Lasserre Hierarchy, Higher Eigenvalues, and Approximation Schemes for Graph Partitioning and Quadratic Integer Programming with PSD Objectives

(Extended Abstract)

Venkatesan Guruswami
Computer Science Department
Carnegie Mellon University
guruswami@cmu.edu

Ali Kemal Sinop
Computer Science Department
Carnegie Mellon University
asinop@cs.cmu.edu

***Abstract*—We present an approximation scheme for optimizing certain Quadratic Integer Programming problems with positive semidefinite objective functions and global linear constraints. This framework includes well known graph problems such as Minimum graph bisection, Edge expansion, Uniform sparsest cut, and Small Set expansion, as well as the Unique Games problem. These problems are notorious for the existence of huge gaps between the known algorithmic results and NP-hardness results. Our algorithm is based on rounding semidefinite programs from the Lasserre hierarchy, and the analysis uses bounds for low-rank approximations of a matrix in Frobenius norm using columns of the matrix.**

For all the above graph problems, we give an algorithm running in time $n^{O(r/\varepsilon^2)}$ with approximation ratio $\frac{1+\varepsilon}{\min\{1,\lambda_r\}}$, where λ_r is the r'th smallest eigenvalue of the normalized graph Laplacian $\mathcal{L}$. In the case of graph bisection and small set expansion, the number of vertices in the cut is within lower-order terms of the stipulated bound. Our results imply $(1+O(\varepsilon))$ factor approximation in time $n^{O(r^*/\varepsilon^2)}$ where r^* is the number of eigenvalues of $\mathcal{L}$ smaller than $1-\varepsilon$. This perhaps gives some indication as to why even showing mere APX-hardness for these problems has been elusive, since the reduction must produce graphs with a slowly growing spectrum (and classes like planar graphs which are known to have such a spectral property often admit good algorithms owing to their nice structure).

For Unique Games, we give a factor $(1+\frac{2+\varepsilon}{\lambda_r})$ approximation for minimizing the number of unsatisfied constraints in $n^{O(r/\varepsilon)}$ time. This improves an earlier bound for solving Unique Games on expanders, and also shows that Lasserre SDPs are powerful enough to solve well-known integrality gap instances for the basic SDP.

We also give an algorithm for independent sets in graphs that performs well when the Laplacian does not have too many eigenvalues bigger than $1+o(1)$.

***Keywords*-Approximation algorithms; graph partitioning; unique games; semidefinite programming.**

I. Introduction

The theory of approximation algorithms has made major strides in the last two decades, pinning down, for many basic optimization problems, the exact (or asymptotic) threshold up to which efficient approximation is possible. Some notorious problems, however, have withstood this wave of progress; for these problems the best known algorithms deliver super-constant approximation ratios, whereas NP-hardness results do not even rule out say a factor 1.1 (or sometimes even a factor $(1+\varepsilon)$ for any constant $\varepsilon > 0$) approximation algorithm. Examples of such problems include graph partitioning problems such as minimum bisection, uniform sparsest cut, and small-set expansion; finding a dense subgraph induced on k vertices; minimum linear arrangement; and constraint satisfaction problems such as minimum CNF deletion or Unique Games.

There has been evidence of three distinct flavors for the hardness of these problems: (i) Ruling out a polynomial time approximation scheme (PTAS) assuming that $\mathsf{NP} \not\subseteq \bigcap_{\varepsilon>0} \mathsf{BPTIME}(2^{n^\varepsilon})$ via quasi-random PCPs [1], [2]; (ii) Inapproximability results within some constant factor assuming average-case hardness of refuting random 3SAT instances [3]; and (iii) Inapproximability within super-constant factors under a strong conjecture on the intractability of the small-set expansion (SSE) problem [4]. While (iii) gives the strongest hardness results, it is conditioned on the conjectured hardness of SSE [5], an assumption that implies the Unique Games conjecture, and arguably does not yet have as much evidence in its support as the complexity assumptions made in (i) or (ii).

In this work, we give a unified algorithm, based on powerful semidefinite programs from the Lasserre hierarchy, for several of these problems, and a broader class of quadratic integer programming problems with linear constraints (more details are in Section I-A below). Our algorithms deliver a good approximation ratio if the eigenvalues of the Laplacian of the underlying graph increase at a reasonable rate. In particular, for all the above graph partitioning problems, we get a $(1+\varepsilon)/\min\{\lambda_r, 1\}$ approximation factor in $n^{O_\varepsilon(r)}$ time, where λ_r is the r'th smallest eigenvalue of the normalized Laplacian (which has eigenvalues in the interval $[0,2]$). Note that if $\lambda_r \geqslant 1-\varepsilon$, then we get a $(1+O(\varepsilon))$ approximation ratio.

Research supported in part by a Packard Fellowship, NSF CCF 0963975, and US-Israel BSF grant 2008293.

Perspective. The direct algorithmic interpretation of our results is simply that one can probably get good approximations for graphs that are pretty "weak-expanders," in that we only require lower bounds on higher eigenvalues rather than on λ_2 as in the case of expanders. In terms of our broader understanding of the complexity of approximating these problems, our results perhaps point to why even showing APX-hardness for these problems has been difficult, as the reduction must produce graphs with a very slowly growing spectrum, with many ($n^{\Omega(1)}$, or even $n^{1-o(1)}$ for near-linear time reductions) small eigenvalues. Trivial examples of such graphs are the disjoint union of many small components (taking the union of r components ensures $\lambda_r = 0$), but these are of course easily handled by working on each component separately. We note that Laplacians of planar graphs, bounded genus graphs, and graphs excluding fixed minors, have many small eigenvalues [6], but these classes are often easier to handle algorithmically due to their rich structure — for example, conductance and edge expansion problems are polynomial time solvable on planar graphs [7]. Also, the recent result of [8] shows that if $\lambda_r = o(1)$ for some $r = n^{\Omega(1)}$, then the graph must have an $n^{1-\Omega(1)}$ sized subset with very few edges leaving it. Speculating somewhat boldly, may be these results suggest that graphs with too many small eigenvalues are also typically not hard instances for these problems. $\square$

Our results also give some explanation for our inability so far to show strong integrality gaps for even 4 rounds of the Lasserre hierarchy for problems which we only know to be hard assuming the Unique Games conjecture (UGC). In fact, it is entirely consistent with current knowledge that just $O(1)$ rounds of the Lasserre hierarchy could refute the UGC, or even give an improvement over the 0.878 performance ratio of the Goemans-Williamson algorithm for Max Cut. Complementing our algorithmic results for graph partitioning problems, it has been recently shown in [9] that even a linear number of rounds of the Lasserre SDP has a constant factor integrality gap for Balanced Separator and Uniform Sparsest Cut.

A. *Summary of results*

Let us now state our specific results. In the statements, we will use `OPT` to denote the optimal value of the respective optimization problem.

Graph partitioning. We begin with results for certain cut/graph partitioning problems. Below λ_p denotes the p'th smallest eigenvalue of the normalized Laplacian $\mathcal{L}$ of the graph G, defined as $\mathcal{L} = D^{-1/2}(I - A)D^{-1/2}$ where A is the adjacency matrix and D is a diagonal matrix with node degrees on the diagonal. (In the stated approximation ratios, λ_r (resp. $2 - \lambda_{n-r}$) should be understood as $\min\{\lambda_r, 1\}$ (resp. $\min\{2 - \lambda_{n-r}, 1\}$), but we don't make this explicit to avoid notational clutter.) The algorithm's running time is $n^{O_\varepsilon(r)}$ in each case. This runtime arises due to solving the standard semidefinite programs (SDP) lifted with $O(r/\varepsilon^2)$ rounds of the Lasserre hierarchy. Our results are shown via an efficient rounding algorithm whose randomized running time is $n^{O(1)}$; the exponential dependence on r is thus limited to solving the SDP. We first state our result for finding the minimum expanding set with given size/volume: here the goal is given a graph $G = (V, E)$ and a target size (resp. volume) μ, to find a subset $U \subset V$ with $|U| = \mu$ (resp. $\mathrm{Vol}(U)$, defined as the sum of degrees of vertices in U, being equal to μ) that minimizes $|E(U, \overline{U})|$, the number of edges leaving U. The result in fact works for edge-weighted graphs.

