Rounding Semidefinite Programming Hierarchies via Global Correlation

Boaz Barak[*]   Prasad Raghavendra[†]   David Steurer[‡]

April 28, 2011

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 that is at most ε worse than the relaxation's value (normalized to lie in $[0,1]$), as long as

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

where k is the alphabet size of $\mathfrak{I}$, $\theta = \operatorname{poly}(\varepsilon/k)$, and $\operatorname{rank}_{\geqslant \theta}(\mathfrak{I})$ 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 rth level of the Lasserre hierarchy, and in some cases r rounds of our program can be evaluated in time $2^{O(r)} \operatorname{poly}(n)$.

---

[*]Microsoft Research New England, Cambridge MA.

[†]Georgia Institute of Technology, Atlanta GA.

[‡]Microsoft Research New England, Cambridge MA.

Contents

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., [GW95, KMS98]). SDP's 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 [SA90], Lovász and Schrijver [LS91], and, later Lasserre [Las01], proposed systematic ways, 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 rth 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 nth level, where the program is guaranteed to find an optimal integral solution. The paper [Lau03] surveys and compares the different hierarchies proposed in the literature, see also the recent survey [CT10].

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., [ARV09]) can be encapsulated in, say, level 3 or 4 of some hierarchies, there have been relatively few results (e.g. [Chl07, BCC⁺10]) 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 [ABLT06, STT07, GMPT07, RS09, KS09]. 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 rate achieved by the basic SDP (or sometimes even simpler algorithms) [Sch08, Tul09].

Semidefinite hierarchies are of particular interest in the case of problems related to Khot's *Unique Games Conjecture* (UGC) [Kho02]. Several works have shown that for a wide variety of problems, the UGC implies that (unless $\mathbf{P} = \mathbf{NP}$) the basic semidefinite program cannot be improved upon by any polynomial-time algorithm [KKMO04, MOO05, Rag08]. Thus in particular the UGC predicts that for all these problems, it will take 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 [RS09, KS09] by a natural (though not strongest possible) SDP hierarchy. 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 [ABS10] gave a $2^{n^{\text{poly}(\varepsilon)}}$-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. We elaborate more on our method in Section 2, but its high level description is that 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 the following:

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 optimal SDP solution as input.*

Results for UNIQUE GAMES constraints. We say that a MAX 2-CSP instance is a UNIQUE GAMES instance if all the relation $\Pi_{i,j}$ have the form that $(a, b) \in \Pi_{i,j}$ iff $a = \pi_{i,j}(b)$ where $\pi_{i,j}$ is a permutation of $[k]$. As mentioned above, the performance of SDP hierarchy on UNIQUE GAMES instances and related problems is of particular interest. We obtain somewhat stronger quantitative results for UNIQUE GAMES instances. Also, as remarked below, our results are "morally stronger" in this case, since it's conceivable that the hardest instances for these types of problems have small threshold rank. 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

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 is about a specific approximation regime for UNIQUE GAMES. Given a UNIQUE GAMES instance with optimal value $1 - \varepsilon$, the goal is to find an assignment with value at least $1/2$.

---

[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.

We also show that in this case a sublinear (and in fact a small root) number of rounds suffice to get such an approximation in the worst case, regardless of the instance's threshold rank. 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 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 $\tau > c/\sqrt{d}$. More generally, graphs where small subsets of vertices have bounded edge-expansion, referred to as *small-set expanders*, 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 $n^{poly(\varepsilon)}$ [ABS10]. Generalizing this result, [Ste10] 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 [ABS10], *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 [FS02, KV05, KS09, RS09]) are all hypercontractive, since they are based on either the noisy Gaussian graph or noisy Boolean cube. In fact, it is conceivable that the SMALL-SET EXPANSION problem is trivial on graphs with large threshold rank, in the sense that we do not know of any example of an instance having, say, $\log^{\omega(1)} n$ 0.99-threshold rank, and objective value smaller than $1/2$. (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, though it could still be the case that the hardest instances are the ones with small threshold rank.) On the other hand, for other 2-CSPs such as LABEL COVER, some natural hard instances have linear threshold-rank. For example this is the case if one considers the natural "clause vs. variable" or "clause vs. clause" 2-CSP obtained from random instances of 3SAT (which is not surprising given that a non-trivial approximation for random 3SAT requires $\Omega(n)$ levels of the Lasserre hierarchy [Sch08]).

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 [KS09, RS09, KPS10]. Second, in the proof of Theorem 1.3, we don't need to utilize the constraints on all $\binom{n}{r}$ r-sized subsets of n variables, but rather sufficiently many random sets suffice. As a result, we can implement our r-round algorithm in time $2^{O(r)} \mathrm{poly}(n)$.

1.2 Related works

Subspace enumeration algorithms. For UNIQUE GAMES and related problems, previous works [KT07, Kol10, ABS10] used *subspace enumeration* to give algorithms with similar running time to Theorem 1.3 in the case that the threshold rank of the *label extended graph*

of the instance is small. This is known to be a stronger condition 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} [ABS10]. These subspace enumeration algorithms also only applied to *nearly satisfiable* instances (whose objective value is close to 1), and so did not give guarantees comparable to Theorems 1.1 and 1.2. As mentioned below in Section 2, SDP-based algorithms have some robustness advantages over spectral techniques. SDP hierarchies are also easily shown to yield polynomial-time approximation scheme for 2CSPs whose constraint graphs can have very *high* threshold rank such as bounded tree width graphs and regular planar graphs (or more generally any *hyperfinite* family of graphs, see e.g. [HKNO09] and the references therein).

Approximation schemes for (pseudo) dense CSP's. For general 2CSP's, several works gave polynomial-time approximation schemes for *dense* and *pseudo-dense* instances [FK99, ACOH$^+$10, COCF10]. Our work generalizes these results, since pseudo-density is a stronger condition than having a constraint graph of low threshold rank. Furthermore, for an ε-approximation the degree of the instance needed by these works is exponential in $\frac{1}{\varepsilon}$, while the results of this work apply even on random graphs of degree $poly(1/\varepsilon)$.

Analyzing SDP hierarchy. Using very different methods, Chlamtac [Chl07] and Bhaskara et al [BCC$^+$10] gave LP/SDP-hierarchy based algorithms for hypergraph coloring and the densest subgraph problem respectively. As mentioned above, several works gave *lower bounds* for LP/SDP hierarchies. In particular [RS09, KS09] 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 [Sch08] proved that achieving a non-trivial approximation for 3SAT on random instances requires $\Omega(n)$ rounds in the Lasserre hierarchy, while Tulsiani [Tul09] showed that Lasserre lower bounds are preserved under common types of NP-hardness reductions.

In a concurrent and independent work, Guruswami and Sinop [GS11] gave results very similar to ours. They also use the Lasserre hierarchy to get an approximation scheme with similar performance to our Theorem 1.1 for 2-CSPs, and in fact even consider generalizations involving additional (approximate) global linear constraints. They also get essentially the same results for UNIQUE GAMES as our Theorem 1.3. Furthermore, their rounding algorithm is the same as ours. However, there are some differences both in results and the proof. First, although [GS11] use a notion similar to our local-to-global correlation, they view 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, they work with the threshold rank of the *label extended graph*, as opposed to the constraint graph as is the case here (however for binary alphabet these two graphs coincide). Their analysis relies on the full power of the Lasserre hierarchy, whereas we show that a weaker hierarchy is sufficient in the UNIQUE GAMES case, and can even be done faster (i.e., $\exp(r) \, poly(n)$ vs $n^{O(r)}$).

2   Our techniques

We now describe, on a very high and imprecise level, the ideas behind our rounding algorithm and its analysis. A semidefinite programming relaxation of an optimization problem yields

a set of vectors $v_1, \ldots, v_n$ satisfying certain conditions and achieving some objective value c. The goal of a rounding algorithm is to transform this set of vectors into, say, a $+1/-1$ solution, satisfying the same conditions and achieving value c' that is close to c. At a very high level, our main result is that if these vectors have some non-trivial *global correlation*, then a good rounding can be achieved with a non-trivially small number of hierarchy rounds. Our second observation is that in several cases, the vectors corresponding to a *good* SDP solution can be shown to have significant mass inside some low-dimensional subspace, and that implies a lower bound on their global correlation. Below we elaborate on what we mean, using the Max Cut problem (which is a special case of Unique Games) as an illustrative example. Our result for Max Cut is worked out in more detail in Section 4.

Rounding SDP's using a small basis. The SDP solution for Max Cut problem consists of a sequence $\mathcal{V} = v_1, \ldots, v_n$ of unit vectors, and the objective value is the expectation of $(1 - \langle v_i, v_j \rangle)/2$ over all edges $\{i, j\}$ in the input graph. 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.

Local to global correlation. From the above discussion, our goal of rounding SDP hierarchies is reduced to finding a small number of basis vectors $v_{i_1}, \ldots, v_{i_r}$ such that every (or at least most) other vector in the solution $\mathcal{V}$ has very large projection into their span. But, why should such vectors exist? We show that we can assume they exist if the original Max Cut instance has small *threshold rank*. The latter is a condition that, as mentioned above, holds for many natural families of instances, including the canonical "hard instances" that are known to fool the GW algorithm— the noisy sphere and noisy Gaussian graphs [FS02, RS09]. The key concept behind our proof is the notion of *local vs global correlations*. It is a very well known property of expander graphs that random edges behave similarly to pairs of independently chosen vertices with respect to some tests. Specifically, if G is an n-vertex expander in the sense that the normalized adjacency matrix A_G's second largest eigenvalue is at most ε, and f is a bounded function mapping vertices to numbers, then we know that $\mathbb{E}_{i,j}[|f(i) - f(j)|^2] \in (1 \pm O(\varepsilon)) \mathbb{E}_{i \sim j}[|f(i) - f(j)|^2])$, where the former expectation is over pairs of vertices and the latter is over pairs connected by an edge. In other words, expander graphs imply that if f is *locally correlated* over the edges of an expander graph, then it is also *globally correlated*. In fact, this is easily shown to hold even if f maps vertices not into numbers but into vectors— i.e., if $v_1, \ldots, v_n$ are unit vectors that are locally correlated over the edges of G then they are also globally correlated. Indeed, this property of expanders has

been used in the work of [AKK⁺08], who showed that the basic SDP program for UNIQUE GAMES can be successfully rounded if the input graph is an expander.

Our starting point is to observe that a somewhat similar, though much weaker condition holds even when the graph has at most, say, $r/100$ eigenvalues larger than ε. In this case it's possible to show that,say, if (*) $\mathbb{E}_{i \sim j} \langle v_i, v_j \rangle^2 > 100 \sqrt{e}$, then (**) $\mathbb{E}_{i,j} \langle v_i, v_j \rangle^2 \geqslant 1/r$ (see Section 6). If r is super-constant, the condition (**) does not seem a-priori useful for obtaining a good integral solution. Indeed, the standard integrality gap example of MAX CUT is a graph with fairly small (polylogarithmic) number of large eigenvalues, but no good integral solution. However, (**) does imply that we can find at least one vector v_{i_1} such that $\mathbb{E}_j \langle v_{i_1}, v_j \rangle^2 \geqslant 1/r$. We can now replace each vector $v \in \mathcal{V}$ with its projection into the orthogonal space to v_{i_1} and continue until we either get stuck or 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 Span$\{v_{i_1}, \ldots, v_{i_r}\}$, in which case we can successfully round the solution. The only way we can get stuck is if at some point we get that (*) is violated. Now, in the case of MAX CUT, if (*) was violated initially, then the value of the SDP would be about $1/2$, which is trivial to round by just taking a random cut. To show that we can easily round even when (*) is violated at some later point in the process, it's useful to switch to the distribution view of SDP hierarchies.

