Small-Set Expansion in Shortcode Graph and the 2-to-2 Conjecture

Boaz Barak * Pravesh K. Kothari † David Steurer ‡

September 12, 2019

Abstract

Dinur, Khot, Kindler, Minzer and Safra [DKK+16] recently showed that the (imperfect completeness variant of) Khot's 2 to 2 games conjecture follows from a combinatorial hypothesis on the soundness of a certain "Grassmanian agreement tester". In this work, we show that hypothesis of Dinur et al follows from a conjecture we call the "Inverse Shortcode Hypothesis" characterizing the non-expanding sets of the degree-two shortcode graph. We also show the latter conjecture is equivalent to a characterization of the non-expanding sets in the Grassman graph, as hypothesized by a follow-up paper of Dinur et al [DKK+18].

Following our work, Khot, Minzer and Safra [KMS18] proved the "Inverse Shortcode Hypothesis". Combining their proof with our result and the reduction of [DKK+16], completes the proof of the 2 to 2 conjecture with imperfect completeness. Moreover, we believe that the shortcode graph provides a useful view of both the hypothesis and the reduction, and might be useful in extending it further.

*Harvard University, b@boazbarak.org. Supported by NSF awards CCF 1565264 and CNS 1618026, and the Simons Foundation. Part of the work done while the author visited Weizmann Institute during Spring 2017.

†Princeton University and IAS kothari@cs.princeton.edu. Work done while the author visited Weizmann Institute in March 2017.

‡ETH Zurich dsteurer@cs.cornell.edu. Work done while the author was a member of the Institute for Advanced Study, Princeton.

1 Introduction

In [Kho02], Subhash Khot put forward a family of conjectures known as the "d-to-d games conjectures". A binary constraint $P(x_1, x_2)$ where x_is take values in alphabet Σ is said to be *d-to-d* if for every value to x_1, there are exactly d values for x_2 that satisfy P and vice-versa. For any d, the "d-to-d games conjecture" roughly says that for every $\varepsilon > 0$, there is some finite alphabet Σ such that it is NP-hard to distinguish, given a constraint satisfaction problem with d-to-d constraints, whether it is possible to satisfy at least $1 - \varepsilon$ fraction of the constraints, or if every assignment satisfies at most ε fraction of the constraints. [1] The case of $d = 1$ corresponds to the more famous *Unique Games Conjecture*, but until recently there was no constant d for which the corresponding d-to-d conjecture was known to be true.

Dinur, Khot, Kindler, Minzer, and Safra [DKK+16], building on ideas of Khot, Minzer and Safra [KMS17], recently initiated an approach towards proving the 2-to-2 conjecture, based on a certain combinatorial hypothesis positing the soundness of the "Grassmann agreement test". In this work we show that their hypothesis follows from a certain natural hypothesis characterizing the structure of non-expanding sets in the degree two shortcode graph [BGH+15]. Following our work, Khot, Minzer and Safra [KMS18] proved the latter hypothesis thus completing the proof of the 2-to-2 games conjecture. This has several implications to hardness of approximation including improving on the NP-hardness of approximation for Vertex Cover along with a host of other improved NP-hardness results. Perhaps more importantly, this also gives a strong evidence for the truth of the Unique Games Conjecture itself. We defer to [DKK+16, DKK+18, KMS18] for a detailed discussion on the importance of the 2-to-2 games conjecture, as well as the reduction of this conjecture to showing the soundness of the Grassmann agreement tester.

1.1 Our Results

Our main result reduces the task of proving the "Grassmann agreement hypothesis" of Dinur et al [DKK+16, Hypothesis 3.6] to characterizing the structure of non-expanding sets in the associated Grassmann graph.

- We show that the Grassmann agreement hypothesis [DKK+16, Hypothesis 3.6] follows from the Grasmann Expansion Hypothesis [DKK+18, Hypothesis 1.7].
- We describe the related Shortcode test and the associated agreement and expansion hypothesis and relate them to the Grassmann versions above.

The above, combined with the work of [DKK+16, KMS18], suffices to prove the 2-to-2 conjecture. However we note that it is possible to directly obtain a proof of the 2-to-2 conjecture (see the recent exposition at [BCS18]) using the "Inverse Shortcode Hypothesis" without going through the Grassmann graph at all. We think the shortcode view provides a natural way to understand the reduction and suggests potential extensions, see Section 1.6.

[1] For $d > 1$, the conjectures are often stated in their *perfect completeness variant*, where we replace $1 - \varepsilon$ with 1 in the first case. In this work (as well as all the line of works following [KMS17]), we refer to the imperfect completeness version as stated above.

1.2 Grassmann Graph and DKKMS Consistency Test

To state our results formally, we need to define the Grassman and shortcode graphs, which we now do. The Grassmann graph $\mathcal{G}(\ell, n)$ with parameters ℓ, n has vertices given by all ℓ-dimensional subspaces (denoted by $\mathcal{V}_\ell$) of $\mathbb{F}_2^n$. Two subspaces V, V' of $\mathbb{F}_2^n$ have an edge between them if $\dim(V \cap V') = \ell - 1$.

Let $\mathsf{LIN}(\mathbb{F}_2^n)$ be the set of all linear functions $\mathbb{F}_2^n \to \mathbb{F}_2$. For every $f \in \mathsf{LIN}(F_2^n)$, let F_f be the map that assign to every $V \in \mathcal{V}_\ell$, $F_f(V) = f_{|V}$ the restriction of the linear function f to the subspace V. Let $\mathsf{LIN}(\ell, n) = \{ F_f \mid f \in \mathsf{LIN}(\mathbb{F}_2^n) \}$ be the set of all such maps.

The Grassmann Consistency test is a two-query test for $\mathsf{LIN}(\ell, n)$ described below:

Test 1: Grassmann Consistency Test

Given: a map F from $\mathcal{V}_\ell \to \mathsf{LIN}(F_2^\ell)$ that maps any $V \in \mathcal{V}_\ell$ to $F(V)$ a linear function on V.

Operation:

1. Pick an edge (V, V') of $\mathcal{G}(\ell, n)$ uniformly at random.
2. Receive $F(V), F(V') \in \mathsf{LIN}(\ell, n)$.
3. Accept if $F(V)_{V \cap V'} = F(V')_{V \cap V'}$ otherwise reject.

