Rounding Semidefinite Programming Hierarchies via Global Correlation

Boaz Barak Prasad Raghavendra David Steurer

Abstract— We show a new way to round vector solutions of semidefinite programming (SDP) hierarchies into integral solutions, based on a connection between these hierarchies and the spectrum of the input graph. We demonstrate the utility of our method by providing a new SDP-hierarchy based algorithm for constraint satisfaction problems with 2-variable constraints (2-CSP's).

More concretely, we show for every 2-CSP instance $\mathfrak{I}$, a rounding algorithm for r rounds of the Lasserre SDP hierarchy for $\mathfrak{I}$ that obtains an integral solution which is at most ε worse than the relaxation's value (normalized to lie in $[0,1]$), as long as

$$r > k \cdot \mathrm{rank}_{\geqslant\theta}(\mathfrak{I})/\,\mathrm{poly}(\varepsilon)\,,$$

where k is the alphabet size of $\mathfrak{I}$, $\theta = \mathrm{poly}(\varepsilon/k)$, and $\mathrm{rank}_{\geqslant\theta}(\mathfrak{I})$ denotes the number of eigenvalues larger than θ in the normalized adjacency matrix of the constraint graph of $\mathfrak{I}$.

In the case that $\mathfrak{I}$ is a Unique Games instance, the threshold θ is only a polynomial in ε, and is independent of the alphabet size. Also in this case, we can give a non-trivial bound on the number of rounds for *every* instance. In particular our result yields an SDP-hierarchy based algorithm that matches the performance of the recent subexponential algorithm of Arora, Barak and Steurer (FOCS 2010) in the worst case, but runs faster on a natural family of instances, thus further restricting the set of possible hard instances for Khot's Unique Games Conjecture.

Our algorithm actually requires less than the $n^{O(r)}$ constraints specified by the r^{th} level of the Lasserre hierarchy, and in some cases r rounds of our program can be evaluated in time $2^{O(r)}\,\mathrm{poly}(n)$.

1. Introduction

This paper is concerned with hierarchies of semi-definite programs (SDP's). Semidefinite programs are an extremely useful tool in algorithms and in particular approximation algorithms (e.g., [15], [18]). Approximation algorithms based on SDP's typically involve finding an integral (say 0/1) solution for some optimization problem, by using convex programming to find a fractional/high-dimensional solution and then *rounding* it into an integral solution. Sherali and Adams [34], Lovász and Schrijver [28], and, later Lasserre [26], proposed systematic techniques, known as *hierarchies*, to make this convex relaxation tighter, thus ensuring that the fractional solution is closer to an integral one. These hierarchies are parameterized by a number r, called the *level* or *number of rounds* of the hierarchy. Given a program on n variables, optimizing over the r^{th} level of the hierarchy can be done in time $n^{O(r)}$. The gap between integral and fractional solutions decreases with r, and reaches zero at the n^{th} level. The paper [27] surveys and compares the different hierarchies proposed in the literature, (see recent survey [10]).

These semidefinite programming hierarchies have been of some interest in recent years, since they provide natural candidate algorithms for many computational problems. In particular, whenever the basic semidefinite or linear program provides a suboptimal approximation factor, it makes sense to ask how many rounds of the hierarchy are required to significantly improve upon this factor. Unfortunately, taking advantage of these hierarchies has often been difficult, and while some algorithms (e.g., [5]) can be encapsulated in, say, level 3 or 4 of some hierarchies, there have been relatively few results (e.g. [9], [8]) that use higher levels to obtain new algorithmic results. In fact, there has been more success in showing that high levels of hierarchies *do not* help for many computational problems [3], [33], [14], [31], [22]. In particular for 3SAT and several other NP-hard problems, it is known that it takes $\Omega(n)$ rounds of the strongest SDP hierarchy (i.e., Lasserre) to improve upon the approximation ratio achieved by the basic SDP (or sometimes even simpler algorithms) [32], [36].

Semidefinite hierarchies are of particular interest in the case of problems related to Khot's *Unique Games Conjecture* (UGC) [19]. Several works have shown that for a wide variety of problems, the UGC implies that (unless **P** = **NP**) the basic semidefinite program cannot be improved upon by any polynomial-time algorithm [20], [29], [30]. Thus in particular

Microsoft Research New England, Cambridge MA.
Georgia Institute of Technology, Atlanta GA.
Microsoft Research New England, Cambridge MA.

the UGC predicts that for all these problems, it will require a super-constant (and in fact polynomial, under widely believed assumptions) number of hierarchy rounds to improve upon the basic SDP. Investigating this prediction, particularly for the Unique Games problem itself and other related problems such as Max Cut, Sparsest Cut and Small-Set Expansion, has been the focus of several works, and it is known that at least $(\log\log n)^{\Omega(1)}$ rounds are required for a non-trivial approximation by a natural (though not strongest possible) SDP hierarchy [31], [22]. However, no non-trivial upper bound was known prior to the current work, and so it was conceivable that these lower bounds can be improved to $\Omega(n)$.

Recently, Arora, Barak and Steurer [2] gave a $2^{n^{\mathrm{poly}(\varepsilon)}}$-time algorithm for solving the Unique Games and Small-Set Expansion problems (where ε is the completeness parameter, see below). However, their algorithm did not use semidefinite programming hierarchies, and so does not immediately imply an upper bound on the number of rounds needed.

1.1. Our results

Our main contribution is a new method to analyze and round SDP hierarchies. The high level description is that it uses *global correlations* inside the high-dimensional SDP solution, combined with the hierarchy constraints, to obtain a better rounding of this solution into an integral one. We believe this method can be of general utility, and in particular we use it here to give new algorithms for approximating constraint satisfaction problem on two-variable constraints (2-CSP's), that run faster than the previously known algorithms for a natural family of instances. To state our results we need the notion of a *threshold rank*. *Threshold rank of graphs and 2-CSPs.* The *τ-threshold rank* of a regular graph G, denoted $\mathrm{rank}_{\geqslant\tau}(G)$, is the number of eigenvalues of the normalized adjacency matrix of G that are larger than τ.[1] An instance $\mathfrak{I}$ of a Max 2-Csp problem consists of a regular graph $G_{\mathfrak{I}}$, known as the *constraint graph* of $\mathfrak{I}$ over a vertex set $[n]$, where every edge (i, j) in the graph is labeled with a relation $\Pi_{i,j} \subseteq [k]\times[k]$ (k is known as the *alphabet size* of $\mathfrak{I}$). The *value* of an assignment $x \in [k]^n$ to the variables of $\mathfrak{I}$, denoted $\mathrm{val}_{\mathfrak{I}}(x)$, is equal to the probability that $(x_i, x_j) \in \Pi_{i,j}$ where (i, j) is a random edge in $G_{\mathfrak{I}}$. The *objective value* of $\mathfrak{I}$ is the maximum $\mathrm{val}_{\mathfrak{I}}(x)$ over all assignments x. We say that $\mathfrak{I}$ is *c-satisfiable* if $\mathfrak{I}$'s objective value is at least c. We define $\mathrm{rank}_{\geqslant\tau}(\mathfrak{I}) = \mathrm{rank}_{\geqslant\tau}(G_{\mathfrak{I}})$. Our main result is,