Distribution view of SDP's. Another, often beneficial way to view SDP hierarchies is as providing *distribution* on integral solutions (see Section 3.2). In this view, for every set of $r + 2$ vertices $i_1, \ldots, i_r, i_{r+1}, i_{r+2}$, the SDP hierarchy provides a distribution $X_{i_1}, \ldots, X_{i_r}$ over ± 1. Moreover, we require that distributions on overlapping sets will be consistent, and that the for every two variables i, j the expectation $E[X_i X_j]$ will equal the inner product $\langle v_i, v_j \rangle$ of the corresponding vectors. The challenge in rounding the SDP is that there is not necessarily a way to sample simultaneously the random variables $X_1, \ldots, X_n$ in some consistent way. The projection of a vector v into the span of $v_{i_1}, \ldots, v_{i_r}$ turns out to capture (an appropriate notion of) the *mutual information* between the variable X_{i_1} and the variables $X_{i_1}, \ldots, X_{i_r}$. Looked at from this viewpoint, our rounding algorithm involves choosing an assignment from the distribution for the basis vertices, and *conditioning* on its value. As long as (**) holds, we can find a random variable X_i such that conditioning on X_i will significantly decrease the entropy of the remaining variables. When we get stuck and (*) is violated, it means that for a typical edge $i \sim j$, the random variables X_i and X_j are close to being *statistically independent*. This means that just sampling each X_i independently will give approximately the same value on a typical constraint.

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 [KT07, Kol10, ABS10, Ste10] 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.

Remark 2.1. Theorem 1.3 considers a different parameter than Theorems 1.1 and 1.2. The latter two results consider threshold ranks for a small (i.e., close to 0) threshold τ, and achieve a very good approximation. In contrast, Theorem 1.1 considers threshold τ that is close to 1, but only achieve a rough approximation (corresponding to the approximation guarantee relevant to the unique games conjecture). This is also manifested in some technical differences in the proofs.

Organization

We begin by fixing notation and a few formal definitions in the next section. For the purpose of exposition, we first present an algorithm for MAX CUT on low-rank graphs using the Lasserre hierarchy in Section 4. Following this, the general algorithm for 2-CSPs on low-rank graphs is presented in Section 5. The connection between local and global correlations in low-rank graphs that is central to our algorithms, is outlined in Section 6. To implement our general approach in a hierarchy weaker than Lasserre hierarchy, we outline an argument to obtain low-rank approximation to any set of vectors in Section 7. The final section (Section 8) of the paper is devoted to subexponential time algorithm for Unique Games.

3   Preliminaries

We will use capital letters X, Y to denote random variables, and lower-case letters to denote assignments to these random variables.

For a real-valued random variable X, let $\mathrm{Var}[X]$ denote its variance. In this work, we will use random variables taking values over a range $[k] = \{1 \ldots k\}$. For a random variable X over $[k]$, and $a \in [k]$, let X_a denote the indicator of the event that $X = a$. We define the variance of X to be,

$$\mathrm{Var}[X] \stackrel{\mathrm{def}}{=} \sum_{a \in [k]} \mathrm{Var}[X_a] = 1 - \mathsf{CP}(X). \tag{3.1}$$

The *collision probability* of X is defined as

$$\mathsf{CP}(X) \stackrel{\mathrm{def}}{=} \mathop{\mathbb{P}}_{XX'} \{X = X'\},$$

where X' is an independent copy of X (so that the sequence X, X' is i.i.d.). It is easy to see that the variance and collision probability are related by,

$$\mathsf{CP}(X) = 1 - \mathrm{Var}[X].$$

For two jointly-distributed random variables X, Y, let $[X \mid Y = y]$ denote the random variable X conditioned on the event that $Y = y$. If it is clear from the context, we write $(X|y)$ for $(X|Y = y)$. We will denote by $\mathbb{E}_{\{Y\}} \mathrm{Var}[X|Y]$ the following quantity,

$$\mathop{\mathbb{E}}_{\{Y\}} \mathrm{Var}[X|Y] = \mathop{\mathbb{E}}_{y} \left[\mathrm{Var}[(X|Y = y)] \right].$$

3.1 Unique Games

Definition 3.1. 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,

$$\text{maximize} \quad \Pr_{(i,j) \in E} \left\{ \pi_{ij}(\ell(i)) = \ell(j) \right\}$$

3.2 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 $\{\mu_T\}_{T \subseteq V,\ |T| \leq m+2}$ is *consistent* if for all $T, T' \subseteq V$ with $|T|, |T'| \leq m + 2$, 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 $\mathbb{P}\left\{ X_i^{(S \cup \{i,j\})} = X_j^{(S \cup \{i,j\})} \mid X_S^{(S \cup \{i,j\})} \right\}$.

3.3 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 $|S| \leq m+2$ 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 $|S \cup T| \leq m + 2$, 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 $|S| \leq m$ and local assignment $x_S \in [k]^S$, the matrix $\left\{ \text{Cov}(X_{ia}, X_{jb} \mid X_S = x_S) \right\}_{i,j \in V,\, a,b \in [k]}$ is positive semidefinite.

---

[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 $\{X_i \mid X_S = x_S\}$.

4 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,

$$\mathop{\mathbb{P}}_{\mu_{ij}}[X_i \neq X_j] = \|v_i - v_j\|^2 .$$

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 $\text{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]$. Using this measure, the statistical distance between independent sampling $(\mu_i \times \mu_j)$ and correlated sampling (μ_{ij}) is given by

$$\|\mu_{ij} - \mu_i \times \mu_j\|_1 \leq |\text{Cov}(X_i, X_j)| .$$

(See Lemma 5.1 for a more general version of the above 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:

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

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,

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

9

The above bound is proven in a more general setting in Lemma 5.2. As all random variables involved have variance at most 1, we can rewrite the above expression as,

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

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 6):

Lemma 4.1. *Let $v_1, \ldots, v_n$ be vectors in the unit ball. Suppose that the vectors are correlated across the edges of a regular n-vertex graph G,*

$$\underset{ij \sim G}{\mathbb{E}} \langle v_i, v_j \rangle \geqslant \rho \,.$$

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

$$\underset{i,j \in V}{\mathbb{E}} |\langle v_i, v_j \rangle| \geqslant \Omega(\rho) / \operatorname{rank}_{\geqslant \Omega(\rho)}(G) \,.$$

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,

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

and $\mathbb{E}_i[\|v_i\|^2] \leqslant 1$. If the graph G is low-rank, then by Lemma 4.1 we get a lower bound on the global correlation of the vectors v_i, namely

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

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 details of this argument are presented in Theorem 5.6.

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 $\mathbb{P}_{(i,j,\Pi) \sim \mathfrak{I}} \big\{ (x_i, x_j) \in \Pi \big\}$.)

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.)

Let $X_1, \ldots, X_n$ be r-local random variables 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.

For two random variables X and X' with the same range, we denote their *statistical distance*,

$$\| \{X\} - \{X'\} \|_1 \overset{\text{def}}{=} \sum_x \big| \mathbb{P}\{X = x\} - \mathbb{P}\{X' = x\} \big|.$$

Independent Sampling and Pairwise Correlation. The following lemma shows that the statistical difference between independent sampling and correlated sampling is explained by local correlation.

Lemma 5.1. *For any two vertices $i, j \in V$,*

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

Proof. Under the distribution $\{X_i X_j\}$, the event $\{X_i = a, X_j = b\}$ has probability $\mathbb{E}\, X_{ia} X_{jb}$. On the other hand, under the product distribution $\{X_i\}\{X_j\}$, this event has probability $\mathbb{E}\, X_{ia} \mathbb{E}\, X_{jb}$. Hence, the difference of these probabilities is equal to $\mathbb{E}\, X_{ia} X_{jb} - \mathbb{E}\, X_{ia} \mathbb{E}\, X_{jb} = \mathrm{Cov}(X_{ia}, X_{jb})$. $\square$

Conditional Variance and Pairwise Correlation. The following lemma shows that conditioning on a variable X_j decreases the variance of a variable X_i by the correlation of the variables X_{ia} and X_{jb}.

Lemma 5.2. *For any two vertices $i, j \in V$,*

$$\mathrm{Var}\, X_i - \mathop{\mathbb{E}}_{\{X_j\}} \mathrm{Var}\Big[X_i \,\big|\, X_j\Big] \geq \tfrac{1}{k} \sum_{a,b \in [k]} \mathop{\mathbb{E}}_{\{X_{ia} X_{jb}\}} \mathrm{Cov}(X_{ia}, X_{jb})^2 / \mathrm{Var}\, X_{jb}$$

Proof. If we condition on X_{jb}, the variance of X_{ia} decreases by $\mathrm{Cov}(X_{ia}, X_{jb})^2 / \mathrm{Var}\, X_{jb}$ (Lemma C.2). Thus, the variance of X_i deceases by $\sum_a \mathrm{Cov}(X_{ia}, X_{jb})^2 / \mathrm{Var}\, X_{jb}$. Hence, there exists b_0 such that conditioning on X_{jb_0} causes a variance decrement of at least $\tfrac{1}{k} \sum_{a,b} \mathrm{Cov}(X_{ia}, X_{jb})^2 / \mathrm{Var}\, X_{jb}$. Since the variance is non-increasing under conditioning, the variance of X_i decreases by at least this amount when we condition on X_j. $\square$

---

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

Pairwise Correlations and Inner Products. The previous paragraphs were about two different notions of pairwise correlation. On the one hand, $\| \{X_i X_j\} - \{X_i\}\{X_j\} \|_1$ and on the other hand, $\mathrm{Var}\, X_i - \mathbb{E}_{\{X_j\}} \mathrm{Var}[X_i \mid X_j]$. The following lemma relates these two notions of pairwise correlations and shows they can be approximated by inner products of vectors.

Lemma 5.3. *Suppose that the matrix $\left(\mathrm{Cov}(X_{ia}, X_{jb})\right)_{i \in V, a \in [k]}$ is positive semidefinite. Then, there exists vectors $v_1, \ldots, v_n$ in the unit ball such that for all vertices $i, j \in V$,*

$$\frac{1}{k^2}\Big(\sum_{(a,b)\in[k]^2} \big| \mathrm{Cov}(X_{ia}, X_{jb}) \big| \Big)^2 \leq \langle v_i, v_j \rangle \leq \frac{1}{k} \sum_{(a,b)\in[k]^2} \frac{1}{2}\big(\tfrac{1}{\mathrm{Var}\, X_{ia}} + \tfrac{1}{\mathrm{Var}\, X_{jb}} \big) \mathrm{Cov}(X_{ia}, X_{jb})^2 \,.$$

Proof. Let $\{u_{ia}\}$ be the collection of vectors such that $\langle u_{ia}, u_{jb} \rangle = \mathrm{Cov}(X_{ia}, X_{jb})$. Note that $\|u_{ia}\|^2 = \mathrm{Var}\, X_{ia}$. Define $v_i := k^{-1/2} \sum_a \overline{u_{ia}^{\otimes 2}}/\|u_{ia}\|$. (Here, $\bar{x}$ denote the unit vector in direction x.) The inner product of v_i and v_j is equal to

$$\langle v_i, v_j \rangle = \frac{1}{k} \sum_{a,b} \frac{1}{\sqrt{\mathrm{Var}\, X_{ia} \, \mathrm{Var}\, X_{jb}}} \mathrm{Cov}(X_{ia}, X_{jb})^2 \,.$$

Using the inequality between arithmetic mean and geometric mean, we have $(\mathrm{Var}\, X_{ia} \, \mathrm{Var}\, X_{jb})^{-1/2} \leq (1/\mathrm{Var}\, X_{ia} + 1/\mathrm{Var}\, X_{jb})/2$, which implies the desired upper bound on the inner product $\langle v_i, v_j \rangle$.

