Rounding Lasserre SDPs using column selection and spectrum-based approximation schemes for graph partitioning and Quadratic IPs

Venkatesan Guruswami [*]
guruswami@cmu.edu
Computer Science Department,
Carnegie Mellon University,
Pittsburgh, PA.

Ali Kemal Sinop [†]
asinop@cs.cmu.edu
School of Mathematics,
Institute for Advanced Study,
Princeton, NJ.

December 9, 2013

Abstract

We present an approximation scheme for minimizing 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, 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 is the number of eigenvalues of $\mathcal{L}$ smaller than $1 - \varepsilon$ (for variants of sparsest cut, $\lambda_{r^*} \geq \mathrm{OPT}/\epsilon$ also suffices, and as OPT is usually $o(1)$ on interesting instances of these problems, this requirement on r^* is typically weaker). 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, improving upon an earlier bound for solving Unique Games on expanders. We also give an algorithm for independent sets in graphs that performs well when the Laplacian does not have too many eigenvalues bigger than $1 + o(1)$.

1   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. Prominent 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; ordering problems such as feedback arc set and 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 $\mathsf{NP} \not\subset \bigcap_{\varepsilon>0} \mathsf{BPTIME}(2^{n^\varepsilon})$ via quasi-random PCPs [Kho06, AMS11]; (ii) Inapproximability results within some constant factor assuming average-case hardness of refuting random 3SAT instances [Fei02]; and (iii) Inapproximability within super-constant factors under a strong conjecture on the intractability of the small-set expansion (SSE) problem [RST12]. While (iii) gives the strongest hardness results, it is conditioned on the conjectured hardness of SSE [RS10], 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).

Hierarchies of convex relaxations In many cases, including all constraint satisfaction problems and various graph partitioning problems, the best algorithms are based on fairly simple semi-definite programming (SDP) relaxations. The UGC foretells that for these problems, no tighter relaxation than these simple SDPs will yield a better approximation ratio in the worst-case. A natural question thus is to understand the power and limitations of potentially stronger SDP relaxations, for example those from various *hierarchies* of relaxations. These hierarchies are parameterized by an integer r (called rounds/levels) which capture higher order correlations between (roughly r-tuples of) variables (the basic SDP captures only pairwise correlations, and certain extensions like triangle inequalities pose constraints on triples). Larger the r, tighter the relaxation. The optimum of n'th level of the hierarchy, where n is the number of variables in the underlying integer program, usually equals the integral optimum.

There are several hierarchies of relaxations that have been studied in the literature, such as Sherali-Adams hierarchy of linear programs [SA90], the Lovász-Schrijver hierarchy [LS91], a "mixed" hierarchy combining Sherali-Adams linear programs with the base level SDP, and the Lasserre hierarchy [Las02] (see [CT11] for a recent survey focusing on their use in approximate combinatorial optimization). Of these hierarchies, the most powerful one is the Lasserre hierarchy (see [Lau03] for a comparison), and therefore holds the most potential for new breakthroughs in approximation algorithms. Arguably, Lasserre SDPs pose the currently strongest known threat to the Unique Games conjecture, as even the possibility of the 4'th level of Lasserre SDP relaxation improving upon the Goemans-Williamson 0.878 approximation factor for Max Cut has *not* been ruled out. Recently, it has also been shown that $O(1)$ rounds of the Lasserre hierarchy are able to solve all candidate gap instances of Unique Games [BBH$^+$12]. (On the other hand, for some of the weaker hierarchies, integrality gaps for super-constant rounds *are* known for various Unique-Games hard problems [KS09, RS09].)

In light of the above, the power and limitations of the Lasserre hierarchy merit further investigation. There has been a fair bit of recent interest in Lasserre hierarchy based approximation

algorithms [CS08, KMN11, GS11, BRS11, RT12, AG11, GS13].

In this work, we give an algorithmic framework, based on rounding semidefinite programs from the Lasserre hierarchy via column-based low-rank approximation, for several of these problems, as well as a broader class of quadratic integer programming problems with linear constraints (more details are in Section 1.1 below). Our algorithms deliver a good approximation ratio when the eigenvalues of the Laplacian matrix of the underlying graph increase at a reasonable rate. For example, in the case of constraint satisfaction type 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 $\lambda_r \geq 1 - \varepsilon$, then we get a $(1 + O(\varepsilon))$ approximation ratio.

Remark 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 [KLPT11], 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 [PP93]. Also, the recent result of [ABS10] 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.

1.1 Summary of results

Let us now state our specific results informally.

Arity 2 **constraint satisfaction problems (2-CSPs)** We begin with results for arity two constraint satisfaction type problems on graphs. For simplicity, we state the results for unweighted graphs — the body of the paper handles weighted graphs. Below λ_p denotes the p'th smallest eigenvalue of the normalized Laplacian $\mathcal{L}$ of the graph G, defined as $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$ 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 $O_\varepsilon(r)$ rounds of the Lasserre hierarchy. Our results are shown via an efficient rounding algorithm whose runtime is $n^{O(1)}$; the exponential dependence on r is thus limited to solving the SDP.

- MAXIMUM CUT AND MINIMUM UNCUT: Given a graph G on n vertices with a partition leaving at most b many edges uncut, we can find a partition that leaves at most $\frac{1+\varepsilon}{2-\lambda_{n-r}}b$ many edges uncut. (We can also get an approximation guarantee of $(1 + \frac{2+\varepsilon}{\lambda_r})$ for Minimum Uncut as a special case of our result for Unique Games.)

- MINIMUM (MAXIMUM) BISECTION: Given a graph G on n vertices with a bisection (partition into two equal parts) cutting (uncutting) at most b edges, we can find a near-bisection, with each side having $\frac{n}{2} \pm O_\varepsilon(\sqrt{n})$ vertices, that cuts at most $\frac{1+\varepsilon}{\lambda_r}b$ ("uncuts" at most $\frac{1+\varepsilon}{2-\lambda_{n-r}}b$) edges respectively.

- UNIQUE GAMES: Given a Unique Games instance with constraint graph $G = (V, E)$, label set $[k]$, and bijective constraints π_e for each edge, if the optimum assignment $\sigma : V \to [k]$ fails to satisfy η of the constraints, we can find an assignment that fails to satisfy at most $\eta \left(1 + \frac{2+\varepsilon}{\lambda_r} \right)$ of the constraints. Note that we work with the minimization version of Unique Games, which makes our approximation guarantees stronger.

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 [Hås01, TSSW00] implies a factor $(5/4 - \varepsilon)$ hardness for this problem (and indeed for the special case of Minimum Uncut).

PSD Quadratic Integer Programs In addition to the above 2-CSP 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.

- QIP WITH PSD COSTS: Given a PSD matrix $A \in \mathbb{R}^{(V \times [k]) \times (V \times [k])}$, consider the problem of finding $\mathbf{x} \in \{0, 1\}^{V \times [k]}$ minimizing $\mathbf{x}^T A \mathbf{x}$ subject to: (i) exactly one of $\{\mathbf{x}_u(i)\}_{i \in [k]}$ equals 1 for each u, and (ii) the linear constraints $B\mathbf{x} \geq c$. We find such an $\mathbf{x}$ with $\mathbf{x}^T L \mathbf{x} \leq \frac{1+\varepsilon}{\min\{1, \lambda_r(\mathcal{A})\}}$ where $\mathcal{A} = \text{diag}(A)^{-1/2} \cdot A \cdot \text{diag}(A)^{-1/2}$.