Theorem 1.1. *There is a constant c such that for every $\varepsilon > 0$, and every* Max 2-Csp *instance $\mathfrak{I}$ with objective value v, alphabet size k the following holds: the objective value* $\mathrm{sdpopt}(\mathfrak{I})$ *of the r-round Lasserre hierarchy for $r \geqslant k \cdot \mathrm{rank}_{\geqslant\tau}(\mathfrak{I})/\varepsilon^c$ is within ε of the objective value v of $\mathfrak{I}$, i.e.,* $\mathrm{sdpopt}(\mathfrak{I}) \leqslant v + \varepsilon$. *Moreover, there exists a polynomial time rounding scheme that finds an assignment x satisfying* $\mathrm{val}_{\mathfrak{I}}(x) > v - \varepsilon$ *given an optimal SDP solution as input.*

Results for Unique Games *constraints.*

We obtain quantitatively stronger results for the important special case of Max 2-Csp, the Unique Games problem. A Max 2-Csp instance is a Unique Games instance if all the relations $\Pi_{i,j}$ have the form $(a, b) \in \Pi_{i,j}$ iff $a = \pi_{i,j}(b)$ where $\pi_{i,j}$ is a permutation of $[k]$. First, we show that for Unique Games instances the threshold τ in Theorem 1.1 does not need to depend on the alphabet size. Namely, we prove the following.

Theorem 1.2. *There is an algorithm, based on rounding r rounds of the Lasserre hierarchy and a constant c, such that for every $\varepsilon > 0$ and input* Unique Games *instance $\mathfrak{I}$ with objective value v, alphabet size k, satisfying $\mathrm{rank}_{\geqslant\tau}(\mathfrak{I}) \leqslant \varepsilon^c r/k$, where $\tau = \varepsilon^c$, the algorithm outputs an assignment x satisfying $\mathrm{val}_{\mathfrak{I}}(x) > v - \varepsilon$.*

The Unique Games Conjecture concerns the approximability of the Unique Games problem in a specific regime, namely, given a Unique Games instance with optimal value $1 - \varepsilon$, the goal is to find an assignment with value at least $1/2$. We also show that a sublinear number of rounds suffice to get such an approximation in the worst case, regardless of the threshold rank of the instance. Moreover, we also show that such an approximation can be obtained in a number of rounds that depends on the τ-threshold rank for τ that is close to 1 (as opposed to the small value of τ needed for

[1] In this paper we only consider regular undirected graphs, although we allow non-negative weights and/or parallel edges. Every such graph can be identified with its normalized adjacency matrix, whose $(i, j)^{th}$ entry is proportional to the weight of the edge (i, j), with all row and column sums equalling one. Similarly, we restrict our attentions to 2-CSP's whose constraint graphs are regular. However, our definitions and results can be appropriately generalized for non-regular graphs and 2-CSPs as well.

Theorems 1.1 and 1.2).

Theorem 1.3. *There is an algorithm, based on rounding r rounds of the Lasserre hierarchy and a constant c, such that for every $\varepsilon > 0$ and input* Unique Games *instance $\mathfrak{I}$ with objective value $1-\varepsilon$ and alphabet size k, satisfying $r \geqslant ck \cdot \min\{n^{c\varepsilon^{1/3}}, \mathrm{rank}_{\geqslant 1-c\varepsilon}(\mathfrak{I})\}$, the algorithm outputs an assignment x satisfying $\mathrm{val}_{\mathfrak{I}}(x) > 1/2$.*

Examples of graphs with small threshold rank. Many interesting graph families have small τ-threshold rank for some small constant τ. Random degree d graphs have τ-threshold equal to 1 for any $\tau > c/\sqrt{d}$ for an absolute constant c. More generally, the class of *small-set expanders* – graphs where the expansion of small subsets of vertices is lower-bounded – also have small threshold rank. For instance, if every set of size $o(n)$ expands by at least $poly(\varepsilon)$ in a graph G, then $\mathrm{rank}_{1-\varepsilon}(G)$ is at most $n^{poly(\varepsilon)}$ [2]. Generalizing this result, [35] showed that if in a graph G every set of size $o(n)$ vertices has near-perfect expansion, then it implies upper bounds on $\mathrm{rank}_\tau(G)$ for τ close to 0.

Also, as noted in [2], *hypercontractive* graphs (i.e., graphs whose 2 to 4 operator norm is bounded) have at most polylogarithmic τ-threshold rank for every constant $\tau > 0$. For several 2-CSP's such as Max Cut, Unique Games, Small-Set Expansion, Sparsest Cut, the constraint graphs for the canonical "problematic instances" (i.e., integrality gap examples [12], [23], [22], [31]) are all hypercontractive, since they are based on either the noisy Gaussian graph or noisy Boolean cube.

Hard instances for our algorithms are graphs with a large threshold rank. For the Unique Games and Max Cut problems it is trivial to construct instances with large threshold rank by taking many disjoint copies of the same instance. For other 2-CSPs such as Label Cover, several natural hard instances have linear threshold-rank. For example, the natural "clause vs. variable" or "clause vs. clause" 2-CSPs obtained from random instances of 3SAT have linear threshold rank. This is not surprising since a non-trivial approximation for random 3SAT requires $\Omega(n)$ levels of the Lasserre hierarchy [32].

However, for the Small-Set Expansion problem all the existing constructions are based on the noisy boolean cube or the noisy Gaussian graph, and hence have only polylogarithmic threshold rank. A subsequent work addressing this issue, exhibits small set expanders with almost-polynomial threshold rank [6].

Algorithm efficiency. Our algorithm actually does not require the full power of the Lasserre hierarchy. First, we can use the relaxed variant with *approximate* constraints studied in [22], [31], [21]. Second, the proof of Theorem 1.3 can be carried out without utilizing the constraints on all $\binom{n}{r}$ r-sized subsets of n variables, but rather just sufficiently many random sets. As a result, our r-round algorithm can be implemented in time $2^{O(r)}\,\mathrm{poly}(n)$. Due to lack of space, the details of this improvement are deferred to the full version [7].