On the other hand, by Cauchy–Schwartz,

$$\Big(\sum_{a,b} \big| \mathrm{Cov}(X_{ia}, X_{jb}) \big| \Big)^2 \leq \sum_{a,b} \sqrt{\mathrm{Var}\, X_{ia} \, \mathrm{Var}\, X_{jb}} \cdot \sum_{a,b} \frac{1}{\sqrt{\mathrm{Var}\, X_{ia} \, \mathrm{Var}\, X_{jb}}} \mathrm{Cov}(X_{ia}, X_{jb}) \,.$$

Since $\sum_a \mathrm{Var}\, X_{ia} \leq \sum_a \mathbb{E}\, X_{ia}^2 = 1$, we have $\sum_a \sqrt{\mathrm{Var}\, X_{ia}} \leq \sqrt{k}$ for all vertices $i \in V$ (by Cauchy–Schwartz). Therefore,

$$\Big(\sum_{a,b} \big| \mathrm{Cov}(X_{ia}, X_{jb}) \big| \Big)^2 \leq k \sum_{a,b} \frac{1}{\sqrt{\mathrm{Var}\, X_{ia} \, \mathrm{Var}\, X_{jb}}} \mathrm{Cov}(X_{ia}, X_{jb}) \,,$$

which gives the desired lower bound on the inner product $\langle v_i, v_j \rangle$. It remains to argue that the vectors $v_1, \ldots, v_n$ are contained in the unit ball. Since $\mathrm{Cov}(X_{ia}, X_{ib})^2 \leq \mathrm{Var}\, X_{ia} \, \mathrm{Var}\, X_{ib}$, we can upper bound $\|v_i\|^2 \leq k^{-1} \sum_{a,b} \sqrt{\mathrm{Var}\, X_{ia} \, \mathrm{Var}\, X_{ib}} \leq 1$ (using $\sum_a \sqrt{\mathrm{Var}\, X_{ia}} \leq \sqrt{k}$). $\square$

Local Correlation vs Global Correlation on Low-Rank Graphs. The following lemma shows that local correlation (correlation across edges of a graph) implies global correlation (correlation between random vertices) if the graph has low threshold rank. (Proof in Section 6.)

Lemma (Restatement of Lemma 4.1)**.** *Let $v_1, \ldots, v_n$ be vectors in the unit ball. Suppose that the vectors are correlated across the edges of a regular n-vertex graph G,*

$$\mathbb{E}_{ij\sim G} \langle v_i, v_j \rangle \geq \rho \,.$$

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

$$\mathbb{E}_{i,j\in V} |\langle v_i, v_j \rangle| \geq \Omega(\rho)/\mathrm{rank}_{\geq\Omega(\rho)}(G) \,.$$

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

12

Putting Things Together. The following lemma shows that either independent sampling is statistically close to correlated sampling across edges of a graph or the typical variance of a vertex decreases non-trivially by conditioning on a random vertex.

Lemma 5.4. *Let G be a regular n-vertex graph and ε be the expected statistical distance between independent and correlated sampling across the edges of G,*

$$\varepsilon = \mathop{\mathbb{E}}_{ij \sim G} \left\| \{X_i X_j\} - \{X_i\}\{X_j\} \right\|_1$$

Further, suppose that the matrix $\left(\mathrm{Cov}(X_{ia}, X_{jb})\right)_{i \in V, a \in [k]}$ is positive semidefinite. Then, conditioning on a random vertex decreases the variances by

$$\mathop{\mathbb{E}}_{i,j \in V} \mathop{\mathbb{E}}_{\{X_j\}} \mathrm{Var}\left[X_i \mid X_j\right] \leqslant \mathop{\mathbb{E}}_{i \in V} \mathrm{Var}\, X_i - \Omega(\varepsilon^2/k)/\mathrm{rank}_{\geqslant \Omega(\varepsilon/k)^2}(G)\,.$$

Proof. Let $v_1, \ldots, v_n$ be the vectors constructed in Lemma 5.3. By Lemma 5.3 and Lemma 5.1, the local correlation of these vectors is at least

$$\mathop{\mathbb{E}}_{ij \sim G} \langle v_i, v_j \rangle \geqslant \tfrac{1}{k^2} \mathop{\mathbb{E}}_{ij \sim G} \left\| \{X_i X_j\} - \{X_i\}\{X_j\} \right\|_1^2 \geqslant \varepsilon^2/k^2\,.$$

(The last step also uses Cauchy–Schwartz.) Hence, Lemma 4.1 implies the following lower bound on the global correlation of these vectors,

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

Lemma 5.3 and Lemma 5.2 allows us to relate the expected decrement of the variances to the global correlation of the vectors $v_1, \ldots, v_n$,

$$\mathop{\mathbb{E}}_{i,j \in V} \left[\mathrm{Var}\, X_i - \mathop{\mathbb{E}}_{\{X_j\}} \mathrm{Var}[X_i \mid X_j] \right] \geqslant k \cdot \mathop{\mathbb{E}}_{i,j \in V} |\langle v_i, v_j \rangle|\,,$$

which gives the desired upper bound on $\mathbb{E}_{i,j \in V} \mathbb{E}_{\{X_j\}} \mathrm{Var}[X_i \mid X_j]$. $\qquad\square$

The following lemma asserts that if the constraint graph has low threshold rank then there exists a partial assignment x_S to a small set S of vertices such that independent sampling conditioned on this assignment x_S gives almost the same value as correlated sampling (without conditioning on the assignment x_S).

Algorithm 5.5 (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.

Theorem 5.6. *Let $X_1, \ldots, X_n$ be r-local random variables and let $X_1', \ldots, X_n'$ be the random variables produced by [Algorithm 5.5](#) on input $X_1, \ldots, X_n$. Suppose that the matrices $(\mathrm{Cov}(X_{ia}, X_{jb} \mid X_S = x_S))_{i \in V, a \in [k]}$ are positive semidefinite for every set $S \subseteq V$ with $|S| \leqslant r$ and local assignment $x_S \in [k]^S$. Then, if $r \gg O(k/\varepsilon^4) \cdot \mathrm{rank}_{\Omega(\varepsilon/k)^2}(G)$,*

$$\underset{ij \sim G}{\mathbb{E}} \left\| \{X_i X_j\} - \{X_i' X_j'\} \right\|_1 \leqslant \varepsilon.$$

Proof. Let us define ε_m,

$$\varepsilon_m = \underset{S \in V^m}{\mathbb{E}} \underset{\{X_S\}}{\mathbb{E}} \underset{ij \sim G}{\mathbb{E}} \left\| \{X_i X_j \mid X_S\} - \{X_i \mid X_S\}\{X_j \mid X_S\} \right\|_1.$$

To prove the current theorem it is enough to show that $\mathbb{E}_{m \in [r]} \varepsilon_m \leqslant \varepsilon$. For $m \leqslant r$, define a non-negative potential Φ_m as follows

$$\Phi_m := \underset{S \in V^m}{\mathbb{E}} \underset{\{X_S\}}{\mathbb{E}} \underset{i \in V}{\mathbb{E}} \mathrm{Var}(X_i \mid X_S).$$

Let $m \in [r]$. Suppose $\varepsilon_m \geqslant \varepsilon/2$. Then,

$$\underset{S \in V^m, \{X_S\}}{\mathbb{P}} \left\{ \underset{ij \sim G}{\mathbb{E}} \left\| \{X_i X_j \mid X_S\} - \{X_i \mid X_S\}\{X_j \mid X_S\} \right\|_1 \geqslant \varepsilon/2 \right\} \geqslant \varepsilon/2.$$

Therefore, by [Lemma 5.4](#),

$$\underset{S \in V^m, \{X_S\}}{\mathbb{E}} \left[\underset{i \in V}{\mathbb{E}} \mathrm{Var}[X_i \mid X_S] - \underset{i,j \in V}{\mathbb{E}} \mathrm{Var}[X_i \mid X_S, X_j] \right] \geqslant \varepsilon/2 \cdot \Omega(\varepsilon^2/k)/\mathrm{rank}_{\geqslant \Omega(\varepsilon/k)^2}(G).$$

In other words, $\Phi_{m+1} \leqslant \Phi_m - \Omega(\varepsilon^3/k)/\mathrm{rank}_{\geqslant \Omega(\varepsilon/k)^2}(G)$. Since $1 \geqslant \Phi_1 \geqslant \ldots \geqslant \Phi_r \geqslant 0$, it follows that there are at most $O(k/\varepsilon^3) \cdot \mathrm{rank}_{\geqslant \Omega(\varepsilon/k)^2}(G)$ indices $m \in [r]$ such that $\varepsilon_m \geqslant \varepsilon/2$. Therefore, if $r \gg O(k/\varepsilon^4) \cdot \mathrm{rank}_{\geqslant \Omega(\varepsilon/k)^2}(G)$, we have

$$\underset{m \in [r]}{\mathbb{E}} \varepsilon_m \leqslant \varepsilon/2 + \frac{1}{r} \cdot O(k/\varepsilon^3) \cdot \mathrm{rank}_{\geqslant \Omega(\varepsilon/k)^2}(G) \leqslant \varepsilon.$$

Finally, by the triangle inequality,

$$\begin{aligned}
&\underset{ij \sim G}{\mathbb{E}} \left\| \{X_i X_j\} - \{X_i' X_j'\} \right\|_1 \\
&= \underset{ij \sim G}{\mathbb{E}} \left\| \left(\underset{m \in [r]}{\mathbb{E}} \underset{S \in V^m}{\mathbb{E}} \underset{\{X_S\}}{\mathbb{E}} \{X_i X_j \mid X_S\} \right) - \left(\underset{m \in [r]}{\mathbb{E}} \underset{S \in V^m}{\mathbb{E}} \underset{\{X_S\}}{\mathbb{E}} \{X_i \mid X_S\}\{X_j \mid X_S\} \right) \right\|_1 \\
&\leqslant \underset{ij \sim G}{\mathbb{E}} \underset{m \in [r]}{\mathbb{E}} \underset{S \in V^m}{\mathbb{E}} \underset{\{X_S\}}{\mathbb{E}} \left\| \{X_i X_j \mid X_S\} - \{X_i \mid X_S\}\{X_j \mid X_S\} \right\|_1 = \underset{m \in [r]}{\mathbb{E}} \varepsilon_m \leqslant \varepsilon. \qquad \square
\end{aligned}$$

The following theorem directly implies [Theorem 1.1](#).

Theorem 5.7. *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 5.5](#) (Propagation Sampling) outputs an assignment with expected value at least $\sigma - \varepsilon$ for $\mathfrak{I}$.*

Proof. An optimal r-round Lasserre solution gives rise to r-local random variables $X_1, \ldots, X_n$ over $[k]$. Let X_{ia} be the indicator variable of the event $X_i = a$. The matrices $\{\mathrm{Cov}(X_{ia}, X_{jb} \mid X_S = x_S)\}_{i,j \in V, a,b \in [k]}$ are positive semidefinite for all sets $S \subseteq V$ with $|S| \leqslant r$ and local assignments $x_S \in [k]^S$. Furthermore, the Lasserre solution satisfies

$$\underset{(i,j,\Pi) \sim \mathfrak{I}}{\mathbb{E}} \underset{\{X_i X_j\}}{\mathbb{P}} \left\{ (X_i, X_j) \in \Pi \right\} = \sigma.$$

Let $X'_1, \ldots, X'_n$ be the jointly-distributed (global) random variables in Theorem 5.6. By Theorem 5.6, we can estimate the expected value of the assignment $X'_1, \ldots, X'_n$ as

$$
\mathop{\mathbb{E}}_{X'_1,\ldots,X'_n} \mathop{\mathbb{P}}_{(i,j,\Pi)\sim\mathfrak{I}} \left\{ (X'_i, X'_j) \in \Pi \right\} = \mathop{\mathbb{E}}_{(i,j,\Pi)\sim\mathfrak{I}} \mathop{\mathbb{P}}_{\{X'_i X'_j\}} \left\{ (X'_i, X'_j) \in \Pi \right\}
$$
$$
\geqslant \mathop{\mathbb{E}}_{(i,j,\Pi)\sim\mathfrak{I}} \mathop{\mathbb{P}}_{\{X_i X_j\}} \left\{ (X_i, X_j) \in \Pi \right\} - \tfrac{1}{2} \mathop{\mathbb{E}}_{ij\sim G} \left\| \{X_i X_j\} - \{X'_i X'_j\} \right\|_1
$$
$$
\geqslant \sigma - \varepsilon . \qquad \square
$$

5.1 Special case of Unique Games

The following lemma is a version of Lemma 5.3 tailored towards Unique Games. The advantage of this version of the lemma is that the bounds are independent of the alphabet size k.

