Pseudorandom Sets in Grassmann Graph have Near-Perfect Expansion

Subhash Khot
Courant Institute of Mathematical Sciences
New York University
New York, USA
khot@cims.nyu.edu

Dor Minzer
Department of Computer Science
Tel Aviv University
Tel Aviv, Israel
minzer.dor@gmail.com

Muli Safra
Department of Computer Science
Tel Aviv University
Tel Aviv, Israel
muli.safra@gmail.com

***Abstract*—We prove that pseudorandom sets in the Grassmann graph have near-perfect expansion. This completes the last missing piece of the proof of the 2-to-2-Games Conjecture (albeit with imperfect completeness).**

The Grassmann graph has induced subgraphs that are themselves isomorphic to Grassmann graphs of lower orders. A set of vertices is called pseudorandom if its density within all such subgraphs (of constant order) is at most slightly higher than its density in the entire graph.

We prove that pseudorandom sets have almost no edges within them. Namely, their edge-expansion is very close to 1.

***Keywords*-PCP, 2-to-2 Games, Unique Games Conjecture, Grassmann Graph.**

I. Introduction

This paper [1] completes the proof of the 2-to-2 Games Conjecture (albeit with imperfect completeness) proposed by Dinur, Kindler, and the authors of this paper [34], [12], [13], along with contributions from Barak, Kothari, and Steurer [8] and Moshkovitz and the authors of this paper [40], [33]. The 2-to-2 Games Conjecture has several implications towards approximability of NP-hard problems (and more widely to computational complexity, optimization, and combinatorics). Its proof is completed by proving a certain combinatorial hypothesis proposed in [13] regarding the expansion properties of Grassmann graph. While the main focus of this paper is on the combinatorial hypothesis (stated as Theorem I.12), we present the broader context here for reader's benefit.

A. PCPs, Vertex Cover, Independent Set

An approximation algorithm for an NP-hard problem is an efficient algorithm that computes a solution that is guaranteed to be within a certain multiplicative factor of the optimum (known as the approximation factor). It turns out that for several NP-hard problems, even computing an approximate solution, within a certain multiplicative factor of the optimum, remains an NP-hard problem (known as the hardness factor). The complementary study of approximation algorithms and hardness of approximation aims at characterizing precise approximation threshold for NP-hard problems of interest, i.e. the threshold at which the approximation factor and the hardness factor (essentially) match.

The hardness of approximation results build on a foundational result known as the Probabilistically Checkable Proofs (PCP) Theorem [17], [3], [2]. The theorem can be viewed from a hardness viewpoint as well as from a proof checking viewpoint. From the hardness viewpoint, it states that there exists an absolute constant $\beta < 1$ such that, given a 3SAT formula ϕ, it is NP-hard to distinguish whether it is satisfiable or whether it is at most β-satisfiable (i.e. no assignment satisfies more than β fraction of the clauses). From the proof checking viewpoint, it states that every NP-statement has a polynomial size proof that can be checked efficiently by a probabilistic verifier that reads only a constant number of bits from the proof. The verifier is complete and sound in the sense that a correct proof of a correct statement is accepted with probability 1 and any proof of an incorrect statement is accepted with probability at most, say $\frac{1}{2}$.

The equivalence between the hardness and the proof checking viewpoints, though not difficult to see, has led to many illuminating insights and strong hardness results over the last three decades. The proof checking viewpoint (whose roots go back to the work on interactive proofs) played a decisive role in the discovery of the PCP Theorem. However, for the sake of uniformity and ease of presentation, we adopt the hardness viewpoint here. We refer the reader to the surveys [1], [25], [49], [30] for an overview of the extensive and influential body work on PCPs and hardness results. While we know, by now, optimal hardness results for a handful of problems, e.g. 3SAT, Clique, Set Cover [23], [24], [16], for a vast majority of problems there remains a significant gap between the best known approximation factor and the best known hardness factor.

One such problem is Vertex Cover. Given a graph $G = (V, E)$, a subset $C \subseteq V$ is called a vertex cover if for every edge $e = (u, v) \in E$, either u or v is in C. Finding a minimum vertex cover is a well known NP-hard problem [27]. It admits a simple 2-approximation, namely an efficient algorithm that outputs a cover C of size at most twice the minimum. The algorithm picks an arbitrary edge

[1] A full version of this paper is available at [35].

2575-8454/18/$31.00 ©2018 IEEE
DOI 10.1109/FOCS.2018.00062

IEEE computer society

$e = (u, v) \in E$, adds both u, v to C, removes all edges that are incident on either u or v, and repeats this step. Whether there exists an algorithm with approximation factor strictly below 2 is among the flagship questions in approximability. Surprisingly, there is now good reason to believe that no such algorithm exists, i.e. it is conceivable that approximating Vertex Cover within factor $2 - \varepsilon$ is NP-hard![2]

An independent set in a graph is complement of a vertex cover, i.e. a subset of vertices that contains no edge inside. Towards proving hardness results for Vertex Cover, it is more convenient to prove hardness results for the Independent Set problem (in the special case where the concern is independent sets of *linear* size). Let $0 < s < c \leqslant \frac{1}{2}$ be constants and let $\mathsf{GapIS}(c, s)$ be a "promise gap problem" where an n-vertex graph is given with the promise that either it contains an independent set of size cn or contains no independent set of size sn and the algorithmic task is to distinguish between the two cases. It is easy to see that if $\mathsf{GapIS}(c, s)$ is NP-hard, then approximating Vertex Cover within factor $\frac{1-s}{1-c}$ is NP-hard. Thus, to prove $2 - \varepsilon$ hardness result for Vertex Cover, it is sufficient (and turns out necessary)[3] to prove hardness of $\mathsf{GapIS}(c, s)$ where $s \to 0$ and $c \to \frac{1}{2}$.

There is a simple but nice connection between hardness of the Independent Set problem and PCPs that is worth pointing out. As noted in [10], the hardness of $\mathsf{GapIS}(c, s)$ is equivalent to a PCP with "zero free bits", namely a PCP where the verifier has completeness c, soundness s, and has exactly one accepting answer to her queries (and thus "knows" the answer before reading the queries!). This seems contradictory: if the verifier knows the answer(s) beforehand, why wouldn't she be able to construct the correct proof by herself? The subtle point is that the PCP has imperfect completeness (i.e. $c < 1$), so even for a correct proof of a correct statement, only a fraction c of the answers that the verifier "knows" are actually correct and the verifier cannot tell which c fraction of the answers (among polynomially many) are correct. Indeed, the task of determining which of the answers are correct amounts to finding an independent set in a related graph (hence the connection).