It is easy to see the following completeness of the Grassmann graph test.

Fact 1.1 (Completeness)**.** *Suppose $F \in \mathsf{LIN}(\ell, n)$. Then, F passes the Grassman Consistency test with probability* 1*.*

The DKKMS hypothesis conjectures a precise version of soundness of the Grassmann Consistency Test.

Hypothesis 1.2 (DKKMS Soundness Hypothesis)**.** *For every $\delta > 0$, there exists $\varepsilon > 0$, and an integer $r > 0$ such that following holds for sufficiently large $n \gg \ell$.*

Let $F : \mathcal{V}_\ell \to \mathsf{LIN}(\mathbb{F}_2^\ell)$ such that $\mathbb{P}_{(V,V') \sim \mathcal{G}(\ell,n)}[F(V)_{V \cap V'} = F(V')_{V \cap V'}] \geqslant \delta$. Then, there exist subspaces $Q, W \subseteq \mathbb{F}_2^n$ of dimensions r and $n - r$ respectively and a $f \in \mathsf{LIN}(\mathbb{F}_2^n)$ such that

$$\mathop{\mathbb{P}}_{V \sim \mathcal{V}_\ell, Q \subseteq V \subseteq W}[F(V) = f_V] \geqslant \varepsilon.$$

1.3 Shortcode Graph and Consistency Test

We now define the closely related *Degree 2 Shortcode* graph and a immediate analog of the Grassmann consistency test on this graph. For parameters ℓ, n as before, the vertices of the degree 2 shortcode graph $\mathcal{S}_{\ell,n}$ are elements of $\mathsf{Mat}_{\ell,n}$, that is, all matrices on $\mathbb{F}_2$ with dimensions $\ell \times n$. Two vertices M_1 and M_2 have an edge between them if $M_1 - M_2$ is a rank 1 matrix over the field $\mathbb{F}_2$. The 2 query codeword test on this graph is entirely analogous to the one above for the Grassmann graph:

Test 2: Degree 2 Shortcode Consistency Test

Given: a map F from $\mathsf{Mat}_{\ell,n} \to \mathbb{F}_2^\ell$.

Operation:

1. Pick $M_1 \sim \mathsf{Mat}_{\ell,n}$ and a rank 1 matrix $ab^\top$ for vectors $a \in \mathbb{F}_2^\ell$, $b \in \mathbb{F}_2^n$ all uniformly at

random from their respective domains. Let $M_2 = M_1 + ab^\top$.

2. Receive $F(M_1), F(M_2) \in \mathbb{F}_2^\ell$.
3. Accept if $F(M_2) \in \{F(M_1), F(M_1) + a\}$.

Just as the Grassmann consistency test, the above shortcode consistency test is "2-to-2" constraint and the following completeness is easy to establish.

Fact 1.3 (Completeness). *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be any affine linear function. Let $F = F_f : \mathsf{Mat}_{\ell,n} \to \mathbb{F}_2^\ell$ be the map that evaluates f on each row the input matrix. Then, F passes the shortcode consistency test with probability 1.*

The analogous soundness hypothesis can now be stated as:

Hypothesis 1.4 (Degree 2 Shortcode Soundness Hypothesis). *For every $\delta > 0$, there exists $\varepsilon > 0$, and an integer $r > 0$ such that following holds for sufficiently large $n \gg \ell$.*

Let $F : \mathsf{Mat}_{\ell,n} \to \mathbb{F}_2^\ell$ such that $\mathbb{P}_{M\sim\mathsf{Mat}_{\ell,n},a\sim\mathbb{F}_2^\ell,b\sim\mathbb{F}_2^n}[F(M+ab^\top) \in \{F(M), F(M)+a\}] \geqslant \delta$. Then, there exists linear constraints (q_i, t_i) and (r_i, s_i) for $i \leqslant r$ and a $z \in \mathbb{F}_2^n, u \in \mathbb{F}_2^\ell$ such that

$$\mathop{\mathbb{P}}_{M\sim\mathsf{Mat}_{\ell,n}}[F(M) = Mz + u \mid Mq_i = t_i, r_i^\top M = s_i \ \forall i \leqslant r] \geqslant \varepsilon.$$

1.4 Soundness vs Small-Set Expansion in Grasmann/Shortcode Graphs

Recall that for a regular graph G, the expansion of a set S of vertices is the probability, that a random walk beginning at a uniformly random vertex in S steps out of S. That is, $\Phi_G(S) = \mathbb{P}_{v\sim S, v'\sim v}[v' \notin S]$.

The DKKMS Soundness Hypothesis implies a natural characterization small non-expanding sets in the $\mathcal{G}(\ell, n)$ noted below as Hypothesis 1.6. Similarly, the degree 2 shortcode soundness hypothesis implies a natural characterization of non-expanding sets in $\mathcal{S}_{\ell,n}$. We include a brief overview of the argument here and refer the reader to the more extensive commentary in Section 1.3 of [DKK+16] for further details.

Suppose $A_1, A_2, \ldots, A_r$ are "non-expanding" sets that cover a constant fraction of vertices in $\mathcal{G}(\ell, n)$. We construct a labeling strategy F by choosing r uniformly random linear functions $f_i : \mathbb{F}_2^n \to \mathbb{F}_2$ and setting $F(V) = f_i$ if $V \sim A_i$ and $F(V)$ is a random linear function otherwise. Clearly, F doesn't agree with a single linear function on significantly more than $1/r$ fraction of the vertices in $\mathcal{V}_\ell$. On the other hand, if A_is are sufficiently non-expanding, then, a random edge will lie inside one of the A_is with a non-trivially large probability and thus F will satisfy the Grassmann consistency test. In this, case, we will hope that there are subspaces Q, W of constant dimension and co-dimension, respectively such that restricting to subspaces $V \in \mathcal{V}_\ell(Q, W)$ (where $\mathcal{V}_\ell(Q, W)$ is the subset $V \in \mathcal{V}_\ell$ such that $V \subseteq W$) implies that $F(V) = f_V$ for some fixed global linear function f. This can happen in the above example for F only if there are Q, W as above such that one of the $\frac{|A_i \cap \mathcal{V}_\ell(Q,W)|}{|\mathcal{V}_\ell(Q,W)|}$ is $\Omega(1)$ (i.e. independent of ℓ, n). Thus, Hypothesis 1.2 forces that the non-expanding sets A_i to be "structured" (in the sense of having a large density inside $\mathcal{V}_\ell(Q, W)$ for some Q, W of constant dimension and co-dimension, respectively.) This can be interpreted as saying that the non-expansion of any set of vertices in $\mathcal{G}(\ell, n)$ can be "explained" away by a more than typical density in one of the canonical non-expanding sets (i.e., those that contain a subspace Q and are contained inside a subspace W of constant dimension and co-dimension, respectively.)