Theorem 1 (Graph Bisection and Small Set Expansion)**.** *Given $0 < \varepsilon < 1$, positive integer r, a target size (resp. target volue) μ, there exists an algorithm, running in time $n^{O_\varepsilon(r)}$ to find a set $U \subseteq V$ such that the total weight of edges it cuts is at most*

$$\frac{1+\varepsilon}{\lambda_r}\texttt{OPT}$$

and its size (resp. its volume) is within $1 \pm \varepsilon$ factor of μ.

We can also handle boundary conditions stipulating that U must contain some subset F of nodes, and avoid some other disjoint subset B of nodes.

As a corollary of Theorem 1, we obtain the following. The terminology used for the variants of graph partitioning is standard, and detailed definitions and problem-specific theorem statements can be found in the full version [10].

Corollary 2 (Graph Partitioning)**.** *Given $0 < \varepsilon < 1$, positive integer r, there exists a factor*

$$\frac{1+\varepsilon}{\lambda_r}$$

approximation algorithm, running in time $n^{O_\varepsilon(r)}$ for the following problems.

- *Uniform Sparsest Cut,*
- *Edge Expansion,*
- *Normalized Cut,*
- *Minimum Conductance,*
- *Their natural ℓ-way extensions where the vertices of the graph must be partitioned into ℓ parts.*

As in Theorem 1, we can also handle boundary conditions.

PSD Quadratic Integer Programs. In addition to the above cut problems, our method applies more abstractly to the class of minimization quadratic integer programs (QIP) with positive semidefinite (PSD) cost functions and arbitrary linear constraints. In particular, given a PSD matrix $A \in \mathbb{R}^{(V\times[k])\times(V\times[k])}$, consider the problem of finding $\widetilde{x} \in \{0,1\}^{V\times[k]}$ minimizing $\widetilde{x}^T A \widetilde{x}$ subject to: (i) exactly one of $\{\widetilde{x}_u(i)\}_{i\in[k]}$ equals 1 for each u, and (ii) the linear constraints $B\widetilde{x} \geqslant c$. We find such an $\widetilde{x}$ with

$$\widetilde{x}^T L \widetilde{x} \leqslant (1+\varepsilon)/\min\{1, \lambda_r(\mathcal{A})\}$$

where $\mathcal{A} = \mathrm{diag}(A)^{-1/2} \cdot A \cdot \mathrm{diag}(A)^{-1/2}$. For the formal statement, and the guarantee on the linear constraints, see the full version [10].

Unique Games. We next state our result for Unique Games.

Theorem 3. *Given any instance of Unique Games over a domain size k on a constraint graph G, for all $\varepsilon \in (0,1)$ and positive integer r, there exists an algorithm to find a labeling that violates at most*

$$\left(1 + \frac{2+\varepsilon}{\lambda_r(\mathcal{L})}\right) \mathtt{OPT}$$

fraction of constraints in time $(nk)^{O(r/\varepsilon)}$, where $\mathtt{OPT}$ is the total weight of unsatisfied constraints in the optimal labeling.

In the special case of maximum cut / minimum uncut, the bound can be improved to

$$\min\left(1 + \frac{2+\varepsilon}{\lambda_r(\mathcal{L})}, \frac{1+\varepsilon}{2-\lambda_{n-r}(\mathcal{L})}\right) \mathtt{OPT}.$$

Furthermore any ℓ-way section constraints (for e.g. maximum bisection) can be imposed, in which case all such constraints will be satisfied within a factor of $1 \pm o(1)$.

Note that in the case of Unique Games, we are only able to get a weaker $\approx 1 + 2/\lambda_r$ approximation factor, which is always larger than 2. In this context, it is interesting to note that minimizing the number of unsatisfied constraints in Unique Games *is* known to be APX-hard; for example, the known NP-hardness for approximating Max Cut [11], [12] implies a factor $(5/4 - \varepsilon)$ hardness for this problem (and indeed for the special case of Minimum Uncut).

Remark 1 (UG on expanders)**.** Arora *et al* [13] showed that Unique Games is easy on expanders, and gave an $O(\frac{\log(1/\mathtt{OPT})}{\lambda_2})$ approximation to the problem of minimizing the number of unsatisfied constraints, where $\mathtt{OPT}$ is the fraction of unsatisfied constraints in the optimal solution. For the subclass of "linear" Unique Games, they achieved an approximation ratio of $O(1/\lambda_2)$ without any dependence on $\mathtt{OPT}$. A factor $O(1/\lambda_2)$ approximation ratio was achieved for general Unique Games instances by Makarychev and Makarychev [14] (assuming λ_2 is large enough, they also get a $O(1/h_G)$ approximation where h_G is the Cheeger constant). Our result achieves an approximation factor of $O(1/\lambda_r)$, if one is allowed $n^{O(r)}$ time.

For instances of ΓMAX2LIN, the paper [13] also gives an $n^{O(r)}$ time algorithm that satisfies all but a fraction $O(\mathtt{OPT}/z_r(G))$ of constraints, where $z_r(G)$ is the value of the r-round Lasserre SDP relaxation of Sparsest Cut on G. For $r = 1$, $z_1(G) = \lambda_2$. But the growth rate of $z_r(G)$, eg. its relation to the Laplacian spectrum, was not known. □

Remark 2 (SDP gap instances)**.** Our algorithm also shows that the Khot-Vishnoi UG gap instance for the basic SDP [15], has $O(1)$ integrality gap for the lifted SDP corresponding to $\mathrm{poly}(\log n)$ rounds of Lasserre hierarchy. In particular, these instances admit quasi-polynomial time constant factor approximations. This latter result was earlier shown by Kolla [16] using spectral techniques. Our result shows that strong enough SDPs also suffice to tackle these instances. □

Remark 3 (Sub-exponential algorithm for Unique Games)**.** In a similar vein to the above remark, applying the ABS graph decomposition [8] to split the graph into components with at most n^{ε} small eigenvalues while cutting very few edges, one also gets that $n^{\varepsilon^{\Omega(1)}}$ rounds of the Lasserre hierarchy suffice to well-approximate Unique Games on instances with at most ε fraction unsatisfied constraints. □

Independent Set in graphs. We also give a rounding algorithm for the natural Lasserre SDP for independent set in graphs.

Theorem 4. *Given $0 < \varepsilon < 1$, positive integer r, a graph G with $d_{\max} \geqslant 3$, there exists an algorithm to find an independent set $I \subseteq V$ such that*

1) *If $\lambda_{n-r} > 1 + \frac{1}{4d_{\max}}$, then $|I| \geqslant \frac{(1-\varepsilon)\cdot\mathtt{OPT}}{2d_{\max}} \frac{2-\lambda_{n-r}(\mathcal{L})}{\lambda_{n-r}(\mathcal{L})-1}$,*
2) *Else $|I| = \mathtt{OPT}$,*

in time $n^{O\left(\frac{r}{\varepsilon^2}\right)}$.

Note that above theorem implies that on any sparse graph, $O(r)$-round Lasserre relaxation of the independent set problem is *integral* where r is the number of eigenvalues of G larger than $1 + \frac{1}{4d_{\max}}$.

For reasons of space, in this extended abstract we only sketch our results for Minimum Graph Bisection (the result of Theorem 1 for the case of finding a non-expanding set of *size* $\mu \pm o(\mu)$) and Unique Games. This should give a flavor of the key ideas behind our methods. The detailed proofs of all the results can be found in the full version of this paper [10].

B. Our Techniques

Our results follow a unified approach, based on a SDP relaxation of the underlying integer program. The SDP is chosen from the Lasserre hierarchy [17], and its solution has vectors $x_T(\sigma)$ corresponding to local assignments to every subset $T \subset V$ of at most r' vertices. (Such an SDP is said to belong to r' *rounds* of the Lasserre hierarchy.) The vectors satisfy dot product constraints corresponding to consistency of pairs of these local assignments.