Independent Set in Graphs We give a rounding algorithm for the natural Lasserre SDP for independent set in graphs. Here, our result gives an algorithm running in $n^{O(r/\varepsilon^2)}$ time algorithm that finds an independent set of size $\approx \frac{n}{2d_{\max}} \frac{1}{\lambda_{n-r}-1}$ where $\lambda_{n-r} > 1$ is the r'th largest eigenvalue of the graph's normalized Laplacian. Thus even exact independent set is easy for graphs for which the number of eigenvalues greater than $\approx 1 + \frac{1}{4d_{\max}}$ is small.

Graph partitioning Finally we consider various variants of the sparsest cut problem on graphs, including those with constraints on the size of the parts. Our guarantees for these problems are stronger: In order to obtain $(1 + \varepsilon)$-factor approximation, we only require $\lambda_r \geq \frac{1}{\varepsilon}\mathsf{OPT}$ (in these problems, usually $\mathsf{OPT} \leq o(1)$) as opposed to $\lambda_r \geq 1 - \varepsilon$.

- NON-UNIFORM SPARSEST CUT: Given graphs G (capacity graph) and H (demand graph), find a subset of nodes, U, which minimize the ratio of capacities cut and demands cut by U. For λ_r being the rth generalized eigenvalue for the Laplacian matrices of graphs G and H, provided that $\lambda_r \geq (1 + 1/\kappa)\mathsf{OPT}$, our algorithm finds a solution whose cost is at most

$$(1 + \kappa)\mathsf{OPT}$$

  for any $\kappa > 0$.

- For the choice of a normalized clique as our demand graph H, we can handle various graph partitioning problems which involve minimizing ratio of cut cost with **size** of partitions. For each problem below, we can find a non-empty set $U \subsetneq V$, whose value is at most $(1+\kappa)$ OPT when $\lambda_r \geq (1 + 1/\kappa)\mathsf{OPT}$ (here λ_r is the rth smallest eigenvalue of Laplacian matrix).

4

- UNIFORM SPARSEST CUT: Minimize the number of edges in the cut $(U, V \setminus U)$ divided by $|U||V \setminus U|$.
- EDGE EXPANSION: Minimize the number of edges leaving U divided by the number of nodes in U, where U is the smaller side of the cut.

We also have similar guarantees for the problem of BALANCED SEPARATOR.

- For the choice of a weighted clique as our demand graph H, where the demand between u and v is the geometric mean of their degrees, we can handle various graph partitioning problems which involve minimizing ratio of cut cost with **volume** of partitions. For each problem below, we can find a non-empty set $U \subsetneq V$, whose value is at most $(1 + \kappa) \operatorname{OPT}$ when $\lambda_r \geq (1 + 1/\kappa)\mathsf{OPT}$ (here λ_r is the rth smallest eigenvalue of **normalized** Laplacian matrix).

  - NORMALIZED CUT: The number of edges in the cut $(U, V \setminus U)$ divided by the product of the volumes of U and $V \setminus U$.
  - CONDUCTANCE: The fraction of edges incident on U that leave U where U is the side of the cut with smaller volume.

We also have similar guarantees for the problem of SMALL SET EXPANSION.

Faster Implementation Using Propagation Rounding Framework. All the algorithms presented in this paper fall under the propagation rounding framework as introduced in [GS12a]. Therefore they can be implemented in time $2^{O_\varepsilon(r)} n^{O_\varepsilon(1)}$ as opposed to $n^{O_\varepsilon(r)}$. Throughout the paper, after each formal theorem statement concerning the approximation performance of our algorithms, we remark on the runtime that can be obtained using our faster solver.

Before discussing our techniques, we make some remarks on a few related guarantees for these problems.

Remark 1.1 (Subspace enumeration)**.** We note that for conductance (and related problems with quotient objectives mentioned above), it is possible to get a $O(\kappa)$ approximation in $n^{O(r)}$ time (provided that $\lambda_r \geq \Omega(\mathsf{OPT}/\kappa)$) by searching for a good cut in the r-dimensional eigenspace corresponding to the r smallest eigenvalues as in [ABS10] and then running the cut improvement algorithm from [AL08]. It is not clear, however, if such methods can give a $1 + \kappa$ factor approximation. Further, this method does not apply in the case of non-uniform sparsest cut. □

Remark 1.2 (UG on expanders)**.** [AKK$^+$08] showed that Unique Games is easy on expanders, and gave an $O(\frac{\log(1/\mathsf{OPT})}{\lambda_2})$ approximation to the problem of minimizing the number of unsatisfied constraints, where 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 OPT. A factor $O(1/\lambda_2)$ approximation ratio was achieved for general Unique Games instances by [MM10] (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, [AKK$^+$08] also give an $n^{O(r)}$ time algorithm that satisfies all but a fraction $O(\mathsf{OPT}/z_r(G))$ 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 1.3 (Unique Games via ABS Graph Decomposition)**.** We can use the graph decomposition of [ABS10] to split the graph into components with at most n^ε small eigenvalues while cutting very few edges, and show 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.

Remark 1.4 (SDP gap instances)**.** Our algorithm also shows that the Khot-Vishnoi UG gap instance for the basic SDP [KV05] 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 is already known and was shown by [Kol10] using spectral techniques. Our result shows that strong enough SDPs also suffice to tackle these instances. As mentioned earlier, it has been shown that $O(1)$ rounds suffice for almost all known strong integrality gaps [BBH$^+$12, OZ13, DMN13]. However these proofs are non-algorithmic – they construct a dual certificate. An interesting open problem is whether these claims can be established algorithmically by a rounding algorithm. □

1.2 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 [Las02], 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. (See Section 2.1 for a formal description.)

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 [AKK$^+$08]. 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 $O(n^5)$ 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 $u \in V \setminus S$ (either independently or by choosing a global threshold), 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 $\sum_{u,v \in V} L_{uv}\langle X_u, X_v\rangle = \mathrm{Tr}(X^T X L) \le \eta$, 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 $\sum_u \|X_S^\perp X_u\|^2 = \mathrm{Tr}(X^T X_S^\perp X)$. (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 [DV06, DR10, BDMI11, GS12b]. It is known by the recent works [BDMI11, GS12b][1] that one can pick r/ε columns S such that $\mathrm{Tr}(X^T X_S^\perp X)$ 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 $\mathrm{Tr}(X^T X L) \le \eta$, we deduce the desired approximation ratio for our algorithm.

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 [Oli10]; saying something

---

[1]In fact our work [GS12b] was motivated by the analysis in this paper.

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 Y_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 by [AKK[+]08] 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.

1.3   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 [KPS10]. It is interesting to contrast this with our positive result. The error needed in the constraints for the construction in [KPS10] 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.

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 [Sch08]. 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 [Tul09].

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 [ARV09], and Chlamtac used the third level of the Lasserre hierarchy to get improvements for coloring 3-colorable graphs [Chl07]. 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 [CS08]. Karlin, Mathieu, and Nguyen show that $1/\varepsilon$ rounds of Lasserre SDP gives a $(1 + \varepsilon)$ approximation to the Knapsack problem [KMN11].

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 $O(1/\gamma^2)$ 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 [Rag08]. 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 [CT11].

In an independent work, Barak, Raghavendra, and Steurer [BRS11] consider the above-mentioned mixed hierarchy, and extend the local propagation rounding of [AKK[+]08] to these SDPs in a manner similar to our work. 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

---

[2]It is known that $\lambda_{r \cdot k n^\delta}(\mathcal{L}(\widehat{G})) \geq \delta \lambda_r(\mathcal{L}(G))$ [ABS10], but this large multiplicative n^δ slack makes this ineffective for $r = n^{o(1)}$.

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, and also a factor $O(1/\lambda_r)$ approximation for minimizing the number of unsatisfied constraints in time *exponential* in k.

1.4 Differences with Conference Versions

This paper is a combination of two conference papers [GS11, GS13] with some additions:

- We unify and generalize the seed selection theorems in Corollary 5.3.

