbf{x}, \mathbf{x}'}[\mathbf{x}, \mathbf{x}' \in R_J] \approx (\varepsilon_J)^2$, and $\sum_{J \neq \emptyset} (\varepsilon_J)^2 \geq \alpha(1 - \ell^{-\Omega(1)})$. In particular, there is some $J \neq \emptyset$ for which $\varepsilon_J \geq \alpha(1 - \ell^{-\Omega(1)})$.

Let us make a few remarks.

- Item 1 implies that conditioned on T' accepting, the queried inputs \mathbf{x}, \mathbf{x}' must whp come from the support of the same non-empty collection of functions $\{f_j\}_{j \in J}$. Since both sides of the inequality are roughly 1 this fully explains the success probability of T' .

Item 2 further claims that the probability that both \mathbf{x} and \mathbf{x}' are chosen in a set R_J is as if they were independent samples, and deduces a nearly tight quantitative lower bound on the possible sizes of the sets R_J .

- We remark that in the informal version we only claimed that

$$\sum_i \Pr_{\mathbf{x}, \mathbf{x}'}[\mathbf{x}, \mathbf{x}' \in S_i \mid T' \text{ accepts on } \mathbf{x}, \mathbf{x}']$$

is at least $\geq 1 - O(\rho_0) = 1 - \ell^{-\Omega(1)}$. Since $\cup S_i = \cup R_J$ this follows from the above. However, it is possibly weaker as discussed above, and this would not enable finding any i for which approximately $|S_i| \geq \alpha |\mathbf{X}|$.

We now turn to explain the proof of this theorem. We begin by describing a straightforward way to prove this theorem and where it fails. Suppose one carries out the following iterative algorithm. Choose a function g_1 that is ρ_0 -supported by the largest fraction of tuples, and let S_1 be the set of tuples ρ_0 -supporting it. By Theorem 1.2, S_1 consists of $\Omega(\alpha^6)$ of the tuples. We can now “erase” F on S_1 (simply by replacing F 's value on those tuples by random values) and repeat. If T' accepts the new F with high enough probability (above some threshold, say ε_0), we find g_2 and S_2 and continue. Since the S_i 's are essentially disjoint we will halt after at most $1/\varepsilon_0^6$ steps. At this time, if T' accepts on \mathbf{x} and \mathbf{x}' then each of them must support some g_i (except with small probability) since the iterative procedure terminated. We now need to rule out the case where \mathbf{x} supports g_1 and say \mathbf{x}' does not support g_1 (but supports say, g_2).

However, this need not hold. It is possible to have a large portion of the tuples support exactly one of g_1, g_2 while nearly supporting the other one (say the distance between $F(\mathbf{x})$ and $\bar{g}_1(\mathbf{x})$ is r which falls within the support threshold, and between $F(\mathbf{x})$ and $\bar{g}_2(\mathbf{x})$ is $r + 1$ which falls outside the support threshold). Two such tuples \mathbf{x}, \mathbf{x}' might easily cause T' to accept.

The essence of the problem is that contrary to the “usual scenario” in locally testing of codes, the direct product encoding does not have a large enough distance between distinct legal codewords (f^ℓ, g^ℓ may be close for $f \neq g$).

In our solution we manage to gather the functions g_i into clusters, such that each cluster has a representative f_j and a radius $\rho_0 \leq \rho_j \leq 2\rho_0$, and we set $S_j = \text{supp}_{\rho_j}(f_j)$. The S_j 's enjoy the property that their boundaries are nearly empty, where the boundary is the set of tuples \mathbf{x} for which $F(\mathbf{x})$ disagrees with $\bar{f}_j(\mathbf{x})$ on $u \in (\rho_j \ell, \rho_j \ell + \delta \ell)$ coordinates. This eliminates the aforementioned obstacle and allows us to complete the proof. The proof of the second item relies on the fact that the transition matrix underlying our test has a large spectral gap to show that $\Pr[\mathbf{x}, \mathbf{x}' \in R_J] \approx (\varepsilon_J)^2$.

The proof for this theorem appears in the full version of the paper.

6 Tightness of Parameters

We would like to claim that our main Theorem 5.1 is tight in the sense that with stronger parameters the theorem does not hold.

Lemma 6.1 *There exists a function $F : \mathbf{X} \rightarrow \Sigma^\ell$ which T' accepts with probability $\Omega(m/\ell)$ and such that for any $f : X \rightarrow \Sigma$ the fraction of tuples \mathbf{x} on which $F(\mathbf{x}) \stackrel{7/8}{\approx} \vec{f}(\mathbf{x})$ is at most $\ell/|X|$.*

In particular this implies the following constraints on the parameters of Theorem 5.1:

- Since $m \geq 1$, (otherwise the test T' is meaningless), then any variant of Theorem 5.1 does not hold if $\varepsilon = O(1/\ell)$. In particular, one cannot hope for an exponentially small ε .
- If $m = \Theta(\ell)$, then Theorem 5.1 cannot hold with arbitrarily small ε , since the acceptance probability of T' on F is $\Omega(1)$ with this choice of m . We comment that [7] raised the question of whether passing the consistency test with non-negligible probability imply non-negligible correlation with a direct-product function. This example shows that in their specific parameter setting ($m = \ell/2$) both our test and their test fail to test such a correlation.

The proof can be found in the full version.

Acknowledgement. We would like to thank Avi Wigderson for helpful discussions.

References

- [1] S. Arora and M. Sudan. Improved low degree testing and its applications. *Combinatorial*, pages 365–426.
- [2] I. Dinur, E. Fischer, G. Kindler, R. Raz, and S. Safra. PCP characterizations of NP: Towards a polynomially-small error-probability. In *Proc. 31st ACM Symp. on Theory of Computing*, 1999.
- [3] I. Dinur and O. Reingold. Assignment testers: Towards combinatorial proofs of the PCP theorem. In *Proceedings of the 45th Symposium on Foundations of Computer Science (FOCS)*, 2004.
- [4] U. Feige and J. Kilian. Two prover protocols—low error at affordable rates. In *Proc. 26th ACM Symp. on Theory of Computing*, pages 172–183, 1994.
- [5] O. Goldreich and S. Safra. A combinatorial consistency lemma with application to proving the PCP theorem. In *RANDOM: International Workshop on Randomization and Approximation Techniques in Computer Science*. LNCS, 1997.
- [6] R. Impagliazzo, R. Jaiswal, and V. Kabanets. Approximately list-decoding direct product codes and uniform hardness amplification. In *Proc. 47th IEEE Symp. on Foundations of Computer Science*, pages 187–196, 2006.
- [7] R. Impagliazzo, R. Jaiswal, V. Kabanets, and A. Wigderson. Uniform direct product theorems: Simplified, optimized, and derandomized. In *Proc. 40th ACM Symp. on Theory of Computing*, 2008. to appear.
- [8] R. O’Donnell and V. Guruswami. Lecture notes from a course on: the PCP theorem and hardness of approximation. 2005.
- [9] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998.
- [10] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proc. 29th ACM Symp. on Theory of Computing*, pages 475–484, 1997.
- [11] R. Rubinfeld and M. Sudan. Testing polynomial functions efficiently and over rational domains. In *Proc. 3rd Annual ACM-SIAM Symp. on Discrete Algorithms*, pages 23–32, 1992.