Given an optimal solution to the Lasserre SDP, we give a rounding method based on local propagation, similar to the rounding algorithm for Unique Games on expanders in [13]. We first find an appropriate subset S of r' nodes (called the *seed* nodes). One could simply try all such subsets in $n^{r'}$ time, though there is an $n^{O(1)}$ time algorithm to locate the set S as well. Then for each assignment f to nodes in S, we randomly extend the assignment to all nodes by assigning, for each $u \in V \setminus S$ independently, a random value from u's marginal distribution based on $x_{S\cup\{u\}}$ conditioned on the assignment f to S.

After arithmetizing the performance of the rounding algorithm, and making a simple but crucial observation that lets us pass from higher order Lasserre vectors to vectors corresponding to single vertices, the core step in the analysis is the following: Given vectors $\{X_v \in \mathbb{R}^\Upsilon\}_{v\in V}$ and an upper bound on a positive semidefinite (PSD) quadratic form $\sum_{u,v\in V} L_{uv}\langle X_u, X_v\rangle = \mathsf{Tr}(X^T X L) \leqslant \eta$, place an upper bound on the sum of the squared distance of X_u from the span of $\{X_s\}_{s\in S}$, i.e., the quantity $\sum_u \|X_S^\perp X_u\|^2 = \mathsf{Tr}(X^T X_S^\perp X)$. (Here $X \in \mathbb{R}^{\Upsilon\times V}$ is the matrix with columns $\{X_v : v \in V\}$.)

We relate the above question to the problem of *column-selection* for low-rank approximations to a matrix, studied in many recent works [18], [19], [20], [21]. It is known by the recent works [20], [21][1] that one can pick r/ε columns S such that $\mathsf{Tr}(X^T X_S^\perp X)$ is at most $1/(1-\varepsilon)$ times the error of the best rank-r approximation to X in Frobenius norm, which equals $\sum_{i>r}\sigma_i$ where the σ_i's are the eigenvalues of $X^T X$ in decreasing order. Combining this with the upper bound $\mathsf{Tr}(X^T X L) \leqslant \eta$, we deduce an approximation ratio of $\left(1 + \frac{1+\varepsilon}{\lambda_{r+1}}\right)$ for our algorithm. Also, the independent rounding of each u implies, by standard Chernoff-bounds, that any linear constraint (such as a balance condition) is met up to lower order deviations.

Note that the above gives an approximation ratio $\approx 1 + 1/\lambda_r$, which always exceeds 2. To get our improved $(1 + \varepsilon)/\lambda_r$ guarantee, we need a more refined analysis, based on iterated application of column selection along with some other ideas.

For Unique Games, a direct application of our framework for quadratic IPs would require relating the spectrum of the constraint graph G of the Unique Games instance to that of the lifted graph $\widehat{G}$. There are such results known for random lifts, for instance [22]; saying something in the case of arbitrary lifts, however, seems very difficult.[2] We therefore resort to an indirect approach, based on *embedding* the set of k vectors $\{x_u(i)\}_{i\in[k]}$ for a vertex into a single vector X_u with some nice distance preserving properties that enables us to relate quadratic forms on the lifted graph to a proxy form on the base constraint graph. This idea was also used in [13] for the analysis of their algorithm on expanders, where they used an embedding based on non-linear tensoring. In our case, we need the embedding to also preserve distances from certain higher-dimensional subspaces (in addition to preserving pairwise distances); this favors an embedding that is as "linear" as possible, which we obtain by passing to a tensor product space.

C. Related work on Lasserre SDPs in approximation

The Lasserre SDPs seem very powerful, and as mentioned earlier, for problems shown to be hard assuming the UGC (such as beating Goemans-Williamson for Max Cut), integrality gaps are not known even for a small constant number of rounds. A gap instance for Unique Games is known if the Lasserre constraints are only *approximately* satisfied [23]. It is interesting to contrast this with our positive result. The error needed in the constraints for the construction in [23] is $r/(\log\log n)^c$ for some $c < 1$, where n is the number of vertices and r the number of rounds. Our analysis requires the Lasserre consistency constraints are met exactly. In the full version [10], we present an algorithm that produces such valid Lasserre SDP solutions in time $(kn)^{O(r)}O(\log(1/\varepsilon_0))$ with an additive error of ε_0 in linear constraints, and an objective value at most ε_0 more than optimal.

Strong Lasserre integrality gaps have been constructed for certain approximation problems that *are* known to be NP-hard. Schoenebeck proved a strong negative result that even $\Omega(n)$ rounds of the Lasserre hierarchy has an integrality gap ≈ 2 for Max 3-LIN [24]. Via reductions from this result, Tulsiani showed gap instances for Max k-CSP (for $\Omega(n)$ rounds), and instances with $n^{1-o(1)}$ gap for $\approx 2^{\sqrt{\log n}}$ rounds for the Independent Set and Chromatic Numbers [25].

In terms of algorithmic results, even few rounds of Lasserre is already as strong as the SDPs used to obtain the best known approximation algorithms for several problems — for example, 3 rounds of Lasserre is enough to capture the ARV SDP relaxation for Sparsest Cut [26], and Chlamtac used the third level of the Lasserre hierarchy to get improvements for coloring 3-colorable graphs [27]. In terms of positive results that use a larger (growing) number of Lasserre rounds, we are aware of only two results. Chlamtac and Singh used $O(1/\gamma^2)$ rounds of Lasserre hierarchy to find an independent set of size $\Omega(n^{\gamma^2/8})$ in 3-uniform hypergraphs with an independent set of size γn [28]. Karlin, Mathieu, and Nguyen show that $1/\varepsilon$ rounds of Lasserre SDP gives a $(1+\varepsilon)$ approximation to the Knapsack problem [29].

However, there are *mixed* hierarchies, which are weaker than Lasserre and based on combining an LP characterized by local distributions (from the Sherali-Adams hierarchy) with a simple SDP, that have been used for several approximation algorithms. For instance, for the above-mentioned result on independent sets in 3-uniform hypergraphs, an $n^{\Omega(\gamma^2)}$ sized independent set can be found with $O(1/\gamma^2)$

[1]In fact our work [21] was motivated by the analysis in this paper.

[2]It is known that $\lambda_{r\cdot kn^\delta}(\mathcal{L}(\widehat{G})) \geqslant \delta\lambda_r(\mathcal{L}(G))$ [8], but this large multiplicative n^δ slack makes this ineffective for $r = n^{o(1)}$.

levels from the mixed hierarchy. Raghavendra's result states that for every constraint satisfaction problem, assuming the Unique Games conjecture, the best approximation ratio is achieved by a small number of levels from the mixed hierarchy [30]. For further information and references on the use of SDP and LP hierarchies in approximation algorithms, we point the reader to the excellent book chapter [31].

In an independent work, Barak, Raghavendra, and Steurer [32] also extend the local propagation rounding of [13] to Lasserre SDPs Their analysis methods are rather different from ours. Instead of column-based low-rank matrix approximation, they use the graph spectrum to infer global correlation amongst the SDP vectors from local correlation, and use it to iteratively to argue that a random seed set works well in the rounding. Their main result is an *additive approximation for Max 2-CSPs*. Translating to the terminology used in this paper, given a 2CSP instance over domain size k with optimal value (fraction of satisfied constraints) equal to v, they give an algorithm to find an assignment with value $v - O\left(k\sqrt{1-\lambda_r}\right)$ based on $r' \gg kr$ rounds of the mixed hierarchy. (Here λ_r is the r'th smallest eigenvalue of the normalized Laplacian of the *constraint* graph; note though that λ_r needs to be fairly close to 1 for the bound to kick in.) For the special case of Unique Games, they get the better performance of $v - O\left(\sqrt[4]{1-\lambda_r}\right)$ which doesn't degrade with k. They also get a factor $O(1/\lambda_r)$ approximation for minimizing the number of unsatisfied constraints. Compared to our $(1+(2+\varepsilon)/\lambda_r)$ approximation, for this result they use the weaker "mixed" Sherali-Adams SDP which enables them to achieve a runtime for solving the SDP that has a $2^{O(r)}$ type dependence on the number of rounds r instead of our $n^{O(r)}$ bound. However, their runtime has an *exponential* dependence on the number of labels k.