To formally state the Grassmann Expansion Hypothesis, we define the special non-expanding sets (referred to as "zoom-in" and "zoom-outs" in [DKK+18]):

Definition 1.5 (Nice Sets in Grassmann Graph)**.** A subset $S \subseteq \mathcal{V}_\ell$ of vertices in $\mathcal{G}(\ell, n)$ is said to be r-nice if there are subspace Q, W of $\mathbb{F}_2^n$ of dimension and co-dimension r_1, r_2 respectively such that $r_1 + r_2 = r$ and $S = \{V \subseteq \mathcal{V}_\ell \mid Q \subseteq V \subseteq W\}$.

Hypothesis 1.6 (Grassmann Expansion Hypothesis)**.** *For every $\eta > 0$, there exists δ, r depending only on η such that if $S \subseteq \mathcal{V}_\ell$ satisfies $\Phi_{\mathcal{G}(\ell,n)}(S) < \eta$, then, there are subspaces Q, W over $\mathbb{F}_2^n$ of dimension and co-dimension r_1, r_2 satisfying $r_1 + r_2 \leqslant r$ respectively, such that $\mathbb{P}_{V:Q\subseteq V\subseteq W}[V \in S] \geqslant \delta$.*

Analogously, we can define nice sets in the degree 2 shortcode graph and state the expansion hypothesis. We call Q, a *right* affine subspace of matrices in $\mathsf{Mat}_{\ell,n}$ if there are pairs (q_i, t_i) and every $M \in Q$ satisfies $Mq_i = t_i$. We define a *left* affine subspace analogously.

Definition 1.7 (Nice Sets in Degree 2 Shortcode Graph)**.** A subset $S \subseteq \mathcal{S}_{\ell,n}$ is said to be r-nice if it is an intersection of a left and right affine subspace in $\mathsf{Mat}_{\ell,n}$ with sum of the dimensions r.

Hypothesis 1.8 (Inverse Shortcode Hypothesis)**.** *For every $\eta > 0$, there exist δ, r depending only on η such that for every subset $S \subseteq \mathsf{Mat}_{\ell,n}$, if $\mathbb{P}_{M\sim S, a\sim\mathbb{F}_2^\ell, b\sim\mathbb{F}_2^n}[M + ab^\top \in S] \geqslant \eta$, then, there exists an r-nice set $\mathcal{T} \subseteq \mathcal{S}_{\ell,n}$ such that $|S \cap \mathcal{T}| \geqslant \delta|\mathcal{T}|$.*

While Hypotheses 1.2 and 1.4 posit soundness of a specific "code-word consistency" test associated with the Grassmann/Shortcode graphs, Hypotheses 1.6 and 1.8 ask for a purely graph theoretic property: a characterization of non-expanding sets in $\mathcal{G}(\ell, n)$ and $\mathcal{S}_{\ell,n}$. As such, it appears easier to attack and [DKK+16] thus suggested understanding the structure of non-expanding sets in $\mathcal{G}(\ell, n)$ as a natural first step. As we show in this note, proving Hypothesis 1.8 is in fact enough to show Hypothesis 1.2. In a follow up work [KMS18], this result was used in to complete the proof of the DKKMS soundness hypothesis.

1.5 Our Results

We are now ready to state our main results formally.

First, we show that the soundness of the shortcode consistency test follows from the expansion hypothesis for the shortcode graph.

Theorem 1.9. *The degree 2 Shortcode Expansion Hypothesis 1.8 implies the Degree 2 Shortcode Soundness Hypothesis 1.4.*

Second, we show that the soundness hypothesis for the shortcode consistency test implies the soundness hypothesis for the Grassmann consistency test. This reduces the DKKMS soundness hypothesis to establishing the expansion hypothesis for the Shortcode graph.

Theorem 1.10. *The degree 2 Shortcode Soundness Hypothesis implies the Grassmann Soundness Hypothesis 1.2.*

Finally, we relate the expansion hypothesis of the Grassmann graph to the expansion hypothesis for the degree 2 shortcode graph.

Theorem 1.11. *The Grassmann Expansion Hypothesis (Hypothesis 1.6) is equivalent to the Inverse Shortcode Hypothesis (Hypothesis 1.8).*

1.6 Discussion

Working with the shortcode consistency test (and consequently, the shortcode expansion hypothesis) makes an approach to proving Hypothesis 1.2 somewhat more tractable. This is because unlike the Grassmann graph, Degree 2 shortcode graph is a Cayley graph on $\mathsf{Mat}_{\ell,n}$ under the group operation of $\mathbb{F}_2$-addition with the set of all rank 1 matrices forming the set of generators. Thus studying expansion of sets of vertices can be approached via powerful methods from Fourier analysis. Indeed, this is the route taken by the recent breakthrough [KMS18] that proves the shortcode expansion hypothesis and completes the proof of the 2-to-2 games conjecture (with imperfect completeness).

Perhaps equally importantly, the shortcode consistency test suggests immediate extensions (*higher degree shortcode graphs*) that provide a natural path to proving the Unique Games Conjecture. We discuss this approach here.

First, the Grassmann/shortcode consistency tests as stated above are "2-to-2" tests. That is, for any reply for the first query, there are two admissible replies for the other query. However, it is simple to modify the tests and make them *unique* or "1-to-1" at the cost of making the completeness 1/2 instead of 1. For concreteness, we describe this simple modification below.

Test 3: Unique Degree 2 Shortcode Consistency Test

Given: a map F from $\mathsf{Mat}_{\ell,n} \to \mathbb{F}_2^\ell$.

Operation:

1. Pick $M_1 \sim \mathsf{Mat}_{\ell,n}$ and a rank 1 matrix $ab^\top$ for vectors $a \in \mathbb{F}_2^\ell$, $b \in \mathbb{F}_2^n$ all uniformly at random from their respective domains. Let $M_2 = M_1 + ab^\top$.
2. Receive $F(M_1), F(M_2) \in \mathbb{F}_2^\ell$.
3. Accept if $F(M_2) = F(M_1)$.

