Direct Product Testing With Nearly Identical Sets

Dana Moshkovitz *

November 25, 2014

Abstract

In this work we analyze a direct product test in which each of two provers receives a subset of size $n$ of a ground set $U$, $|U| = \Theta(n)$, and the two subsets intersect in about $(1 - \delta)n$ elements. We show that if each of the provers provides labels to the $n$ elements it received, and the labels of the two provers agree in the intersection between the subsets with non-negligible probability, then the answers of the provers must correspond to a certain global assignment to the elements of $U$.

While previous results only worked for intersection of size at most $n/2$, in our model the questions and expected answers of the two provers are nearly identical. This is related to a recent construction of a unique games instance (ECCC TR14-142) where this setup arises at the “outer verifier” level.

Our main tool is a hypercontractive bound on the Bernoulli-Laplace model (aka a slice of the Boolean hypercube), from which we can deduce a “small set expansion”-type lemma. We then use ideas from a recent work of the author about “fortification” to reduce the case of large intersection to the already studied case of smaller intersection.

1 Direct Product Testing

Let $U$ and $R$ be finite sets, let $n \geq 1$ be a natural number, and let $0 \leq \alpha \leq 1$. We define the following two prover game. The provers are asked to agree on an assignment $f : U \rightarrow R$ ahead of time. Then a verifier picks at random two subsets $S, S' \subseteq U$ that intersect in $\alpha n$ elements on average. The verifier sends one subset to each prover. The prover is supposed to report what $f(x)$ is for every $x$ in the subset it got, but is not necessarily truthful. The verifier checks that the assignments reported by the provers agree on the intersection between the two sets. Let $S$ be the family of sized-$n$ subsets of $U$.

Direct Product Test($\alpha$):

1. The verifier picks uniformly at random a set $S \in S$.

2. The verifier picks a correlated set $S' \in S$ as follows: for $1 \sim \text{Poisson}((1 - \alpha)n)$ times, switch an element from $S$ with a uniform element outside of $S$.

*dmoshkov@csail.mit.edu. Department of Electrical Engineering and Computer Science, MIT. This material is based upon work supported by the National Science Foundation under Grant Number 1218547.

1 Several other choices of a correlated $S'$ have been used in other works. The one we use follows a convention from probability literature.
3. The verifier sends $S$ to one prover, and sends $S'$ to the other prover. Each prover responds with $n$ elements in $R$ that are supposed to be assignments to the $n$ elements in the set the prover got.

4. The verifier checks that the two provers agree on all the assignments of the elements in $S \cap S'$.

If the provers indeed agree on $f : U \to R$ and respond accordingly, then the verifier always accepts. A direct product theorem shows that if a strategy of the provers makes the verifier accept with large probability, then the strategy necessarily gives rise to some global assignment $f : U \to R$ (the exact meaning of giving rise to a global assignment will be discussed shortly).

Many works prove direct product testing theorems [6, 3, 1, 8, 5]. Their main motivation is hardness amplification for PCP (see also [14, 2, 12] for applications). The idea is that the sets $S, S'$ can span roughly $n$ tests conducted on the global assignment $f$ rather than one, thereby amplifying the soundness of the PCP test.

The different works proving direct product testing theorems differ in the size of the intersection they consider (determined by $\alpha$), and in the success probability $\varepsilon$ of the provers from which they can deduce a global assignment. The theorems become more challenging to prove as $\alpha$ increases (more intersection between the provers) and $\varepsilon$ decreases (the test succeeds less often). The frontier of research today is $\alpha = \frac{1}{2}$ and $\varepsilon = \exp(-\Omega(n))$ [5].

The focus of the current work is the “early unique” case of $\alpha = 1 - \delta$ for a possibly very small $\delta > 0$. In this case the sets that the two provers receive are nearly identical. This is a case that was not addressed in any previous work. The nearly identical questions to the provers make it easier for the provers to coordinate answers on the intersection. Indeed, if the provers had exactly identical subsets, they could simply decide on a different assignment for every subset ahead of time, and we could not have possibly expected a direct product testing theorem. The nearly unique setup is related to a recent construction of a unique games instance [9], where the setup arises at the “outer verifier” level.