For 2CSPs, our results only apply to a restricted class (corresponding to PSD quadratic forms)[3], but we get approximation-scheme style *multiplicative* guarantees for the harder *minimization objective*, and can handle *global linear constraints*. (Also, for Unique Games, as mentioned above our algorithm has running time *polynomial* in the number of labels k, but a worse dependence on r.) Our approach enables us to get approximation-scheme style guarantees for several notorious graph partitioning problems that have eluded even APX-hardness.

II. Lasserre hierarchy of semidefinite programs

We present the formal definitions of the Lasserre family of SDP relaxations [17], tailored to the setting of the problems we are interested in, where the goal is to assign to each vertex/variable from a set V a label from $[k] = \{1, 2, \ldots, k\}$.

[3]Though it we are willing to tolerate some form of additive approximation, one can apply the result to any QIP after adding diagonal terms to the quadratic form's matrix to make it PSD.

Definition 5 (Lasserre vector set)**.** *Given a set of variables V and a set $[k] = \{1, 2, \ldots, k\}$ of labels, and an integer $r \geqslant 0$, a vector set x is said to satisfy r-levels of Lasserre hierarchy constraints on k labels, denoted*

$$x \in \mathbf{Lasserre}^{(r)}(V \times [k]) \ ,$$

if it satisfies the following conditions:

1) *For each set $S \in \binom{V}{\leqslant r+1}$, there exists a function $x_S : [k]^S \to \mathbb{R}^{\Upsilon}$ that associates a vector of some finite dimension Υ with each possible labeling of S. We use $x_S(f)$ to denote the vector associated with the labeling $f \in [k]^S$. For singletons $u \in V$, we will use $x_u(i)$ and $x_u(i^u)$ for $i \in [k]$ interchangeably.*
   For $f \in [k]^S$ and $v \in S$, we use $f(v)$ as the label v receives from f. Also given sets S with labeling $f \in [k]^S$ and T with labeling $g \in [k]^T$ such that f and g agree on $S \cap T$, we use $f \circ g$ to denote the labeling of $S \cup T$ consistent with f and g: If $u \in S$, $(f \circ g)(u) = f(u)$ and vice versa.
2) $\|x_\emptyset\|^2 = 1$.
3) $\langle x_S(f), x_T(g)\rangle = 0$ *if there exists $u \in S \cap T$ such that $f(u) \neq g(u)$.*
4) $\langle x_S(f), x_T(g)\rangle = \langle x_A(f'), x_B(g')\rangle$ *if $S \cup T = A \cup B$ and $f \circ g = f' \circ g'$.*
5) *For any $u \in V$, $\sum_{j \in [k]} \|x_u(j)\|^2 = \|x_\emptyset\|^2$.*
6) *(implied by above constraints) For any $S \in \binom{V}{\leqslant r+1}$, $u \in S$ and $f \in [k]^{S \setminus \{u\}}$, $\sum_{g \in [k]^u} x_S(f \circ g) = x_{S \setminus \{u\}}(f)$.*

We will use $\mathcal{X}(i)$ to denote a matrix of size $\Upsilon \times n$, $\mathcal{X}(i) \in \mathbb{R}^{\Upsilon \times V}$ whose columns are the vectors $\{x_u(i)\}_{u \in V}$.

We now add linear constraints to the SDP formulation.

Definition 6 (Linear constraints in Lasserre SDPs)**.** *Given a matrix $B = [b_1 \ \ldots \ b_\ell] \in \mathbb{R}^{(V \times [k]) \times \ell}$ and a vector $c = (c_1, \ldots, c_\ell)^T \in \mathbb{R}^\ell$, $x \in \mathbf{Lasserre}^{(r)}(V \times [k])$, is said to satisfy linear constraints $\{(b_i, c_i)\}_{i=1}^{\ell}$ if the following holds for all $i \in [\ell]$:*

For all subsets $S \in \binom{V}{\leqslant r}$ and $f \in [k]^V$,

$$\sum_{u \in V, g \in [k]^u} \langle x_S(f), x_u(g)\rangle b_i(u, g) \leqslant c_i \langle x_S(f), x_\emptyset\rangle \ ,$$

which is equivalent to

$$\sum_{u \in V, g \in [k]^u} \|x_{S \cup \{u\}}(f \circ g)\|^2 b_i(u, g) \leqslant c_i \|x_S(f)\|^2 .$$

We denote the set of such x as $x \in \mathbf{Lasserre}^{(r)}(V \times [k], B^{\leqslant c})$.

Remark 4 (Convenient matrix notation)**.** One common expression we will use throughout this paper is the following.

For matrices $X \in \mathbb{R}^{\Upsilon \times V}$ and $M \in \mathbb{R}^{V \times V}$:

$$\mathsf{Tr}(X^T X M) = \sum_{u,v \in V} M_{u,v} \langle X_u, X_v \rangle \ .$$

Note that if M is positive semidefinite (denoted $M \succeq 0$), then $\mathsf{Tr}(X^T X M) \geqslant 0$.[4]

Also, if L is Laplacian matrix of an undirected graph $G = (V, E)$, we have

$$\mathsf{Tr}(X^T X L) = \sum_{e=\{u,v\} \in E} \|X_u - X_v\|^2$$

where X_u denotes the column of X corresponding to $u \in V$. $\square$

The analysis of our rounding algorithm will involve projections on certain subspaces, which we define next.

Definition 7 (Projection operators)**.** *Given $x \in \mathbf{Lasserre}^{(r)}(V \times [k])$, we define $\Pi : \binom{V}{\leqslant r+1} \to \mathbb{R}^{\Upsilon \times \Upsilon}$ as the projection matrix onto the span of $\{x_S(f)\}_{f \in [k]^S}$ for given S:*

$$\Pi_S \triangleq \sum_{f \in [k]^S} \overline{x_S(f)} \cdot \overline{x_S(f)}^T \ .$$

(Here $\overline{x_S(f)}$ is the unit vector in the direction of $x_S(f)$ if $x_S(f)$ is nonzero, and 0 otherwise.)

Similarly we define $P : \binom{V}{\leqslant r+1} \to \mathbb{R}^{\Upsilon \times \Upsilon}$ as the matrix corresponding to projection onto the span of $\{x_v(f)\}_{v \in S, f \in [k]}$: $P_S \triangleq \sum_{v \in S, f \in [k]} \overline{x_v(f)} \cdot \overline{x_v(f)}^T$.

We will denote by $\Pi_S^\perp = I - \Pi_S$ and $P_S^\perp = I - P_S$ the projection matrices onto the respective orthogonal complements, where I denotes the identity matrix of appropriate dimension.

III. Case Study: Approximating Minimum Bisection

All our algorithmic results follow a unified method (except small set expansion on irregular graphs and unique games, both of which we treat separately). In this section, we will illustrate the main ideas involved in our work in a simplified setting, by working out progressively better approximation ratios for the following basic, well-studied problem: Given as input a graph $G = (V, E)$ and an integer size parameter μ, find a subset $U \subset V$ with $|U| = \mu$ that minimizes the number of edges between U and $V \setminus U$, denoted $\Gamma_G(U)$. The special case when $\mu = |V|/2$ and we want to partition the vertex set into two equal parts is the minimum bisection problem. We will loosely refer to the general μ case also as minimum bisection.[5]

[4] The use of this inequality in various places is the reason why our analysis only works for minimizing PSD quadratic forms.

[5] We will be interested in finding a set of size $\mu \pm o(\mu)$, so we avoid the terminology *Balanced Separator* which typically refers to the variant where $\Omega(n)$ slack is allowed in the set size.