Test 4: Unique Degree 3 Shortcode Consistency Test

Given: a map F from $\mathsf{Ten}_{\ell,m,n} \to \mathbb{F}_2^\ell$.

Operation:

1. Pick $T_1 \sim \mathsf{Ten}_{\ell,m,n}$ and a rank 1 tensor $a \otimes b \otimes c$ for vectors $a \in \mathbb{F}_2^\ell$, $b \in \mathbb{F}_2^m$ and $c \in \mathbb{F}_2^n$ all uniformly at random from their respective domains. Let $T_2 = T_1 + a \otimes b \otimes c$.
2. Receive $F(T_1), F(T_2) \in \mathbb{F}_2^\ell$.
3. Accept if $F(T_2) = F(T_1)$.

It is easy to check that the any strategy that passes the 2-to-2 test can be modified to obtain a success probability of 1/2 in passing the "unique" test above (see proof of Lemma 2.2 below). This is one of the several ways that the NP hardness of "2-to-2" games implies the NP hardness of $(1/2, \varepsilon)$-unique games - that is, distingushing between instances where at least 1/2 the constraints are satisfiable from those where at most ε fraction of constraints are satisfiable.

A natural strategy, thus, to try to show NP hardness of $(1-\varepsilon, \varepsilon)$-unique games is to use some variant of the shortcode consistency test above that has completeness $1-\varepsilon$ instead of 1/2. Indeed,

the degree 2 shortcode consistency test suggests natural analogs with higher completeness - by moving to higher degree shortcode graphs. For concreteness, consider the following test on degree 3 shortcode graphs, where it is easy to argue a completeness of 3/4.

Let $\mathsf{Ten}_{\ell,m,n}$ be the set of all $\ell \times m \times n$ tensors over $\mathbb{F}_2$. Recall that a rank 1 tensor is defined by 3 vectors $a \in \mathbb{F}_2^\ell$, $b \in \mathbb{F}_2^m$ and $c \in \mathbb{F}_2^n$ and can be written as $a \otimes b \otimes c$.

To see why there's a natural analog of the strategy in case of the degree 2 shortcode consistency test that gives a completeness of 3/4, we show:

Lemma 1.12 (Completeness). *Let $y \in \mathbb{F}_2^m$ and $z \in \mathbb{F}_2^n$. Let $F_f : \mathsf{Ten}_{\ell,m,n} \to \mathbb{F}_2^\ell$ be the map that assigns to any tensor T, the value $F(T)_i = \sum_{j,k} T(i,j,k) y_j z_k$. Then, F_f passes the test with probability 3/4.*

Proof. Let T, T' be such that $T - T'$ is rank 1 tensor. Then, F_f passes the test only if $F_f(T - T') = 0$. If $T - T' = a \otimes b \otimes c$, then $F_f(T - T') = \langle b, y \rangle \cdot \langle c, z \rangle a$. Since b, c are independently chosen in the test, the probability that $F_f(T - T') = 0$ is 3/4. □

Thus, the degree 3 shortcode consistency test gives a natural analog of the degree 2 shortcode consistency test with higher completeness. Indeed, degree r version gives a test with completeness of $1 - 2^{-r}$ as expected. One can also frame expansion hypotheses similar to the ones for the degree 2 case that posit a characterization of the non-expanding sets in higher degree shortcode graphs.

While our current efforts to compose this test with the "outer-PCP" in order to get a reduction to Unique Games problem (with higher completeness) have not succeeded, it seems a natural avenue for launching an attack on the UGC.[2]

2 Small-Set-Expansion vs Soundness

In this section, we establish that the inverse shortcode hypothesis (Hypothesis 1.8) implies the soundness of the degree 2 shortcode consistency test 1.4.

From 2-to-2 to Unique Tests. For the sake of exposition, it will be easier to work with Test 1.6, the "unique" version of the degree 2 shortcode consistency test. Thus, we restate the soundness hypothesis for Test 1.6 and show that it is enough to establish Hypothesis 1.4.

Hypothesis 2.1 (Soundness of Test 1.6). *For every $\eta > 0$, there exists $\delta > 0$, and an integer $r > 0$ such that following holds for sufficiently large $n \gg \ell$.*

Let $F : \mathsf{Mat}_{\ell,n} \to \mathbb{F}_2^\ell$ such that $\mathbb{P}_{M \sim \mathsf{Mat}_{\ell,n}, a \sim \mathbb{F}_2^\ell, b \sim \mathbb{F}_2^n}[F(M + ab^\top) = F(M)] \geqslant \eta$. Then, there exists linear constraints (q_i, t_i) and (r_i, s_i) for $i \leqslant r$ and a $z \in \mathbb{F}_2^n, u \in \mathbb{F}_2^\ell$ such that

$$\mathbb{P}_{M \sim \mathsf{Mat}_{\ell,n}} [F(M) = Mz + u \mid Mq_i = t_i, r_i^\top M = s_i \; \forall i \leqslant r] \geqslant \delta.$$

We first show that Hypothesis 2.1 implies Hypothesis 1.4.

Lemma 2.2. *Hypothesis 2.1 implies Hypothesis 1.4.*

[2]There are indeed very serious obstacles that must be overcome before carrying this out. Specifically, the reduction of [DKK+16] uses a careful interplay between *smoothness* properties of the outer PCP and *efficiency* or "blow up" properties of the test (i.e., the number of potential queries by the verifier as a function of the number of honest strategies). The tensor based test has too much of a blowup to be able to be simply "plugged in" in the outer PCP used by [DKK+16].

Proof. Let F be the labeling strategy for Test 1.3. We will first obtain a good labeling strategy for Test 1.6 by modifying F slightly.

Choose h uniformly at random from $\mathbb{F}_2^n$. For any $M \in \mathsf{Mat}_{\ell,n}$, let $G(M) = F(M) + Mh$. We claim that if F passes the Test 1.3 with probability η, then G passes Test 1.6 with probability at least $\eta/2$.

To see this, take any M, M' such that $M \sim M'$ in $\mathcal{S}_{\ell,n}$. That is, $M - M' = ab^\top$ for vectors a, b. We will argue that $G(M) = G(M')$ with probability $1/2$. This will imply that in expectation over the choice of h, G satisfies at least $1/2$ the constraints satisfied by F in Test 1.3 completing the proof.