- We present approximation algorithms with guarantees similar to uniform sparsest cut for balanced separator and similar problems in Section 7.2 (these work when λ_r exceeds OPT by a constant factor).

- We present an analysis of the partial coloring algorithm for 3-colorable graphs of [AG11] and approximating general 2-CSPs of [BRS11] using our framework in Appendix A.

- As mentioned earlier, we also state the improved $2^{O_\varepsilon(r)} n^{O_\varepsilon(1)}$ type runtimes obtained via combination with the faster solver in [GS12a].

1.5 Organization

We first introduce the basic notations and definitions (including Lasserre SDPs) in Section 2. To illustrate the main ideas in our work, we present them in a self-contained way for a simplified setting in Section 3, by developing an algorithm for the Minimum Bisection problem. In Section 4, we describe the two rounding schemes we use and analyze them. In Section 5, we relate their performance to the column based matrix reconstruction problem and prove the main technical theorems for analysis.

The rest of our results are proved in remaining Sections: quadratic integer programming in Section 6.1; maximum cut and unique games in Section 6.2; independent set in Section 6.3; and non-uniform sparsest cut along with related problems in Section 7.

In Appendices A.1 and A.2, we show how to analyze two different rounding algorithms, the approximation algorithm for general 2-CSPs from [BRS11] and the partial coloring algorithm for 3-colorable graphs from [AG11] using our column selection framework.

2 Preliminaries

We now formally define the notation and terminology that will be useful to us in the paper.

Sets Given set A and positive integer k, we use $\binom{A}{k}$ (resp. $\binom{A}{\leq k}$) to denote the set of all possible size k (resp. size at most k) subsets of A. We use $\mathbb{R}_+$ to denote the set of non-negative reals.

Euclidean Space Given row set B, we use $\mathbb{R}^B$ to denote the set of real vectors where each row (axis) is associated with an element of B. For any vector $\mathrm{X} \in \mathbb{R}^B$, its coordinate at axis $b \in B$ is denoted by $\mathrm{X}(b)$. Let $\|\mathrm{X}\|_p$ be its pth norm with $\|\mathrm{X}\| \stackrel{\text{def}}{=} \|\mathrm{X}\|_2$, and X^T be its transpose. Finally for any $\mathrm{X}, \mathrm{Y} \in \mathbb{R}^B$, let $\langle \mathrm{X}, \mathrm{Y} \rangle = \mathrm{X}^T \mathrm{Y}$ be their inner product $\sum_{b \in B} \mathrm{X}_b \mathrm{Y}_b$.

Matrices Given row set R and column set C, we use $\mathbb{R}^{R,C}$ [3] to denote the set of real matrices whose rows and columns are associated with elements of R and C, respectively. Given matrix $X \in \mathbb{R}^{R,C}$, for any $r \in R, c \in C$, we will use $X_{r,c} \in \mathbb{R}$ to denote entry of X at row r and column c. For convenience, we use $X_c \in \mathbb{R}^R$ to denote the vector corresponding to the column c of X. Likewise given subset of columns of X, $S \subseteq C$, we use $X_S \in \mathbb{R}^{R,S}$ to denote the matrix corresponding to the columns S of X. Given matrix X, we use $\|X\|_F$, $\mathrm{Tr}(X)$ and X^T to denote Frobenius norm of X, its trace and transpose. Finally we use X^{Π} and $\mathrm{X}^{\perp}$ to denote the projection matrices onto the column span of X, $\mathrm{span}(\mathrm{X})$, and its orthogonal complement. For a matrix A, we denote by $\mathrm{null}(A)$ its right nullspace, i.e., the set of vectors x for which $Ax = 0$.

We will use $\mathbb{S}^C$ and $\mathbb{S}_+^C$ to denote the set of symmetric and positive semidefinite matrices, respectively.

Matrix Width For any matrix $A \in \mathbb{R}^{B,C}$, we will refer to the maximum number of non-zero entries among rows of A as the **width** of A.

Positive Semidefinite (PSD) Ordering Given a symmetric matrix $\mathrm{X} \in \mathbb{S}^A$, we say X is a PSD matrix (i.e. $\mathrm{X} \in \mathbb{S}_+^A$), denoted by $\mathrm{X} \succeq 0$, iff $\mathrm{Y}^T \mathrm{X} \mathrm{Y} \geq 0$ for all $\mathrm{Y} \in \mathbb{R}^A$.

Remark 2.1 (Convenient matrix notation). One common expression we will use throughout this paper is the following. For matrices $\mathrm{X} = [\mathrm{X}_u] \in \mathbb{R}^{\Upsilon \times R_0}$ and $M \in \mathbb{R}^{R_1 \times R_1}$ with $R_1 \subseteq R_0$ [4]:

$$\mathrm{Tr}(\mathrm{X}^T \mathrm{X} M) = \sum_{\substack{u \in R_0 \\ v \in R_0}} M_{u,v} \langle \mathrm{X}_u, \mathrm{X}_v \rangle \ .$$

Note that if M is positive semidefinite, i.e. $M \succeq 0$, then $\mathrm{Tr}(\mathrm{X}^T \mathrm{X} M) \geq 0$. [5] $\qquad \square$

Eigenvalues Given symmetric matrix $M \in \mathbb{S}^A$, for any integer $i \leq |A|$, we define its ith smallest and largest eigenvalues of M as the following, respectively:

$$
\begin{aligned}
\lambda_i(M) &\stackrel{\text{def}}{=} \max_{\mathrm{rank}(Z) \leq i-1} \min_{\mathrm{W} \perp \mathrm{Z}, \mathrm{W} \neq 0} \frac{\mathrm{W}^T M \mathrm{W}}{\mathrm{W}^T \mathrm{W}}, \\
\sigma_i(M) &\stackrel{\text{def}}{=} \min_{\mathrm{rank}(Z) \leq i-1} \max_{\mathrm{W} \perp \mathrm{Z}, \mathrm{W} \neq 0} \frac{\mathrm{W}^T M \mathrm{W}}{\mathrm{W}^T \mathrm{W}}.
\end{aligned}
\tag{1}
$$

---

[3] We chose this notation over the conventional one ($\mathbb{R}^{R \times C}$) so as to prevent ambiguity when the rows, R, or columns, C, are Cartesian products themselves.

[4] If M corresponds to a proper principal minor of $\mathrm{X}^T \mathrm{X}$, we assume M is padded with enough 0's before multiplication.

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

Generalized Eigenvalues Given $L \in \mathbb{S}^A$ and $M \in \mathbb{S}_+^A$ with $\mathrm{null}(M) \subseteq \mathrm{null}(L)$, for any integer $i \leq \mathrm{rank}(M)$, we define i^{th} smallest generalized eigenvalue of L and M as the following:

$$\lambda_i(L, M) \stackrel{\mathrm{def}}{=} \max_{\mathrm{rank}(Z) \leq i-1} \min_{M\mathrm{W} \perp \mathrm{Z}, M\mathrm{W} \neq 0} \frac{\mathrm{W}^T L \mathrm{W}}{\mathrm{W}^T M \mathrm{W}} \tag{2}$$

Observe that, for r being the nullity of M, $\lambda_i(L, M) = \lambda_{i+r}((M^\dagger)^{1/2} \, L \, (M^\dagger)^{1/2})$.

Graphs We assume all graphs are simple, undirected and edge-weighted with non-negative weights. We associate each graph $G = (V, C)$ with its edge weight function of the form $C : \binom{V}{2} \to \mathbb{R}_+$, where we use $C_{u,v}$ to denote the weight of edge between u and v.