For simplicity we will assume G is unweighted and d-regular, however all our results hold for *any weighted undirected graph*. We can formulate this problem as a binary integer programming problem as follows:

$$\min_{\widetilde{x} \in \{0,1\}^{V \times [2]}} \sum_{e=\{u,v\} \in E} (\widetilde{x}_u(1) - \widetilde{x}_v(1))^2, \tag{1}$$

$$\text{subject to} \sum_u \widetilde{x}_u(1) = \mu; \ \forall u, \widetilde{x}_u(1) + \widetilde{x}_u(2) = 1; \tag{2}$$

If we let L be the Laplacian matrix for G, we can rewrite the objective as $\eta \triangleq \widetilde{x}(1)^T L \widetilde{x}(1)$. We will denote by $\mathcal{L} = \frac{1}{d} L$ the normalized Laplacian of G.

Note that the above is a quadratic integer programming (QIP) problem with linear constraints. The somewhat peculiar formulation is in anticipation of the Lasserre semidefinite programming relaxation for this problem, which we describe below.

A. Lasserre relaxation for Minimum Bisection

Let b be the vector on $V \times [2]$ with $b_v(1) = 1$ and $b_v(2) = 0$ for every $v \in V$. For an integer $r' \geqslant 0$, the r'-round Lasserre SDP for Minimum Bisection consists of finding $x \in \mathbf{Lasserre}^{(r')}(V \times [k], b^{=\mu})$ that minimizes the objective function

$$\sum_{e=\{u,v\} \in E(G)} \|x_u(1) - x_v(1)\|^2 \ . \tag{3}$$

It is easy to see that this is indeed a relaxation of our original QIP formulation (1).

B. Main theorem on rounding

Let x be an (optimal) solution to the above r'-round Lasserre SDP. We will always use η in this section to refer to the objective value of x, i.e., $\eta = \sum_{e=\{u,v\} \in E(G)} \|x_u(1) - x_v(1)\|^2$.

Our ultimate goal in this section is to give an algorithm to round the SDP solution x to a good cut U of size very close to μ, and prove the below theorem.

Theorem 8. *For all $r \geqslant 1$ and $\varepsilon > 0$, there exists $r' = O\left(\frac{r}{\varepsilon^2}\right)$, such that given $x \in \mathbf{Lasserre}^{(r')}(V \times [k], b^{=\mu})$ with objective value (3) equal to η, one can find in randomized $n^{O(1)}$ time, a set $U \subseteq V$ satisfying the following two properties w.h.p:*

1) $\Gamma_G(U) \leqslant \frac{1+\varepsilon}{\min\{1, \lambda_{r+1}(\mathcal{L})\}} \eta$.
2) $\mu(1 - o(1)) = \mu - O\left(\sqrt{\mu \log(1/\varepsilon)}\right) \leqslant |U| \leqslant \mu + O\left(\sqrt{\mu \log(1/\varepsilon)}\right) = \mu(1 + o(1))$.

Since one can solve the Lasserre relaxation in $n^{O(r')}$ time we get the result claimed in the introduction: an $n^{O(r/\varepsilon^2)}$ time factor $(1+\varepsilon)/\min\{\lambda_r, 1\}$ approximation algorithm; the formal theorem, for general (non-regular, weighted) graphs appears in the full version. Note that if $t = \arg\min_r\{r \mid$

$\lambda_r(\mathcal{L}) \geqslant 1-\varepsilon/2\}$, then this gives an $n^{O_\varepsilon(t)}$ time algorithm for approximating minimum bisection to within a $(1+\varepsilon)$ factor, provided we allow $O(\sqrt{n})$ imbalance.

C. The rounding algorithm

Recall that the solution $x \in \mathbf{Lasserre}^{(r')}(V \times [k], b^{=\mu})$ contains a vector $x_T(f)$ for each $T \in \binom{V}{\leqslant r'}$ and every possible labeling of T, $f \in [2]^T$ of T. Our approach to round x to a solution $\widetilde{x}$ to the integer program (1) is similar to the label propagation approach used in [13].

Consider fixing a set of r' nodes, $S \in \binom{V}{r'}$, and assigning a label $f(s)$ to every $s \in S$ by choosing $f \in [2]^S$ with probability $\|x_S(f)\|^2$. (The best choice of S can be found by brute-forcing over all of $\binom{V}{r'}$, since solving the Lasserre SDP takes $n^{O(r')}$ time anyway. But there is also a faster method to find a good S, as mentioned in Theorem 11.) Conditional on choosing a specific labeling f to S, we propagate the labeling to other nodes as follows: Independently for each $u \in V$, choose $i \in [2]$ and assign $\widetilde{x}_u(i) \leftarrow 1$ with probability

$$\mathsf{Pr}\left[\widetilde{x}_u(i)=1\right] = \frac{\|x_{S\cup\{u\}}(f \circ i^u)\|^2}{\|x_S(f)\|^2} = \frac{\langle \overline{x_S(f)}, x_u(i)\rangle}{\|x_S(f)\|}.$$

Observe that if $u \in S$, label of u will always be $f(u)$. Finally, output $U = \{u \mid \widetilde{x}_u(1)=1\}$ as the cut. Below Π_S denotes the projection matrix from Definition 7.

Lemma 9. *For the above rounding procedure, the size of the cut produced $\Gamma_G(U)$ satisfies*

$$\mathbb{E}\left[\Gamma_G(U)\right] = \eta + \sum_{(u,v)\in E} \langle \Pi_S^{\perp} x_u(1), \Pi_S^{\perp} x_v(1)\rangle \ . \qquad (4)$$

Proof: Note that for $u \neq v$, and $i, j \in [2]$,

$$\begin{aligned}\mathsf{Pr}\left[\widetilde{x}_u(i)=1 \wedge \widetilde{x}_v(j)=1\right] \\ &= \sum_f \|x_S(f)\|^2 \frac{\langle \overline{x_S(f)}, x_u(i)\rangle}{\|x_S(f)\|} \frac{\langle \overline{x_S(f)}, x_v(j)\rangle}{\|x_S(f)\|} \\ &= \sum_f \langle \overline{x_S(f)}, x_u(i)\rangle \langle \overline{x_S(f)}, x_v(j)\rangle.\end{aligned}$$

Since $\{\overline{x_S(f)}\}_f$ is an orthonormal basis, the above expression can be written as the inner product of *projections* of $x_u(i)$ and $x_v(j)$ onto the span of $\{x_S(f)\}_{f\in[2]^S}$, which we denote by Π_S. Let us now calculate the expected number $\Gamma_G(U)$ of edges cut by this rounding. It is slightly more convenient to treat edges $e = \{u, v\}$ as two directed edges (u, v) and (v, u), and count directed edges (u, v) with $u \in U$ and $v \in V \setminus U$ in the cut. Therefore,

$$\mathbb{E}\left[\Gamma_G(U)\right] = \sum_{(u,v)\in E} \langle \Pi_S x_u(1), \Pi_S x_v(2)\rangle$$

which is equal to $\eta + \sum_{(u,v)\in E} \langle \Pi_S^{\perp} x_u(1), \Pi_S^{\perp} x_v(1)\rangle$. ■

Note that the matrix Π_S depends on vectors $x_S(f)$ which are hard to control because we do not have any constraint relating $x_S(f)$ to a known matrix. The main driving force behind all our results is the following fact, which follows since given any $u \in S$ and $i \in [2]$, $x_u(i) = \sum_{f: f(u)=i} x_S(f)$ by Lasserre constraints.

Observation 10. *For all $S \in \binom{V}{r'}$,*

$$\operatorname{span}\left(\{x_S(f)\}_{f\in[2]^S}\right) \supseteq \operatorname{span}\left(\{x_u(i)\}_{u\in S, i\in[2]}\right) \ .$$

Equivalently for P_S being the projection matrix onto span of $\{x_u(i)\}_{u\in S,i\in[2]}$, $P_S \preceq \Pi_S$.

Thus we will try to upper bound the term in (4) by replacing $\Pi_S^{\perp}$ with $P_S^{\perp}$, but we cannot directly perform this switch: $\langle P_S^{\perp} x_u(i), P_S^{\perp} x_v(j)\rangle$ might be negative while $\Pi_S^{\perp} x_u(i) = 0$.