Next we discuss the sense in which a successful strategy gives rise to a global assignment. To motivate our definition we consider two examples:

**Example 1.1** (list decoding example). For the provers to succeed with probability $p \leq \frac{1}{2}$, the provers can pick several global assignments $f_1, \ldots, f_{1/p} : U \to R$; and answer according to $f_i$ in fraction $p$ of the subsets.

This example shows that for $p \leq \frac{1}{2}$ we are in the “list decoding” regime; there might not be a single global assignment, but rather a set of about $1/p$ global assignments.

**Example 1.2** (local correlation example). The provers can use the intersection between the subsets to agree on assignments locally. They can agree on a different assignment $f_x : U \to R$ for every $x \in U$; then, given a subset of $U$, pick a random element $x$ in the subset, and assign all $n$ elements according to $f_x$. Since the intersection is of size $\alpha n$ on average, the provers have probability about $\alpha$ to agree, even though there are $|U|$ different global assignments, each agreeing only with a small fraction of the subsets.

Because of this example, we aim for a global assignment that agrees with the strategy of the provers on a large fraction of the subsets among those that pass through many common points.

Fix a strategy for the provers in the direct product game. For a small set of points $C \subseteq X$, $|C| < n$, let $S_C$ be those sets in $S$ that contain $C$. 

2
Definition 1.1 (Strategy list decoding). We say that a strategy of a prover in response to questions \( S \) has an \((l, \epsilon, k)\)-list decoding \( f_1, \ldots, f_l : X \rightarrow R \) if there exists a partition \( S = S_1 \cup \ldots \cup S_l \), where for \( 1 \leq i \leq l \), for every subset \( S \in S_i \), the strategy is consistent with \( f_i \) for at least \( 1 - \epsilon \) fraction of the elements in \( S \).

The following was proved by Dinur and Steurer [5]:

Lemma 1.1 (Direct product test [5]). Suppose \( \alpha \leq 1/2 \), and one considers Direct Product Test(\( \alpha \)) conditioned on \( |S \cap S'| = \alpha n \). For any strategy for the provers that the verifier accepts with probability at least \((1 - \eta)^{\alpha n/2} \), there exists \( C \subseteq X \), \(|C| \leq (3/2)\alpha n \), where the strategy of each of the provers on \( S_C \) has an \((R^{\lceil C \rceil}, O(\eta))\)-list decoding.

In this work we prove the following theorem:

Theorem 1 (Direct product test with nearly identical subsets). Let \( 0 < r \leq 1/2 \) such that \( n \leq r |U| / 2 \). Let \( \delta' = \rho \delta / (1 + \rho \delta) \). There exists \( \text{err} \) that is exponentially small in \( \text{poly}(\epsilon)n \), such that for any strategy for the provers that the verifier in Direct Product Test(\( 1 - \delta \)) accepts with probability at least

\[
2 \left( (1 - \eta)^{\alpha n/2} + \text{err} \right)^{\delta'} + \text{err},
\]

there exists \( C \subseteq X \), \(|C| \leq 3rn/2 \), where \( S_C \) has an \((R^{\lceil C \rceil}, O(\eta))\)-list decoding.

1.1 Our Technique

The proof of Theorem 1 uses a couple of ingredients: (i) A hypercontractive inequality for the Bernoulli-Laplace model and a “Small Sets Partitions” lemma. (ii) An idea from [12, 13] that is used to reduce the case of very large intersection to the case of lower intersection that Lemma 1.1 deals with. In this section we discuss those techniques and the contributions of this work.