Håstad, building on the works of Bellare, Goldreich, and Sudan [10] and Raz [46], proved that $\mathsf{GapIS}(\frac{1}{4} - \varepsilon, \frac{1}{8} + \varepsilon)$ is NP-hard, implying a hardness factor $\frac{7}{6} - \varepsilon \approx 1.16$ for Vertex Cover. Dinur and Safra [15] proved that $\mathsf{GapIS}(p - \varepsilon, 4p^3 - 3p^4 + \varepsilon)$ is NP-hard for $p \leqslant \frac{3-\sqrt{5}}{2}$, implying a hardness factor $10\sqrt{5} - 21 - \varepsilon \approx 1.36$ for Vertex Cover. Their paper introduced the Biased Long Code and used, for the first time, theorems from Analysis of Boolean functions [4], and implicitly, the notion of 2-to-2 Games. Analysis of Boolean functions has since become ubiquitous in study of PCP constructions. The hardness result of [15] remained the best hardness result for Vertex Cover (until the present work).

As noted, the quest towards proving optimal hardness results was stalled after the remarkable but relatively few successes. The Unique Games Conjecture was introduced in [29] as a plausible avenue to make further progress and moreover presented a hardness result for the Min-2SAT-Deletion problem as a demonstration. In addition, motivated by the Dinur-Safra paper (where the 2-to-2 Games appeared implicitly), the d-to-d Conjecture for $d \geqslant 2$ was also introduced in [29], and it was shown to imply that $\mathsf{GapIS}(1 - 2^{-1/d} - \varepsilon, \varepsilon)$ is NP-hard. For $d = 2$, this would give a hardness factor of $\sqrt{2} - \varepsilon \approx 1.42$ for Vertex Cover, an improvement over Dinur-Safra result.[5] Subsequently, in [39], it was shown that the Unique Games Conjecture implies NP-hardness of $\mathsf{GapIS}(\frac{1}{2} - \varepsilon, \varepsilon)$ and hence hardness of approximating Vertex Cover within $2 - \varepsilon$.

B. The Unique Games Conjecture

For the purposes of this paper, it suffices to define the Unique Games as the following computational problem. Let $\mathbb{F}_2^\ell$ denote the ℓ-dimensional vector space over the binary field $\mathbb{F}_2$, considered as an additive group with the $\oplus$ operation.

Definition I.1. *An instance $\mathcal{U}$ of the $\mathsf{UniqueGame}[\mathbb{F}_2^\ell]$ problem consists of n variables $x_1, \ldots, x_n$ taking values over (the alphabet) $\mathbb{F}_2^\ell$ and m constraints $C_1, \ldots, C_m$ where each constraint C_i is a linear equation of form $x_{i_1} \oplus x_{i_2} = b_i$ and $b_i \in \mathbb{F}_2^\ell$. Let $\mathsf{OPT}(\mathcal{U})$ denote the maximum fraction of the constraints that can be satisfied by any assignment to the instance.*

The term "unique" refers to the specific nature of the constraints: for every assignment to the variable x_{i_1}, there is a unique assignment to the variable x_{i_2} that satisfies the constraint and vice versa (the Unique Games problem was studied earlier by Feige and Lovász in the context of parallel repetition [18]). For constants $0 < s < c < 1$, let $\mathsf{GapUG}[\mathbb{F}_2^\ell](c, s)$ be the gap-version where the instance $\mathcal{U}$ of the $\mathsf{UniqueGame}[\mathbb{F}_2^\ell]$ problem is promised to have either $\mathsf{OPT}(\mathcal{U}) \geqslant c$ or $\mathsf{OPT}(\mathcal{U}) \leqslant s$. The Unique Games Conjecture states that[6].

[2] Unless stated otherwise, $\varepsilon > 0$ will denote an arbitrarily small constant and the statements are meant to hold for every such ε.

[3] One notes that if a graph contains an independent set of fractional size $c \geqslant \frac{1}{2} + \varepsilon$, then by obtaining a 2-approximation for Vertex Cover and taking the complement, one finds an independent set of fractional size $\geqslant 2\varepsilon$. Therefore, the best (c, s) hardness gap one can hope for Independent Set is $c = \frac{1}{2} - \varepsilon$, $s = \varepsilon$.

[4] Namely , Friedgut's Junta Theorem [19] and the Russo-Margulis Lemma [47], [43].

[5] For $d \geqslant 3$, while one does not get improvement for Vertex Cover, one does get NP-hardness for $\mathsf{GapIS}(c, s)$ for a fixed c and $s \to 0$, which in authors' opinion is a more fundamental issue.

[6] The original statement in [29] refers to more general constraints. However it follows from [32] that the original conjecture is equivalent to the statement here, i.e. when the constraints are linear equations over the group $\mathbb{F}_2^\ell$.

Conjecture I.2. *For every constant $\varepsilon > 0$, there exists a sufficiently large integer $\ell = \ell(\varepsilon)$ such that* $\mathsf{GapUG}[\mathbb{F}_2^\ell](1-\varepsilon,\ \varepsilon)$ *is* NP*-hard.*

The Unique Games Conjecture has since been established as a prominent open question in theoretical computer science (please see the surveys [50], [30], [28], [31]). The conjecture has facilitated numerous connections among algorithm design, computational complexity, geometry, and analysis. In particular, assuming the conjecture, optimal hardness results are known for Vertex Cover, Max Cut, Max Acyclic Subgraph [39], [32], [21], super-constant hardness results for Sparsest Cut [11], [41], and one can even deduce optimality of generic approximation algorithms for the entire class of Constraint Satisfaction Problems [44]. In spite of this rather large body of work, the correctness of the conjecture itself remains open.