This is simple to see: since F passes the test, $F(M) = F(M')$ or $F(M) - F(M') = a$. WLOG, say the first happens. Observe that G passes the unique test on M, M' if $F(M) + Mh = F(M') + M'h$ or $F(M) - F(M') = (M - M')h = \langle b, h\rangle a$. Since $F(M) = F(M')$, G thus passes if $\langle b, h\rangle = 0$ which happens with probability $1/2$.

□

Expansion to Soundness. We will now show that Hypothesis 1.8 implies Hypothesis 2.1. This completes the proof of Theorem 1.9. A similar argument can be used to directly establish that Hypothesis 1.6 implies Hypothesis 1.2. We do not include it here explicitly. Instead, we relate the expansion and soundness hypothesis for the degree 2 shortcode test to the analogs for the Grassmann test as we believe this could shed light on showing expansion hypotheses for higher degree shortcode tests discussed in the next section.

Lemma 2.3. *Hypothesis 1.8 implies Hypothesis 2.1*

Proof. Let F be the labeling function as in the assumption in Hypothesis 2.1. Then, we know that $\mathbb{P}_{M\sim\mathsf{Mat}_{\ell,n},a\sim\mathbb{F}_2^\ell,b\sim\mathbb{F}_2^n}[F(M) = F(M + ab^\top)] \geqslant \eta$. For any $z \in \{0,1\}^\ell$, let S_z be the set of all matrices M with $F(M) = z$. Then, by an averaging argument, there must be a $z \in \{0,1\}^\ell$ such that $\mathbb{P}_{M\sim S_z,a\sim\mathbb{F}_2^\ell,b\sim\mathbb{F}_2^n}[M + ab^\top \in S_z] \geqslant \eta$.

Apply Hypothesis 1.8 to S_z to obtain r-nice subset Q of $\mathsf{Mat}_{\ell,n}$ such that $|Q \cap S_z| \geqslant \delta|Q|$. Let $Mq = t$ be a affine constraint satisfied by every $M \in Q$. Consider the affine linear strategy $H : \mathsf{Mat}_{\ell,n} \to \mathbb{F}_2^\ell$ that maps any M to $H(M) = Mq + t + z$. Observe that for every $M \in Q$, $H(M) = z$ by this choice. As a result, when $M \sim M'$ are such that $M, M' \subseteq Q$, $\mathbb{P}[H(M) = H(M')] \geqslant \delta$. Thus, H is the "decoded" strategy that satisfies the requirements of Hypothesis 2.1 as required. This completes the proof.

□

3 Relating Grassmann Graphs to Degree 2 Shortcode Graphs

In this section, we show a formal relationship between the Grassmann and the degree Shortcode tests. In particular, we will prove Theorems 1.10 and 1.11.

3.1 A homomorphism from $\mathcal{G}(\ell, n)$ into $\mathcal{S}_{\ell,n}$

Key to the relationship between the two tests is an embedding of the degree 2 shortcode graph $\mathcal{S}_{\ell,n}$ into $\mathsf{Mat}_{\ell,n-\ell}$. We describe this embedding first. As justified in the previous section, it is without loss of generality to work with the "unique" versions of both the tests.

To describe the above embedding, we need the notion of *projection* of a subspace of $\mathbb{F}_2^n$ to a set of coordinates.

Definition 3.1 (Projection of a Subspace)**.** Given a subspace $V \subseteq \mathbb{F}_2^n$, the projection of V to a set of coordinates $S \subseteq [n]$, written as $\mathsf{Proj}_S(V)$ is the subspace of $\mathbb{F}_2^{|S|}$ defined by taking the vectors obtained by keeping only the coordinates indexed by S for every vector $v \in V$.

Let $\mathcal{B} \subseteq \mathbb{F}_2^n$ be the set n-tuples of linearly independent elements of $\mathbb{F}_2^n$, i.e. each $B \in \mathcal{B}$ forms a basis for the vector space $\mathbb{F}_2^n$. We will use B_0 to denote the standard basis $\{e_1, e_2, \ldots, e_n\}$.

We will now describe a class of graph homomorphisms from $\mathcal{G}(\ell, n)$ into $\mathcal{S}_{\ell,n-\ell}$. Each element of this class can be described by a basis B of $\mathbb{F}_2^n$.

For each basis $B \in \mathcal{B}$, let $\mathcal{V}_\ell(B) \subseteq \mathcal{V}_\ell$ be the set of all subspaces $V \in \mathcal{V}_\ell$ such that the projection of V to the first ℓ coordinates when written w.r.t. the basis B is full-dimensional. Our embedding will map each element of $\mathcal{V}_\ell(B)$ into a distinct element of $\mathsf{Mat}_{\ell,n}$ such that the edge structure within $\mathcal{V}_\ell(B)$ in $\mathcal{G}(\ell, n)$ is preserved under this embedding.

Definition 3.2 (Homomorphism from $\mathcal{G}(\ell, n)$ into $\mathcal{S}_{\ell,n}$)**.** Let $\phi = \phi_B : \mathcal{V}_\ell(B) \to Mat_{\ell,n-\ell}$ be defined as follows. Write every vector in the B-basis. For any $V \in \mathcal{V}_\ell(B)$ and for $1 \leqslant i \leqslant \ell$, let v_i be the unique vector in V such that $Proj_{[\ell]}(v_i) = e_i \in \mathbb{F}_2^\ell$. We call $v_1, v_2, \ldots, v_\ell$ to be the *canonical basis* for V.

Define $\phi(V)$ to be the $\ell \times (n-\ell)$ matrix with the i^{th} row given by the projection of v_i on the last $(n-\ell)$ coordinates for each $1 \leqslant i \leqslant \ell$. When the basis B is clear from the context, we will omit the subscript and write ϕ.

It is easy to confirm that ϕ is a bijection from $\mathcal{V}_\ell(B)$ into $\mathsf{Mat}_{\ell,n}$. This is because canonical basis for a subspace V is unique.

Next, we prove some important properties of the homomorphism ϕ that will be useful in the proof of Theorem 1.10.

First, we show that the map ϕ is indeed a homomorphism as promised and thus, preserves edge structure.