Lemma 5.8. *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 $\left(\mathrm{Cov}(X_{ia}, X_{jb})\right)_{i\in V,\, a\in[k]}$ is positive semidefinite. Then, there exists vectors $v_1, \ldots, v_n$ in the unit ball such that for all vertices $i, j \in V$ and permutations π of $[k]$,*

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

The following theorem immediately implies Theorem 1.2. Let $\mathfrak{I}$ be a Unique Games instance with alphabet size k and constraint graph G.

Theorem 5.9. *Let $\varepsilon > 0$ and $r = k \cdot \mathrm{rank}_{\geqslant\Omega(\varepsilon^4)}(G)/\varepsilon^{O(1)}$. Suppose that the r-round Lasserre value of the Unique Games instance $\mathfrak{I}$ is σ. Then, given an optimal r-round Lasserre solution, Algorithm 5.5 (Propagation Sampling) outputs an assignment with expected value at least $\sigma - \varepsilon$ for $\mathfrak{I}$.*

Proof Sketch. Let $X_1, \ldots, X_n$ be r-local random variables over $[k]$ from an optimal r-round Lasserre solution for $\mathfrak{I}$. The local variables satisfy

$$
\mathop{\mathbb{E}}_{(i,j,\pi)\sim\mathfrak{I}} \mathop{\mathbb{P}}_{\{X_i X_j\}} \left\{ X_j = \pi(X_i) \right\} = \sigma .
$$

For a permutation π of $[k]$, we define a modified version of statistical distance,

$$
\| \{X_i X_j\} - \{X_i\}\{X_j\} \|_\pi \overset{\text{def}}{=} \sum_a \big| \mathop{\mathbb{P}}_{\{X_i X_j\}} \left\{ X_j = \pi(X_i) \right\} - \mathop{\mathbb{P}}_{\{X_i\}\{X_j\}} \left\{ X_j = \pi(X_i) \right\} \big| .
$$

The following analog of Lemma 5.1 holds,

$$
\| \{X_i X_j\} - \{X_i\}\{X_j\} \|_\pi = \sum_a |\mathrm{Cov}(X_{ia}, X_{j\pi(a)})| .
$$

Using Lemma 5.8, it is straight-forward to prove a better versions of Lemma 5.4 and Theorem 5.6 for our modified notion of statistical distance. The conclusion is that for $r \geqslant k \cdot \mathrm{rank}_{\geqslant\Omega(\varepsilon^4)}(G)/\varepsilon^4$, Algorithm 5.5 (Propagation Sampling) produces (global) random variables $X'_1, \ldots, X'_n$ such that

$$
\mathop{\mathbb{E}}_{(i,j,\pi\sim\mathfrak{I})} \| \{X_i X_j\} - \{X'_i - X'_j\} \|_\pi \leqslant \varepsilon .
$$

15

Therefore, we can estimate the expected fraction of satisfied constraints as

$$\underset{X'_1,\ldots,X'_n}{\mathbb{E}} \underset{(i,j,\pi)\sim\mathfrak{I}}{\mathbb{P}} \left\{X'_j = \pi(X'_i)\right\} \geqslant \underset{X'_1,\ldots,X'_n}{\mathbb{E}} \underset{(i,j,\pi)\sim\mathfrak{I}}{\mathbb{P}} \left\{X_j = \pi(X_i)\right\} - \underset{(i,j,\pi\sim\mathfrak{I})}{\mathbb{E}} \|\{X_i X_j\} - \{X'_i - X'_j\}\|_\pi$$

$$\geqslant \sigma - \varepsilon \,. \qquad\qquad \square$$

6 Local Correlation implies Global Correlation in 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 6.1. *Suppose there exist vectors $v_1,\ldots,v_n \in \mathbb{R}^n$ such that*

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

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 $(\langle v_i, v_j \rangle)_{i,j\in V}$ represented in the eigenbasis of G, so that

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

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,

$$q = \sum_{r=1}^{m'} p_r \leqslant m' \sum_{r=1}^{n} p_r^2 \leqslant \tfrac{m'}{m} \,.$$

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,

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

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 $\lambda_r \geqslant -C \cdot \varepsilon$. $\square$

Note that Lemma 4.1 follows directly from the previous lemma by picking $C = \frac{(1-\rho/100)}{(1-\rho)}$ and observing that $\mathbb{E}_{i,j\in V} |\langle v_i, v_j \rangle| \geqslant \mathbb{E}_{i,j\in V} |\langle v_i, v_j \rangle|^2$ since $|\langle v_i, v_j \rangle| \leqslant 1$ for all $i, j \in V$

As a converse to Lemma 6.1, the following lemma shows that if a graph has many eigenvalues close to 1, then there exist vectors for the vertices of the graph with high local correlation and low global correlation.

16

Lemma 6.2. *If $\lambda_m \geqslant 1 - \varepsilon$, then there exist vectors $v_1, \ldots, v_n \in \mathbb{R}^m$ such that*

$$\mathop{\mathbb{E}}_{ij \sim G} \langle v_i, v_j \rangle \geqslant 1 - \varepsilon \,,$$

$$\mathop{\mathbb{E}}_{i,j \in V} \langle v_i, v_j \rangle^2 = \tfrac{1}{m} \,,$$

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

Proof. Let $f^{(1)}, \ldots, f^{(m)} \colon V \to \mathbb{R}$ be orthonormal eigenfunctions of G with eigenvalue larger than $1 - \varepsilon$. Consider vectors $v_1, \ldots, v_n \in \mathbb{R}^m$ satisfying $\langle v_i, v_j \rangle = \mathbb{E}_{r \in [m]} f_i^{(r)} f_j^{(r)}$. Since the functions $f^{(r)}$ have norm 1, the typical squared norm of the vectors v_i satisfies

$$\mathop{\mathbb{E}}_{i \in V} \|v_i\|^2 = \mathop{\mathbb{E}}_{r \in [m]} \|f^{(r)}\|^2 = 1 \,.$$

Since the eigenvalues of the eigenfunctions $f^{(r)}$ are larger than $1 - \varepsilon$, we can lower bound the local correlation of the vectors v_i,

$$\mathop{\mathbb{E}}_{ij \sim G} \langle v_i, v_j \rangle = \mathop{\mathbb{E}}_{r \in [m]} \langle f^{(r)}, G f^{(r)} \rangle \geqslant 1 - \varepsilon \,.$$

Finally, since the function $f^{(m)}$ are orthonormal, the global correlation of the vectors v_i is

$$\mathop{\mathbb{E}}_{i,j \in V} \langle v_i, v_j \rangle^2 = \mathop{\mathbb{E}}_{i,j \in V} \mathop{\mathbb{E}}_{r,s \in [m]} f_i^{(r)} f_j^{(r)} f_i^{(s)} f_j^{(s)} = \mathop{\mathbb{E}}_{r,s \in [m]} \langle f^{(r)}, f^{(s)} \rangle^2 = \tfrac{1}{m} \,. \qquad \square$$

Remark 6.3. The condition that there exist vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ with

$$\mathop{\mathbb{E}}_{ij \sim G} \langle v_i, v_j \rangle \geqslant 1 - \varepsilon \,,$$

$$\mathop{\mathbb{E}}_{i,j \in V} \langle v_i, v_j \rangle^2 \leqslant \tfrac{1}{m} \,,$$

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

is equivalent to the condition that there exists a symmetric positive semidefinite matrix $X \in \mathbb{R}^{V \times V}$ such that

$$\operatorname{Tr} GX \geqslant 1 - \varepsilon \,,$$

$$\operatorname{Tr} X^2 \leqslant 1/m \,,$$

$$\operatorname{Tr} X = 1 \,.$$

7 On Low Rank Approximations to Sets of Vectors

Theorem 7.1. *Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be vectors in the unit ball. Then for every $\varepsilon > 0$, there exists a subset $U \subseteq \{v_1, \ldots, v_n\}$ with $|U| \leqslant 1/\varepsilon$ such that $\mathbb{E}_{i,j \in [n]} \|w_i\| \|w_j\| \langle \bar{w}_i, \bar{w}_j \rangle^2 \leqslant \varepsilon$, where w_i is the projection of v_i to the orthogonal complement of the span of U.*

The proof of Theorem 7.1 is by an iterative construction. In each iteration, we will use the following lemma.

Lemma 7.2. *Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be vectors. Then, there exists a unit vector $u \in \{\bar{v}_1, \ldots, \bar{v}_n\}$ such that the vectors $v'_1, \ldots, v'_n$ with $v'_i = v_i - \langle v_i, u \rangle u$ satisfy the following condition,*

$$\mathop{\mathbb{E}}_{i \in [n]} \|v'_i\|^2 \leqslant \mathop{\mathbb{E}}_{i \in [n]} \|v_i\|^2 - \mathop{\mathbb{E}}_{i,j \in [n]} \|v_i\| \|v_j\| \langle \bar{v}_i, \bar{v}_j \rangle^2 \,.$$

Proof. Suppose we pick a random index $j \in [n]$ and choose $u = \bar{v}_j$. In this case, the squared norm of the vectors $v_i' = v_i - \langle v_i, u \rangle u$ equals

$$\|v_i'\|^2 = \|v_i\|^2 - \langle v_i, u \rangle^2 = (1 - \langle \bar{v}_i, \bar{v}_j \rangle^2) \|v_i\|^2 \,.$$

Hence, we can estimate the expected decrease of the typical squared norms for a random vector $u \in \{\bar{v}_1, \ldots, \bar{v}_n\}$.

$$\mathop{\mathbb{E}}_{i \in [n]} \|v_i\|^2 - \mathop{\mathbb{E}}_{u} \mathop{\mathbb{E}}_{i \in [n]} \|v_i'\|^2 = \mathop{\mathbb{E}}_{i,j \in [n]} \left(\tfrac{1}{2} \|v_i\|^2 + \tfrac{1}{2} \|v_j\|^2 \right) \langle \bar{v}_i, \bar{v}_j \rangle^2$$

$$\geqslant \mathop{\mathbb{E}}_{i,j \in [n]} \|v_i\| \, \|v_j\| \, \langle \bar{v}_i, \bar{v}_j \rangle^2$$

It follows that there exists a unit vector $u \in \{\bar{v}_1, \ldots, \bar{v}_n\}$ such that the vectors $v_i' = v_i - \langle v_i, u \rangle u$ have the desired property

$$\mathop{\mathbb{E}}_{i \in [n]} \|v_i'\|^2 \leqslant \mathop{\mathbb{E}}_{i \in [n]} \|v_i\|^2 - \mathop{\mathbb{E}}_{i,j \in [n]} \|v_i\| \, \|v_j\| \, \langle \bar{v}_i, \bar{v}_j \rangle^2 \,. \qquad \square$$

Proof of Theorem 7.1. We can construct the set U in a greedy fashion so as to minimize the total squared norm of the vectors $w_1, \ldots, w_n$ (the projections of the vectors v_i to the orthogonal complement of the span of U). (In fact, we could choose set U randomly.) To make the analysis more convenient, we use the following, slightly different construction.