Over the last decade, several arguments were put forward against the Unique Games Conjecture. We sketch these arguments here to the best of our knowledge.[7]

- There is no known distribution over instances of the Unique Games problem that is plausibly hard (more on this below). In fact, results in [5], [42] showed that the problem is easy on "semi-random" instances (for a rather generous interpretation of the term semi-random), thus indicating otherwise.
- Arora-Barak-Steurer [4] presented an algorithm that solves $(1-\varepsilon)$-satisfiable instances of the Unique Games problem (i.e. finds say $\frac{1}{10}$-satisfying assignment) in time $2^{n^{\varepsilon'}}$ where ε' depends on ε. The algorithm was later incorporated into the Sum-of-Squares framework as well [9]. If the running time of the algorithm is improved so that $\varepsilon' \to 0$ independently of ε, then the Unique Games problem would not be NP-hard.[8] The improvement could arguably come from a quantitative improvement in the connection between the number of large eigenvalues and expansion of small sets in graphs.
- The Arora-Barak-Steuter algorithm and the Unique Games Conjecture together imply that the Unique Games problem has "intermediate complexity" (see Section I-C), a behavior one might not expect for a constraint satisfaction problem.
- No integrality gap instances are known for the Unique Games problem for a constant number, say 10, of rounds of the Sum-of-Squares (a.k.a. Lasserre, Parrilo) hierarchy of semi-definite programming relaxation. Stated differently: arguably, a constant number of rounds of the Sum-of-Squares hierarchy might already qualify as an efficient algorithm for the problem, disproving the Unique Games Conjecture.
- Results in [7] showed that the known integrality gap instances of the problem (that hold for a very basic SDP relaxation) [9] do not survive a few additional rounds of the Sum-of-Squares hierarchy. The authors argued that there is a barrier for the "common" techniques used to construct SDP integrality gaps (in the spirit of the natural proofs barrier for circuit lower bounds).

Given these arguments, the situation was remedied somewhat (from the viewpoint of a believer in the Unique Games Conjecture) in the recent works on "PCPs over reals" [37], [36]. The authors therein construct a candidate explicit family of mildly hard Unique Games instances (the qualifier "candidate" acknowledges that the authors are unable to provide a soundness analysis for the construction). The construction is best viewed as a candidate integrality gap for the problem $\mathsf{GapUG}[\mathbb{F}_2](1-\varepsilon,\ 1-K(\varepsilon)\cdot\varepsilon))$ for $K(\varepsilon)$ rounds of the Sum-of-Squares relaxation. Here the Unique Games problem is over the binary alphabet and $K(\varepsilon) \to \infty$ as $\varepsilon \to 0$, i.e. the gap is (hypothesized to be) super-constant in terms of the fraction of the unsatisfied constraints and for a super-constant number of Sum-of-Squares rounds. One notes that even though the (hypothesized) gap would be mild, it nevertheless would be a big progress. It is known for instance that if the Unique Games problem were shown NP-hard with a mild gap $(1-\varepsilon,\ 1-K(\varepsilon)\cdot\sqrt{\varepsilon})$, the gap could then be boosted by parallel repetition [46], [26], [45] to the gap $(1-\varepsilon',\ \varepsilon')$, proving the Unique Games Conjecture in its fullness.

The context hitherto leads finally to the present line of work: a sequence of papers [34], [12], [13] presented an approach towards proving the related 2-to-2 Games Conjecture[10] (or alternately the Unique Games Conjecture with completeness $\frac{1}{2}$, if reader finds it more convenient to think about). In conjunction with a missing link provided in [8], the approach finally reduced the 2-to-2 Games Conjecture to a concrete combinatorial hypothesis regarding the expansion properties of the Grassmann graph (stated as Theorem I.12). In the present paper, we prove this combinatorial hypothesis, thus completing the proof of the 2-to-2 Games Conjecture (now a theorem).[11] The 2-to-2 Games Theorem gives, among other things, a strong evidence towards the Unique Games Conjecture (in our opinion). All the arguments against the Unique Games Conjecture that we described apply equally well to the 2-to-2 Games Conjecture and in spite of it, the 2-to-2 Games Conjecture, at the end of the day, does happen to be correct!

[7]Since these arguments were not made by us, we are not taking responsibility as to whether these arguments were indeed made or whether these are/were considered pressing arguments. We present these only for reader's benefit.

[8]Under the rather standard hypothesis NP $\not\subseteq \cap_{\gamma>0}$ DTIME (2^{n^γ}).

[9]Qualitatively speaking, there is just one example, from [41]. Lack of qualitatively different examples might itself have been an argument against the Unique Games Conjecture.

[10]Albeit with imperfect completeness; we will ignore this issue henceforth.

[11]As for proving the Unique Games Conjecture itself, in the opinion of the first-named author, the approach in [37], [36] is a more viable approach.

C. The 2-to-2 Games Theorem and its Significance

We now state the 2-to-2 Games Theorem formally, describe its significance, and sketch the developments that led to it. This is followed by the description of the Grassmann graph and its role in these developments.

Definition I.3. *An instance $\mathcal{U}_{2\leftrightarrow 2}$ of the* $\textsf{2-to-2 Game}[\mathbb{F}_2^\ell]$ *problem consists of n variables $x_1, \ldots, x_n$ taking values over (the alphabet) $\mathbb{F}_2^\ell$ and m constraints $C_1, \ldots, C_m$ where each constraint is of the form $T_{ij}x_i \oplus T'_{ij}x_j \in \{b_{ij}, b'_{ij}\}$, T_{ij}, T'_{ij} are $\ell \times \ell$ invertible matrices, and $b_{ij}, b'_{ij} \in \mathbb{F}_2^\ell$. Let $\mathsf{OPT}(\mathcal{U}_{2\leftrightarrow 2})$ denote the maximum fraction of the constraints that can be satisfied by any assignment to the instance.*

The term "2-to-2" refers to the specific nature of the constraints: for every assignment to the variable x_i, there are exactly two assignments to the variable x_j that satisfy the constraint and vice versa. For constants $0 < s < c \leqslant 1$, let $\textsf{Gap 2-to-2}[\mathbb{F}_2^\ell](c, s)$ be the gap-version where the instance $\mathcal{U}_{2\leftrightarrow 2}$ of the $\textsf{2-to-2 Game}[\mathbb{F}_2^\ell]$ problem is promised to have either $\mathsf{OPT}(\mathcal{U}_{2\leftrightarrow 2}) \geqslant c$ or $\mathsf{OPT}(\mathcal{U}_{2\leftrightarrow 2}) \leqslant s$. [12] The 2-to-2 Games Theorem is stated below along with an immediate corollary for the hardness of the Unique Games problem with completeness $\frac{1}{2}$. The latter is obtained by writing each 2-to-2 Game constraint as a pair of Unique Games constraints so that in the completeness case, there is a $\frac{1}{2}(1-\varepsilon)$-satisfying assignment. The completeness can be increased artificially to precisely $\frac{1}{2}$ by adding a small fraction of constraints that are always satisfied.