1.1.1 Small Sets Partition

A “small set expansion” lemma for a graph \( G = (V, E) \) says that any small set of vertices in the graph \( A \subseteq V \), \(|A| \leq \delta |V| \), expands, that is, there can’t be too many edges that stay inside \( A \) relative to the number of edges that touch \( A \) (See Ryan O’Donnell’s book [15] for more details on such theorems and how they are proved for graphs of interest). For this work we prove a “small sets partition” lemma for the graph that corresponds to the direct product test, namely for the graph whose vertex set is the family of all sized-\( n \) subsets of \( U \), and in which edges correspond to tests of the verifier. The lemma says that given two partitions of the sets of vertices into small sets, \( A_1 \cup \ldots \cup A_k = V \) and \( B_1 \cup \ldots \cup B_k = V \), \(|A_i|, |B_i| \leq \delta |V| \) for all \( 1 \leq i \leq k \), there can’t be too many edges between matching \( A_i \) and \( B_i \) over all \( 1 \leq i \leq k \):

Lemma 1.2 (Small sets partitions). Consider two partitions into small sets \( A_1 \cup \ldots \cup A_k = \binom{V}{n} \) and \( B_1 \cup \ldots \cup B_k = \binom{V}{n} \), \( \alpha_i = |A_i| / \binom{V}{n} \), \( \beta_i = |B_i| / \binom{V}{n} \). Then, for any \( 0 < \epsilon < 1 \), if we pick \( S \) uniformly at random from \( \binom{V}{n} \) and \( S' \) by applying Poisson(\( \delta n \)) random transpositions on \( S \), then

\[
\Pr \left[ \exists i, \ S' \in B_i \land S \in A_i \right] \leq \left( 1 / \epsilon \right)^{\frac{k}{2}} \sum_{i=1}^{k} \alpha_i \beta_i + \epsilon^{\delta / \sqrt{2}}.
\]
Note that $\varepsilon$ should be very small in order for $\varepsilon^{\rho^2/2}$ to be small, and, moreover, the sets in the partition should be small for the square-root term in the right hand side to be small as well. The small set partition lemma is proved in Section 2 using a hypercontractive inequality for slices of the hypercube (the Bernoulli-Laplace model) due to Lee-Yau [11] and Gross [7].

1.1.2 Fortification and Small Set Expanders

In [12, 13] the author shows how strong parallel repetition and direct product testing type theorems can be proved fairly easily for natural tests in which the underlying graph is a good extractor (see the definition of fortified games in [13]). In this work we use one of the ideas from these works to reduce the case of large intersection between sets to the already studied case of small intersection between sets. Interestingly, we show that an extractor is not needed for the reduction, only a graph that admits a “small sets partition” lemma as discussed in the previous section (see Section 3). This extends the possible applications of the general fortification paradigm.

1.2 Related Work

Direct product testing can be thought of as a generalization of parallel repetition of two prover games (in the sense that it can be used to simulate parallel repetition [8]). Part of our work is analogous to a recent work of Tulsiani, Wright and Zhou [17] on parallel repetition of projection games on small set expanders (the other part in our work is to argue the relevant small set expansion properties of the relevant graph). The work of Tulsiani et al does not apply to our setting, and we employ different techniques from theirs. More details follow.

Tulsiani et al used a linear algebra approach pioneered by Dinur and Steurer [4] to show that in the small set expander case one can obtain stronger parameters than possible for general projection games (this was already known for expanders [16]). The Dinur-Steurer analysis of parallel repetition relies on Cheeger’s inequality, and the Tulsiani et al result relies on an extension of Cheeger’s inequality to the $k$ largest eigenvalues, which was proved by [10]. Dinur and Steurer were able to convert their parallel repetition [4] to the direct product testing setting [5]. However, there it does not require Cheeger’s inequality and some of the other steps needed in the parallel repetition setting, but does require certain other steps in order to argue about the structure of the strategy.

2 Bernoulli-Laplace and Partitions Into Small Sets

The main technical tool we use in our analysis is a hypercontractive inequality in the Bernoulli-Laplace model, which is the model that corresponds to the random walk on the $n$-sized subsets of $U$ performed by the verifier to pick $S$ and $S'$. The hypercontractive inequality allows us to prove a small sets partition lemma (see Lemma 2.3) for partitions of $S$. In this section we present the inequality and the small sets partition lemma.