1. Let $v_i^{(1)} = v_i$ for all $i \in [n]$.

2. For t from 1 to $1/\varepsilon$, construct vectors $u^{(t)} \in \mathbb{R}^n$ and $v_1^{(t+1)}, \ldots, v_n^{(t+1)} \in \mathbb{R}^n$ as follows:

   (a) Using Lemma 7.2, pick a unit vector $u^{(t)} \in \{\bar{v}_1^{(t)}, \ldots, \bar{v}_n^{(t)}\}$ such that the vectors $v_i^{(t+1)} = v_i^{(t)} - \langle v_i^{(t)}, u^{(t)} \rangle u^{(t)}$ satisfy the condition

$$\mathop{\mathbb{E}}_{i \in [n]} \|v_i^{(t+1)}\|^2 \leqslant \mathop{\mathbb{E}}_{i \in [n]} \|v_i^{(t)}\|^2 - \mathop{\mathbb{E}}_{i,j \in [n]} \|v_i^{(t)}\| \, \|v_j^{(t)}\| \, \langle v_i^{(t)}, v_j^{(t)} \rangle^2 \,.$$

Notice that the vectors $v_1^{(t)}, \ldots, v_n^{(t)}$ are the projections of the vector $v_1, \ldots, v_n$ into the orthogonal complement of the span of the vectors $u^{(1)}, \ldots, u^{(t-1)}$. Let U be the set of all indices j such that $u^{(t)} = \bar{v}_j^{(t)}$ for some $t \in \{1, \ldots, 1/\varepsilon\}$. We can verify that the vectors $u^{(1)}, \ldots, u^{(1/\varepsilon)}$ are an orthonormal basis of the span of U. Let $w_1, \ldots, w_n$ be the projections of the vectors $v_1, \ldots, v_n$ into the orthogonal complement of the span of U (so that $w_i = v_i^{(1/\varepsilon)}$). Since the vectors $w_1, \ldots, w_n$ are projections of the vectors $v_1^{(t)}, \ldots, v_n^{(t)}$ for all $t \in 1, \ldots, 1/\varepsilon$, it follows that

$$\mathop{\mathbb{E}}_{i,j \in [n]} \|w_i\| \, \|w_j\| \, \langle \bar{w}_i, \bar{w}_j \rangle^2 \leqslant \mathop{\mathbb{E}}_{i,j \in [n]} \|v_i^{(t)}\| \, \|v_j^{(t)}\| \, \langle v_i^{(t)}, v_j^{(t)} \rangle^2 \,.$$

Hence, we can bound the typical squared norm of the vectors w_i,

$$\mathop{\mathbb{E}}_{i \in [n]} \|w_i\|^2 \leqslant \mathop{\mathbb{E}}_{i \in [n]} \|v_i\|^2 - \tfrac{1}{\varepsilon} \mathop{\mathbb{E}}_{i,j \in [n]} \|w_i\| \, \|w_j\| \, \langle \bar{w}_i, \bar{w}_j \rangle^2 \,.$$

Since the left-hand side is nonnegative and $\mathbb{E}_{i \in [n]} \|v_i\|^2 \leqslant 1$, it follows that $\mathbb{E}_{i,j \in [n]} \|w_i\| \, \|w_j\| \, \langle \bar{w}_i, \bar{w}_j \rangle^2 \leqslant \varepsilon$, as desired. $\qquad \square$

For our applications it will sometimes be convenient to associate different subspace with subsets U of vectors (in Theorem 7.1, we associate the span of vectors in U with the subset U).

Theorem 7.3. *Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be vectors in the unit ball. For every subset $U \subseteq V$, let Q_U be the projector on some subspace orthogonal to the span of U. (Note that Q_U is not necessarily the projector on the orthogonal complement of the span of U.) Then for every $\varepsilon > 0$, there exists a subset $U \subseteq \{v_1, \ldots, v_n\}$ with $|U| \leqslant 1/\varepsilon$ such that $\mathbb{E}_{i,j \in [n]} \|w_i\| \|w_j\| \langle \bar{w}_i, \bar{w}_j \rangle^2 \leqslant \varepsilon$, where $w_i = Q_U v_i$.*

Proof. We use the same construction as in the proof of Theorem 7.1. The only difference is that we define $v_i^{(t+1)} = P_{U^{(t)}} v_i$ (instead of $v_i^{(t+1)} = v_i^{(t)} - \langle v_i^{(t)}, u^{(t)} \rangle u^{(t)}$). Here, $U^{(t)}$ is the set of all indices j such that $u^{(t')} = \bar{v}_j^{(t')}$ for some $t' \leqslant t$. The proof is still applies to this modifies construction because $\|v_i^{(t+1)}\| \leqslant \|v_i^{(t)} - \langle v_i^{(t)}, u^{(t)} \rangle u^{(t)}\|$ (which is the only fact used about these vectors). $\qquad \square$

8   Rounding SDP Solutions to Unique Games

In this section, we will present a subexponential time algorithm for Unique Games based on a SDP hierarchy, namely the simple SDP augmented with Sherali-Adams hierarchy. This hierarchy of relaxations weaker than the Lasserre hierarchy was studied in some earlier works [RS09, KS09]. Roughly speaking, the mth round relaxation in this hierarchy corresponds to the basic semidefinite program, along with all valid constraints on at most m vectors. Formally, the variables in the mth round relaxation for Unique Games consists of

- A collection of "local distributions" $\{\mu_T\}_{T \subseteq V, |T| \leqslant m}$. Each distribution μ_T is over local assignments $x_T \in [k]^T$.

- A set of vectors $\mathcal{V} = \{v_{ia}\} i \in V, a \in [k]$ with k orthogonal vectors for every vertex $i \in V$.

The constraint of the SDP relaxation ensure that the inner products of the vectors are consistent with the corresponding local distributions, i.e., for all $S \subseteq V, |S| \leqslant m$ $i, j \in S$ and $a, b \in [k]$,

$$\mathbb{P}_{\mu_S}\left[X_i = a \wedge X_j = b\right] = \langle v_{ia}, v_{jb} \rangle.$$

The objective value of the SDP corresponds to minimizing the number of violated constraints,

$$\text{Minimize} \quad E_{i,j \in E}\left[\sum_{a \in [k]} \|v_{ia} - v_{j\pi_{ij}(a)}\|^2\right].$$

8.1   Propagation Rounding

Let $\{\mu_T\}_{T \subseteq V, |T| \leqslant m}$ be a set of consistent local distributions over assignments. For a subset of vertices S, the distribution $\mu^{|S}$ over global assignments is sampled as follows:

1. Sample a assignment $x_S \in [k]^S$ for S according to its local distribution μ_S.

2. 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.

The above procedure will be referred to as *propagation rounding* and the set S of vertices will be called the *seed* vertices.

The following lemma implies that if the seed vertices S nearly determine the values of a set of vertices T, then the assignment output by the propagation rounding has a distribution similar to the local distribution μ_T that is part of the LP/SDP solution (hence gets close to the SDP value).

Lemma 8.1. *For a set $S \subseteq V$ $|S| = m - t$, let $\mu^{|S}$ denote the distribution over global assignments $x \in [k]^V$ output by propagation rounding with S as the seed vertex set. Then, for every subset T with $|T| \leqslant t$ we have*

$$\left\| \mu_T - \mu_T^{|S} \right\| \leqslant \sum_{t \in T} \mathrm{Var}[X_t | X_S].$$

Proof. Consider the following experiment,

1. Sample a assignment $x_S \in [k]^S$ for S according to its local distribution μ_S,

$$x_S \sim \mu_S.$$

2. Sample an assignment $y_T \in [k]^T$ according to the local distribution for $S \cup T$ conditioned on the assignment x_S for S,

$$y_T \sim \mu_{S \cup T} \mid x_S.$$

3. For every vertex $t \in T$, sample a label $x_t \in [k]$ according to the local distribution for $S \cup \{t\}$ conditioned on the assignment x_S for S,

$$x_t \sim \mu_{S,t} \mid x_S.$$

Clearly the distribution of y_T is μ_T, while the distribution of x_T is $\mu_T^{|S}$. For any $t \in T$, the coordinates x_t and y_t are independent samples from $\mu_{S,t} \mid x_S$. Therefore we have,

$$\mathbb{P}[x_t \neq y_t | x_S] = 1 - \mathrm{CP}(\{x_t | x_S\}) = \mathrm{Var}\left[(X_t | x_S)\right].$$

By a union bound we get,

$$\mathbb{P}[x_T \neq y_T | x_S] = \sum_{t \in T} \mathrm{Var}\left[(X_t | x_S)\right].$$

Averaging over the different choices of x_S,

$$\mathbb{P}[x_T \neq y_T] = \underset{x_S}{\mathbb{E}}\left[\sum_{t \in T} \mathrm{Var}\left[(X_t | x_S)\right]\right] = \sum_{t \in T} \mathrm{Var}\left[X_t | X_S\right].$$

Therefore, the statistical distance between the distributions μ_T and $\mu_T^{|S}$ associated with y_T and x_T is almost $\sum_{t \in T} \mathrm{Var}\left[X_t | X_S\right]$. $\qquad \square$

8.2 Unique Games on Low Rank Graphs

Let G be an instance of unique games whose constraint graph G has low threshold rank. Let $\mathcal{V} = \{v_{ia}\}_{i \in V, a \leqslant [k]}$ be an SDP solution for G, and let $\{\mu_S\}_{S \subseteq V, |S| \leqslant m}$ denote the associated set of locla distributions. Let $X_1, \ldots, X_n$ denote the associated m-local random variables. The main result of this section shows that there exists a small set of seed vertices fixing whose value determines the value of almost every other vertex. Formally, we show the following

Lemma 8.2. *For every integer m, there exists a subset of vertices $S \subseteq V$ of size $|S| = k^2 m$ such that*

$$\mathbb{E}_i [\mathrm{Var}[X_i|X_S]] \leqslant O\left(\frac{\eta}{\lambda_m}\right)$$

To this end, we will relate conditioning a random variable X_i on a set X_S, to projecting the SDP vectors corresponding to the variable X_i in to the span of the vectors corresponding to X_S. This analogy is formalized in the following lemma.

Lemma 8.3. *Let $X_1, X_2, \ldots, X_r$ be random variables with range $[k]$ with a joint distribution μ associated with them. For each $i \in [r]$, $a \in [k]$, let X_{ia} be the indicator of the event that $X_i = a$. Let us suppose there exists vectors $\{v_{ia}\}_{i \in [r], a \in [k]}$ such that*

$$\langle v_{ia}, v_{jb} \rangle = \mathbb{P}_{X_i X_j} \{X_i = a, X_j = b\} = \mathbb{E}[X_{ia}X_{jb}].$$

Then, for every subset $S \subseteq [r]$ we have (1) $\mathrm{Var}\,[X_{ia}|X_S] \leqslant \|P_S v_{ia}\|^2$. and (2) $\mathrm{Var}\,[X_i|X_S] \leqslant \sum_{a \in [k]} \|P_S v_{ia}\|^2$. where P_S is the projector of $\mathbb{R}^n$ in to the space orthogonal to the span of $\{v_{jb}\}_{j \in S, b \in [k]}$.