1.2. Related works

Subspace enumeration algorithms. For Unique Games and related problems, previous works [25], [24], [2] used *subspace enumeration* to give algorithms with similar running time to Theorem 1.3 on instances where the threshold rank of the *label extended graph* is small. This is known to be a stricter requirement on the instances than bounding the threshold rank of the constraint graph. The only known bound on the $1-\varepsilon$ threshold rank of the label extended graph in terms of the $1-\varepsilon$ threshold rank of the constraint graph loses a factor of about n^ε [2]. These subspace enumeration algorithms are applicable only to *nearly satisfiable* instances (whose objective value is close to 1), and therefore do not yield guarantees comparable to Theorems 1.1 and 1.2. Moreover, SDP-based algorithms are more robust and malleable than spectral techniques. For instance, it is easy to see that SDP hierarchies yield polynomial-time approximation scheme for 2CSPs whose constraint graphs are bounded tree width graphs or regular planar graphs (or more generally any *hyperfinite* family of graphs, see e.g. [17] and the references therein), although these classes of graphs could have high threshold degree.

Approximation schemes for (pseudo) dense CSP's. Previously, polynomial-time approximation schemes have been designed for general 2CSP's on *dense* and *pseudo-dense* instances [13], [1], [11]. This work generalizes these results, since pseudo-density is a stricter condition than having a constraint graph of

low threshold rank. Furthermore, for an ε-approximation, these polynomial time approximation schemes require the degree of the instance to be exponential in $\frac{1}{\varepsilon}$, while our results hold even on random graphs of degree $poly(1/\varepsilon)$.

Analyzing SDP hierarchy. Using very different techniques, Chlamtac [9] and Bhaskara et al [8] gave LP/SDP-hierarchy based algorithms for graph coloring and the densest subgraph problem respectively. As mentioned above, several works gave *lower bounds* for LP/SDP hierarchies. In particular [31], [22] showed that approximation such as those achieved in Theorem 1.3 for Unique Games problem require $\log\log^{\Omega(1)} n$ rounds of a relaxed variant of the Lasserre hierarchy. This relaxed variant captures our hierarchy as well. Schoenebeck [32] proved that achieving a non-trivial approximation for 3SAT on random instances requires $\Omega(n)$ rounds in the Lasserre hierarchy, while Tulsiani [36] showed that Lasserre lower bounds are preserved under common types of NP-hardness reductions.

In a concurrent and independent work, Guruswami and Sinop [16] obtained very similar results as this work. Using the Lasserre hierarchy, they obtain an approximation scheme with similar performance guarantees as the one in Theorem 1.1 for 2-CSPs, and in fact even consider generalizations involving additional (approximate) global linear constraints. Moreover, they obtain essentially the same results as Theorem 1.3 for the case of Unique Games. Furthermore, the rounding scheme in [16] is identical to ours. However, there are several differences both in results and the proof. First, although Guruswami and Sinop [16] use a notion similar to local-to-global correlation used here, they formalize it differently, and interestingly relate it to the problem of column selection for low rank approximations of matrices. Also, apart from the special case of unique constraints, [16] use a bound on the threshold rank of the *label extended graph*, as opposed to the constraint graph. The analysis in [16] relies on the full power of the Lasserre hierarchy, whereas we show that a weaker hierarchy is sufficient in the case of Unique Games, and it can be implemented more efficiently (i.e., $\exp(r)\,\mathrm{poly}(n)$ vs $n^{O(r)}$).

2. Preliminaries

We will use capital letters X, Y to denote random variables, and lower-case letters to denote assignments to these random variables. $\mathsf{CP}(X)$ denotes the collision probability of a random variable. For a random variable X and an element a in its domain, X_a will denote the indicator variable that equals 1 if $X = a$ and equals 0 otherwise. For a random variable X with range $[k]$, we define the *variance* of X as $\mathrm{Var}[X] \stackrel{\text{def}}{=} \sum_{a\in[k]} \mathrm{Var}[X_{1a}] = 1 - \mathsf{CP}(X)$, where $\mathsf{CP}(X)$ denotes the collision probability of X.

Unique games. An instance of Unique Games consists of a graph $G = (V, E)$, a label set $[k] = \{1, \ldots, k\}$ and a bijection $\pi_{ij} \colon [k] \to [k]$ for every edge $(i, j) \in E$. A labelling $\ell \colon V \to [k]$ is said to *satisfy* an edge (i, j) if $\pi_{ij}(\ell(i)) = \ell(j)$. The goal is to find a labeling $\ell \colon V \to [k]$ that satisfies the maximum number of edges namely, maximize $\mathbb{P}_{(i,j)\in E}\left\{\pi_{ij}(\ell(i)) = \ell(j)\right\}$

Local distributions.. Let $V = [n]$ be a set of vertices and let $[k]$ be a set of labels. An *m-local distribution* is a distribution μ^T over the set of assignments $[k]^T$ of the vertices of some set $T \subseteq V$ of size at most $m + 2$. (The choice of $m + 2$ is immaterial but will be convenient later on.) A collection of m-local distributions $\{\mu_T\}_{T\subseteq V,\ |T|\leqslant m+2}$ is *consistent* if for all $T, T' \subseteq V$ with $|T|, |T'| \leqslant m + 2$, the distributions μ_T and $\mu_{T'}$ are consistent on their intersection $T \cap T'$. We sometimes will view these distributions as random variables, hence writing $X_i^{(T)}$ for the random variable over $[k]$ that is distributed according to the label that $\mu^{T\cup\{i\}}$ assigns to i, and refer to a collection $X_1, \ldots, X_n$ of *m-local random variables*. However, we stress that these are not necessarily jointly distributed random variables, but rather for any subset of at most $m + 2$ of them, one can find a sample space on which they are jointly distributed. For succinctness, we omit the superscript for variables $X_i^{(S)}$ whenever it is clear from the context. For example, we will use $\{X_i \mid X_S\}$ is short for the random variable obtained by conditioning $X_i^{(S\cup\{i\})}$ on the variables $\{X_j^{(S\cup\{i\})}\}_{j\in S}$;[2] and use $\mathbb{P}\left\{X_i = X_j \mid X_S\right\}$ is short for the $[0, 1]$-valued random variable $\mathbb{P}\left\{X_i^{(S\cup\{i,j\})} = X_j^{(S\cup\{i,j\})} \mid X_S^{(S\cup\{i,j\})}\right\}$.

[2] Strictly speaking, the range of the random variable $\{X_i \mid X_S\}$ are random variables with range $[k]$. For every possible value x_S for X_S, one obtains a $[k]$-valued random variable $\{X_i \mid X_S = x_S\}$.