The Bernoulli-Laplace model is as follows: Given a set $U$ and a parameter $n$, consider the graph whose vertex set is $\binom{U}{n}$, the family of subsets of size $n$ of $U$. A random walk is defined on the graph, where starting at a set $S \subseteq U$, one makes a random transposition, picking a random element $x \in S$ and switching it with a random element $x' \in U - S$. Let $K$ be the transition matrix of this random walk, and let its Laplacian be $L = I - K$. The noise operator $H_t$ induced
by this random walk is defined by applying a number of random transpositions determined by a Poisson random variable whose mean is $t$:

**Definition 2.1** (Noise operator).

$$H_t \doteq e^{-tL} = e^{-t} \sum_{l=0}^{\infty} \frac{t^l}{l!} K^l.$$ 

Lee and Yau [11] analyzed the log Sobolev constant of this Markov chain and determined that it is:

$$\rho = \Theta \left( 1/ \left( \log \frac{|U|^2}{n(|U| - n)} \right) \right).$$

This implies via Gross’s work [7] the following hypercontractive inequality (where $\|g\|_p = (\mathbb{E}_x [g(x)]^p)^{1/p}$):

**Lemma 2.1** (Hypercontractivity for Bernoulli-Laplace [11, 7]). For every $t \geq 0$, $1 \leq p \leq q \leq \infty$, satisfying $\sqrt{\frac{p-1}{q-1}} \geq e^{-pt/n}$, for any $f$ defined over $\binom{U}{n}$, we have

$$\|H_t f\|_q \leq \|f\|_p.$$

Using Hölder inequality, we can deduce a two-function version of Lemma 2.1:

**Lemma 2.2** (Two function hypercontractivity for Bernoulli-Laplace). For every $t \geq 0$, $1 \leq p \leq q \leq \infty$, satisfying $\sqrt{\frac{p-1}{q-1}} \geq e^{-pt/n}$, for any $f, g$ defined over $\binom{U}{n}$, we have

$$\langle H_t f; g \rangle \leq \|f\|_p \|g\|_{q/(q-1)}.$$

We will further deduce a “small set expansion” type conclusion from Lemma 2.2. The lemma upper bounds the probability that correlated $S, S'$ are in matching parts of the partitions in the case where the sets of the partitions are small.

Note that we work with partitions into small sets rather than with two small sets $A, B \subseteq \binom{U}{n}$, because we might have sets of varying sizes, and if $B$ contains $A$ as well as its boundary, we might get that the probability that $S \in A$ and $S' \in B$ is upper bounded by the probability that $S \in A$, or that the probability that $S' \in B$ given that $S \in A$ is upper bounded by 1, which is a trivial statement.

**Lemma 2.3** (Small sets partitions). Consider two partitions into small sets $A_1 \cup \ldots \cup A_k = \binom{U}{n}$ and $B_1 \cup \ldots \cup B_k = \binom{U}{n}$, $\alpha_i \doteq |A_i| / \binom{|U|}{n}$, $\beta_i \doteq |B_i| / \binom{|U|}{n}$. Then, for any $0 < \varepsilon < 1$, if we pick $S$ uniformly at random from $\binom{U}{n}$ and $S'$ by applying Poisson$(\delta n)$ random transpositions on $S$, then

$$\Pr \left[ \exists i, S' \in B_i \land S \in A_i \right] \leq \sqrt{k} \prod_{i=1}^{k} \alpha_i \beta_i + e^{\delta^2 / 2}.$$

**Proof.** Let $f_i = 1_{B_i}$ be the indicator function of $B_i$, and let $g_i = 1_{A_i}$ be the indicator function of $A_i$. Take $t = \delta n$, $q = 2$ and $p$ so that $\sqrt{\frac{p-1}{q-1}} = e^{-pt/n}$. Note that $e^{-pt/n} \leq 1 - \frac{p^2 \delta}{1 + \rho t}$, so
\[ p \leq 2(1 - \frac{\rho \delta}{1 + p \delta}). \] By Lemma 2.2 and since \( \frac{2}{p} \geq 1/(1 - \frac{\rho \delta}{1 + p \delta}) = 1 + \rho \delta \) and \( 1/p - 1/q \geq \rho \delta/2, \)

\[ \Pr [\exists i, S \in A_i \land S' \in B_i] = \sum_{i=1}^{k} (H_i f_i, g_i) \]
\[ \leq \sum_{i=1}^{k} \|f_i\|_p \|g_i\|_q/(q-1) \]
\[ = \sum_{i=1}^{k} \beta_i^{1/p} \alpha_i^{1-1/q} \]