Proof. Let us suppose $v_{ia} = \sum_{j \in S, b \in [k]} c_{jb}v_{jb} + P_S v_{ia}$. Define a random variable C_S as follows,

$$C_S \stackrel{\text{def}}{=} \sum_{j \in S, b \in [k]} c_{jb}X_{jb}.$$

Note that on fixing the values $\{X_j\}_{j \in S}$, the random variable C_S is fixed.

By the definition of variance of a real random variable we have the following inequality.

$$\mathrm{Var}\,[(X_{ia}|x_S)] = \min_C \mathbb{E}[(X_{ia} - C)^2|x_S] \leqslant \mathbb{E}_{X_{ia}|x_S} [(X_{ia} - C_S)^2|x_S].$$

Averaging the above inequality over the settings of x_S, we get

$$\mathrm{Var}[X_{ia}|X_S] = \mathbb{E}_{x_S} \mathrm{Var}(\{X_{ia}|x_S\}) \leqslant \mathbb{E}_{x_S} \mathbb{E}_{X_{ia}|x_S} [(X_{ia} - C_S)^2|x_S] = \mathbb{E}_\mu\left[(X_{ia} - \sum_{j \in S, b \in [k]} c_{jb}X_{jb})^2\right]. \tag{8.1}$$

Note that the second moments of the random variables $\{X_{ia}\}_{i \in [r], a \in [k]}$ match with the corresponding inner products of vectors $\{v_{ia}\}_{i \in [r], a \in [k]}$. Hence,

$$\mathbb{E}_\mu\left[(X_{ia} - \sum_{j \in S, b \in [k]} c_{jb}X_{jb})^2\right] = \|v_{ia} - \sum_{j \in S, b \in [k]} c_{jb}v_{jb}\|^2 = \|P_S v_{ia}\|^2. \tag{8.2}$$

The claim (1) follows from (8.1) and (8.2).

The claim (2) follows from (1) and the definition of variance of a random variable taking values over $[k]$. □

Proof of Lemma 8.2. For a subset $T \subseteq \mathcal{V} = \{v_{ia}\}_{i \in V, a \in [k]}$, let S_T be the set of vertices associated with it namely,

$$S_T = \{i \in V \mid \exists b \in [k], v_{ib} \in T\},$$

Let Q_T denote the projector on to the subspace orthogonal to span of $\{v_{ia} | i \in S_T, a \in [k]\}$. In particular, Q_T is a projector on to a subspace orthogonal to T for all $T \subseteq \mathcal{V}$.

Apply Theorem 7.3 on the set of vectors $\mathcal{V} = \{v_{ia}\}_{i \in V, a \in [k]}$ with the projectors Q_T for a subset $T \subseteq \mathcal{V}$. Theorem 7.3 implies that there exists a choice of $T \subseteq \mathcal{V}$ of size $|T| = k^2 m$ such that if $u_{ia} = Q_T v_{ia}$ then,

$$\mathop{\mathbb{E}}_{i,j,a,b} \|u_{ia}\| \|u_{jb}\| \langle \bar{u}_{ia}, \bar{u}_{jb} \rangle^2 \leqslant \frac{1}{k^2 m} \tag{8.3}$$

Let $S_T \subseteq V$ be the vertex set associated with the set of vectors T. We will drop the subscript and refer to this set as S. Let P_S denote the projector in to the space orthogonal to the span of vectors $\{v_{ia}\}_{i \in S, a \in [k]}$.

Let us fix some notation: $u_{ia} \stackrel{\text{def}}{=} P_S v_{ia}$, $\tilde{u}_{ia} \stackrel{\text{def}}{=} \|u_{ia}\| \bar{u}_{ia}^{\otimes 2} \otimes \bar{v}_{ia}$. As for each $i \in V$, the vectors $\{v_{ia}\}_{a \in [k]}$ are orthogonal to each other, the set of vectors $\{\tilde{u}_{ia}\}_{a \in [k]}$ are orthogonal to each other too. For each vertex $i \in V$, we can associate a vector U_i defined as $U_i \stackrel{\text{def}}{=} \sum_{a \in [k]} \tilde{u}_{ia}$.

From (8.3), we get the following bound on the average correlation of vectors $\{\tilde{u}_{ia}\}_{i \in V, a \in [k]}$,

$$\mathop{\mathbb{E}}_{i,j,a,b} \langle \tilde{u}_{ia}, \tilde{u}_{jb} \rangle \leqslant \mathop{\mathbb{E}}_{i,j,a,b} \|u_{ia}\| \|u_{jb}\| \langle \bar{u}_{ia}, \bar{u}_{jb} \rangle^2 \leqslant \frac{1}{k^2 m} \tag{8.4}$$

Using the low global correlation between vectors $\{\tilde{u}_{ia}\}_{i \in V, a \in [k]}$ ((8.3)), we bound the global correlation between the vectors $\{U_i\}_{i \in V}$ as shown below,

$$\mathop{\mathbb{E}}_{i,j \in V} \left[\langle U_i, U_j \rangle \right] = \mathop{\mathbb{E}}_{i,j \in V} \sum_{a,b \in [k]} \langle \tilde{u}_{ia}, \tilde{u}_{jb} \rangle = k^2 \mathop{\mathbb{E}}_{i,j \in V, a,b \in [k]} \langle \tilde{u}_{ia}, \tilde{u}_{jb} \rangle \leqslant \frac{1}{m} .$$

From Lemma 8.5, the low global correlation of vectors $\{U_i\}_{i \in V}$ implies that their squared length is small, i.e.,

$$\mathop{\mathbb{E}}_{i \in V} \left[\|U_i\|^2 \right] \leqslant O\left(\frac{\eta}{\lambda_m} \right) .$$

Notice that

$$\|U_i\|^2 = \sum_{a \in [k]} \|\tilde{u}_{ia}\|^2 = \sum_{a \in [k]} \|P_s v_{ia}\|^2 .$$

By Lemma 8.3, this implies that

$$\mathop{\mathbb{E}}_{i \in V} \mathrm{Var}[X_i | X_S] \leqslant O\left(\frac{\eta}{\lambda_m} \right) .$$

$\square$

Lemma 8.4 (High Local Correlation). *If $\mathcal{V}$ is an SDP solution to unique games with value $1 - \eta$, i.e.,*

$$\mathop{\mathbb{E}}_{(i,j) \in E} \sum_{a \in [k]} \|v_{ia} - v_{j\pi_{ij}(a)}\|^2 \leqslant \eta ,$$

then

$$\mathop{\mathbb{E}}_{(i,j) \in E} \|U_i - U_j\|^2 \leqslant 3\eta .$$

22

We defer the proof to Appendix B.

Lemma 8.5 (Local Correlation $\longrightarrow$ Global Correlation). *If the vectors $\{U_i\}_{i \in V}$ satisfy,*

$$\mathbb{E}_i \left[\|U_i\|^2 \right] \geqslant \frac{4\eta}{\lambda_m} \,,$$

then the average correlation among the vectors $\{U_i\}_{i \in V}$ is at least $1/m$, i.e.,

$$\mathbb{E}_{i,j \in V} \langle U_i, U_j \rangle \geqslant \frac{1}{m} \,.$$

Proof. By Lemma 8.4, the vectors $\{U_i\}$ satisfy

$$\mathbb{E}_{(i,j) \in E} \|U_i - U_j\|^2 \leqslant 3\eta \,.$$

This implies that,

$$\mathbb{E}_{(i,j) \in E} \langle U_i, U_j \rangle \geqslant \mathbb{E}_i \|U_i\|^2 - \frac{3}{2}\eta \,.$$

Let $\mathbb{E}_i \|U_i\|^2 = C \geqslant 4\eta/\lambda_m$. Normalize the vectors U_i so as to make their average squared length equal to 1. The resulting vectors have correlation at least $(1 - \eta/C) \geqslant 1 - \lambda_m/2$. By Lemma 6.1, this implies that $\mathbb{E}_{i,j \in V} \langle U_i, U_j \rangle^2 \geqslant \frac{1}{m}$. Since $\|U_i\| \leqslant 1$ for all $i \in V$, we get

$$\mathbb{E}_{i,j \in V} \langle U_i, U_j \rangle \geqslant \mathbb{E}_{i,j \in V} \langle U_i, U_j \rangle \geqslant \frac{1}{m} \,.$$

$\square$

8.3 Wrapping Up

Our main result about UNIQUE GAMES (Theorem 1.3) is a direct consequence of Theorem 8.6 and Theorem 8.7 presented here.

Theorem 8.6. *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 mth smallest eigen value of the Laplacian of the constraint graph Γ.*

Proof. The algorithm proceeds by solving the $k^2 m + 2$-round Lasserre SDP for the given instance. Starting with the SDP solution, the algorithm runs the propagation rounding algorithm starting from every possible *seed set S* of size $|S| = k^2 m$.

By Lemma 8.2, there exists one such set S for which we have,

$$\mathbb{E}_i \left[\text{Var}[X_i | X_S] \right] \leqslant O\left(\frac{\eta}{\lambda_m} \right) \,. \tag{8.5}$$

Let $\mu^{|S}$ denote the distribution over global assignments output by the propagation rounding scheme. For an edge (i, j), let μ_{ij} denote the local distribution over $[k]^2$ suggested by the SDP solution. From Lemma 8.1, the statistical distance between μ_{ij} and $\mu_{ij}^{|S}$ is at most

$$\|\mu_{ij}^{|S} - \mu_{ij}\|_1 \leqslant \text{Var}[X_i | X_S] + \text{Var}[X_j | X_S] \,.$$

Hence for every edge (i, j),

$$\mathop{\mathbb{P}}_{\mu_{ij}^{|S}}[x_i = \pi_{ij}(x_j)] \geqslant \mathop{\mathbb{P}}_{\mu_{ij}}[x_i = \pi_{ij}(x_j)] - \mathrm{Var}[X_i|X_S] - \mathrm{Var}[X_j|X_S].$$

Averaging over all the edges we see that,

$$\mathop{\mathbb{E}}_{i,j \in E} \mathop{\mathbb{P}}_{\mu_{ij}^{|S}}[x_i = \pi_{ij}(x_j)] \quad \geqslant \quad \mathop{\mathbb{E}}_{i,j \in E} \left[\mathop{\mathbb{P}}_{\mu_{ij}}[x_i = \pi_{ij}(x_j)] \right] - \mathop{\mathbb{E}}_{i,j \in E} \mathrm{Var}[X_i|X_S] - \mathop{\mathbb{E}}_{i,j \in E} \mathrm{Var}[X_j|X_S]$$

$$\geqslant \quad \mathop{\mathbb{E}}_{i,j \in E} \left[\mathop{\mathbb{P}}_{\mu_{ij}}[x_i = \pi_{ij}(x_j)] \right] - 2 \mathop{\mathbb{E}}_{i \in V} \mathrm{Var}[X_i|X_S]$$

$$\geqslant \quad \mathsf{Val}(\mathcal{V}) - 2 \mathop{\mathbb{E}}_{i \in V} \mathrm{Var}[X_i|X_S]$$

where $\mathsf{Val}(\mathcal{V})$ is the SDP objective value of the solution $\mathcal{V}$. Along with (8.5), this implies that the algorithm on the choice of the appropriate seed set S would find a solution with value at least $1 - \eta - O(\frac{\eta}{\lambda_m})$. $\square$

Theorem 8.7. *There exists an algorithm that given a* Unique Games *instance* Γ *with vertex set* $[n]$, *label set* $[k]$, *and optimal value* $1 - \varepsilon$, *finds an assignment with value at least* $1/2$ *by rounding an* $k^2 \cdot n^{O(\varepsilon^{1/3})}$-*round Lasserre solution.*

Proof sketch. The proof follows by combining our propagation rounding and the decomposition theorem of [ABS10]. The latter result allows us to partition the input graph into disjoint components each with $1 - c\varepsilon$ rank at most $n^{O(\varepsilon^{1/3})}$ by removing at most 0.01 fraction of the edges in our input graph. An SDP solution for the input graph induces a solution for each of the components, and hence we can round the solution for each component separately using propagation rounding. $\square$

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 $\log \log^{\Omega(1)} n$ [RS09, KS09], and so a natural question, with obvious relevance to the unique games conjecture, is which bound is closer to the truth. The fact that our algorithm's running time for r rounds is only $2^{O(r)}$ (as opposed to $n^{O(r)}$), challenges the interpretation of lower bounds in the range $[\omega(1), (\log n)]$ as corresponding to super-polynomial running time, and so provides further motivation to the question of whether the current hierarchy lower bounds can be improved further.