Adjacency, Laplacian and Incidence Matrices Given an edge-weighted graph $G = (V, C)$ with no self loops, we define its adjacency matrix, $A_G \in \mathbb{S}^V$; degree matrix D_G; Laplacian matrix, $L_G \in \mathbb{S}^V$; and edge-node incidence matrix, $\Gamma_G \in \mathbb{R}^{\binom{V}{2}, V}$, as:

$$(A_G)_{u,v} = C_{u,v}.$$

$$(D_G)_{u,v} = \begin{cases} \sum_w C_{u,w} & \text{if } u = v, \\ 0 & \text{else} \end{cases}$$

$$L_G = D_G - A_G, \quad \text{and}$$

$$(\Gamma_G)_{uv,w} = \begin{cases} \sqrt{C_{uv}} & \text{if } w = \min(u, v), \\ -\sqrt{C_{uv}} & \text{if } w = \max(u, v), \\ 0 & \text{else.} \end{cases}$$

Observe that: (i) Γ_G has width-2. (ii) $L_G = \Gamma_G^T \Gamma_G$, hence $L_G \in \mathbb{S}_+^V$. (iii) For any $X = [X_u] \in \mathbb{R}^{\Upsilon, V}$, $\mathrm{Tr}\left[X^T X L_G\right] = \mathrm{Tr}\left[X^T X \Gamma_G^T \Gamma_G\right] = \|X\Gamma_G^T\|_F^2 = \sum_{u<v} C_{u,v} \|X_u - X_v\|^2$. (iv) $\lambda_1(L_G) = 0$. (v) If G is connected, then $\lambda_2(L_G) > 0$.

The normalized Lapalcian matrix $\mathcal{L}_G$ is defined as $\mathcal{L}_G = D^{-1/2} L_G D^{-1/2}$. We will often omit the subscript G from these matrices as the graph will be clear from the context.

2.1 Lasserre Hierarchy of Semidefinite Programming Relaxations

We present the formal definitions of the family of SDP relaxations, tailored to the setting of the problems we are interested in, where the goal is to assign to each vertex/variable of a set V a label from $[k] = \{0, 1, \ldots, k-1\}$. One can see that this relaxation is equivalent to the SDP hierarchy as given in [Las02] by basic inclusion-exclusion (for a formal proof, see the decomposition theorem in [KMN11]).

Definition 2.2 (SDP Relaxation). *Given a set of variables V, a set $[k] = \{0, 1, 2, \ldots, k\}$ of labels and an integer $r \geq 0$, $\mathrm{X} = (\mathrm{X}_S(f) \in \mathbb{R}^\Upsilon)$ is said to satisfy r^{th} moment constraints on k labels, denoted by*

$$\mathrm{X} \in \mathrm{Moment}_r(V, k),$$

if it satisfies the following conditions:

1. *For each set $S \in \binom{V}{\leq r+1}$ and $f : S \to [k]$, there exists a function $\mathrm{X}_S : [k]^S \to \mathbb{R}^\Upsilon$ that associates a vector of some finite dimension Υ with each possible labeling of S. We use $\mathrm{X}_S(f)$ to denote the vector associated with the labeling $f \in [k]^S$. For singletons $u \in V$, we will use $\mathrm{X}_u(i)$ and $\mathrm{X}_u(i^u)$ for $i \in [k]$ interchangeably. Similarly, when $k = 2$, we will use X_u instead of $\mathrm{X}_u(1)$.*

    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. *Let $[k]^\emptyset = \{\top\}$ where $\top$ denotes the (only) labeling of empty set with $\mathrm{X}_\emptyset(\top) = \mathrm{X}_\emptyset$.*

3. *$\langle \mathrm{X}_S(f), \mathrm{X}_T(g) \rangle = 0$ if there exists $u \in S \cap T$ such that $f(u) \neq g(u)$.*

4. *$\langle \mathrm{X}_S(f), \mathrm{X}_T(g) \rangle = \langle \mathrm{X}_A(f'), \mathrm{X}_B(g') \rangle$ 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 $S \in \binom{V}{\leq r+1}$, $u \in S$ and $f \in [k]^{S \setminus \{u\}}$, $\sum_{g \in [k]^u} \mathrm{X}_S(f \circ g) = \mathrm{X}_{S \setminus \{u\}}(f)$.*

It was shown in [Las02] that one can handle any polynomial constraints in the following way:

Definition 2.3 (Inequality Constraints). *Given a degree-d polynomial constraint of the form*

$$z(\mathbf{x}) = \sum_{S \in \binom{V}{\leq d}, f : S \to [k]} z_f \prod_{u \in S} \mathbf{x}_u(f(u)) \geq 0,$$

we say $\mathrm{X} \in \mathrm{Moment}_r(V, k)$ satisfies z:

$$\mathrm{X} * z \in \mathrm{Moment}_r(V, k),$$

if there exists $\mathrm{Y} = [\mathrm{Y}_S(f)] \in \mathrm{Moment}_{r-d}(V, k)$ such that

$$\|\mathrm{Y}_T(g)\|^2 = \sum_{S \in \binom{V}{\leq d}, f : S \to [k]} z_S \|\mathrm{X}_{S \cup T}(f \circ g)\|^2$$

for any $g : T \to [k]$ with $|T| \leq r - d$.

In the case of equality constraints, there is a simpler way to enforce which is also equivalent:

Definition 2.4 (Equality Constraints). *Given a degree-d polynomial constraint of the form $z(\mathbf{x}) = 0$, we say $\mathrm{X} \in \mathrm{Moment}_r(V, k)$ satisfies z if*

$$\mathrm{X} \cdot z = \sum_{f : S \to [k]} z_f \mathrm{X}_S(f) = 0.$$

Claim 2.5. *Given a degree-d polynomial z, for any $\mathrm{X} \in \mathrm{Moment}_r(V, k)$, $\{\mathrm{X} * z, \mathrm{X} * -z\} \subset \mathrm{Moment}_{r-d}(V, k)$ if and only if $\mathrm{X} \cdot z = 0$.*

Definition 2.6 (Conditioning). *Given $S \subseteq [n]$ and $f \in [k]^S$ with $\mathrm{X}_S(f) \neq 0$, we define the vectors conditioned on f as follows. For any $T \subseteq [n]$ and $g \in [k]^T$, the vector $\mathrm{X}_{T|f}(g)$ is given by:*

$$\mathrm{X}_{T|f}(g) \stackrel{\mathrm{def}}{=} \frac{\mathrm{X}_{S \cup T}(f \circ g)}{\|\mathrm{X}_S(f)\|}.$$

We will use $\mathrm{X}_{|f} = [\mathrm{X}_{T|f}(g)]$ to denote the corresponding matrix.

11

Formally the conditional vectors $X_{T|f}(g)$ correspond to relaxations of respective indicator variables. Thus such vectors behave exactly in the same way with non-conditional vectors. Some of these properties are given in the following easy claim, whose proof we skip.

Proposition 2.7. *Given* $X \in \text{Moment}_r(V, k)$, *for any* $f \in [k]^S$ *with* $X_S(f) \neq 0$, *the following are true:*

(a) $X_{|f} \in \text{Moment}_{r-|S|}(V, k)$.

(b) *If* $X * z \in \text{Moment}_{r-d}(V, k)$ *for some degree-d polynomial* z, *then* $X_{|f} * z \in \text{Moment}_{r-|S|-d}(V, k)$.

(c) *For any* $g \in [k]^T$, $\left(X_{|f}\right)_{|g} = X_{|f \circ g}$.

Assume that some labeling $f_0 \in [k]^{S_0}$ to S_0 has been fixed, and we further sample a labeling f to S with probability $\|X_{S|f_0}(f)\|^2$ (i.e., from the conditional probability distribution of labelings to S given labeling f_0 to S_0). The following defines a projection matrix which captures the effect of further conditioning according to the labeling to S. For a nonzero vector v, we denote by $\overline{v}$ the unit vector in the direction of v.