First let us bound the contribution from \( i \)'s where \( \beta_i \leq \varepsilon: \)
\[ \sum_{i: \beta_i \leq \varepsilon} \beta_i^{1/p} \alpha_i^{1-1/q} \leq \varepsilon \rho \delta/2 \cdot \sum_{i=1}^{k} \sqrt{\alpha_i \beta_i} \]

Cauchy-Schwarz \[ \leq \varepsilon \rho \delta/2 \sqrt{\sum_{i=1}^{k} \alpha_i} \cdot \sqrt{\sum_{i=1}^{k} \beta_i} \]
\[ \leq \varepsilon \rho \delta/2 \]

There are less than \( 1/\varepsilon \) many \( i \)'s where \( \beta_i > \varepsilon \). We bound the contribution from such \( i \)'s using concavity:
\[ \sum_{i: \beta_i > \varepsilon} \beta_i^{1/p} \alpha_i^{1-1/q} \leq \sum_{i: \beta_i > \varepsilon} \sqrt{\alpha_i \beta_i} \]
\[ \leq \sqrt{(1/\varepsilon) \cdot \sum_{i: \beta_i > \varepsilon} \alpha_i \beta_i} \]
\[ \leq \sqrt{(1/\varepsilon) \cdot \sum_{i=1}^{k} \alpha_i \beta_i} \]

The claim follows.

\[ \square \]

3 Analysis of Direct Product Test with Nearly Identical Sets

In this section we’ll prove Theorem 1. For our analysis we will use a generalization of Lemma 1.1 that was also proved by Dinur-Steurer [5]. The generalization considers tests that compare their subsets only on a random part of their intersection.

Lemma 3.1 (Direct product test [5]). Suppose \( r \leq 1/2; r/2 \leq \alpha \leq r; \beta \leq r/2. \) One considers a direct product test that picks at random two sized-\( n \) subsets of \( U \) that intersect in \( (\alpha + \beta)n \) points, and compares their assignments on \( \alpha \) random points of the intersection.

For any strategy for the provers that the verifier accepts with probability at least \( (1 - \eta)^{r^2/2} \), there exists \( C \subseteq X, |C| \leq (\alpha + \beta)n, \) where the strategy of each of the provers on \( SC \) has an \((R|C|, O(\eta))-list decoding.\)

Theorem 1 follows from Lemma 3.1, as well as Lemma 3.2 below, which relates the acceptance probabilities in two tests:
\[ T_1 = T_1(\delta) \] is Direct Product Test \((1-\delta)\) that compares subsets that intersect in \(\approx (1-\delta)n\) elements.

\[ T_2 = T_2(\alpha, \beta) \] compares subsets that intersect in \((\alpha+\beta)n\) elements on random \(\alpha n\) elements.

Note that \(T_1(\delta)\) is the test we wish to analyze, while \(T_2(\alpha, \beta)\) is the test that Lemma 3.1 analyzes.

We use the small sets partition lemma to relate the acceptance probabilities of the two tests:

**Lemma 3.2** (Reduction lemma). Assume that \(n = \gamma |U|\). Let \(\gamma < \alpha < 1, 0 < \delta < 1\) and \(\epsilon > 0\) such that \(\alpha + \gamma + \epsilon < 1 - \delta\). Let \(\delta' = \rho \delta / (1 + \rho \delta)\). For any strategy for the provers, there exist \(\alpha', \beta'\) such that \(\alpha' = (\alpha - \gamma)/(1 - \gamma)\), \(\alpha' + \beta' \leq \alpha + \epsilon\), and

\[
\Pr \left[ T_1(\delta) \text{ accepts} \right] - \text{err} \leq 2 \left( \Pr \left[ T_2(\alpha', \beta') \text{ accepts} \right] + \text{err} \right)^{\delta'},
\]

where \(\text{err}\) is exponentially small in \(\text{poly}(\epsilon)n\).