Lasserre hierarchy. Let U be a Unique Games instance with constraint graph $G = (V, E)$, label set $[k] = \{1, \ldots, k\}$, and bisections $\{\pi_{ij}\}_{ij \in E}$. An *m-round Lasserre solution* consists of m-local random variables $X_1, \ldots, X_n$ and vectors $v_{S,\alpha}$ for all vertex sets $S \subseteq V$ with $|S| \leqslant m + 2$ and all local assignments $\alpha \in [k]^S$. A Lasserre solution is *feasible* if the local random variables are consistent with the vectors, in the sense that for all $S, T \subseteq V$ and $\alpha \in [k]^S, \beta \in [k]^T$ with $|S \cup T| \leqslant m + 2$, we have $\langle v_{S,\alpha}, v_{T,\beta} \rangle = \mathbb{P}\{X_S = \alpha,\ X_T = \beta\}$ .The objective is to maximize the following expression $\mathbb{E}_{ij \in E} \mathbb{P}\left\{X_j = \pi_{ij}(X_i)\right\}$ .An important consequence of the existence of the vectors $v_{S,\alpha}$ is that for every set $S \subseteq V$ with $|S| \leqslant m$ and local assignment $x_S \in [k]^S$, the matrix $\left\{\mathrm{Cov}(X_{ia}, X_{jb} \mid X_S = x_S)\right\}_{i,j \in V, a,b \in [k]}$ is positive semidefinite.

3. Warmup – MaxCut Example

For the sake of exposition, we first present an algorithm for the Max Cut problem on low-rank graphs. In the Max Cut problem, the input consists of a graph $G = (V, E)$ and the goal is to find a cut $S \cup \bar{S} = V$ of the vertices that maximizes the number of edges crossing, i.e., maximizes $|E(S, \bar{S})|$.

The Goemans-Williamson SDP relaxation for the problem assigns a unit vector v_i for every vertex $i \in V$, so as to maximize the average squared length $E_{i,j \in E}\|v_i - v_j\|^2$ of the edges. Formally, the SDP relaxation is given by,

$$\text{maximize} \ \mathop{\mathbb{E}}_{i,j \in E} \|v_i - v_j\|^2 \ \text{ subject to } \|v_i\|^2 = 1 \ \forall i \in V$$

Stronger SDP relaxations produced by hierarchies such as Sherali-Adams and Lasserre hierarchy also yield probability distributions over local assignments. More precisely, given a m-round Lasserre SDP solution, it can be associated with a set of m-local random variables $X_1, \ldots, X_n$ taking values in $\{-1, 1\}$. For an edge (i, j), its contribution to the SDP objective value $(\|v_i - v_j\|^2)$ is equal to the probability that the edge (i, j) is cut under the distribution of local assignments μ_{ij}, namely, $\mathbb{P}_{\mu_{ij}}[X_i \neq X_j] = \|v_i - v_j\|^2$. Consequently, in order to obtain a cut with value *close* to the SDP objective, it is sufficient to jointly sample $X_1, \ldots, X_n$, such that on every edge (i, j) the distribution of X_i and X_j is *close* to the corresponding local distribution μ_{ij}. However, the variables $X_1, \ldots, X_n$ are not jointly distributed, and hence cannot all be sampled together.

As a first attempt, let us suppose we sample each X_i *independently* from its associated marginal μ_i. If on most edges (i, j), the distribution of the resulting samples X_i, X_j is close to μ_{ij}, then we are done. On an edge (i, j), the local distribution μ_{ij} is *far* from the independent sampling distribution $\mu_i \times \mu_j$ only if the random variables X_i, X_j are *correlated*. Henceforth, these correlations across the edges would be refered to as "local correlations". A natural measure for correlations that we will utilize here is defined as $\mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]$. Using this measure, the statistical distance between independent sampling $(\mu_i \times \mu_j)$ and correlated sampling (μ_{ij}) is given by

$$\|\mu_{ij} - \mu_i \times \mu_j\|_1 \leqslant |\mathrm{Cov}(X_i, X_j)|\,. \tag{3.1}$$

(See Lemma 5.3 for a more general version of the above bound).

Proof: Under the distribution $\{X_i X_j\}$, the event $\{X_i = a, X_j = b\}$ has probability $\mathbb{E}(a - X_i)(b - X_j)/4$. On the other hand, under the product distribution $\{X_i\}\{X_j\}$, this event has probability $\mathbb{E}(a - X_i)\,\mathbb{E}(b - X_j)/4$. Hence, the difference of these probabilities is equal to $\frac{1}{4}(\mathbb{E}\,X_i X_j - \mathbb{E}\,X_i\,\mathbb{E}\,X_j) = |\,\mathrm{Cov}(X_i, X_j)|/4$. Summing up over the four different assignments yields the desired bound. ∎

On the flip side, the existence of correlations makes the problem of sampling $X_1, \ldots, X_n$ easier! If two variables X_i, X_j are correlated, then sampling/fixing the value of X_i reduces the uncertainty in the value of X_j. More precisely, conditioning on the value of X_i reduces the variance of X_j as shown below:

$$\mathop{\mathbb{E}}_{\{X_i\}} \mathrm{Var}[X_j | X_i] = \mathrm{Var}[X_j] - \frac{1}{\mathrm{Var}[X_i]} \left[\mathrm{Cov}(X_i, X_j)\right]^2 . \tag{3.2}$$

Proof: Set $A = X_i - \mathbb{E}\,X_i$ and $B = X_j - \mathbb{E}\,X_j$. The random variables A, B have expectation zero, and have the same variance and covariance as X_i, X_j. Set $\tilde{B} = B/(\mathbb{E}[B^2])^{1/2}$.

The set of random variables $\{1, \tilde{B}\}$ is an orthonormal basis for the subspace of functions of B. Let $\rho = \mathbb{E}\,A\tilde{B}$. Then, $\rho\tilde{B}$ is the orthogonal projection of A to the subspace of functions of B. (Here, we use the assumption $\mathbb{E}\,A = 0$.) Hence, using the previous lemma,

$$\mathop{\mathbb{E}}_{\{B\}} \mathrm{Var}[A \mid B] = \mathbb{E}\,A^2 - \mathbb{E}(\rho\tilde{B})^2 = \mathbb{E}\,A^2 - \rho^2\,,$$

which is the desired identity because $\mathbb{E}A^2 = \operatorname{Var} A$ and $\rho^2 = \operatorname{Cov}(A,B)^2/\operatorname{Var} B$. ∎

Therefore, if we pick an $i \in V$ at random and fix its value then the expected decrease in the variance of all the other variables is given by,

$$\mathop{\mathbb{E}}_{i\in V,\{X_i\}}\left[\mathop{\mathbb{E}}_{j\in V}\operatorname{Var}[X_j|X_i]\right] - \mathop{\mathbb{E}}_{j\in V}\operatorname{Var}[X_j]$$
$$= \mathop{\mathbb{E}}_{i,j\in V}\operatorname{Cov}(X_i,X_j)^2 \cdot \frac{1}{2}\left(\frac{1}{\operatorname{Var}[X_i]} + \frac{1}{\operatorname{Var}[X_j]}\right).$$

The above bound is proven in a more general setting in Lemma 5.4. As all random variables involved have variance at most 1, we can rewrite the above expression as,

$$\mathop{\mathbb{E}}_{i\in V,\{X_i\}}\left[\mathop{\mathbb{E}}_{j\in V}\operatorname{Var}[X_j|X_i]\right] - \mathop{\mathbb{E}}_{j\in V}\operatorname{Var}[X_j] \geqslant \mathop{\mathbb{E}}_{i,j\in V}|\operatorname{Cov}(X_i,X_j)|^2.$$

The decrease in the variance is directly related to the *global correlations* between random pairs of vertices $i, j \in V$.