Lemma 3.3 (ϕ is a homomorphism)**.** *For $\phi = \phi_B$ defined above and any $V, V' \in \mathcal{V}_\ell(B)$, $V \sim V'$ in $\mathcal{G}(\ell, n)$ iff $\phi(V) \sim \phi(V')$ in $\mathcal{S}_{\ell,n}$.*

Proof. Let $u \in GF(2)^\ell, v \in GF(2)^{n-\ell}$ be arbitrary non-zero vectors that define a rank 1 matrix $uv^\top$. Consider the matrix $M = M_V + uv^\top$. Then, $M \in Mat_{\ell,n-\ell}$ and thus $\phi^{-1}(M) = W \in \mathcal{V}_\ell(B)$. We claim that $dim(W \cap V) = \ell - 1$. Suppose $b_1, b_2, \ldots, b_\ell$ are the rows of M_V. Then, the rows of M are given by $b_i + u_i v$. Thus, W is spanned by $(e_i, b_i + u_i v)$ where e_i is the i^{th} standard basis element on the first ℓ coordinates and the notation $(e_i, b_i + u_i v)$ indicates the concatenation of the vectors in the ordered pair to get a n dimensional vector. In particular, every element of W can be written as $\sum_{i \leqslant \ell} \lambda_i (e_i, b_i) + (\sum_{i \leqslant \ell} \lambda_i u_i) v$ and any such vector is contained in V if $(\sum_{i \leqslant \ell} \lambda_i u_i)$ implying that $dim(V \cap W) = dim(V) - 1 = \ell - 1$.

On the other hand, let V' be a subspace in $\mathcal{V}_\ell(B)$ such that $V' \sim V$ and let M_V and $M_{V'}$ be the matrices obtained via the map ϕ. Then, M_V and $M_{V'}$ must differ in at least one row, say, WLOG, the last row of M_V and $M_{V'}$ are (e_ℓ, v) and (e_ℓ, v') respectively. Notice that since the vector with e_ℓ in the first ℓ coordinates is unique in V, V', neither of $(e_\ell, v), (e_\ell, v')$ belong to the intersection $V \cap V'$. Further, for every vector $z \in V$, either z or $z + (e_\ell, v)$ must be contained in the intersection $V \cap V'$ (as the extra linear equation that $V \cap V'$ satisfies over and above V is satisfied by exactly one of z and $z + (e_\ell, v)$. Thus, by letting $b'_i = b_i + (\ell, v) + (\ell, v')$ to every one of the canonical basis elements b_i of V that are not in $V \cap V'$, we get a set of elements that are all 1) contained in V' 2) $Proj_{[\ell]} b'_i = e_i$ for every i. This then has to be the canonical basis of $M_{V'}$ (by uniqueness of the canonical basis)

and further, the corresponding $M_{V'}$ can be written as $1_S(w + w')^\top$ where S is the set of i such that b_i is not in $V \cap V'$. □

Next, we want to argue that expansion of sets is preserved up to constant factors under the map ϕ. Towards this, we first show that $\mathcal{V}_\ell(B)$ contains a fraction of the vertices of $\mathcal{G}(\ell, n)$ as we next show.

Lemma 3.4 (Projections of Subspaces)**.** *Let $V \sim \mathcal{V}_\ell$ for $\ell \leqslant \sqrt{n}/2$. Then, $\mathbb{P}[\dim \mathsf{Proj}_{[\ell]}(V) = \ell] \geqslant 0.288$ for large enough n and $\ell = \omega(1)$.*

Further, let $V \in \mathcal{V}_\ell(B)$ for some B. Then, at least $1/2$ fraction of the neighbors of V in $\mathcal{G}(\ell, n)$ are contained in $\mathcal{V}_\ell(B)$.

Proof. We can sample a random subspace of ℓ dimension as follows: Choose ℓ uniformly random and independent points from $GF(2)^n$. If they are linearly independent, let V be the subspace spanned by them.

We can estimate the probability that the sampled points are linearly independent as: $\Pi_{i=0}^{\ell-1}(1 - 2^{-n+i}) \geqslant 1 - 2^{-n}2^{\ell^2}$.

Next, we estimate the probability that the projection to first ℓ coordinates of the sampled vectors is linearly independent. By a similar reasoning as above, this probability is at least $\Pi_{i=0}^{\ell-1}(1 - 2^{-\ell+i}) \approx 0.289$ (the limit of this product for large ℓ.)

By a union bound, thus, a random subspace has a full dimensional projection on S with probability at least $0.289 - 2^{-n/2}$ for any $\ell < \sqrt{n}/2$.

For the remaining part, assume that $B = B_0$ - the standard basis. Notice that a random neighbor of V can be sampled as follows: choose a uniformly random basis for V, say $v_1, v_2, \ldots, v_\ell$. Replace v_ℓ by a uniformly random vector v'_ℓoutside of V in $\mathbb{F}_2^n$. Since $V \in \mathcal{V}_\ell(B)$, the projection of V to the first ℓ coordinates is linearly independent. V' would thus satisfy the same property whenever $v_{\ell'}$ is such that the projection of v'_ℓ to the first ℓ coordinates is not in the span of the projection to the first ℓ coordinates of $v_1, v_2, \ldots, v_{\ell-1}$. The chance of this happening is exactly $1/2$. This completes the proof. □

As a consequence of above, we can now obtain that the preimages of non-expanding sets under ϕ are non-expanding in $\mathcal{G}(\ell, n)$.

Lemma 3.5. *Let $T \subseteq \mathsf{Mat}_{\ell,n}$ be a subset satisfying $\mathbb{P}_{M\sim T, M'\sim M}[M' \in T] = \eta$. Then, $\phi^{-1}(T)$ satisfies: $\mathbb{P}_{V\sim\phi^{-1}(T), V'\sim V}[V' \in \phi^{-1}(T)] \geqslant \eta/2$.*

Proof. Let B the basis used to construct ϕ. Then, $\phi(T) \subseteq \mathcal{V}_\ell(B)$. By Lemma 3.4, $1/2$ the neighbors of $\phi(T)$ are contained in $\mathcal{V}_\ell(B)$. By assumption, η fraction of these neighbors are contained inside T. This finishes the proof. □

Via a similar application of Lemma 3.4, we can establish an appropriate converse.