**Proof.** Fix a strategy for the provers. We consider randomized choices of sets \(S, S'\) in a direct product test such that \(S, S'\) contain a set \(C \subseteq U\) of size \(\alpha'n\). The first choice approximates the choice of sets in \(T_1\), while the second choice approximates the choice of sets in \(T_2\).

**Choice 1:**

1. Pick uniformly at random \(C \subseteq U\) of size \(\alpha'n\).
2. Pick uniformly at random \(S \subseteq U\) of size \(n\) that contains \(C\).
3. Pick \(S' \subseteq U\) by repeatedly, for \(\lambda\) times where \(\lambda \sim \text{Poisson}(\delta n)\), switching an element from \(S - C\) with an element outside \(S\).

Note that in \(T_1\) there is a low probability \(\text{err}\) that the intersection of \(S, S'\) is smaller than \(\alpha'n\), while with choice 1 this cannot happen.

**Choice 2:**

1. Pick uniformly at random \(C \subseteq U\) of size \(\alpha'n\).
2. Pick uniformly and independently at random \(S, S' \subseteq U\) of size \(n\) that contain \(C\).

Since \(n = \gamma |U|\), the set \(S'\) has expected intersection size \(\alpha'n + (1 - \alpha')\gamma n = \alpha n\) with \(S\). With choice 2, the probability that the intersection is of size larger than \((\alpha + \epsilon)n\) is at most \(\text{err}\).

Let \(T_1', T_2'\) be the direct product testers that pick their sets \(S, S'\) according to choices 1 and 2, respectively, and compare \(S, S'\) on \(C\). We have that

\[
\Pr \left[ T_1' \text{ accepts} \right] \geq \Pr \left[ T_1(\delta) \text{ accepts} \right] - \text{err},
\]

and that there exists \(\beta'\) such that \(\alpha' + \beta' \leq \alpha + \epsilon\) for which

\[
\Pr \left[ T_2(\alpha', \beta') \text{ accepts} \right] \geq \Pr \left[ T_2' \text{ accepts} \right] - \text{err}.
\]

Next we’ll relate the probabilities that \(T_1', T_2'\) accept.

For a fixed \(C \subseteq U\) of size \((\alpha - \gamma)n\), let \(S_C\) be all the \(n\)-sized subsets of \(U\) that contain \(C\). Partition \(S_C\) in two ways, according to the assignments the two provers give \(C\). That is, for
every $\vec{a} \in R^{|C|}$, let $A_{\vec{a}}$ denote all the sets in $S_C$ such that the first prover assigns the elements in $C$ the assignments $\vec{a}$. Similarly, let $B_{\vec{a}}$ all the sets in $S_C$ such that the second prover assigns the elements in $C$ the assignment $\vec{a}$. Let $\alpha_\vec{a} = |A_{\vec{a}}| / |S_C|$, and let $\beta_\vec{a} = |B_{\vec{a}}| / |S_C|$. Note that the acceptance probability of $T'_2$ is precisely

$$\Pr[T_2' \text{ accepts}] = \mathbb{E}_C \left[ \sum_{\vec{a}} \alpha_\vec{a} \beta_\vec{a} \right].$$

Meanwhile, the acceptance probability of $T'_1$ is

$$\Pr[T_1' \text{ accepts}] = \mathbb{E}_C \left[ \Pr_{\vec{s}, \vec{s}' \in S_C} [\exists \vec{a}, S \in A_{\vec{a}} \land S' \in B_{\vec{a}}] \right].$$

From Lemma 2.3 with $\varepsilon = \Pr[T_2' \text{ accepts}]^{1+r}$,

$$\Pr[T_1' \text{ accepts}] \leq 2\Pr[T'_2 \text{ accepts}]^{r'}.$$

The lemma follows. \hfill \Box

Theorem 1 follows from Lemma 3.2 and Lemma 3.1 by taking $\alpha = r$.

References