Theorem I.4. *For every constant $\varepsilon > 0$, there exists a sufficiently large integer $\ell = \ell(\varepsilon)$ such that* $\textsf{Gap 2-to-2}[\mathbb{F}_2^\ell](1-\varepsilon,\ \varepsilon)$ *is* NP*-hard.*

Theorem I.5. *For every constant $\varepsilon > 0$, there exists a sufficiently large integer $\ell = \ell(\varepsilon)$ such that* $\textsf{Gap UG}[\mathbb{F}_2^\ell](\frac{1}{2},\ \varepsilon)$ *is* NP*-hard.*

1) Implications of the 2-to-2 Games Theorem: We now summarize the main implications of the 2-to-2 Games Theorem (with imperfect completeness; some of these implications depend on its specific proof obtained in the present and previous works). As before, $\varepsilon > 0$ denotes constant that can be taken as arbitrarily small.

- **Hardness Results**

The following results were already known based on the 2-to-2 Games Conjecture (as indicated in the references; perfect completeness in the last two results if 2-to-2 Games Conjecture holds with perfect completeness). These represent a big progress, in our opinion, on flagship problems in approximability.

  - [38]: $\textsf{Gap Max Cut}\left(\frac{1}{2} + \Omega(\varepsilon), \frac{1}{2} + \frac{\varepsilon}{\log(1/\varepsilon)}\right)$ is NP-hard. This is optimal up to the constant in the Ω-notation.
  - [29]: $\textsf{Gap Independent Set}\left(1 - \frac{1}{\sqrt{2}} - \varepsilon,\ \varepsilon\right)$ is NP-hard and as a corollary, Vertex Cover is NP-hard to approximate within a factor strictly less than $\sqrt{2}$. Between these two implications, the "correct gap-location" (arbitrarily low soundness) for the Independent Set problem is more fundamental.
  - [14]: It is NP-hard to distinguish whether a graph has four disjoint independent sets of (relative) size $\frac{1}{4} - \varepsilon$ each (and hence is almost 4-colorable) or whether there is no independent set of (relative) size ε (and hence is not almost $\left(\frac{1}{\varepsilon}\right)$-colorable).
  - [22]: It is NP-hard to properly color (using k colors) more than a fraction $1 - \frac{1}{k} + O(\frac{\log k}{k^2})$ edges of an almost k–colorable graph. This is optimal up to the constant in the O-notation.

- **Integrality Gaps, Plausibly Hard Distributions, Cart before the Horse**

The present line of work gives a reduction from the $\textsf{3Lin}$ problem to the 2-to-2 Game problem and subsequently to the Unique Games problem with completeness $\frac{1}{2}$ and to the graph theoretic problems mentioned above. Denoting any of these problems by $\mathcal{P}$, the reduction can be used

  - To "translate" an integrality gap instance of the $\textsf{3Lin}$ problem (see below) to an integrality gap instance of the problem $\mathcal{P}$ (the idea to use a reduction to construct integrality gap was used earlier in [41]).
  - To "translate" a distribution over $\textsf{3Lin}$ instances that is plausibly hard (e.g. random instances with appropriate parameters) to a distribution over $\mathcal{P}$ instances that is plausibly hard.

In both cases, we do not know an alternate construction, i.e. without having to go through a NP-hardness reduction (and lack of any construction so far was an argument against the Unique Games Conjecture as discussed before). On the other hand, "logically", integrality gap construction (and maybe construction of a plausibly hard distribution as well) ought to *precede* an NP-hardness reduction. We find this "cart before the horse" phenomenon quite interesting.

- **Sum-of-Squares Integrality Gaps with Perfect Completeness**
  - If one concerns integrality gap (say up to a poly-

[12]Comments regarding the original formulation of this conjecture in [29]: (1) It was proposed with *perfect completeness*, i.e. stating that $\textsf{Gap 2-to-2}[\mathbb{F}_2^\ell](1, \varepsilon)$ is NP-hard. However, as far as this paper is considered, we view the issue of perfect versus imperfect completeness as being relatively minor. (2) It was proposed with more general constraints (rather than the special case with linear structure described herein) and with "2-to-1" constraints (rather than with "2-to-2" constraints described herein; the conjecture was referred to as the 2-to-1 Conjecture). Both these are non-issues however: the current and preceding works [34], [12], [13] now prove the conjecture *with* linear structure and the constraints are easily reinterpreted as being 2-to-1 constraints (hence proving the 2-to-1 Conjecture as well, which in any case is morally equivalent to the 2-to-2 Conjecture).

nomial number of rounds of the Sum-of-Squares relaxation), the previous result for graph coloring holds with perfect completeness. I.e. there is a graph along with an SDP solution such that (a) the SDP solution pretends as if the graph is 4-colorable whereas (b) in actuality, the graph has no independent set of size ε.

  - Integrality gap (say up to a polynomial number of rounds of the Sum-of-Squares relaxation) for the 2-to-2 Games problem holds with perfect completeness and soundness ε.

  These results are a consequence of the integrality gap known for the $\mathsf{3Lin}$ problem with perfect completeness [20], [48]. The integrality gap instance for $\mathsf{3Lin}$ can be translated via the reduction as remarked above.

- **Intermediate Complexity Theorem**
  Barak [6] pointed out that Theorem I.5, along with the Arora-Barak-Steurer algorithm and the (rather standard) hypothesis $\mathsf{NP} \not\subseteq \bigcap_{\gamma>0} \mathsf{DTIME}(2^{n^\gamma})$, implies the Intermediate Complexity Theorem:
  For every $\varepsilon > 0$, there exist $\varepsilon' > \varepsilon'' > 0$ and integer ℓ such that $\varepsilon' \to 0$ as $\varepsilon \to 0$ and the promise constraint satisfaction problem $\mathsf{GapUG}[\mathbb{F}_2^\ell](\frac{1}{2}, \varepsilon)$ on n variables can be solved in time $2^{n^{\varepsilon'}}$ but not in time $2^{n^{\varepsilon''}}$.
  This is perhaps surprising (and was perhaps cited as an argument against the Unique Games Conjecture). The past experience (e.g. the Dichotomy Conjecture/Theorem, the Exponential Time Hypothesis, near-linear sized PCPs etc) perhaps suggested (and if so, incorrectly as it turns out) that time complexity of n-variable CSPs ought to be either polynomial or truly exponential, i.e. $2^{\Omega(n)}$ amounting to a brute-force search over all assignments.