D. Factor $1 + \frac{1}{\lambda_r}$ approximation of cut value

Our first bound is by directly upper bounding (4) in terms of $\|\Pi_S^{\perp} x_u(i)\|^2 \leqslant \|P_S^{\perp} x_u(i)\|^2$. Using Cauchy-Schwarz and Arithmetic-Geometric Mean inequalities, (4) implies that the expected number of edges cut is upper bounded by

$$\eta + d \sum_u \|\Pi_S^{\perp} x_u(1)\|^2 \leqslant \eta + d \sum_u \|P_S^{\perp} x_u(1)\|^2 \ .$$

Now define $X_u \triangleq x_u(1)$, and let $X \in \mathbb{R}^{\Upsilon \times V}$ be the matrix with columns X_u. By (3), we have the objective value $\eta = \mathsf{Tr}(X^T X L)$. Let $X_S^{\Pi} \triangleq \sum_{u\in S} \overline{X_u X_u}^T$ be the projection matrix onto the span of $\{X_u\}_{u\in S}$. Since this set is a subset of $\{x_u(i)\}_{u\in S, i\in[2]}$, we have $X_S^{\Pi} \preceq P_S$. Therefore, we can bound (5) further as

$$\mathbb{E}\left[\text{number of edges cut}\right] \leqslant \eta + d \sum_u \|X_S^{\perp} X_u\|^2. \qquad (5)$$

To get the best upper bound, we want to pick $S \in \binom{V}{r'}$ to minimize $\sum_{u\in V} \|X_S^{\perp} X_u\|^2$. It is a well known fact that among *all* projection matrices M of rank r' (not necessarily restricted to projection onto columns of X), the minimum value of $\sum_u \|M^{\perp} X_u\|^2 = \mathsf{Tr}(X^T M^{\perp} X)$ is achieved by matrix M projecting onto the space of the largest r' singular vectors of X. Further, this minimum value equals $\sum_{i\geqslant r'+1} \sigma_i$ where $\sigma_i = \sigma_i(X)$ denotes the *squared* i^{th} largest singular value of X (equivalently $\sigma_i(X)$ is the i^{th} largest eigenvalue of $X^T X$). Hence $\mathsf{Tr}(X^T X_S^{\perp} X) \geqslant \sum_{i\geqslant r'+1} \sigma_i$ for every choice of S. The following theorem from [21] shows the existence of S which comes close to this lower bound:

Theorem 11. *[21] For every real matrix X with column set V, and positive integers $r \leqslant r'$, we have*

$$\delta_{r'}(X) \triangleq \min_{S\in\binom{V}{r'}} \mathsf{Tr}(X^T X_S^{\perp} X) \leqslant \frac{r'+1}{r'-r+1}\Big(\sum_{i\geqslant r+1} \sigma_i\Big) \ .$$

In particular, for all $\varepsilon \in (0,1)$, $\delta_{r/\varepsilon} \leqslant \frac{1}{1-\varepsilon}\left(\sum_{i\geqslant r+1}\sigma_i\right)$. Further one can find a set $S \in \binom{V}{r'}$ achieving the claimed bounds in deterministic $O(rn^4)$ time.

Remark 5. Prior to our paper [21], it was shown in [20] that $\delta_{r(2+\varepsilon)/\varepsilon} \leqslant (1+\varepsilon)\left(\sum_{i\geqslant r+1}\sigma_i\right)$. The improvement in the bound on r' from $2r/\varepsilon$ to r/ε to achieve $(1+\varepsilon)$ approximation is not of major significance to our application, but since the tight bound is now available, we decided to state and use it. □

Remark 6 (Running time of our algorithms)**.** If the Lasserre SDP can be solved faster than $n^{O(r')}$ time, perhaps in $\exp(O(r'))n^c$ time for some absolute constant c, then the fact that we can find S deterministically in only $O(n^5)$ time would lead to a similar runtime for the overall algorithm. □

Picking the subset $\mathcal{S}^* \in \binom{V}{r'}$ that achieves the bound (5)) guaranteed by Theorem 11, we have

$$\mathsf{Tr}(X^T X^{\perp}_{\mathcal{S}^*} X) = \delta_{\frac{r}{\varepsilon}}(X) \leqslant (1-\varepsilon)^{-1}\sum_{i>r}\sigma_i \ .$$

In order to relate this quantity to the SDP objective value $\eta = \mathsf{Tr}(X^T XL)$, we use the fact that $\mathsf{Tr}(X^T XL)$ is minimized when eigenvectors of $X^T X$ and L are matched in reverse order: i^{th} largest eigenvector of $X^T X$ corresponds to i^{th} smallest eigenvector of L. Letting $0 = \lambda_1(\mathcal{L}) \leqslant \lambda_2(\mathcal{L}) \leqslant \ldots \leqslant \lambda_n(\mathcal{L}) \leqslant 2$ be the eigenvalues of normalized graph Laplacian matrix, $\mathcal{L} = \frac{1}{d}L$, we have

$$\begin{aligned}\frac{\eta}{d} &= \frac{1}{d}\mathsf{Tr}(X^T XL) \geqslant \sum_i \sigma_i(X)\lambda_i(\mathcal{L}) \\ &\geqslant \sum_{i\geqslant r+1}\sigma_i(X)\lambda_{r+1}(\mathcal{L}) \geqslant (1-\varepsilon)\lambda_{r+1}(\mathcal{L})\delta_{\frac{r}{\varepsilon}}(X).\end{aligned}$$

Plugging this into (5), we can conclude our first bound:

Theorem 12. *For all positive integers r and $\varepsilon \in (0,1)$, given SDP solution $x \in \mathbf{Lasserre}^{(\lceil r/\varepsilon\rceil)}(V\times[k], b^{=\mu})$, the rounding algorithm given in Section III-C cuts at most*

$$\left(1+\frac{1}{(1-\varepsilon)\lambda_{r+1}(\mathcal{L})}\right)\sum_{e=(u,v)\in E}\|x_u(1)-x_v(1)\|^2$$

edges in expectation. In particular, the algorithm cuts at most a factor $\left(1+\frac{1}{(1-\varepsilon)\lambda_{r+1}(\mathcal{L})}\right)$ more edges than the SDP objective value of the solution x. [6]

Note that $\lambda_n(\mathcal{L}) \leqslant 2$, hence even if we use n-rounds of Lasserre relaxation, for which x is an integral solution, we can only show an upper bound $\geqslant \frac{3}{2}$. Although this is too weak by itself for our purposes, this bound will be crucial to obtain our final bound.

[6] We will later argue that the cut will also meet the balance requirement up to $o(\mu)$ vertices.

E. Improved analysis and factor $\frac{1}{\lambda_r}$ approximation on cut value

First notice that (4) can be written as

$$\mathbb{E}\,[\text{number of edges cut}] = \mathsf{Tr}(X^T\Pi^{\perp}_S X)+\mathsf{Tr}(X^T\Pi_S XL)\ . \tag{6}$$

If value of this expression is larger than $\frac{\eta}{(1-\varepsilon)\lambda_{r+1}} + \eta\varepsilon$, then value of $\mathsf{Tr}(X^T\Pi_S XL)$ has to be larger than $\varepsilon\eta$ due to the bound we proved on $\mathsf{Tr}(X^T\Pi^{\perp}_S X)$. Consider choosing another subset T that achieves the bound $\delta_r(\Pi^{\perp}_S X)$. The crucial observation is that distances between neighboring nodes on vectors $\Pi^{\perp}_S X$ has decreased by an additive factor of $\eta\varepsilon$,

$$\mathsf{Tr}(X^T\Pi^{\perp}_S XL) = \mathsf{Tr}(X^T XL) - \mathsf{Tr}(X^T\Pi_S XL) < \eta(1-\varepsilon)$$

so that $\mathsf{Tr}(X^T\Pi^{\perp}_{S\cup T}X) < (1-\varepsilon)\frac{\eta}{(1-\varepsilon)\lambda_{r+1}}$. Now, if we run the rounding algorithm with $S\cup T$ as the seed set, and (6) with $S\cup T$ in place of S is larger than $\frac{\eta}{(1-\varepsilon)\lambda_{r+1}}+\eta\varepsilon$, then $\mathsf{Tr}(X^T\Pi_{S\cup T}XL) > 2\varepsilon\eta$. Hence

$$\begin{aligned}\mathsf{Tr}(X^T\Pi^{\perp}_{S\cup T}XL) &\leqslant \mathsf{Tr}(X^T XL) - \mathsf{Tr}(X^T\Pi_{S\cup T}XL) \\ &< \eta(1-2\varepsilon)\ .\end{aligned} \tag{7}$$

Picking another set T', we will have $\mathsf{Tr}(X^T\Pi^{\perp}_{S\cup T\cup T'}X) < (1-2\varepsilon)\frac{\eta}{(1-\varepsilon)\lambda_{r+1}}$. Continuing this process, if the quantity (6) is not upper bounded by $\frac{\eta}{(1-\varepsilon)\lambda_{r+1}}+\eta\varepsilon$ after $\lceil\frac{1}{\varepsilon}\rceil$ many such iterations, then the total projection distance becomes

$$\mathsf{Tr}(X^T\Pi^{\perp}_{S\cup T\cup\ldots}X) < (1-\lceil 1/\varepsilon\rceil\varepsilon)\frac{\eta}{(1-\varepsilon)\lambda_{r+1}} \leqslant 0$$

which is a contradiction. For formal statement and proof in a more general setting, see full version.