Recall that, the failure of independent sampling yields a lower bound on the average local correlations on the edges namely, $E_{i,j\in E}|\operatorname{Cov}(X_i,X_j)|$. The crucial observation is that if the graph G is a *good expander* in a suitable sense, then these local correlations translate in to non-negligible global correlations. Formally, we show the following (in Section 4):

Lemma 3.1. *Let $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$ be vectors in the unit ball. Suppose that the vectors are correlated across the edges of a regular n-vertex graph G,*

$$\mathop{\mathbb{E}}_{ij\sim G}\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle \geqslant \rho.$$

Then, the global correlation of the vectors is lower bounded by

$$\mathop{\mathbb{E}}_{i,j\in V}|\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle| \geqslant \Omega(\rho)/\operatorname{rank}_{\geqslant\Omega(\rho)}(G).$$

where $\operatorname{rank}_{\geqslant\rho}(G)$ is the number of eigenvalues of adjacency matrix of G that are larger than ρ.

As random variables X_i arise from the solution to a SDP, the matrix $\left(\operatorname{Cov}(X_i,X_j)\right)_{i,j\in V}$ is positive semidefinite, i.e., there exists vectors u_i such that $\langle u_i,u_j\rangle = \operatorname{Cov}(X_i,X_j)\ \forall i,j\in V$. Let us consider the vectors $v_i = u_i^{\otimes 2}$. Suppose the local correlation $\mathbb{E}_{i,j\in E}|\operatorname{Cov}(X_i,X_j)|$ is at least ε then we have,

$$\mathop{\mathbb{E}}_{i,j\in E}\langle v_i,v_j\rangle = \mathop{\mathbb{E}}_{i,j\in E}|\operatorname{Cov}(X_i,X_j)|^2 \geqslant \varepsilon^2,$$

and $\mathbb{E}_i[\|v_i\|^2] \leqslant 1$. If the graph G is low-rank, then by Lemma 3.1 we get a lower bound on the global correlation of the vectors v_i, namely

$$\mathop{\mathbb{E}}_{i,j\in V}|\operatorname{Cov}(X_i,X_j)|^2 = \mathop{\mathbb{E}}_{i,j\in V}\langle v_i,v_j\rangle \geqslant \Omega(\varepsilon^2)/\operatorname{rank}_{\geqslant\varepsilon^2}(G).$$

Summarizing, if the independent sampling is on average ε-far from correlated sampling over the edges, then conditioning on the value of a random vertex $i \in V$ reduces the average variance by $\varepsilon^2/\operatorname{rank}_{\geqslant\varepsilon^2}(G)$ in expectation. The same argument can now be applied on the variables obtained after conditioning on i. In fact, starting with an SDP solution to m-round Lasserre hierarchy, the local distributions remain consistent and their covariance matrices remain semidefinite as long as we condition on at most $m-2$ vertices. Observe that average variance is at most 1. Hence, after at most $\operatorname{rank}_{\geqslant\varepsilon^2}(G)/\varepsilon^2$ steps, the independent sampling distribution will be within average distance ε from the correlated sampling on the edges.

The rounding scheme will iteratively condition the SDP solution on the value of a vertex for at most $\operatorname{rank}_{\geqslant\varepsilon^2}(G)/\varepsilon^2$ iterations. In each iteration, a uniformly randomly chosen vertex X_i is fixed to a value sampled from its corresponding marginal distribution. When the global correlation is less than ε after conditioning, the algorithm just rounds the solution by independent sampling, and outputs the assignment.

Rounding SDP's using a small basis. There is an alternate way to view the above rounding scheme. Note that in the case that the vectors $v_1,\ldots,v_n$ are one dimensional unit vectors (i.e., $v_i \in \{\pm 1\}$), $\mathcal{V}$ exactly corresponds to a cut in the graph, and the objective value measures the fraction of edges cut. Now, suppose that you could find r vectors $v_{i_1},\ldots,v_{i_r} \in \mathcal{V}$, whom we'll call the *basis vectors*, such that every other $v \in \mathcal{V}$ has some significant projection ρ into the span of $v_{i_1},\ldots,v_{i_r}$. That is, if we let P be the projection operator corresponding to this space, then for every $v \in \mathcal{V}$, $\|Pv\|_2 \geqslant \rho$. It turns out that in this case, if ρ is sufficiently close to 1 and the vector solution $\mathcal{V}$ satisfied $r+2$ rounds of an appropriate SDP hierarchy, then we can round $\mathcal{V}$ to achieve a very good cut. The intuition behind this is the following: the constraints of $r+2$ hierarchy rounds allow us to essentially assume without loss of generality that the vectors $v_{i_1},\ldots,v_{i_r}$ are one-dimensional. That is, after applying an appropriate rotation, we can think of each one

of them as a vector of the form $(\pm 1, 0, \ldots, 0)$. Moreover, our assumption implies that every other vector in v has a magnitude of at least ρ in its first coordinate. Now one can show that simply rounding each vector to the sign of its first coordinate will result in a ± 1 assignment to the vertices corresponding to a good cut.

Hence, the problem of rounding reduces to finding a small basis of vectors $v_{i_1}, \ldots, v_{i_r}$, such that the SDP solution has a significant projection or lies almost completely in their span. Again, if the local correlation on the edges $\mathbb{E}_{i\sim j}\langle v_i, v_j\rangle^2$ is tiny, then independent sampling yields a good assignment. Otherwise by Lemma 3.1, the global correlation $\mathbb{E}_{i,j}\langle v_i, v_j\rangle^2$ is non-negligible ($\Omega(\varepsilon^2)/\mathrm{rank}_{\geqslant\varepsilon^2}(G)$). In particular, there exists at least one vector v_{i_1} such that $\mathbb{E}_j\langle v_{i_1}, v_j\rangle^2 \geqslant \Omega(\varepsilon^2)/\mathrm{rank}_{\geqslant\varepsilon^2}(G)$. We can now replace each vector $v \in \mathcal{V}$ with its projection into the orthogonal space to v_{i_1} and continue. Eventually, either the local correlation becomes negligible making independent sampling a good rounding, or we find a basis $v_{i_1}, \ldots, v_{i_r}$ such that (almost all) vectors $v \in \mathcal{V}$ have most of their mass in $\mathrm{Span}\{v_{i_1}, \ldots, v_{i_r}\}$, in which case we can successfully round the solution.