With the exception of the Small-Set Expansion problem, we do not know how to translate algorithms for Unique Games into other computational problems. We hope that our ideas will help in combining the [ABS10] subexponential algorithm for Unique Games with SDP-based method to make progress on other Unique Games-hard computational problems. Indeed, Arora and Ge (personal communication) recently used the ideas of this work to obtain improved algorithms for 3-coloring on some interesting families of instances. A concrete open question along similar lines is whether one can get an algorithm for the Max Cut problem with approximation factor ε better than the factor of the Goemans-Williamson algorithm that runs in time $\exp(n^{\mathrm{poly}(\varepsilon)})$.

For general 2-CSPs, we know that some instances will require a large number of hierarchy rounds, but it's interesting to see whether there is any clean characterization of the instances

on which SDP hierarchies do well, encompassing, say, both low threshold rank graphs and planar graphs. Another interesting question is to find the right generalization of the low threshold rank condition to k-CSPs for $k > 2$.

References

[ABLT06]   Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis, *Proving integrality gaps without knowing the linear program*, Theory of Computing **2** (2006), no. 1, 19–51. 1

[ABS10]   Sanjeev Arora, Boaz Barak, and David Steurer, *Subexponential algorithms for unique games and related problems*, FOCS, 2010, pp. 563–572. 1, 3, 4, 6, 24

[ACOH+10] Noga Alon, Amin Coja-Oghlan, Hiêp Hàn, Mihyun Kang, Vojtech Rödl, and Mathias Schacht, *Quasi-randomness and algorithmic regularity for graphs with general degree distributions*, SIAM J. Comput. **39** (2010), no. 6, 2336–2362. 4

[AKK+08]   Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K. Vishnoi, *Unique games on expanding constraint graphs are easy*, STOC, 2008, pp. 21–28. 6, 26

[ARV09]   Sanjeev Arora, Satish Rao, and Umesh V. Vazirani, *Expander flows, geometric embeddings and graph partitioning*, J. ACM **56** (2009), no. 2. 1

[BCC+10]   Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan, *Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph*, STOC, 2010, pp. 201–210. 1, 4

[Chl07]   Eden Chlamtac, *Approximation algorithms using hierarchies of semidefinite programming relaxations*, FOCS, 2007, pp. 691–701. 1, 4

[COCF10]   Amin Coja-Oghlan, Colin Cooper, and Alan M. Frieze, *An efficient sparse regularity concept*, SIAM J. Discrete Math. **23** (2010), no. 4, 2000–2034. 4

[CT10]   Eden Chlamtac and Madhur Tulsiani, *Convex relaxations and integrality gaps*, 2010, Chapter in Handbook on Semidefinite, Cone and Polynomial Optimization. 1

[FK99]   Alan M. Frieze and Ravi Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), no. 2, 175–220. 4

[FS02]   Uriel Feige and Gideon Schechtman, *On the optimality of the random hyperplane rounding technique for MAX CUT*, Random Struct. Algorithms **20** (2002), no. 3, 403–440. 3, 5

[GMPT07]   Konstantinos Georgiou, Avner Magen, Toniann Pitassi, and Iannis Tourlakis, *Integrality gaps of $2 - o(1)$ for vertex cover sdps in the Lovász–Schrijver hierarchy*, FOCS, 2007, pp. 702–712. 1

[GS11]   Venkatesan Guruswami and Ali Kemal Sinop, *Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with psd objectives*, 2011, Manuscript. 4

[GW95]   Michel X. Goemans and David P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM **42** (1995), no. 6, 1115–1145. 1

[HKNO09]   Avinatan Hassidim, Jonathan A. Kelner, Huy N. Nguyen, and Krzysztof Onak, *Local graph partitions for approximation and testing*, FOCS, 2009, pp. 22–31. 4

[Kho02]   Subhash Khot, *On the power of unique 2-prover 1-round games*, STOC, 2002, pp. 767–775. 1

[KKMO04]   Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell, *Optimal inapproximability results for Max-Cut and other 2-variable CSPs?*, FOCS, 2004, pp. 146–154. 1

[KMS98]   David R. Karger, Rajeev Motwani, and Madhu Sudan, *Approximate graph coloring by semidefinite programming*, J. ACM **45** (1998), no. 2, 246–265. 1

[Kol10]   Alexandra Kolla, *Spectral algorithms for unique games*, IEEE Conference on Computational Complexity, 2010, pp. 122–130. 3, 6

[KPS10]   Subhash Khot, Preyas Popat, and Rishi Saket, *Approximate lasserre integrality gap for unique games*, APPROX-RANDOM, 2010, pp. 298–311. 3

[KS09]   Subhash Khot and Rishi Saket, *SDP integrality gaps with local ℓ_1-embeddability*, FOCS, 2009, pp. 565–574. 1, 3, 4, 19, 24

[KT07]   Alexandra Kolla and Madhur Tulsiani, *Playing random and expanding unique games*, Unpublished manuscript available from the authors' webpages, to appear in journal version of [AKK$^+$08], 2007. 3, 6

[KV05]   Subhash Khot and Nisheeth K. Vishnoi, *The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1*, FOCS, 2005, pp. 53–62. 3

[Las01]   Jean B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization **11** (2001), no. 3, 796–817. 1

[Lau03]   Monique Laurent, *A comparison of the sherali-adams, lovász-schrijver, and lasserre relaxations for 0-1 programming*, Math. Oper. Res. **28** (2003), no. 3, 470–496. 1

[LS91]   L. Lovász and A. Schrijver, *Cones of matrices and set-functions and 0-1 optimization*, SIAM J. Optim. **1** (1991), no. 2, 166–190. 1

[MOO05]   Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences invariance and optimality*, FOCS, 2005, pp. 21–30. 1

[Rag08]   Prasad Raghavendra, *Optimal algorithms and inapproximability results for every CSP?*, STOC, 2008, pp. 245–254. 1

[RS09]   Prasad Raghavendra and David Steurer, *Integrality gaps for strong SDP relaxations of unique games*, FOCS, 2009, pp. 575–585. 1, 3, 4, 5, 19, 24

[SA90]   Hanif D. Sherali and Warren P. Adams, *A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems*, SIAM J. Discrete Math. **3** (1990), no. 3, 411–430. MR MR1061981 (91k:90116) 1

[Sch08]   Grant Schoenebeck, *Linear level Lasserre lower bounds for certain k-CSPs*, FOCS, 2008, pp. 593–602. 1, 3, 4

[Ste10]   David Steurer, *Subexponential algorithms for d-to-1 two-prover games and for certifying almost perfect expansion*, Manuscript, 2010. 3, 6

[STT07]   Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani, *A linear round lower bound for lovasz-schrijver sdp relaxations of vertex cover*, IEEE Conference on Computational Complexity, 2007, pp. 205–216. 1

[Tul09]   Madhur Tulsiani, *CSP gaps and reductions in the Lasserre hierarchy*, STOC, 2009, pp. 303–312. 1, 4

A   Faster Algorithms for SDP hierarchies

In this section, we argue that our rounding algorithm also works with weaker SDP hierarchies. We will show that for these weaker hierarchies, a near-optimal m-round solution can be computed in time $2^{O(r)} \operatorname{poly}(n)$. Due to the equivalence of optimization and separation, it is enough to describe a separation oracle with running time $2^{O(r)} \operatorname{poly}(n)$. Given a collection of vectors $\{v_{ia}\}$, the separation oracle either has to output a good assignment or it has to output a valid linear constraint violated by the inner products of the input vectors.

We argue that such a separation oracle can easily be extracted from our rounding algorithm. Our rounding algorithm for Unique Games first selects a set S of roughly m vertices, then samples an assignment x_S for these vertices, and finally samples labels x_i for the remaining vertices from the local distributions conditioned on the event x_S. The selection of the set S depends only on the SDP vectors $\{v_{ia}\}$ but not on the local distributions (which are not known to the separation oracle).

Hence, given vectors $\{v_{ia}\}$, our separation oracle can simply work as follows:

1. Select a vertex subset using Theorem 7.1 based on the given vectors $\{v_{ia}\}$.

2. Using linear programming, find local distributions that are as consistent as possible with the inner products of the vectors $\{v_{ia}\}$. If these local distributions match the inner products sufficiently closely, then our propagation rounding algorithm will succeed. On the other hand, if the local distributions do not match the inner products closely enough, then we can find a valid linear constraints that is violated by the inner product of the given vectors. (This separating linear constraint can be obtained from the dual solution of the linear program that was used to find the best local distributions.)

B Omitted proofs from Section 5 and Section 8

This appendix contains the proofs for some omitted proofs.

Lemma B.1 (High Local Correlation). *([Lemma 8.4](#) restated) If $\mathcal{V}$ is an SDP solution to unique games with value $1 - \eta$, i.e.,*

$$\mathop{\mathbb{E}}_{(i,j)\in E} \sum_{a\in[k]} \|v_{ia} - v_{j\pi_{ij}(a)}\|^2 \leq \eta\,,$$

then

$$\mathop{\mathbb{E}}_{(i,j)\in E} \|U_i - U_j\|^2 \leq 3\eta\,, \tag{B.1}$$

Proof. Observe that the vectors u_{ia} are projections of v_{ia} and projections shrinks distances, which implies that the $\{u_{ia}\}$ vectors are correlated across constraints of the UNIQUE GAMES instance,

$$\mathop{\mathbb{E}}_{(i,j)\in E} \sum_{a\in[k]} \|u_{ia} - u_{j\pi_{ij}(a)}\|^2 \leq \mathop{\mathbb{E}}_{(i,j)\in E} \sum_{a\in[k]} \|v_{ia} - v_{j\pi_{ij}(a)}\|^2 \leq \eta\,. \tag{B.2}$$

Let $\tilde{u}_{ia} = \|u_{ia}\|\bar{u}_{ia}^{\otimes 2} \otimes \bar{v}_{ia}$. Notice that $\langle \tilde{u}_{ia}, \tilde{u}_{ib}\rangle = 0$ for distinct labels $a, b \in [k]$. We claim that the $\tilde{u}_{ia}$ vectors are also correlated across constraints of the UNIQUE GAMES instance,

Claim B.2.

$$\mathop{\mathbb{E}}_{(i,j)\in E} \sum_{a\in[k]} \|\tilde{u}_{ia} - \tilde{u}_{j\pi_{ij}(a)}\|^2 \leq 3\eta \tag{B.3}$$

Proof. The following identity relates the distance of vectors to differences in their norms and the distance of the corresponding unit vectors, $\|x - y\|^2 = (\|x\| - \|y\|)^2 + \|x\|\,\|y\|\,\|\bar{x} - \bar{y}\|^2$. Since $\|\tilde{u}_{ia}\| = \|u_{ia}\|$, we get

$$\|\tilde{u}_{ia} - \tilde{u}_{j\pi_{ij}(a)}\|^2 = \left(\|u_{ia}\| - \|u_{j\pi_{ij}(a)}\|\right)^2 + \|u_{ia}\|\,\|u_{j\pi_{ij}(a)}\| \left\|\bar{u}_{ia}^{\otimes 2} \otimes \bar{v}_{ia} - \bar{u}_{j\pi_{ij}(a)}^{\otimes 2} \otimes \bar{v}_{j\pi_{ij}(a)}\right\|^2$$