Notation 2.8. *Given* $f_0 \in [k]^{S_0}$ *and* $S \subseteq [n]$, *let*

$$\Pi_{S|f_0} \stackrel{\text{def}}{=} \sum_{f : X_{S|f_0}(f) \neq 0} \overline{X_{S|f_0}(f)} \cdot \overline{X_{S|f_0}(f)}^T.$$

Similarly let $\Pi_S^{\perp} \stackrel{\text{def}}{=} I - \Pi_S$ *where* I *is the identity matrix of the appropriate dimension.*

2.2 Overview of Our Rounding

Consider any solution satisfying $(r + 2)^{th}$ moment constraints. For any set of r variables, S, SDP solution gives us a distribution over $f : S \rightarrow [k]$. Moreover, conditioned on having labelled S according to f, we know that the conditional vectors as in Definition 2.6 satisfies 2nd moment constraints.

This suggests a natural approach for rounding:

1. Fix some set of "seeds", S.

2. Choose a labeling for S, $f : S \rightarrow [k]$ with probability $\|X_S(f)\|^2$.

3. For every node $u \in V$, choose a label such that the marginal probability of assigning label i to u is $\|X_{u|f}(i)\|^2$.

A simple way to choose labelings so as to satisfy the marginal probability condition is by sampling labels independently at random with respective probabilities. This is not the only way for choosing labels, and indeed our rounding for Non-Uniform Sparsest Cut problem uses a different rounding scheme.

For graph partitioning type problems, we can upper bound the probability of not satisfying an edge constraint with the sum of SDP value and the (co-)variance of associated variables (see Section 2.1 for formal proof).

Our main observation is that, the expected variance can be related to a highly geometric quantity involving only the "singleton" vectors: The expected variance on $\mathbf{x}_u(i)$ is upper bounded by the squared distance of $X_u(i)$ to the span of vectors in the seed set.

In particular, the total variance is now upper bounded by how well the vectors in seed set approximate the singleton matrix in Frobenius norm. Using the column selection methods from

12

from [GS12b, BDMI11], we can finally upper bound the total variance of labeling from such distribution by the sum of largest r eigenvalues of associated Gram matrix. Finally by using trace inequalities, we can lower bound this sum in terms of associated constraint graph's spectra.

2.3 Faithful Rounding

All our rounding algorithms will follow the same outline. We will assume that we are given a subset of **seeds**, $S \subset V$, with corresponding vectors $(\mathrm{X}_S(f) \mid f : S \rightarrow [k])$ satisfying some r rounds of Lasserre Hierachy constraints on labels $[k]$. We proceed by choosing a seed labelling, $f : S \rightarrow [k]$, and then **propagating** f to other nodes in such a way that the probability of some $u \in V$ receiving label $j \in [k]$ is equal to $\|\mathrm{X}_{u|f}(j)\|^2$, as in Definition 2.6.

More formally, if we use $\mathcal{X}_u(j)$ to denote the random indicator variable of $u \in V$ receiving label $j \in [k]$, then
$$\mathbb{E}\big[\mathcal{X}_u(j) \mid f \text{ is chosen}\big] = \|\mathrm{X}_{u|f}(j)\|^2.$$

When $[k] = \{0, 1\}$, we will drop j and simply use $\mathcal{X}_u = \mathcal{X}_u(1)$.

Note that any such rounding algorithm will be **faithful** [CDK12]. In particular, for any lifted constraint of the form $\langle a, \mathrm{X} \rangle \leq b$,
$$\mathbb{E}\Big[\sum_{u,j} a_{u,j} \mathcal{X}_u(j)\Big] = \sum_{u,j} a_{u,j} \|\mathrm{X}_{u|f}(j)\|^2 \geq b.$$

Combining this with Proposition 2.7, we can see that such the output of such rounding procedure will satisfy each linear constraint **in expectation**. Furthermore, when labelings are chosen independently of each other, we can use usual concentration bounds, such as Chernoff's inequality, to conclude that w.h.p. linear constraints will be satisfied within some negligible error.

3 Case Study: 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.[6]

For simplicity we will assume G is unweighted and d-regular, however all our results given in Section 6.1 are for any weighted undirected graph G. We can formulate this problem as a binary integer programming problem as follows:

$$\min_{\mathbf{x}} \sum_{e=\{u,v\}\in E} (\mathbf{x}_u - \mathbf{x}_v)^2, \tag{3}$$
$$\text{st} \sum_u \mathbf{x}_u = \mu; \quad \mathbf{x} \in \{0,1\}^V.$$

---

[6]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. In such case, we can do better using threshold based rounding (see Theorem 7.8 in Section 7.1).

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

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

3.1 SDP Relaxation for Minimum Bisection

We consider the following relaxation of eq. (3):

$$\min \quad \sum_{e=\{u,v\}\in E} \|X_u - X_v\|^2 \qquad (4)$$

$$\text{st} \quad \sum_u X_u = \mu X_\emptyset,$$

$$\|X_\emptyset\|^2 = 1,$$

$$X = [X_S(f)] \in \text{Moment}_r(V, 2).$$

It is easy to see that this is indeed a relaxation of our original QIP formulation (3).

3.2 Main Theorem on Rounding

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

Our ultimate goal in this section is to give an algorithm to round the SDP solution X to a good $\mathbf{x} \in \{0,1\}^V$ of size very close to μ, and prove the below theorem.

Theorem 3.1. *For all $r \geq 1$ and $\varepsilon > 0$, by solving $O(r/\varepsilon^2)$-rounds of Lasserre Hierachy relaxation, we can find $x \in \{0,1\}^V$ with the following properties:*

1. $x^T L x \leq \frac{1+\varepsilon}{\min(1,\lambda_{r+1}(\mathcal{L}))} \eta.$

2. $\left| \|x\|_1 - \mu \right| \leq o(1) = O\left(\sqrt{\mu \log(1/\varepsilon)} \right).$

Since one can solve the Lasserre Hierachy 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 directly follows from Theorem 6.1 in Section 6.1. Note that if $t = \text{argmin}_r\{r \mid \lambda_r(\mathcal{L}) \geq 1 - \varepsilon/2\}$, 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.

3.3 The Rounding Algorithm

Recall that the solution $X = [X_S(f)]$ contains a vector $x_T(f)$ for each $T \in \binom{V}{\leq r'}$ and every possible labeling of T, $f \in \{0,1\}^T$ of T. Our approach to round X to $\mathbf{x} \in \{0,1\}$ which is an approximate solution of the integer program eq. (3) is similar to the label propagation approach used in [AKK$^+$08].

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 $f \in \{0,1\}^S$ 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 SDP takes $n^{O(r')}$ time anyway. But there is also a faster method to find

14

a good S, as mentioned in Theorem 5.2.) Conditional on choosing a specific labeling f to S, we propagate the labeling to other nodes as follows: Independently for each $u \in V$, assign $\mathcal{x}_u \leftarrow 1$ with probability $\|X_{u|f}\|^2$ and $\mathcal{x}_u \leftarrow 0$ otherwise. Observe that if $u \in S$, then $\mathcal{x}_u = f(u)$ always. Finally, output $\mathcal{x}$. Below Π_S denotes the projection matrix from Notation 2.8.

Lemma 3.2. *For the above rounding procedure, the size of the cut produced, $\mathcal{x}^T L \mathcal{x}$, satisfies:*

$$\mathbb{E}\left[\mathcal{x}^T L \mathcal{x}\right] = \eta + \sum_{(u,v)\in E} \langle \Pi_{\bar{S}}^{\perp} X_u, \Pi_{\bar{S}}^{\perp} X_v \rangle = \eta + \mathrm{Tr}\left(X^T \Pi_{\bar{S}}^{\perp} X A\right). \tag{5}$$

Here A is the adjacency matrix of G.