- **Evidence towards the Unique Games Conjecture**
  $\mathsf{GapUG}(\frac{1}{2}, \varepsilon)$ is NP-hard, i.e. a weaker form of the Unique Games Conjecture holds with completeness $\frac{1}{2}$. As far as the authors know (and we have consulted the algorithmic experts), the known algorithmic attacks on the Unique Games problem work equally well whether the completeness is ≈ 1 or whether it is $\frac{1}{2}$. Thus, the implication that $\mathsf{GapUG}(\frac{1}{2}, \ \varepsilon)$ is NP-hard is a compelling evidence, in our opinion, that the known algorithmic attacks are (far) short of disproving the Unique Games Conjecture.
  Moreover, as remarked before, all the arguments against the Unique Games Conjecture, sketched in Section I-B, apply equally well to its weaker form with completeness $\frac{1}{2}$. In spite of all these arguments, the $\mathsf{GapUG}(\frac{1}{2}, \varepsilon)$ problem, at the end of the day, does happen to be NP-hard, circumventing all the arguments mentioned!

2) Works Leading to the 2-to-2 Games Theorem: As noted, the 2-to-2 Games Theorem is proven in a sequence of works [34], [12], [13] and completed in the present work. In addition, contributions from [40], [8], [33] have been crucial in the overall proof. We summarize these developments below. At a high level, the proof involves the chain of implications (all these are now theorems):

$$\begin{array}{c} [13] \\ \textsf{Grassmann Expansion Hypothesis} \\ \Big\Downarrow \ [8] \\ \textsf{Linearity Testing Hypothesis} \\ \Big\Downarrow \ [34][12] \\ \textsf{2 to 2 Games Conjecture.} \end{array}$$

- The Grassmann graphs and their potential application to the 2-to-2 Game problem were proposed in [34]. The contributions therein were: (1) introducing a certain linearity testing primitive based on the Grassmann graph/code (2) using a certain "sub-code covering" property of this code analogous to a similar property of the Hadamard code, previously introduced in [40] (3) proposing a reduction to the 2-to-2 Game problem (4) proposing a Weak Linearity Testing Hypothesis and showing that it implies, via the reduction, a Weak 2-to-2 Games Conjecture. We do not elaborate on the qualifier "weak" here. It refers to a rather awkward variant that is nevertheless quite natural and useful as far as application to Independent Set and Vertex Cover is concerned, which was the main motivation in [34].
- In [12], the authors formulated a Linearity Testing Hypothesis (the "right" and clean formulation in hindsight) and showed that it implied, via a reduction that is very similar to that in [34], the 2-to-2 Games Conjecture (with imperfect completeness).
- In [34], [12], it was already clear that the connectivity and expansion properties of the Grassmann graph would be crucial towards proving the Linearity Testing Hypotheses therein. In [13], the authors proposed (let's call it) Grassmann Expansion Hypothesis (stated as Theorem I.12 in the present paper), and argued that it would at least be necessary towards proving the Linearity Testing Hypothesis. The authors presented a Fourier analytic framework and a preliminary set of results (for the first and second Fourier levels) towards proving the Grassmann Expansion Hypothesis.
- Barak, Kothari, and Steurer [8] proved that the Grassmann Expansion Hypothesis (almost immediately) implies the Linearity Testing hypothesis. While simple, this link is nevertheless important and was missed by the authors of [13].
- Finally, the Grassmann Expansion Hypothesis is proved in the present paper, stated as Theorem I.12. A similar Johnson Expansion Hypothesis is proved in [33] and the

technical insight therein has been useful in the present paper.

We now discuss the Grassmann Linearity Testing Hypothesis, the Grassmann Expansion Hypothesis, and indicate how the latter implies the former. We do not attempt here an overview of how the Linearity Testing Hypothesis is used to analyze a PCP/reduction and prove the 2-to-2 Games Conjecture. The reader is instead referred to the introductory section of the paper [12].

D. The Grassmann Graph/Code/Test and Linearity Testing Hypothesis

In the following, one thinks of the parameter ℓ as a sufficiently large integer and (after fixing it) the parameter k as a sufficiently large integer.

Definition I.6. *The Grassmann graph $\mathsf{Gr}_{k,\ell}$ is defined as follows. Its vertex set consists of all ℓ-dimensional subspaces L of $\mathbb{F}_2^k$ and (L, L') is an edge if and only if $\mathsf{dim}(L \cap L') = \ell - 1$.*

Associated with the Grassmann graph is the Grassmann code that encodes linear functions $f \colon \mathbb{F}_2^k \to \mathbb{F}_2$. The encoding of a linear function f is given by a word/table $F[\cdot]$ that assigns to each vertex L of the graph, the restriction of f to L, i.e. $F[L] = f|_L$. Since there are 2^ℓ linear functions on an ℓ-dimensional space, the alphabet for the encoding has size 2^ℓ. The Grassmann code is equipped with a natural testing primitive that we call the Grassmann Linearity Test: given a word $F[\cdot]$ (not necessarily a codeword), the test picks an edge (L, L') uniformly at random from the graph and checks that $F[L]|_{L\cap L'} = F[L']_{L\cap L'}$, i.e. that the linear functions $F[L]$ and $F[L']$ are consistent on the common intersection of L, L'.

It is observed immediately that the test is a "2-to-2 test" in the sense that for every assignment/answer $F[L]$ there are exactly two answers to $F[L']$ so that the test accepts (and vice versa). This is because a linear function on $L \cap L'$ can be extended to L (and similarly to L') in exactly two ways (and this is how the test eventually leads to hardness of 2-to-2 Games). By design, the test has perfect completeness: if $F[\cdot]$ is a codeword, then the test passes with probability 1 since $F[L], F[L']$ are then restrictions of the same global linear function. The question of interest is what about the soundness of the test? I.e. if a given word $F[\cdot]$ passes the test with (small) probability $\geqslant \delta$, what "decoding" could we infer? Could we infer that the given word $F[\cdot]$ necessarily has good consistency with some codeword (and if so, list-decode)? Based on the past experience (e.g. the Low Degree Test and the Blum-Luby-Rubinfeld Test that are well-understood and are crucial building blocks of PCPs), one is tempted to speculate that the answer is positive, formally stated below. [13] Here δ, ε are thought of as constants independent of the parameters k, ℓ.