Since $\|\bar{x}_1 \otimes \bar{x}_2 - \bar{y}_1 \otimes \bar{y}_2\|^2 \leq \|\bar{x}_1 - \bar{y}_1\|^2 + \|\bar{x}_2 - \bar{y}_2\|^2$, we can further upper bound

$$\|\tilde{u}_{ia} - \tilde{u}_{j\pi_{ij}(a)}\|^2 \leq \left(\|u_{ia}\| - \|u_{j\pi_{ij}(a)}\|\right)^2 + \|u_{ia}\|\,\|u_{j\pi_{ij}(a)}\| \left(2\left\|\bar{u}_{ia} - \bar{u}_{j\pi_{ij}(a)}\right\|^2 + \left\|\bar{v}_{ia} - \bar{v}_{j\pi_{ij}(a)}\right\|^2\right)$$

$$\leq 2\|u_{ia} - u_{j\pi_{ij}(a)}\|^2 + \|v_{ia} - v_{j\pi_{ij}(a)}\|^2\,.$$

(In the last step, we again used the identity $\|x - y\|^2 = (\|x\| - \|y\|)^2 + \|x\|\,\|y\|\,\|\bar{x} - \bar{y}\|^2$. and the fact that $\|u_{ia}\|\,\|u_{j\pi_{ij}(a)}\| \leq \|v_{ia}\|\,\|v_{j\pi_{ij}(a)}\|$.) By averaging over the label set and the edges of the graph, it follows as claimed that

$$\mathop{\mathbb{E}}_{ij\in E} \sum_{a\in[k]} \|\tilde{u}_{ia} - \tilde{u}_{j\pi_{ij}(a)}\|^2 \leq 3\eta\,. \qquad \square$$

To finish the proof of the lemma, we relate the distances $\|U_i - U_j\|^2$ across an edge $ij \in E$ to distances of the vectors $\{\tilde{u}_{ia}\}$ across the constraint π_{ij},

$$\langle U_i, U_j\rangle = \sum_{a,b} \langle \tilde{u}_{ia}, \tilde{u}_{jb}\rangle$$

$$\geq \sum_{a} \langle \tilde{u}_{ia}, \tilde{u}_{j\pi_{ij}(a)}\rangle \quad \text{(using non-negativity of involved inner products)}$$

$$= \sum_a \tfrac{1}{2}\left(\|\tilde{u}_{ia}\|^2 + \|\tilde{u}_{j\pi_{ij}(a)}\|^2 - \|\tilde{u}_{ia} - \tilde{u}_{j\pi_{ij}(a)}\|^2\right)$$

$$= \tfrac{1}{2}\|U_i\|^2 + \tfrac{1}{2}\|U_j\|^2 - \sum_a \tfrac{1}{2}\|\tilde{u}_{ia} - \tilde{u}_{j\pi_{ij}(a)}\|^2 \,.$$

Rearranging gives $\|U_i - U_j\|^2 \leqslant \sum_a \|\tilde{u}_{ia} - \tilde{u}_{j\pi_{ij}(a)}\|^2$. Hence, using Claim B.2,

$$\mathop{\mathbb{E}}_{ij\in E}\|U_i - U_j\|^2 \leqslant \mathop{\mathbb{E}}_{ij\in E} \sum_a \|\tilde{u}_{ia} - \tilde{u}_{j\pi_{ij}(a)}\|^2 \leqslant 3\eta.$$

$\square$

Lemma (Restatement of Lemma 5.8). *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 $\left(\mathrm{Cov}(X_{ia}, X_{jb})\right)_{i\in V,\, a\in[k]}$ is positive semidefinite. Then, there exists vectors $v_1, \ldots, v_n$ in the unit ball such that for all vertices $i, j \in V$ and permutations π of $[k]$,*

$$\left(\sum_{a\in[k]} \left|\mathrm{Cov}(X_{ia}, X_{j\pi(a)})\right|\right)^4 \leqslant \langle v_i, v_j\rangle \leqslant \sum_{(a,b)\in[k]^2} \tfrac{1}{2}\left(\tfrac{1}{\mathrm{Var}\,X_{ia}} + \tfrac{1}{\mathrm{Var}\,X_{jb}}\right)\mathrm{Cov}(X_{ia}, X_{jb})^2 \,.$$

Proof. Let $\{u_{ia}\}$ be the collection of vectors such that $\langle u_{ia}, u_{jb}\rangle = \mathrm{Cov}(X_{ia}, X_{jb})$. Note that $\|u_{ia}\|^2 = \mathrm{Var}\,X_{ia}$. Let $v_{ia} = u_{ia} + \mathbb{E}\,X_{ia}\,v_\emptyset$, where $v_\emptyset$ is a unit vector orthogonal to all vectors u_{ia}. Define $v_i = \sum_a \|u_{ia}\|\bar{u}_{ia}^{\otimes 2} \otimes \bar{v}_{ia}^{\otimes 2}$. (Here, $\bar{x}$ denotes the unit vector in direction x.) Let us first lower bound the inner product of v_i and v_j,

$$\left(\sum_{a\in[k]} \left|\mathrm{Cov}(X_{ia}, X_{j\pi(a)})\right|\right)^4 = \left(\sum_a \|u_{ia}\|\|u_{j\pi(a)}\| |\langle \bar{u}_{ia}, \bar{u}_{j\pi(a)}\rangle|\right)^4$$

$$= \left(\sum_a \|u_{ia}\|\,\|u_{j\pi(a)}\| \cdot |\langle\bar{u}_{ia}, \bar{u}_{j\pi(a)}\rangle| \sqrt{\tfrac{\|u_{ia}\|\,\|u_{j\pi(a)}\|}{\|v_{ia}\|\,\|v_{j\pi(a)}\|}} \cdot \sqrt{\tfrac{\|v_{ia}\|\,\|v_{j\pi(a)}\|}{\|u_{ia}\|\,\|u_{j\pi(a)}\|}}\right)^4$$

$$\leqslant \left(\sum_a \|u_{ia}\|\,\|u_{j\pi(a)}\| \cdot \langle\bar{u}_{ia}, \bar{u}_{j\pi(a)}\rangle^2 \tfrac{\|u_{ia}\|\,\|u_{j\pi(a)}\|}{\|v_{ia}\|\,\|v_{j\pi(a)}\|}\right)^2$$

$$\cdot \left(\sum_a \|u_{ia}\|\,\|u_{j\pi(a)}\| \cdot \tfrac{\|v_{ia}\|\,\|v_{j\pi(a)}\|}{\|u_{ia}\|\,\|u_{j\pi(a)}\|}\right)^2 \quad \text{(using Cauchy–Schwarz)}$$

$$\leqslant \left(\sum_a \|u_{ia}\|\,\|u_{j\pi(a)}\| \cdot |\langle\bar{u}_{ia}, \bar{u}_{j\pi(a)}\rangle\langle\bar{v}_{ia}, \bar{v}_{j\pi(a)}\rangle|\right)^2$$

$$\text{(using } \langle v_{ia}, v_{j\pi(a)}\rangle \geqslant \langle u_{ia}, u_{j\pi(a)}\rangle \text{ and } \textstyle\sum_a \|v_{ia}\|\,\|v_{j\pi(a)}\| \leqslant 1)$$

$$\leqslant \sum_a \|u_{ia}\|\,\|u_{j\pi(a)}\| \cdot \langle\bar{u}_{ia}, \bar{u}_{j\pi(a)}\rangle^2 \langle\bar{v}_{ia}, \bar{v}_{j\pi(a)}\rangle^2$$

$$\text{(using Cauchy–Schwarz and } \sum_a \|u_{ia}\|\,\|u_{j\pi(a)}\| \leqslant 1)$$

$$\leqslant \langle v_i, v_j\rangle \,.$$

On the other hand, we can upper bound the inner product of v_i and v_j,

$$\langle v_i, v_j\rangle = \sum_{a,b} \|u_{ia}\|\,\|u_{jb}\| \cdot \langle\bar{u}_{ia}, \bar{u}_{jb}\rangle^2 \langle\bar{v}_{ia}, \bar{v}_{jb}\rangle^2$$

$$\leqslant \sum_{a,b} \tfrac{1}{2}(\|u_{ia}\|^2 + \|u_{jb}\|^2) \cdot \langle \bar{u}_{ia}, \bar{u}_{jb} \rangle^2$$

$$= \sum_{a,b} \tfrac{1}{2}(\tfrac{1}{\operatorname{Var} X_{ia}} + \tfrac{1}{\operatorname{Var} X_{jb}}) \operatorname{Cov}(X_{ia}, X_{jb}).$$

Finally, the vectors $v_1 \ldots, v_n$ are in the unit ball,

$$\|v_i\|^2 = \sum_{a,b} \|u_{ia}\| \, \|u_{ib}\| \cdot \langle \bar{u}_{ia}, \bar{u}_{ib} \rangle^2 \langle \bar{v}_{ia}, \bar{v}_{ib} \rangle^2 = \sum_a \|u_{ia}\|^2 \leqslant 1.$$

Here, we are using the fact that $\langle v_{ia}, v_{ib} \rangle = 0$ for all distinct $a, b \in [k]$. $\qquad\square$

C Facts about Variance

Lemma C.1. *Let X and Y be jointly-distributed random variables. Assume that Y has finite range. Let Z be the orthogonal projection of the random variable X onto the subspace of functions of the random variable Y. Then,*

$$\mathbb{E}_{\{Y\}} \operatorname{Var}[X \mid Y] = \mathbb{E}\, X^2 - \mathbb{E}\, Z^2.$$

Proof. By construction Z is a function $f(Y)$ of the random variable Y and $X - Z$ is orthogonal to all functions of the variable Y. Hence, $\mathbb{E}[X \mid Y = y] = f(y)$. Therefore, the expected variance of $[X \mid Y]$ is

$$\mathbb{E}_{\{Y\}} \operatorname{Var}[X \mid Y] = \mathbb{E}\, X^2 - \mathbb{E}_{\{Y\}} \left(\mathbb{E}[X \mid Y] \right)^2$$

$$= \mathbb{E}\, X^2 - \mathbb{E}_{\{Y\}} f(Y)$$

which gives the desired identity using $Z = f(Y)$. $\qquad\square$

Lemma C.2. *Let X and Y be as in the previous lemma. Suppose the range of Y has cardinality 2. Then,*

$$\mathbb{E}_{\{Y\}} \operatorname{Var}[X \mid Y] = \operatorname{Var} X - \operatorname{Cov}(X, Y)^2 / \operatorname{Var}(Y).$$

Proof. Without loss of generality, we may assume that $\mathbb{E}\, X = \mathbb{E}\, Y = 0$ and $\mathbb{E}\, Y^2 = 1$. Then, the set of random variables $\{1, Y\}$ is an orthonormal basis for the subspace of functions of Y. Let $\rho = \mathbb{E}\, XY$. Then, ρY is the orthogonal projection of X to the subspace of function of Y. (Here, we use the assumption $\mathbb{E}\, X = 0$.) Hence, using the previous lemma,

$$\mathbb{E}_{\{Y\}} \operatorname{Var}[X \mid Y] = \mathbb{E}\, X^2 - \mathbb{E}(\rho Y)^2 = \mathbb{E}\, X^2 - \rho^2,$$

which is the desired identity because $\mathbb{E}\, X^2 = \operatorname{Var} X$ and $\rho^2 = \operatorname{Cov}(X, Y)^2 / \operatorname{Var} Y$. $\qquad\square$