Threshold rank vs global correlation. Whenever the graph has small number of large eigenvalues, the condition that local correlation implies global correlation holds. This is useful to simulate eigenspace enumeration algorithms such as used by [25], [24], [2], [35] since in the case of Unique Games (and other related problems), a good SDP solution must be locally well correlated. But the notion of local to global correlation is somewhat more general and robust than having small threshold rank. For example, adding $\sqrt{n}$ isolated vertices to a graph will increase correspondingly the number of eigenvectors with value 1, but will actually not change by much the local to global correlation. This captures to a certain extent the fact that SDP-based solutions are more robust than the spectral based algorithms. (A similar example of this phenomenon is that adding a tiny bipartite disjoint graph to the input graph makes the smallest eigenvalue become -1, but does not change by much the value of the Goemans-Williamson SDP.) We hope that this robustness of the SDP-based approach will enable further improvements in the future.

4. Local Correlation implies Global Correlation in Low-Rank Graphs

In this section, we present the crucial ingredient for our algorithm – the translation of local-correlations to non-trivial global correlations (Lemma 3.1) in expanders and more generally low-rank graphs. Let G be a regular graph with vertex set $V = \{1, \ldots, n\}$. We identify G with its normalized adjacency matrix, a symmetric stochastic matrix. Let $\lambda_1 \geqslant \ldots \geqslant \lambda_n \in [-1, 1]$ be the eigenvalues of G in non-increasing order.

The following lemma shows that a violation of the local vs global correlation condition implies that the graph has high threshold rank.

Lemma 4.1. *Suppose there exist vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ such that*

$$\mathop{\mathbb{E}}_{ij\sim G}\langle v_i, v_j\rangle \geqslant 1-\varepsilon\,, \qquad \mathop{\mathbb{E}}_{i,j\in V}\langle v_i, v_j\rangle^2 \leqslant \tfrac{1}{m}\,,$$

$$\mathop{\mathbb{E}}_{i\in V}\|v_i\|^2 = 1\,.$$

Then for all $C > 1$, $\lambda_{(1-1/C)m} \geqslant 1 - C\cdot\varepsilon$. In particular, $\lambda_{m/2} > 1 - 2\varepsilon$.

Proof: Let $X = (x_{r,s})_{r,s\in[n]}$ be the Gram matrix $(\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle)_{i,j\in V}$ represented in the eigenbasis of G, so that

$$\mathop{\mathbb{E}}_{ij\sim G}\langle v_i, v_j\rangle = \sum_{r\in[n]}\lambda_r x_{r,r}\,, \qquad \mathop{\mathbb{E}}_{i,j\in V}\langle v_i, v_j\rangle^2 = \sum_{r,s\in[n]} x_{r,s}^2\,,$$

$$\mathop{\mathbb{E}}_{i\in V}\|v_i\|^2 = \sum_{r\in[n]} x_{r,r}\,.$$

Let m' be the largest index such that $\lambda_{m'} \geqslant 1 - C\cdot\varepsilon$. Notice that the numbers $p_1 = x_{1,1}, \ldots, p_n = x_{n,n}$ form a probability distribution over $r \in [n]$. Let $q = \sum_{i=1}^{m'} p_i$ be the probability of the event $r \leqslant m'$. Using Cauchy–Schwarz, we can bound this probability in terms of m, $q = \sum_{r=1}^{m'} p_r \leqslant m' \sum_{r=1}^{n} p_r^2 \leqslant \frac{m'}{m}$. On the other hand, we can bound the expectation of λ_r with respect to the probability distribution $(p_1, \ldots, _n)$ in terms of this probability q,

$$\begin{aligned} 1-\varepsilon &\leqslant \sum_{r=1}^{n}\lambda_r p_r \leqslant \sum_{r=1}^{m'} p_r + (1 - C\cdot\varepsilon)\sum_{r=m'+1}^{m} p_r \\ &= 1 - (1-q)C\cdot\varepsilon \leqslant 1 - \left(1 - \tfrac{m'}{m}\right)C\cdot\varepsilon\,. \end{aligned}$$

It follows that $m' \geqslant (1 - 1/C)\cdot m$, which gives the desired conclusion that G has at least $(1 - 1/C)\cdot m$ eigenvalues $\lambda_r \geqslant -C\cdot\varepsilon$. ∎

Note that Lemma 3.1 follows directly from the previous lemma by picking C =

$\frac{(1-\rho/100)}{(1-\rho)}$ and observing that $\mathbb{E}_{i,j\in V}|\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle| \geqslant \mathbb{E}_{i,j\in V}|\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle|^2$ since $|\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle| \leqslant 1$ for all $i, j \in V$. A weaker converse for Lemma 3.1 is also true (see full version [7]).

5. General 2-CSP on Low Rank Graphs

Let $\mathfrak{I}$ be a (general) Max 2-Csp instance with variable set $V = [n]$ and label set $[k]$. (We represent $\mathfrak{I}$ as a distribution over triples (i, j, Π), where $i, j \in V$ and $\Pi \subseteq [k] \times [k]$ is an arbitrary binary predicate. The goal is to find an assignment $x \in [k]^V$ that maximizes the probability $\mathbb{P}_{(i,j,\Pi)\sim\mathfrak{I}}\left\{(x_i, x_j) \in \Pi\right\}$.)

For simplicity,[3] we will assume that the constraint graph of $\mathfrak{I}$ is regular, i.e., every variable $i \in V$ appears in the same number of constraints. (Since we allow the constraints to be weighted, the precise condition is that the total weight of the constraints incident to a vertex is the same for every vertex.)

The Lasserre program for $\mathfrak{I}$ gives rise to m-local random variables $X_1, \ldots, X_n$ with range $[k]$. We write X_{ia} to denote the $\{0, 1\}$-indicator of the event $X_i = a$. Notice that $\{X_{ia}\}_{i\in V, a\in[k]}$ are also m-local random variables.

The rounding algorithm and its analysis are a somewhat straight-forward generalization of the algorithm for MaxCut described in Section 3.

Algorithm 5.1 (Propagation Sampling)**.**

Input: r-local random variables $X_1, \ldots, X_n$ over $[k]$

Output(global) distribution over assignments $x \in [k]^V$.

1) Choose $m \in \{1, \ldots, r\}$ at random.
2) Sample a random set of "seed vertices" $S \in V^m$. (Repeated vertices are allowed.)
3) Sample a assignment $x_S \in [k]^S$ for S according to its local distribution $\{X_S\}$.
4) For every other vertex $i \in V \setminus S$, sample a label $x_i \in [k]$ according to the local distribution for $S \cup \{i\}$ conditioned on the assignment x_S for S.