Proof. Note that for $u \neq v$:

$$\mathbb{E}\left[\mathcal{x}_u\right] = \mathbb{E}_f\left[\|X_{u|f}\|^2\right] = \|X_u\|^2,$$

$$\mathbb{E}\left[\mathcal{x}_u \mathcal{x}_v\right] = \sum_f \|X_S(f)\|^2 \|X_{u|f}\|^2 \|X_{v|f}\|^2 = \sum_f \|X_S(f)\|^2 \frac{\langle \overline{X_S(f)}, X_u(i) \rangle}{\|X_S(f)\|} \frac{\langle \overline{X_S(f)}, X_v(j) \rangle}{\|X_S(f)\|}$$

$$= \sum_f \langle \overline{X_S(f)}, X_u \rangle \langle \overline{X_S(f)}, X_v \rangle.$$

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 and X_v onto the span of $\{X_S(f)\}_{f:S\to\{0,1\}}$, which we denote by Π_S. Let us now calculate the expected number 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 $\mathcal{x}_u = 1$ and $\mathcal{x}_v = 0$. Therefore,

$$\mathbb{E}\left[\mathcal{x}^T L \mathcal{x}\right] = \sum_{(u,v)\in E} \mathbb{E}\left[\mathcal{x}_u(1-\mathcal{x}_v)\right] = \sum_{(u,v)\in E} \left(\|X_u\|^2 - \langle \Pi_S X_u, \Pi_S X_v \rangle\right) \tag{6}$$

Π_S is a projection matrix, so $\langle \Pi_S X_u, \Pi_S X_v \rangle = \langle X_u, X_v \rangle - \langle \Pi_{\bar{S}}^{\perp} X_u, \Pi_{\bar{S}}^{\perp} X_v \rangle$. Substituting this in eq. (6):

$$= \sum_{(u,v)\in E} \|X_u\|^2 - \langle X_u, X_v \rangle + \langle \Pi_{\bar{S}}^{\perp} X_u, \Pi_{\bar{S}}^{\perp} X_v \rangle$$

$$= \eta + \sum_{(u,v)\in E} \langle \Pi_{\bar{S}}^{\perp} X_u, \Pi_{\bar{S}}^{\perp} X_v \rangle.$$

$\square$

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 $i \in [k]$, $X_u(i) = \sum_{f:f(u)=i} X_S(f)$ by SDP constraints.

Observation 3.3. *For all $S \in \binom{V}{r'}$,*

$$\mathrm{span}\left(\{X_S(f)\}_{f\in\{0,1\}^S}\right) \supseteq \mathrm{span}\left(\{X_u\}_{u\in S}\right).$$

Equivalently for P_S being the projection matrix onto span of $\{X_u\}_{u\in S}$, $P_S \preceq \Pi_S$.

Thus we will try to upper bound the term in eq. (5) by replacing $\Pi_{\bar{S}}^{\perp}$ with $P_{\bar{S}}^{\perp}$. But we cannot directly perform this switch, as the adjacency matrix A is not PSD.

3.4   Factor $1 + \frac{1}{\lambda_r}$ Approximation of Cut Value

Our first bound is by directly upper bounding eq. (5) in terms of $\|\Pi_{\bar{S}}^{\perp} X_u\|^2 \leq \|P_{\bar{S}}^{\perp} X_u\|^2$. Using Cauchy-Schwarz and Arithmetic-Geometric Mean inequalities, (5) implies that the expected number of edges cut is upper bounded by

$$\eta + \frac{1}{2} \sum_{e=(u,v)\in E} \left(\|\Pi_{\bar{S}}^{\perp} X_u\|^2 + \|\Pi_{\bar{S}}^{\perp} X_v\|^2 \right) = \eta + d \sum_u \|\Pi_{\bar{S}}^{\perp} X_u\|^2 \leq \eta + d \sum_u \|P_{\bar{S}}^{\perp} X_u\|^2 . \tag{7}$$

If we define the matrix $Y \stackrel{\text{def}}{=} [X_u]_{u\in V}$, then:

$$d \sum_u \|P_{\bar{S}}^{\perp} X_u\|^2 = d \operatorname{Tr}(Y^T Y_{\bar{S}}^{\perp} Y) = d \|Y_{\bar{S}}^{\perp} Y\|_F^2.$$

To get the best upper bound, we want to pick $S \in \binom{V}{r'}$ that minimizes $\|Y_{\bar{S}}^{\perp} Y\|_F^2$. It is a well known fact that among *all* projection matrices M of rank r' (not necessarily restricted to projection onto X_u's), the minimum value of $\|M^{\perp} Y\|_F^2$ is achieved by matrix M projecting onto the space of the largest r' singular vectors of Y. Further, this minimum value equals $\sum_{i\geq r'+1} \sigma_i$ where $\sigma_i = \sigma_i(Y^T Y)$ is i^{th} largest eigenvalue of $Y^T Y$. Hence $\|Y_{\bar{S}}^{\perp} Y\|_F^2 \geq \sum_{i\geq r'+1} \sigma_i$ for every choice of S. The following theorem from [GS12b] shows the existence of S which comes close to this lower bound:

Theorem 3.4. *[GS12b] For every real matrix X with column set V, and positive integers $r \leq r'$, we have*

$$\delta_{r'}(X) \stackrel{\text{def}}{=} \min_{S \in \binom{V}{r'}} \operatorname{Tr}(X^T X_{\bar{S}}^{\perp} X) \leq \frac{r'+1}{r'-r+1} \left(\sum_{i\geq r+1} \sigma_i \right) .$$

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

Remark 3.5. Prior to our paper [GS12b], it was shown in [BDMI11] that $\delta_{2r/\varepsilon+r} \leq (1+\varepsilon)\left(\sum_{i\geq r+1} \sigma_i \right)$. The improvement in the bound on r' from $2r/\varepsilon+r$ to $r/\varepsilon+r-1$ 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. $\square$

Picking the subset $S^* \in \binom{V}{r'}$ that achieves the bound guaranteed by Theorem 3.4, we have

$$\|Y_{\bar{S}}^{\perp} Y\|_F^2 \leq (1+\varepsilon) \sum_{i>r} \sigma_i .$$

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

$$\frac{\eta}{d} = \frac{1}{d} \operatorname{Tr}(Y^T Y L) \geq \sum_i \sigma_i \lambda_i(\mathcal{L}) \geq \sum_{i\geq r+1} \sigma_i \lambda_{r+1} \geq (1+\varepsilon)^{-1} \lambda_{r+1}(\mathcal{L}) \|Y_{\bar{S}}^{\perp} Y\|_F^2.$$

Plugging this into eq. (7), we can conclude our first bound:

Theorem 3.6. *The rounding algorithm given in Section 3.3 cuts at most*

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

edges in expectation. In particular, the algorithm cuts at most a factor $1 + \frac{1+\varepsilon}{\lambda_{r+1}(\mathcal{L})}$ more edges than the optimal cut with μ nodes on one side.[7]

Note that $\lambda_n(\mathcal{L}) \leq 2$, hence even if we use n-rounds (in which case X is an integral solution), the smallest upper bound we can show is $\frac{3}{2}$. Although this is too weak by itself for our purposes, this bound will be crucial to obtain our final bound.