Speculation I.7. *For every $\delta > 0$, there exists $\varepsilon > 0$ such that if a table $F[\cdot]$ passes the Grassmann Linearity Test with probability δ, then there exists a global linear function $f \colon \mathbb{F}_2^k \to \mathbb{F}_2$ such that*

$$\Pr_L [F[L] = f|_L] \geqslant \varepsilon.$$

It turns out however that the speculation is false, the key reason being that the Grassmann graph has small sets whose (edge-)expansion is strictly bounded away from 1.

Definition I.8. *Let $G = (V, E)$ be an n-vertex, d-regular graph. For a non-empty set of vertices $S \subseteq V$ with $|S| \leqslant \frac{n}{2}$, its (edge-)expansion is defined as*

$$\Phi(S) = \frac{|E(S, \overline{S})|}{d \cdot |S|},$$

where $E(S, \overline{S})$ denotes the set of edges with one endpoint in S and the other in $\overline{S} = V \setminus S$.

Alternately, $\Phi(S)$ is the probability that selecting a uniformly random vertex in S and moving along a uniformly random edge incident on that vertex, one lands outside S. We will be interested in whether a set S has expansion very close to 1 (near-perfect expansion) or has expansion strictly bounded away from 1.

Counter-example to Speculation I.7: Consider the following construction (it will be clear soon what the sets S_i would be):

1) Let $S_1, \ldots, S_m$ be disjoint subsets of vertices of the Grassmann graph $\mathsf{Gr}_{k,\ell}$, all of equal size, such that their union constitutes a constant α fraction of vertices of the graph.
2) The sets S_i are very small. Specifically, $m = m(k, \ell) \to \infty$ as $k, \ell \to \infty$.
3) Suppose that $\Phi(S_i) \leqslant 1 - \beta$ for every $1 \leqslant i \leqslant m$ for a constant β.
4) For each $1 \leqslant i \leqslant m$, select a global linear function $f_i \colon \mathbb{F}_2^k \to \mathbb{F}_2$ at random.
5) Define $F[L] = f_i|_L$ for every $L \in S_i$. For $L \notin \cup_{i=1}^m S_i$, $F[L]$ is defined at random.

We show that the word/table $F[\cdot]$ passes the Grassmann test with probability $\alpha\beta$, but has negligible consistency with any global linear function. Firstly, since S_i cover α fraction of vertices and each S_i has expansion at most $1-\beta$, the fraction of edges of the Grassmann graph that are inside some S_i is at least $\alpha\beta$. Since on each S_i, the table $F[\cdot]$ is consistent with the global function f_i, the table passes the test for all edges (L, L') that are inside some S_i. Secondly, since the

[13] Moreover, a positive answer would lead to a very straightforward analysis of the PCP/reduction to 2-to-2 Games, avoiding most of the complications in [34], [12].

functions f_i on different pieces S_i are random and unrelated to each other, no single global function has non-negligible consistency with $F[\cdot]$. This completes the description of the counter-example.

How does one get around this counter-example, i.e. reformulate Speculation I.7 so that it is *correct* as well as *sufficient* towards analysis of the PCP/reduction to 2-to-2 Games? With regards to the specific counter-example above, here is a vacuous statement: if we restrict our attention to *only* the subset of vertices in say S_1, then $F[\cdot]$ indeed has full consistency with a global linear function, namely the function f_1. Moreover, as we will see, a canonical example of a small set with expansion strictly bounded away from 1 is $S = \mathsf{Gr}_{k,\ell}[A, B]$ where $A \subseteq B \subseteq \mathbb{F}_2^k$ are subspaces with $\mathsf{dim}(A) + \mathsf{codim}(B) \leqslant r$ and

$$S = \mathsf{Gr}_{k,\ell}[A, B] = \{L \mid A \subseteq L \subseteq B\}.$$

In this case, $\phi(S) = 1-2^{-r}$ which is strictly bounded away from 1 for small integer r (say $r = 4$). These observations motivated the following Linearity Testing Hypothesis in [12].

Hypothesis I.9. *For every constant $\delta > 0$, there exists a constant $\varepsilon > 0$ and an integer r such that for all sufficiently large integers ℓ and (after fixing it) for all sufficiently large integer k, the following holds. If a table $F[\cdot]$ passes the Grassmann Linearity Test with probability δ, then there exist subspaces $A \subseteq B \subseteq \mathbb{F}_2^k$ with $\mathsf{dim}(A) + \mathsf{codim}(B) \leqslant r$ and a linear function $f\colon B \to \mathbb{F}_2$, such that*

$$\Pr_{A\subseteq L\subseteq B} [F[L] = f|_L] \geqslant \varepsilon.$$

In words, while $F[\cdot]$ need not have good consistency with a global linear function on the *entire* graph $\mathsf{Gr}_{k,\ell}$, there must be a *structured subgraph* $\mathsf{Gr}_{k,\ell}[A, B]$ on which it does have good consistency with a global linear function and moreover this subgraph is of constant "co-order", defined as $\mathsf{dim}(A) + \mathsf{codim}(B)$. The Linearity Testing Hypothesis above was shown to be sufficient towards analysis of the PCP/reduction to 2-to-2 Games in [12]. Towards proving the hypothesis itself, the authors (naturally) proposed to study structure of sets with expansion strictly bounded away from 1, formulated the Grassmann Expansion Hypothesis (essentially) characterizing such sets, and then made partial progress towards proving the Expansion Hypothesis [13]. The authors also argued that proving the Expansion Hypothesis is at least necessary towards proving the Linearity Testing Hypothesis. The missing link, namely that it is also sufficient, was provided in [8].