Theorem 13. *For all integers $r \geqslant 1$ and $\varepsilon \in (0,1)$, for an integer $r' = O\left(\frac{r}{\varepsilon^2}\right)$, the following holds. Given an optimal SDP solution $x \in \mathbf{Lasserre}^{(r')}(V\times[k], b^{=\mu})$, the expected number of edges cut by the solution obtained by running Algorithm 1 on the seed set returned by Algorithm 2 is at most $(1+\varepsilon)/\min\{1,\lambda_{r+1}(\mathcal{L})\}$ times the size of the optimal cut with μ nodes on one side. (Here $\lambda_{r+1}(\mathcal{L})$ is the $(r+1)$'th smallest eigenvalue of the normalized Laplacian $\mathcal{L} = \frac{1}{d}L$ of the G.)*

F. Bounding Set Size

We now analyze the balance of the cut, and show that we can ensure that $|U| = \mu \pm o(\mu)$ in addition to $\Gamma_G(U)$ being close to the expected bound of Theorem 13 (and similarly for Theorem 12).

Let $\mathcal{S}^*$ fixed to be $\arg\min_{S\in\binom{V}{r'}}\mathsf{Tr}(X^T X^{\perp}_S X)$. We will show that conditioned on finding cuts with small $\Gamma_G(U)$, the probability that one of them has $|U| \approx \mu$ is bounded away from zero. We can use a simple Markov bound to show that there is a non-zero probability that both cut size and set size are within 3-factor of corresponding bounds. But by exploiting the independence in our rounding algorithm and Lasserre relaxations of linear constraints, we can do much better. Note that in the r'-round Lasserre relaxation, for each $f \in [2]^{\mathcal{S}^*}$, due to the set size constraint in original IP formulation, x satisfies:

$$\sum_u \widetilde{x}_u(1) = \mu \implies \sum_u \langle x_{\mathcal{S}^*}(f), x_u(1) \rangle = \mu \|x_{\mathcal{S}^*}(f)\|^2 \ .$$

This implies that conditioned on the choice of f, the expectation of $\sum_u \widetilde{x}_u(1)$ is μ and events $\widetilde{x}_u(1) = 1$ for various u are independent. Applying the Chernoff bound, we get

$$\Pr_{\widetilde{x}} \left[\left| \sum_u \widetilde{x}_u(1) - \mu \right| \geqslant 2\sqrt{\mu \log \frac{1}{\zeta}} \right] \leqslant o(\zeta) \leqslant \frac{\zeta}{3}.$$

Consider choosing $f \in [2]^{\mathcal{S}^*}$ so that $\mathbb{E}\big[\text{number of edges cut} \,\big|\, f\big] \leqslant \mathbb{E}\big[\text{number of edges cut}\big] \triangleq b$. By Markov inequality, if we pick such an f, $\Pr\big[\text{number of edges cut} \geqslant (1+\zeta)b\big] \leqslant 1 - \frac{\zeta}{2}$, where the probability is over the random propagation once $\mathcal{S}^*$ and f are fixed.

Hence with probability at least $\frac{\zeta}{6}$, the solution $\widetilde{x}$ will yield a cut U with $\Gamma_G(U) \leqslant (1+\zeta)b$ and size $|U|$ in the range $\mu \pm 2\sqrt{\mu \log \frac{1}{\zeta}}$. Taking $\zeta = \varepsilon$ and repeating this procedure $O\left(\varepsilon^{-1} \log n\right)$ times, we get a high probability statement and finish our main Theorem 8 on minimum bisection.

Algorithm 1 Algorithm for labeling in time $O\left(k^{r'} + n\right)$.

Input: $\mathcal{S}^* \subseteq V$ of size at most r', $x \in \mathbf{Lasserre}^{(r')}(V \times [k])$.
Output: $\widetilde{x} \in \{0,1\}^{V \times [k]}$.
Procedure:

1) Choose $f \in [k]^{\mathcal{S}^*}$ with probability $\|x_{\mathcal{S}^*}(f)\|^2$.
2) Label every node $u \in V$ by choosing a label $j \in [k]$ with probability $\frac{\langle x_{x_{\mathcal{S}^*}}(f), x_u(j) \rangle}{\|x_{\mathcal{S}^*}(f)\|^2}$.

IV. Proof Sketch for Theorem 3 (Unique Games)

In this section, we will give a sketch of Theorem 3's proof. Given a Unique Games instance with constraint graph $G = (V,E)$, label set $[k]$, and bijective constraints $\pi_e : [k] \to [k]$ for each edge $e \in E$, we consider the following QIP formulation:

$$\min_{\widetilde{x}} \sum_{e=\{u,v\} \in E} w_e \cdot \frac{1}{2} \sum_{i \in [k]} (\widetilde{x}_u(i) - \widetilde{x}_v(\pi_e(i)))^2,$$

$$\text{subject to} \sum_{i \in [k]} \widetilde{x}_u(i) = 1 \quad \forall u \in V,$$

$$\widetilde{x} \in \{0,1\}^{V \times [k]} \ .$$

Algorithm 2 Algorithm for finding seed set in time $O(n^5)$ deterministically.

Input: Positive integers r, $r' = \frac{r}{\varepsilon^2}$, $x \in \mathbf{Lasserre}^{(r')}(V \times [k])$ and a PSD matrix $L \in \mathbb{R}^{(V \times [k]) \times (V \times [k])}$.
Output: Seed set $\mathcal{S}^* \subseteq V$ of size at most r' satisfying Theorem 13.
Procedure:

1) Let $\mathcal{S}^* \leftarrow \emptyset$.
2) Repeat $1/\varepsilon$ times:
   a) Find new $\frac{r}{\varepsilon}$-many seeds $\widetilde{T} \in \binom{V \times [k]}{r/\varepsilon}$ using deterministic column selection algorithm given in [21] on matrix $\mathrm{diag}(L)^{1/2} \Pi^{\perp}_{\mathcal{S}^*} \mathcal{X}$.
   b) $T \leftarrow \left\{ u \middle| \exists j \in [k] : \ (u,j) \in \widetilde{T} \right\}$.
   c) $\mathcal{S}^* \leftarrow \mathcal{S}^* \bigcup T$.

Let $x \in \mathbf{Lasserre}^{(r')}(V \times [k])$ be the vector solution (for the Lasserre SDP relaxation of the above QIP) satisfying $r' = O(r/\varepsilon)$-levels of Lasserre hierarchy constraints with objective value equal to:

$$\eta \triangleq \frac{1}{2} \sum_{e=\{u,v\} \in E} \sum_f \|x_u(f) - x_v(\pi_e(f))\|^2 \ .$$

For convenience we will assume below that G is unweighted and d-regular. We stress though that our results do not require these assumptions, and the proof in the full version [10] works with general instances.