Our main Theorem of this section, which immediately implies Theorem 1.1, is the following:

Theorem 5.2. *Let $\varepsilon > 0$ and $r = O(k) \cdot \mathrm{rank}_{\geqslant\Omega(\varepsilon/k)^2}(G)/\varepsilon^4$. Suppose that the r-round Lasserre value of the* Max 2-Csp *instance $\mathfrak{I}$ is σ. Then, given an optimal r-round Lasserre solution, Algorithm 5.1 (Propagation Sampling) outputs an assignment with expected value at least $\sigma - \varepsilon$ for $\mathfrak{I}$.*

The proof of the theorem follows the general outline in Section 3. Specifically, show that as long as for a random pair (i, j) of indices there is noticeable correlation between X_i and X_j, conditioning on X_i "makes progress" and reduces the global variance of the system, while once there is no correlation between almost all pairs (i, j) involved in constraints, then we may as well use independent rounding to come up with a solution. The fact that the graph has low rank, allows us to pass between correlations for a random pair (i, j), and correlations for a random pair (i, j) that is involved in a constraint.

To implement this approach, we need to formalize the appropriate measures of correlation, and show an equivalence between them. One measure is the statistical distance between correlated and independent sampling of two variables X_i, X_j. If this measure is low on the edges of the constraint graph, then an independent sampling of the variables would give a good assignment. As in (3.1), we relate this measure to the covariances of the corresponding variables $\{X_{ia}\}, \{X_{jb}\}$.

Lemma 5.3. *] Let X_i and X_j be the two random variables over $[k]$ from the Lasserre solution, then*

$$\left\|\{X_iX_j\} - \{X_i\}\{X_j\}\right\|_1 = \sum_{(a,b)\in[k]^2} \left|\mathrm{Cov}(X_{ia}, X_{jb})\right|.$$

where the left-hand side denotes the ℓ_1 (i.e. total variation) distance between the joint distribution of X_i and X_j and the distribution when they are sampled independently.

We omit the (easy) proof from this extended abstract. Next, we lower-bound the decrease in variance during the conditioning process in terms of the covariances (analogue of (3.2)).

Lemma 5.4. *For any two vertices $i, j \in V$,*

$$\mathrm{Var}\, X_i - \mathop{\mathbb{E}}_{\{X_j\}} \mathrm{Var}\left[X_i \,\middle|\, X_j\right] \geqslant \frac{1}{k} \sum_{a,b\in[k]} \mathop{\mathbb{E}}_{\{X_{ia}X_{jb}\}} \mathrm{Cov}(X_{ia}, X_{jb})^2 / \mathrm{Var}\, X_{jb}$$

We defer the easy proof to the full version [7]. We now have two natural measures of correlation between X_i and X_j— one is the statistical distance between sampling X_i, X_j

[3] If the constraint graph is not regular, all of our results still hold for an appropriate definition of threshold rank.

jointly and independently, and the other is the amount the variance of X_i decreases after conditioning on X_j, with both ways relating to the covariances of X_{ia} and X_{ib}. By using convexity arguments it can be shown that, in our case where the local random variables come from a semidefinite program, both notions can be approximated by inner products of vectors:

Lemma 5.5. *Suppose that the matrix* $\big(\mathrm{Cov}(X_{ia}, X_{jb})\big)_{i\in V,\, a\in[k]}$ *is positive semidefinite. Then, there exists vectors* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ *in the unit ball such that for all vertices* $i, j \in V$*,*

$$\tfrac{1}{k^2}\Big(\sum_{(a,b)\in[k]^2} \big|\mathrm{Cov}(X_{ia}, X_{jb})\big|\Big)^2 \leqslant \langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle \leqslant \tfrac{1}{k}\sum_{(a,b)\in[k]^2} \tfrac{1}{2}\big(\tfrac{1}{\mathrm{Var}\, X_{ia}} + \tfrac{1}{\mathrm{Var}\, X_{jb}}\big)\,\mathrm{Cov}(X_{ia}, X_{jb})^2\,.$$

With the equivalence between these notions of correlations, it is easy to finish the argument using Lemma 3.1. Specifically, Lemma 3.1 implies that if the variables are globally uncorrelated, then they are locally uncorrelated. Therefore, after a small number of conditionings, the independent sampling of the variables yields an assignment whose value is close to that of the SDP solution.

6. Results for Unique Games

As mentioned earlier, in case the CSP is of the Unique Games type, the threshold for the eigenvalues does not need to depend on the alphabet size k. The key technical step is the following variant of Lemma 5.5:

Lemma 6.1. *Let* $X_1, \ldots, X_n$ *be r-local random variables over* $[k]$ *and let* X_{ia} *be the indicator of the event* $X_i = a$*. Suppose that the matrix* $\big(\mathrm{Cov}(X_{ia}, X_{jb})\big)_{i\in V,\, a\in[k]}$ *is positive semidefinite. Then, there exists vectors* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ *in the unit ball such that for all vertices* $i, j \in V$ *and permutations* π *of* $[k]$*,*

$$\Big(\sum_{a\in[k]} \big|\mathrm{Cov}(X_{ia}, X_{j\pi(a)})\big|\Big)^4 \leqslant \langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle \leqslant \sum_{(a,b)\in[k]^2} \tfrac{1}{2}\big(\tfrac{1}{\mathrm{Var}\, X_{ia}} + \tfrac{1}{\mathrm{Var}\, X_{jb}}\big)\,\mathrm{Cov}(X_{ia}, X_{jb})^2\,.$$

In fact, for unique games we can also obtain results when the threshold is a constant *close to* 1 as opposed to close to zero, or equivalently, only requiring that the Laplacian of the constraint graph has few eigenvalues close to 0. Formally, this is stated in the following theorem:

Theorem 6.2. *For every positive integer* m*, there exists an algorithm running in time* $n^{O(mk^2)}$ *that given a unique games instance* Γ *over alphabet* $[k]$ *with value* $1-\eta$*, finds a labelling satisfying* $1 - O(\frac{\eta}{\lambda_m})$ *fraction of the edges. Here* λ_m *is the* m^{th} *smallest eigenvalue of the Laplacian of the constraint graph* Γ*.*

The (omitted) proof follows similar lines to the proof of Theorem 5.2, but is not identical. In particular, at the moment our analysis requires to use a variant of propagation rounding where we enumerate over *all* possible seed sets of a certain size rather than choosing a random one. We can then use this to obtain a subexponential algorithm for *every* instance of Unique Games, thus proving Theorem 1.3. The proof follows by combining Theorem 6.2 and the decomposition theorem of [2] that showed how to break up any graph into pieces of low rank. The main observation is that one does not need to use the decomposition first, and then run the SDP on each part, but rather we can use the decomposition in order to round the SDP solution, by applying propagation rounding to each part independently.