E. Grassmann Expansion Hypothesis

Definition I.10. *Suppose $A \subseteq B \subseteq \mathbb{F}_2^k$ are subspaces. Let $\mathsf{dim}(A) = a, \mathsf{codim}(B) = b$ and think of a, b as small constants (say $a = b = 2$). Then (as introduced before) the subgraph $\mathsf{Gr}_{k,\ell}[A, B]$ is an induced subgraph of $\mathsf{Gr}_{k,\ell}$ induced on precisely the set of vertices L such that $A \subseteq L \subseteq B$. It is easily seen that $\mathsf{Gr}_{k,\ell}[A, B]$ is an isomorphic copy of a lower order Grassmann graph $\mathsf{Gr}_{k-a-b,\ell-a}$. We call $a + b$ as the co-order of $\mathsf{Gr}_{k,\ell}[A, B]$ with respect to $\mathsf{Gr}_{k,\ell}$.*

The sets $\mathsf{Gr}_{k,\ell}[A, B]$ are natural examples of sets in $\mathsf{Gr}_{k,\ell}$ that have expansion strictly bounded away from 1 (when a, b are small constants). Indeed, the expansion of $\mathsf{Gr}_{k,\ell}[A, B]$, when seen as a subset of $\mathsf{Gr}_{k,\ell}$, has expansion precisely $1 - 2^{-(a+b)}$ (up to an error $O(2^{-\ell})$ which is thought of as negligible and ignored for the ease of presentation). The reasoning is as follows. For a vertex $L \in \mathsf{Gr}_{k,\ell}[A, B]$, its random neighbor L' is obtained by picking a random subspace $T \subseteq L$, $\mathsf{dim}(T) = \ell - 1$ and a random point $x \in \mathbb{F}_2^k \setminus L$ and letting $L' = T \oplus \mathsf{Span}(x)$. Now $L' \in \mathsf{Gr}_{k,\ell}[A, B]$ if and only if $A \subseteq T$ and $x \in B$ and these events happen independently with probabilities 2^{-a} and 2^{-b} respectively (up to an error $O(2^{-\ell})$). Thus a random neighbor of a random vertex in $\mathsf{Gr}_{k,\ell}[A, B]$ is also inside it with probability $2^{-(a+b)}$ and hence its expansion is $1-2^{-(a+b)}$. Furthermore, we observe that if $S \subseteq \mathsf{Gr}_{k,\ell}[A, B] \subseteq \mathsf{Gr}_{k,\ell}$ is such that

$$\frac{|S|}{|\mathsf{Gr}_{k,\ell}[A, B]|} = \varepsilon,$$

then $\Phi(S) \leqslant 1 - \varepsilon \cdot 2^{-(a+b)}$. This is because (we skip the easy proof) any set of density ε inside a Grassmann graph has at least ε^2 fraction of the edges inside it (and hence has expansion at most $1 - \varepsilon$). Therefore, a random neighbor of a random vertex in $S \subseteq \mathsf{Gr}_{k,\ell}[A, B]$ lies inside $\mathsf{Gr}_{k,\ell}[A, B]$ with probability $2^{-(a+b)}$ as seen above and then inside S with probability at least ε, justifying the observation. We summarize the overall observation as:

Fact I.11. *(Informal): A subset of constant density inside a constant co-order copy of Grassmann graph inside a Grassmann graph has expansion strictly bounded away from 1.*

(Formal): Let $S \subseteq \mathsf{Gr}_{k,\ell}[A, B] \subseteq \mathsf{Gr}_{k,\ell}$ be such that $\mathsf{dim}(A) = a, \mathsf{codim}(B) = b$ and the density of S inside $\mathsf{Gr}_{k,\ell}[A, B]$ is ε. Then $\Phi(S) \leqslant 1 - \varepsilon \cdot 2^{-(a+b)}$.

The authors of [13] hypothesize, essentially, that the converse of the above fact is true. Informally, their hypothesis is that any set S in the Grassmann graph $\mathsf{Gr}_{k,\ell}$ whose expansion is strictly bounded away from 1 has constant density inside *some* copy of Grassmann graph of constant co-order. A precise statement appears below (now as a theorem and the main result in this paper):

Theorem I.12. *For every constant $0 < \alpha < 1$, there exists a constant $\varepsilon > 0$ and an integer $r \geqslant 0$ such that for all sufficiently large integers ℓ and (after fixing it) for all sufficiently large integers k, the following holds. let $S \subseteq \mathsf{Gr}_{k,\ell}$ be such that $\Phi(S) \leqslant \alpha$. Then there exist subspaces $A \subseteq B \subseteq \mathbb{F}_2^k$*

such that $\mathsf{dim}(A) = a, \mathsf{codim}(B) = b,\ a + b \leqslant r$ *and*

$$\frac{|S \cap \mathsf{Gr}_{k,\ell}[A,B]|}{|\mathsf{Gr}_{k,\ell}[A,B]|} \geqslant \varepsilon.$$

Following [13], Barak-Kothari-Steurer [8] showed that Theorem I.12 (a hypothesis at the time) implies Hypothesis I.9, and by the work of [34], [12] the 2-to-2 Games Conjecture. It remained therefore to prove Theorem I.12. Partial progress towards its proof was already made in [13] where the authors prove the theorem when $\alpha < \frac{7}{8}$, via spectral analysis of the Grassmann graph, introduced therein (the eigenvalues and eigenspaces of the Grassmann graph were known before). Roughly speaking, given a set S with expansion at most $\alpha < 1 - 2^{-(s+1)}$, it is easily observed that the indicator vector of the set $\mathbf{1}_S$ must have a significant projection onto the eigenspace at "level" at most s (s is a constant when α is strictly bounded away from 1). The spectral analysis then attempts to use this projection to deduce the desired structure of S. The approach is worked out in [13] when $s = 2$, corresponding to $\alpha < \frac{7}{8}$. It already requires rather difficult and lengthy case analysis. In principle, the same approach could be extended to higher levels $s \geqslant 3$, but the number of cases to handle seems to explode beyond control. Instead, we are able to argue in a more systematic fashion and avoid the explosion in potential case analysis (easier said than done of course).

We end this section with some remarks on Theorem I.12. Firstly, the subspaces A and B therein are referred to as "zoom-in" and "zoom-out" spaces respectively [34], [12], [13]. This makes sense if one imagines searching for the appropriate subgraph $\mathsf{Gr}_{k,\ell}[A,B]$ where the set S happens to have significant density. Secondly, we note that if S has density $\geqslant \varepsilon$, then the conclusion of the theorem is vacuously true without any need for a zoom-in or a zoom-out (i.e. $a = b = 0, A = \{0\}, B = \mathbb{F}_2^k$), so the theorem is really about "small" sets. Thirdly, our proof gives correct dependence of the required zoom-in-out dimension r on the upper bound on expansion α. For $\alpha < 1 - 2^{-(s+1)}$, one gets a significant projection onto the eigenspace at level at most s and then in our proof, a combined zooim-in-out dimension of at most $r = s$ is needed. This is tight (i.e. a lesser zoom-in-out dimension is not sufficient) since we know that subgraphs $\mathsf{Gr}_{k,\ell}[A,B]$ have expansion $1 - 2^{-(a+b)}$ and the combined zoom-in-out dimension (obviously) $a + b$. Finally, we note that towards proving the theorem, it will be easier to work with the contra-positive: a set S that has very small density inside every copy of the Grassmann graph with constant co-order (such a set will be called pseudorandom) has near-perfect expansion (i.e. very near 1).