3.5 Improved Analysis and Factor $\frac{1}{\lambda_r}$ Approximation on Cut Value

First notice that eq. (5) can be written as:

$$\mathbb{E}\big[\text{number of edges cut}\big] = \text{Tr}(X^T X L) + \text{Tr}(X^T \Pi_{\bar{S}}^{\perp} X A) = d\,\text{Tr}(X^T \Pi_{\bar{S}}^{\perp} X) + \text{Tr}(X^T \Pi_S X L)\,. \quad (8)$$

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

$$\text{Tr}(X^T \Pi_{\bar{S}}^{\perp} X L) = \text{Tr}(X^T X L) - \text{Tr}(X^T \Pi_S X L) < \eta(1-\varepsilon)$$

so that $\text{Tr}(X^T \Pi_{\bar{S}\cup T}^{\perp} X) < (1-\varepsilon)\frac{(1+\varepsilon)\eta}{\lambda_{r+1}}$. Now, if we run the rounding algorithm with $S \cup T$ as the seed set, and (8) with $S \cup T$ in place of S is larger than $\frac{(1+\varepsilon)\eta}{\lambda_{r+1}} + \eta\varepsilon$, then $\text{Tr}(X^T \Pi_{S\cup T} X L) > 2\varepsilon\eta$. Hence

$$\text{Tr}(X^T \Pi_{\bar{S}\cup T}^{\perp} X L) \leq \text{Tr}(X^T X L) - \text{Tr}(X^T \Pi_{S\cup T} X L) < \eta(1-2\varepsilon)\,.$$

Picking another set U, we will have $\text{Tr}(X^T \Pi_{\bar{S}\cup T \cup U}^{\perp} X) < (1-2\varepsilon)\frac{(1+\varepsilon)\eta}{\lambda_{r+1}}$. Continuing this process, if the quantity eq. (8) is not upper bounded by $\frac{(1+\varepsilon)\eta}{\lambda_{r+1}} + \eta\varepsilon$ after $\lceil\frac{1}{\varepsilon}\rceil$ many such iterations, then the total projection distance becomes:

$$0 \leq \text{Tr}(X^T \Pi_{\bar{S}\cup T \cup \ldots}^{\perp} X) < (1 - \lceil 1/\varepsilon\rceil\varepsilon)\frac{(1+\varepsilon)\eta}{\lambda_{r+1}} \leq 0$$

which is a contradiction. For formal statement and proof in a more general setting, see Theorem 5.4 in Section 5.

Theorem 3.7. *The expected number of edges cut by the rounding algorithm from Section 3.3 using the seed selection procedure as described in Section 3.5 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.*

---

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

3.6 Bounding Set Size

We now analyze the balance of the cut, and show how to ensure that $\|x\|_1 = \mu \pm o(\mu)$ in addition to $x^T L x$ being close to the expected bound of Theorem 3.7 (and similarly for Theorem 3.6).

Let $f : \mathcal{S}^* \to \{0, 1\}$ be fixed. We will show that conditioned on finding cuts with small cost, the probability that one of them has size $\approx \mu$ is bounded away from zero. We can then use a simple Markov bound to show that there is a non-zero probability that both cut size and set size are within $O(1)$-factor of corresponding bounds. But by exploiting the independence in our rounding algorithm and Lasserre Hierachy relaxations of linear constraints, we can do much better.

Note that in the r'-round Lasserre Hierachy relaxation, we had:

$$\sum_u \mathrm{X}_u = \mu \mathrm{X}_\emptyset.$$

This means, for each $f \in \{0, 1\}^{\mathcal{S}^*}$:

$$\sum_u \|\mathrm{X}_{u|f}\|^2 = \mu .$$

This implies that conditioned on the choice of f, the expectation of $\sum_u x_u$ is μ and $\mathbf{x}_u$ for various u are all independent. Applying the Chernoff bound, we get

$$\mathrm{Prob}\left[\left| \sum_u x_u - \mu \right| \geq 2\sqrt{\mu \log \frac{1}{\zeta}} \right] \leq o(\zeta) \leq \frac{\zeta}{3}.$$

Consider choosing $f \in \{0, 1\}^{\mathcal{S}^*}$ so that $\mathbb{E}\left[\text{number of edges cut} \,\middle|\, f\right] \leq \mathbb{E}\left[\text{number of edges cut}\right] \stackrel{\text{def}}{=} b$. By Markov inequality, if we pick such an f, $\Pr\left[\text{number of edges cut} \geq (1 + \zeta)b\right] \leq 1 - \frac{\zeta}{2}$, where the probability is over the random propagation once $\mathcal{S}^*$ and f are fixed.

Hence with probability at least $\frac{\zeta}{6}$, the subset found has cut cost $\leq (1 + \zeta)b$ and size 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 Theorem 3.1 on minimum bisection.

4 Analysis of Propagation Rounding

Intuitively, the main quantity that affects our approximation factor is how far $\|\mathrm{X}_{u|f}\|^2$ is from $\{0, 1\}$ in expectation over f. In this section, we will show how to choose seed sets so as to relate this distance to certain spectral properties of the underlying graph. We start with proving some simple properties:

Let $x_{A|f}(g)$ be a random variable over $\{0, 1\}$ with $\mathrm{Prob}\left[x_{A|f}(g) = 1\right] = \|\mathrm{X}_{A|f}(g)\|^2$. Given fixed $f_0 \in [k]^{S_0}$ and subset S, we choose our next conditioning $f : S \to [k]$ by sampling f with probability $\|\mathrm{X}_{S|f_0}(f)\|^2$. We will denote this by $f \sim \|\mathrm{X}_{S|f_0}(f)\|^2$.

Claim 4.1.

$$\mathbb{E}_{f \sim \|\mathrm{X}_S(f)\|^2}\left[\mathbb{E}\left[x_u(j) \mid f \text{ is chosen}\right]\right] = \|\mathrm{X}_u(j)\|^2.$$

Proof.

$$\mathbb{E}_{f \sim \|\mathrm{X}_S(f)\|^2}\left[\mathbb{E}\left[x_u(j) \mid f \text{ is chosen}\right]\right] = \sum_f \|\mathrm{X}_S(f)\|^2 \frac{\langle \mathrm{X}_S(f), \mathrm{X}_u(j)\rangle}{\|\mathrm{X}_S(f)\|^2} = \langle \mathrm{X}_\emptyset, \mathrm{X}_u(j)\rangle = \|\mathrm{X}_u(j)\|^2.$$

$\square$

After having chosen S and f, there are two natural rounding schemes:

1. Independent Rounding: For each $u \in V$, choose $j \in [k]$ with probability $\|X_{u|f}(j)\|^2$. This rounding scheme will be used for unique games type problems as well as independent set.

2. Threshold Rounding: (When $k = 2$) Choose $\tau \in (0, 1]$ uniformly at random. For each $u \in V$, let $x_u \leftarrow 1$ if $\|X_{u|f}(j)\|^2 \geq \tau$ and $x_u \leftarrow 0$ otherwise. We will use this rounding for minimum cut type problems such as sparsest cut, balanced separator, etc.

4.1 Independent Rounding Scheme

As we discussed in Section 2.2, it is easy to show that, under such scheme, linear constraints will be satisfied with high probability within error $\widetilde{O}(\sqrt{n})$ (here $\widetilde{O}$ hides logarithmic terms) using concentration inequalities. Therefore we only need to bound the probability that some constraint on a pair of variables (u, v) is left unsatisfied.