Conclusions

We have shown that $n^{O(\varepsilon^{1/3})}$ rounds of an SDP hierarchy suffice for solving the Unique Games problem on $(1-\varepsilon)$-satisfiable instances. The best lower bound known for the hierarchy we used is $\log\log^{\Omega(1)} n$ [31], [22], and finding the tight bound has obvious relevance to the unique games conjecture.

It is our hope that the correlation based rounding technique presented here, will yield more approximation algorithms based on SDP hierarchies. Indeed, Arora and Ge (personal communication) recently used ideas from this work to obtain improved algorithms for 3-coloring on an interesting families of instances.

References

[1] N. Alon, A. Coja-Oghlan, H. Hàn, M. Kang, V. Rödl, and M. Schacht, "Quasi-randomness and algorithmic regularity for graphs with general degree distributions," *SIAM J. Comput.*, vol. 39, no. 6, pp. 2336–2362, 2010.

[2] S. Arora, B. Barak, and D. Steurer, "Subexponential algorithms for unique games and related problems," in *FOCS*, 2010, pp. 563–572.

[3] S. Arora, B. Bollobás, L. Lovász, and I. Tourlakis, "Proving integrality gaps without knowing the linear program," *Theory of Computing*, vol. 2, no. 1, pp. 19–51, 2006.

[4] S. Arora, S. Khot, A. Kolla, D. Steurer, M. Tulsiani, and N. K. Vishnoi, "Unique games on expanding constraint graphs are easy," in *STOC*, 2008, pp. 21–28.

[5] S. Arora, S. Rao, and U. V. Vazirani, "Expander flows, geometric embeddings and graph partitioning," *J. ACM*, vol. 56, no. 2, 2009.

[6] B. Barak, P. Gopalan, J. Hastad, R. Meka, P. Raghavendra, and D. Steurer, "Making the long code shorter, with applications to unique games conjecture," 2011, manuscript.

[7] B. Barak, P. Raghavendra, and D. Steurer, "Rounding semidefinite programming hierarchies via global correlation," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 18, p. 65, 2011.

[8] A. Bhaskara, M. Charikar, E. Chlamtac, U. Feige, and A. Vijayaraghavan, "Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph," in *STOC*, 2010, pp. 201–210.

[9] E. Chlamtac, "Approximation algorithms using hierarchies of semidefinite programming relaxations," in *FOCS*, 2007, pp. 691–701.

[10] E. Chlamtac and M. Tulsiani, "Convex relaxations and integrality gaps," 2010, chapter in Handbook on Semidefinite, Cone and Polynomial Optimization.

[11] A. Coja-Oghlan, C. Cooper, and A. M. Frieze, "An efficient sparse regularity concept," *SIAM J. Discrete Math.*, vol. 23, no. 4, pp. 2000–2034, 2010.

[12] U. Feige and G. Schechtman, "On the optimality of the random hyperplane rounding technique for MAX CUT," *Random Struct. Algorithms*, vol. 20, no. 3, pp. 403–440, 2002.

[13] A. M. Frieze and R. Kannan, "Quick approximation to matrices and applications," *Combinatorica*, vol. 19, no. 2, pp. 175–220, 1999.

[14] K. Georgiou, A. Magen, T. Pitassi, and I. Tourlakis, "Integrality gaps of $2 - o(1)$ for vertex cover sdps in the Lovász–Schrijver hierarchy," in *FOCS*, 2007, pp. 702–712.

[15] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

[16] V. Guruswami and A. K. Sinop, "Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with psd objectives," fOCS 2011.

[17] A. Hassidim, J. A. Kelner, H. N. Nguyen, and K. Onak, "Local graph partitions for approximation and testing," in *FOCS*, 2009, pp. 22–31.

[18] D. R. Karger, R. Motwani, and M. Sudan, "Approximate graph coloring by semidefinite programming," *J. ACM*, vol. 45, no. 2, pp. 246–265, 1998.

[19] S. Khot, "On the power of unique 2-prover 1-round games," in *STOC*, 2002, pp. 767–775.

[20] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, "Optimal inapproximability results for Max-Cut and other 2-variable CSPs?" in *FOCS*, 2004, pp. 146–154.

[21] S. Khot, P. Popat, and R. Saket, "Approximate lasserre integrality gap for unique games," in *APPROX-RANDOM*, 2010, pp. 298–311.

[22] S. Khot and R. Saket, "SDP integrality gaps with local ℓ_1-embeddability," in *FOCS*, 2009, pp. 565–574.

[23] S. Khot and N. K. Vishnoi, "The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1," in *FOCS*, 2005, pp. 53–62.

[24] A. Kolla, "Spectral algorithms for unique games," in *IEEE CCC*, 2010, pp. 122–130.

[25] A. Kolla and M. Tulsiani, "Playing random and expanding unique games," 2007, unpublished manuscript available from the authors' webpages, to appear in journal version of [4].

[26] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[27] M. Laurent, "A comparison of the sherali-adams, lovász-schrijver, and lasserre relaxations for 0-1 programming," *Math. Oper. Res.*, vol. 28, no. 3, pp. 470–496, 2003.

[28] L. Lovász and A. Schrijver, "Cones of matrices and set-functions and 0-1 optimization," *SIAM J. Optim.*, vol. 1, no. 2, pp. 166–190, 1991.

[29] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, "Noise stability of functions with low influences invariance and optimality," in *FOCS*, 2005, pp. 21–30.

[30] P. Raghavendra, "Optimal algorithms and inapproximability results for every CSP?" in *STOC*, 2008, pp. 245–254.

[31] P. Raghavendra and D. Steurer, "Integrality gaps for strong SDP relaxations of unique games," in *FOCS*, 2009, pp. 575–585.

[32] G. Schoenebeck, "Linear level Lasserre lower bounds for certain k-CSPs," in *FOCS*, 2008, pp. 593–602.

[33] G. Schoenebeck, L. Trevisan, and M. Tulsiani, "A linear round lower bound for lovasz-schrijver sdp relaxations of vertex cover," in *IEEE CCC*, 2007, pp. 205–216.

[34] H. D. Sherali and W. P. Adams, "A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems," *SIAM J. Discrete Math.*, vol. 3, no. 3, pp. 411–430, 1990.

[35] D. Steurer, "Subexponential algorithms for d-to-1 two-prover games and for certifying almost perfect expansion," 2010, manuscript.

[36] M. Tulsiani, "CSP gaps and reductions in the Lasserre hierarchy," in *STOC*, 2009, pp. 303–312.