A straightforward analysis of the rounding procedure Algorithm 1 yields the following bound on number of unsatisfied constraints,

$$\eta + d \sum_u \sum_f \|P_S^{\perp} x_u(f)\|^2, \tag{8}$$

where P_S denotes the projection matrix onto span of vectors $\{x_s(g)\}_{s \in S, g}$ for seed set S chosen using Algorithm 2.

As mentioned in the introduction, bounding this quantity requires the knowledge of lifted graph, $\widehat{G}$. Instead we use an embedding of vectors on $\widehat{G}$ to G. Our embedding is as follows. Assume that the vectors $x_u(f)$ belong to $\mathbb{R}^m$. Let $e_1, e_2, \ldots, e_m \in \mathbb{R}^m$ be the standard basis vectors. Define $X_u \in \mathbb{R}^m \otimes \mathbb{R}^m$ as

$$X_u = \sum_{i=1}^m \sum_{f \in [k]^u} \langle \overline{x_u(f)}, e_i \rangle x_u(f) \otimes e_i \ . \tag{9}$$

For proof of the following theorem, see the full version [10].

Theorem 14 (A useful embedding)**.** *Given vectors $x \in \mathbb{R}^{m \times (V \times [k])}$ with the property that, for any $u \in V$, whenever $f, g \in [k]^u$ are two different labellings of u, $f \neq g$,*

$$\langle x_u(f), x_u(g) \rangle = 0.$$

Then the embedding given in (9) *satisfies the following properties:*

1) *For any* $u \in V$, $\|X_u\|^2 = \sum_f \|x_u(f)\|^2$.
2) *For any* $u, v \in V$ *and any permutation* $\pi \in \mathrm{Sym}([k])$*:*

$$\sum_{i \in [k]} \|x_u(i^u) - x_v(\pi(i)^v)\|^2 \geqslant \frac{1}{2}\|X_u - X_v\|^2.$$

3) *For any set* $S \subseteq V$ *and any node* $u \in V$*, if we let* P_S *be the projection matrix onto the span of* $\{x_s(f)\}_{s \in S, f \in [k]}$*:*

$$\|X_S^{\perp} X_u\|^2 \geqslant \sum_{f \in [k]^u} \|P_S^{\perp} x_u(f)\|^2.$$

With this embedding, we can easily bound (8):

$$d \sum_u \sum_f \|P_S^{\perp} x_u(f)\|^2 \leqslant d \sum_u \|X_S^{\perp} X_u\|^2$$
$$\leqslant \frac{1+\varepsilon}{\lambda_{r+1}} \sum_{e=(u,v) \in E} \|X_u - X_v\|^2.$$

Finally using the fact that $\sum_{e=(u,v)\in E} \|X_u - X_v\|^2$ is bounded by $2\sum_{e=(u,v)\in E}\sum_f \|x_u(f) - x_v(\pi_e(f))\|^2 = 2\eta$, we obtain the bound given in Theorem 3.

Acknowledgments

We thank Sivaraman Balakrishnan, Ravishankar Krishnaswamy and Srivatsan Narayanan for useful discussions.

References

[1] S. Khot, "Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique," *SIAM J. Comput.*, vol. 36, no. 4, pp. 1025–1071, 2006.

[2] C. Ambühl, M. Mastrolilli, and O. Svensson, "Inapproximability results for sparsest cut, optimal linear arrangement, and precedence constrained scheduling," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, 2007, pp. 329–337.

[3] U. Feige, "Relations between average case complexity and approximation complexity," in *STOC*, 2002, pp. 534–543.

[4] P. Raghavendra, D. Steurer, and M. Tulsiani, "Reductions between expansion problems," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 17, p. 172, 2010.

[5] P. Raghavendra and D. Steurer, "Graph expansion and the unique games conjecture," in *Proceedings of the 42nd ACM Symposium on Theory of Computing*, 2010, pp. 755–764.

[6] J. A. Kelner, J. R. Lee, G. N. Price, and S.-H. Teng, "Metric uniformization and spectral bounds for graphs," *CoRR*, vol. abs/1008.3594, 2010, prelim. version in FOCS'09.

[7] J. K. Park and C. A. Phillips, "Finding minimum-quotient cuts in planar graphs," in *STOC*, 1993, pp. 766–775.

[8] S. Arora, B. Barak, and D. Steurer, "Subexponential algorithms for Unique Games and related problems," in *Proceedings of the 51st IEEE Symposium on Foundations of Computer Science*, 2010, pp. 563–572.

[9] V. Guruswami, A. K. Sinop, and Y. Zhou, "Constant factor lasserre gaps for graph partitioning problems," *Manuscript*, July 2011.

[10] V. Guruswami and A. K. Sinop, "Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with PSD objectives," *arXiv:cs.DS/1104.4746*, 2011.

[11] J. Håstad, "Some optimal inapproximability results," *Journal of the ACM*, vol. 48, no. 4, pp. 798–859, 2001.

[12] L. Trevisan, G. B. Sorkin, M. Sudan, and D. P. Williamson, "Gadgets, approximation, and linear programming," *SIAM J. Comput.*, vol. 29, no. 6, pp. 2074–2097, 2000.

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

[14] K. Makarychev and Y. Makarychev, "How to play unique games on expanders," in *WAOA*, ser. Lecture Notes in Computer Science, K. Jansen and R. Solis-Oba, Eds., vol. 6534. Springer, 2010, pp. 190–200.

[15] S. Khot and N. K. Vishnoi, "The Unique Games conjecture, integrality gap for cut problems and embeddability of negative type metrics into l_1," in *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, 2005, pp. 53–62.

[16] A. Kolla, "Spectral algorithms for Unique Games," in *Proceedings of the 25th IEEE Conference on Computational Complexity*, 2010, pp. 122–130.

[17] J. B. Lasserre, "An explicit equivalent positive semidefinite program for nonlinear 0-1 programs," *SIAM Journal on Optimization*, vol. 12, no. 3, pp. 756–769, 2002.

[18] A. Deshpande and S. Vempala, "Adaptive sampling and fast low-rank matrix approximation," in *APPROX-RANDOM*, 2006, pp. 292–303.

[19] A. Deshpande and L. Rademacher, "Efficient volume sampling for row/column subset selection," in *FOCS*, 2010, pp. 329–338.

[20] C. Boutsidis, P. Drineas, and M. Magdon-Ismail, "Near-optimal column-based matrix reconstruction," *CoRR*, vol. abs/1103.0995, 2011.

[21] V. Guruswami and A. K. Sinop, "Optimal column-based low-rank matrix reconstruction," *arXiv:cs.DS/1104.1732*, April 2011.

[22] R. Oliveira, "The spectrum of random k-lifts of large graphs (with possibly large k)," *Journal of Combinatorics*, vol. 1, no. 3-4, pp. 285–306, 2010.

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

[24] G. Schoenebeck, "Linear level Lasserre lower bounds for certain k-CSPs," in *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, 2008, pp. 593–602.

[25] M. Tulsiani, "CSP gaps and reductions in the Lasserre hierarchy," in *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, 2009, pp. 303–312.

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

[27] E. Chlamtac, "Approximation algorithms using hierarchies of semidefinite programming relaxations," in *48th Annual IEEE Symposium on Foundations of Computer Science*, 2007, pp. 691–701.

[28] E. Chlamtac and G. Singh, "Improved approximation guarantees through higher levels of sdp hierarchies," in *APPROX-RANDOM*, 2008, pp. 49–62.

[29] A. R. Karlin, C. Mathieu, and C. T. Nguyen, "Integrality gaps of linear and semi-definite programming relaxations for knapsack," *CoRR*, vol. abs/1007.1283, 2010.

[30] P. Raghavendra, "Optimal algorithms and inapproximability results for every CSP?" in *Proceedings of the 40th ACM Symposium on Theory of Computing*, 2008, pp. 245–254.

[31] E. Chlamtac and M. Tulsiani, "Convex relaxations and integrality gaps," in *Handbook on Semidefinite, Cone and Polynomial Optimization*. Springer, 2011, to appear.

[32] B. Barak, P. Raghavendra, and D. Steurer, "Rounding semidefinite programming hierarchies via global correlation," *arXiv:cs.DS/1104.4680*, April 2011.