The phenomena as in Theorem I.12 occurs also in the Johnson graph and has been analyzed in [33]. In a Johnson graph, the vertices are ℓ-subsets of a k-set and the edges are t-wise intersecting pairs (we are concerned with the case when $t = \lfloor \frac{\ell}{2} \rfloor$). Therein the notion of zoom-out and Fourier analysis are not needed. The Johnson case can informally be seen as a special case of the Grassmann case and the analysis of the former in [33] has been insightful in the analysis of the later in the current paper.

II. Acknowledgement

Our most sincere thanks to Boaz Barak, Irit Dinur, Yuval Filmus, Guy Kindler, Pravesh Kothari, Dana Moshkovitz, Prasad Raghavendra, Ran Raz, and David Steurer for several discussions and collaborations that led to this work.

Author Khot was supported by NSF grant CCF-1422159, Simons Collaboration on Algorithms and Geometry, and Simons Investigator Award. Author Minzer was supported by Clore Fellowship, ISF grant 2013/17, BSF grant 2016414 and Blavatnik fund. Author Safra was supported by ISF grant 2013/17, BSF grant 2016414 and Blavatnik fund.

References

[1] S. Arora and C. Lund. *Approximation Algorithms for NP-hard Problems, editor : D. Hochbaum.* PWS Publishing, 1996.

[2] S. Arora, C. Lund, R. Motawani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[3] S. Arora and S. Safra. Probabilistic checking of proofs : A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

[4] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. *J. ACM*, 62(5):42:1–42:25, 2015.

[5] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K. Vishnoi. Unique games on expanding constraint graphs are easy. In *Proc. ACM Symposium on the Theory of Computing*, pages 21–28, 2008.

[6] Boaz Barak. The intermediate complexity conjecture. Blog post, https://windowsontheory.org/2017/08/09/the-intermediate-complexity-conjecture/.

[7] Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 307–326, 2012.

[8] Boaz Barak, Pravesh Kothari, and David Steurer. Small-set expansion in shortcode graph and the 2-to-1 conjecture. *ECCC Report TR18-077.*

[9] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS*, pages 472–481, 2011.

[10] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs and non-approximability. *Electronic Colloquium on Computational Complexity, Technical Report TR95-024*, 1995.

[11] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *Proc. 20th IEEE Conference on Computational Complexity*, pages 144–153, 2005.

[12] 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)*, 23:198, 2016.

[13] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in Grassmann graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:94, 2017.

[14] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. *SIAM J. Comput.*, 39(3):843–873, 2009.

[15] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Ann. of Math. (2)*, 162(1):439–485, 2005.

[16] U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.

[17] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.

[18] U. Feige and L. Lovász. Two-prover one-round proof systems, their power and their problems. In *Proc. 24th Annual ACM Symposium on Theory of Computing*, pages 733–744, 1992.

[19] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[20] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001.

[21] Venkatesan Guruswami, Rajsekar Manokaran, and Prasad Raghavendra. Beating the random ordering is hard: Inapproximability of maximum acyclic subgraph. In *Proc. Annual IEEE Symposium on Foundations of Computer Science*, pages 573–582, 2008.

[22] Venkatesan Guruswami and Ali Kemal Sinop. Improved inapproximability results for maximum k-colorable subgraph. *Theory of Computing*, 9:413–435, 2013.

[23] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182:105–142, 1999.

[24] J. Håstad. Some optimal inapproximability results. *Journal of ACM*, 48:798–859, 2001.

[25] Johan Håstad. On the Efficient Approximability of Constraint Satisfaction Problems. In *Surveys in Combinatorics*, volume 346, pages 201–222. Cambridge University Press, 2007.

[26] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proc. ACM Symposium on the Theory of Computing*, pages 411–419, 2007.

[27] Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972.

[28] S. Khot. Inapproximability of NP-complete problems, discrete Fourier analysis, and geometry. In *Proc. the International Congress of Mathematicians*, 2010.

[29] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 767–775, 2002.

[30] Subhash Khot. On the unique games conjecture (invited survey). In *IEEE Conference on Computational Complexity*, pages 99–121, 2010.

[31] Subhash Khot. Hardness of approximation. *In Proc. of the International Congress of Mathematicians*, 2014.

[32] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.

[33] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Pseudorandom sets in Johnson graph have near-perfect expansion. *ECCC Report TR18-078*.

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

[35] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:6, 2018.

[36] Subhash Khot and Dana Moshkovitz. *NP*-hardness of approximately solving linear equations over reals. *SIAM J. Comput.*, 42(3):752–791, 2013.

[37] Subhash Khot and Dana Moshkovitz. Candidate hard unique game. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 63–76, 2016.

[38] Subhash Khot and Ryan O'Donnell. SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory of Computing*, 5(1):83–117, 2009.

[39] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within $2-\epsilon$. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008.

[40] Subhash Khot and Muli Safra. A two prover one round game with strong soundness. In *FOCS*, pages 648–657, 2011.

[41] Subhash Khot and Nisheeth K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into l_1. *J. ACM*, 62(1):8:1–8:39, 2015.

[42] Alexandra Kolla, Konstantin Makarychev, and Yury Makarychev. How to play unique games against a semi-random adversary: Study of semi-random models of unique games. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 443–452, 2011.

[43] G.A. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problemy Peredaci Informacii*, 10(2):101–108, 1974.

[44] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 245–254, 2008.

[45] Anup Rao. Parallel repetition in projection games and a concentration bound. In *Proc. ACM Symposium on the Theory of Computing*, pages 1–10, 2008.

[46] R. Raz. A parallel repetition theorem. *SIAM J. of Computing*, 27(3):763–803, 1998.

[47] L. Russo. An approximate zero-one law. *Z. Wahrsch. Verw. Gebiete*, 61(1):129–139, 1982.

[48] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 593–602, 2008.

[49] L. Trevisan. Inapproximability of combinatorial optimization problems. *Optimisation Combinatiore 2, (Vangelis Paschos, Editor), Hermes*, 2005.

[50] Luca Trevisan. On Khot's unique games conjecture. *Bull. Amer. Math. Soc. (N.S.)*, 49(1):91–111, 2012.