Lemma 3.6. *Let $S \subseteq \mathcal{V}_\ell$ be a subset satisfying $\mathbb{P}_{V\sim S, V'\sim V}[V' \in S] \geqslant \eta$. Then, for a uniformly random choice of basis B for $\mathbb{F}_2^n$, $\mathbb{E}_B |\phi(S \cap \mathcal{V}_\ell(B)| = \Omega(|S|)$ and $\mathbb{P}_{M,M'\sim\phi(S\cap\mathcal{V}_\ell(B)), M'\sim M}[M' \in \phi(S \cap \mathcal{V}_\ell(B)] \geqslant \Omega(\eta)$.*

Finally, we show that r-nice sets in $\mathcal{G}\ell, n$ get mapped to r-nice sets in $\mathcal{S}(\ell, n)$ and vice-versa.

Lemma 3.7. *Let $S \subseteq \mathcal{V}_\ell$ be an r-nice set in $\mathcal{G}(\ell, n)$. Then, $\phi_B(S \cap \mathcal{V}_\ell(B))$ is an r-nice set in $\mathcal{S}_{\ell,n}$. Conversely, if $T \subseteq \mathsf{Mat}_{\ell,n}$ is an r-nice set in $\mathcal{S}_{\ell,n}$ then $\phi^{-1}(T) = Q \cap \mathcal{V}_\ell(B)$ for some r-nice set Q in $\mathcal{G}(\ell, n)$.*

Proof. WLOG, assume that $B = B_0$. We will assume that $S \subseteq \mathcal{V}_\ell$ is the set of all subspaces in $\mathcal{V}_\ell$ contained in a subspace W of co-dimension r. The general case is analogous. Equivalently, if $w_1, w_2, \ldots, w_r$ form a basis for W, then, for every $V \in S \cap \mathcal{V}_\ell(B)$ and ever $v \in V$ $\langle v, w_i \rangle = 0$ for every i.

Consider the canonical basis $v_1, v_2, \ldots, v_\ell$ for V - recall that this means that the projection of v_i to the first ℓ coordinates equal e_i. Thus, for every i, we can write $v_i = (e_i, v_i')$ for some vectors v_i' of $n - \ell$ dimensions.

Then, $\phi(V)$ is the matrix M_V with rows v_i' by our construction. In particular, this means that the M_V satisfies the constrain: $M_V \cdot w_i = t_i$ where t_i is the vector with jth coordinate equal to $\langle e_j, w_i \rangle$. Thus, we have shown that for every $V \in S$, $\phi(V)$ satisfies a set of r affine linear equations.

Conversely, observe that if any M satisfies the affine linear equation $M_V w_i = t_i$ as above, the set of all (e_i, u_i) for $i \leqslant \ell$ where u_i is the ith row of M_V, must span a subspace in S. This yields that $\phi(S \cap \mathcal{V}_\ell(B))$ is an r-nice set.

The converse follows from entirely similar ideas. Suppose $T \subseteq \mathsf{Mat}_{\ell,n}$ is an r-nice set. WLOG, we restrict to the case where T is the set of all matrices satisfying linear constraints $Mq_i = t_i$ for some choice of r linearly independent constraints (q_i, t_i). Letting $u_1, u_2, \ldots, u_\ell$ be the rows of M, this implies that every vector v in the span of (e_i, u_i) for $i \leqslant \ell$ satisfies the linear equation $\langle q, v \rangle = 0$ where $q = (q_i, t_i(1), t_i(2), \ldots, t_i(\ell))$. This immediately yields that $\phi^{-1}(M)$ is contained in a subspace W of co-dimension r. Conversely, it is easy to check that for every subspace V of dimension ℓ contained in $W \cap \mathcal{V}_\ell(B)$, $\phi(V)$ satisfies the r affine linear constraints above.

This completes the proof.

□

3.2 Shortcode Test vs Grassmann Test

We now employ the homomorphism constructed in the previous subsection to relate the soundness and expansion hypothesis in shortcode and Grassmann tests.

First, we show that the soundness hypothesis for degree 2 shortcode consistency test implies the soundness hypothesis for the Grassmann consistency test and complete the proof of Theorem 1.10.

Lemma 3.8. *The degree 2 shortcode soundness hypothesis (Hypothesis 2.1) implies the Grassmann soundness hypothesis (Hypothesis 1.2).*

Proof. Let F be the assumed labeling strategy in Hypothesis 1.2. We will construct a labeling strategy for $\mathcal{S}_{\ell,n}$ from G so that we can apply the conclusion of 2.1. We will first choose an embedding of the type we constructed before in order to construct G.

Let $B \sim \mathcal{B}$ be chosen uniformly at random and let $\phi = \phi_B$ as in the previous subsection. For any $V \in \mathcal{V}_\ell(B)$, let $F(V) = f$, a linear function restricted to V. Let $v_1, v_2, \ldots, v_\ell$ be the canonical basis for V, i.e., the projection of v_i to the first ℓ coordinates (when written in basis B) equals e_i for every i. Set $G(\phi(V)) = z$ where $z_i = f(v_i)$. Since ϕ is a onto, this defines a labeling strategy for all of $\mathsf{Mat}_{\ell,n}$.

Next, we claim that if F passes the Grassmann consistency test with probability η then G passes the degree 2 shortcode consistency test with probability $\Omega(\eta)$.

Before going on to the proof of this claim, observe that this completes the proof of the lemma. To see this, we first apply Hypothesis 2.1 to conclude that there's an r-nice set Q in $\mathcal{S}_{\ell,n}$ and an affine function defined by $z \in \mathbb{F}_2^{n-\ell}, u \in \mathbb{F}_2^\ell$ such that the labeling strategy $H(M) = Mz + u$ passes the degree 2 shortcode consistency test with probability δ for all M in Q. It it easy to construct the an analogous linear strategy for the Grassmann consistency test: For any $V \in \mathcal{V}_\ell(B)$ with

the canonical basis $v_1, v_2, \ldots, v_\ell$ defined above, set $f(v_i) = u_i + \langle v_i, z\rangle$. Extend f linearly to the span of all such vectors. Finally, extend f to all vectors by taking any linear extension. From Lemma 3.4, 1/2 the neighbors of vertices in $\phi^{-1}(Q)$ are contained in $\mathcal{V}_\ell(B)$. From Lemma **??**, $\phi^{-1}(Q) = \mathcal{F} \cap \mathcal{V}_\ell(B)$ for some r-nice set $\mathcal{F}$ in $\mathcal{G}(\ell, n)$. Finally, by an argument similar to the one in Lemma 3.4, $|\mathcal{F} \cap \mathcal{V}_\ell(B)| \geqslant \Omega(|\mathcal{F}|)$ with high probability over the draw of B. Combining the above three observations yileds that f passes the Grassmann consistency test when restricted to the nice set $\mathcal{F}$ with probability $\Omega(\delta)$.

We now complete the proof of the claim. This follows immediately if we show that for any $V \sim V'$ chosen from $\mathcal{V}_\ell(B)$, $\mathbb{P}_{V\sim V', V, V' \in \mathcal{V}_\ell(B)}[F(V)_{|V} = F(V')_{|V'}] \geqslant 0.07(\eta - 2^{-n+\ell})$.

Without loss of generality, we assume that B is the standard basis $\{e_1, e_2, \ldots, e_n\}$. First, notice that $\mathrm{Span}\{V \cup V'\}$ is of dimension $\ell + 1$ for for all but $2^{-n+\ell}$ fraction of pairs $V \sim V'$. Thus, we can assume that $\mathbb{P}_{V\sim V' | \dim \mathrm{Span}\{V\cup V'\} = \ell+1}[F(V)_{|V} = F(V')_{|V'}] \geqslant \eta - 2^{-n+\ell}$.

Let $C = B^{-1}$, the basis change matrix corresponding to B and let C_i be the i^{th} row of C and let $C_{[\ell]}$ be the matrix formed by taking the first ℓ rows of C. Fix $V \sim V'$ for some $V, V' \in \mathcal{V}_\ell$. Assume now that $\mathrm{Span}\{V \cup V'\}$ is of dimension $\ell + 1$. Let $v_1, v_2, \ldots, v_{\ell-1}$ be a basis for $V \cap V'$. Let $V = \{V \cap V' \cup w_1\}$ and $V' = \{V \cap V' \cup w_2\}$ for some w_1, w_2 that linearly independent of each other and of any vector in $V \cap V'$. We estimate the probability that $V, V' \in \mathcal{V}_\ell(B)$. Then, this is the probability that $v_1, v_2, \ldots, v_{\ell-1}, w_1, w_2$ are mapped by $C^{[\ell]}$ into $a_1, a_2, \ldots, a_{\ell-1}, a_\ell, a_{\ell+1}$ respectively, satisfying $a_\ell, a_{\ell+1} \notin \mathrm{Span}\{a_i \mid i \leqslant \ell - 1\}$. It is easy to check that the probability of this over the random choice of B is at least $0.288 * 1/4 > 0.07$. This proves the claim.

By taking n large enough (compared to ℓ), this probability can be made larger than, say, 0.06η (say). This finishes the proof.

□

Next, we show that the Grassmann Expansion Hypothesis (Hypothesis 1.6) is equivalent to the Inverse Shortcode Hypothesis (Hypothesis 1.8) and complete the proof of Theorem 1.11.

Lemma 3.9. *The Grassmann Expansion Hypothesis (Hypothesis 1.6) is equivalent to the Inverse Shortcode Hypothesis (Hypothesis 1.8).*

Proof. First, we show that Hypothesis 1.6 implies Hypothesis 1.8.

Let $S \subseteq \mathsf{Mat}_{\ell,n}$ be such that $\mathbb{P}_{M\sim S, a\in\mathbb{F}_2^\ell, b\in\mathbb{F}_2^n}[M + ab^\top \in S] = \eta$. Then, by Lemma 3.5, $\phi_B^{-1}(S)$ has an expansion of $\Omega(\eta)$ in $\mathcal{G}(\ell, n)$.

Applying the Grassmann expansion hypothesis (Hypothesis 1.6), we know that there exists a r-nice set $\mathcal{F}$ in $\mathcal{G}(\ell, n)$ such that $|\mathcal{F} \cap \phi_B^{-1}(S)| \geqslant \delta|\mathcal{F}|$. Further, since $\phi_B^{-1}(S) \subseteq \mathcal{V}_\ell(B)$, we must have: $|(\mathcal{F} \cap \mathcal{V}_\ell(B)) \cap \phi_B^{-1}(S)| \geqslant \delta|\mathcal{F} \cap \phi_B^{-1}(S)|$. To finish, observe that by Lemma 3.7, $\phi(\mathcal{F} \cap \phi_B^{-1}(S))$ is an r-nice set, say $\mathcal{Q}$ in $\mathcal{S}_{\ell,n}$. This, show that $|S \cap \mathcal{Q}| \geqslant \delta|\mathcal{Q}|$ completing the proof.

The proof of the other direction, that is, Hypothesis 1.8 implies Hypothesis 1.6, is analogous and relies on the use of Lemma 3.6. □

References

[BCS18] Mitali Bafna, Chi-Ning Chou, and Zhao Song, *An exposition of dinur-khot-kindler-minzer-safra proof for the 2-to-2 games conjecture*, http://boazbarak.org/dkkmsnotes.pdf. 1

[BGH+15] Boaz Barak, Parikshit Gopalan, Johan Håstad, Raghu Meka, Prasad Raghavendra, and David Steurer, *Making the long code shorter*, SIAM J. Comput. **44** (2015), no. 5, 1287–1324. MR 3416138 1

[DKK+16] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra, *Towards a proof of the 2-to-1 games conjecture?*, Electronic Colloquium on Computational Complexity (ECCC) **23** (2016), 198. 1, 3, 4, 6

[DKK+18] _____, *On non-optimally expanding sets in grassmann graphs*, STOC **24** (2018), 94. 1, 3

[Kho02] Subhash Khot, *On the power of unique 2-prover 1-round games*, Proceedings of the 17th Annual IEEE Conference on Computational Complexity, Montréal, Québec, Canada, May 21-24, 2002, 2002, p. 25. 1

[KMS17] Subhash Khot, Dor Minzer, and Muli Safra, *On independent sets, 2-to-2 games, and grassmann graphs*, Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017, 2017, pp. 576–589. 1

[KMS18] _____, *Pseudorandom sets in grassmann graph have near-perfect expansion*, Electronic Colloquium on Computational Complexity (ECCC) **25** (2018), 6. 1, 4, 5