Claim 4.2. *Given $u \neq v \in V$ and $i, j \in [k]$,*

$$\mathbb{E}_f\big[x_u(i)x_v(j)\big] = X_u(i)^T \Pi_S X_v(j).$$

Proof.

$$\mathbb{E}_f\big[x_u(i)x_v(j)\big] = \sum_f \|X_S(f)\|^2 \overline{\langle \overline{X_S(f)}, X_u(i)\rangle} \langle \overline{X_S(f)}, X_v(j)\rangle = \sum_f \langle \overline{X_S(f)}, X_u(i)\rangle \langle \overline{X_S(f)}, X_v(j)\rangle$$

$$= X_u(i)^T \Big(\sum_f \overline{X_S(f)} \cdot \overline{X_S(f)}^T \Big) X_v(j) = X_u(i)^T \Pi_S X_v(j).$$

In the last identity, we used Notation 2.8. $\qquad\square$

Lemma 4.3. *Given set of seeds S, for any symmetric matrix A of appropriate dimensions:*

$$\mathbb{E}\Big[\sum_{u,v}\sum_{i,j} x_u(i)x_v(j)A_{(u,i),(v,j)}\Big] = \mathrm{Tr}(X^T\Pi_S X A) + \mathrm{Tr}(X^T\Pi_S^\perp X \operatorname{diag}(A)).$$

Proof. Fix u and i. Since Π_S is a projection matrix, we have: $\mathbb{E}\big[x_u(i)\big] = \|X_u(i)\|^2 = X_u(i)^T\Pi_S X_u(i) + X_u(i)^T\Pi_S^\perp X_u(i)$. For $A^o \overset{\text{def}}{=} A - \operatorname{diag}(A)$ (the off-diagonal entries of A):

$$\mathbb{E}\big[x^T A x\big] = \mathbb{E}\big[x^T \operatorname{diag}(A) x\big] + \mathbb{E}\big[x^T A^o x\big]$$
$$= \mathrm{Tr}(X^T\Pi_S^\perp X \operatorname{diag}(A)) + \mathrm{Tr}(X^T\Pi_S X \operatorname{diag}(A)) + \mathrm{Tr}(X^T\Pi_S X A^o)$$
$$= \mathrm{Tr}(X^T\Pi_S^\perp X \operatorname{diag}(A)) + \mathrm{Tr}(X^T\Pi_S X A)$$

where we used Claim 4.2 in the first step. $\qquad\square$

Observe that Π is a matrix that depends on the higher-order vectors $\{X_S(f)\}$. This makes it very difficult to reason about Π matrix. In Lemma 4.4, we will relate the matrix Π_S to projection matrices onto the linear subspaces of $\{X_S(f)\}_f$'s span:

Lemma 4.4. *For any $S \subseteq [n]$, let Y be a matrix whose columns are linear combinations of vectors from the set $\{X_S(g)\}_g$. Then:*

1. *Π_S is a projection matrix, i.e. $\Pi_S^2 = \Pi_S$;*

2. $\Pi_S^{\perp}$ is the projection matrix onto the orthogonal complement of $\Pi_S's$ row span, i.e. $\Pi_S^{\perp}\Pi_S = 0$;

3. Π_S dominates Y^{Π}, i.e. $\Pi_S \succeq Y^{\Pi}$ where Y^{Π} is the projection matrix onto the span of Y.

Proof. Denote $\Pi = \Pi_S$ for notational simplicity. By construction, Π is symmetric and $\Pi \succeq 0$. We will show that $\Pi^2 = \Pi$ which implies Π is a projection matrix:

$$\Pi^2 = \sum_{g,h} \frac{1}{\|\mathrm{X}(g)\|^2 \mathrm{X}(h)\|^2}\mathrm{X}(g)\mathrm{X}(g)^T\mathrm{X}(h)\mathrm{X}(h) = \sum_g \overline{\mathrm{X}(g)\mathrm{X}(g)}^T = \Pi.$$

Since Y's columns are linear combinations of Π, this means $\Pi Y = \Pi$. Using the fact that $0 \preceq Y^{\Pi} \preceq I$, we obtain $Y^{\Pi} = \Pi Y^{\Pi}\Pi \preceq \Pi \cdot I \cdot \Pi = \Pi$. $\qquad\square$

4.2 Threshold Based Rounding Scheme

We will use such rounding for problems that involve *minimizing* the number of edges cut under various constraints involving quantities like partition size or the number of edges cut in some other "demand" graph. Our first quantity of interest is the probability of separating pair (u, v).

Claim 4.5. $\left|\|\mathrm{X}_{u|f}\|^2 - \|\mathrm{X}_{v|f}\|^2\right| \leq \|\mathrm{X}_{u|f} - \mathrm{X}_{v|f}\|^2.$

Proof.

$$\left|\|\mathrm{X}_{u|f}\|^2 - \|\mathrm{X}_{v|f}\|^2\right| = \left|\|\mathrm{X}_{uv|f}(10)\|^2 - \|\mathrm{X}_{uv|f}(01)\|^2\right|$$
$$\leq \|\mathrm{X}_{uv|f}(10)\|^2 + \|\mathrm{X}_{uv|f}(01)\|^2 = \|\mathrm{X}_{u|f} - \mathrm{X}_{v|f}\|^2. \qquad\square$$

Lemma 4.6. *For fixed f, let τ be chosen uniformly at random from $(0, 1]$. For all $u \in V$, if we let x_u be the indicator of $\|\mathrm{X}_{u|f}\|^2 \geq \tau$, then the probability of $x_u \neq x_v$ is bounded as follows:*

$$\mathrm{Prob}_\tau\left[x_u \neq x_v \mid f\right] \leq \|\mathrm{X}_{u|f} - \mathrm{X}_{v|f}\|^2.$$

If f is chosen with probability $\|\mathrm{X}(f)\|^2$, then:

$$\mathrm{Prob}_{f,\tau}\left[x_u \neq x_v\right] \leq \|\mathrm{X}_u - \mathrm{X}_v\|^2.$$

Proof. First part follows from Claim 4.5. Second part follows from Proposition 2.7(c). $\qquad\square$

For (non-uniform) sparsest cut problem, we need the following lower bound:

Claim 4.7. *Suppose the threshold, τ, is chosen uniformly at random from $(0, 1]$. Then, the probability of $x_u \neq x_v$ over random f is:*

$$\mathrm{Prob}_{f,\tau}\left[x_u \neq x_v \mid f\right] \geq \|\mathrm{X}_u - \mathrm{X}_v\|^2 - \|\Pi_S^{\perp}(\mathrm{X}_u - \mathrm{X}_v)\|^2.$$

Proof.

$$\mathrm{Prob}\left[x_u \neq x_v \mid f\right] = \mathbb{E}_f\left[\left|\langle\overline{\mathrm{X}_S(f)}, \mathrm{X}_u\rangle - \langle\overline{\mathrm{X}_S(f)}, \mathrm{X}_u\rangle\right|\right] \geq \mathbb{E}_f\left[\left(\langle\overline{\mathrm{X}_S(f)}, \mathrm{X}_u\rangle - \langle\overline{\mathrm{X}_S(f)}, \mathrm{X}_u\rangle\right)^2\right]$$
$$= \|\Pi_S(\mathrm{X}_u - \mathrm{X}_v)\|^2.$$

Using the fact that Π_S is a projection matrix (see Lemma 4.4) with orthogonal complement $\Pi_S^{\perp}$ finishes the proof. $\qquad\square$

5 Choosing A Good Seed Set by Column Selection

As we have seen in Sections 3 and 4, the performance depends on minimizing $\mathrm{Tr}(X^T \Pi_S^\perp X B)$. In this section, we will relate this quantity to the spectrum of objective matrix via column based matrix reconstruction. First, we need a generalized form of von Neumann's Trace Inequality [HJ90] for relating the approximation factor to the (generalized) spectrum.

Theorem 5.1 (Generalized Trace Inequality). *Let $A, B \in \mathbb{S}_+^C$ be two matrices with $B = \Gamma^T \Gamma$ for some $\Gamma \in \mathbb{R}^{R,C}$ such that $\mathrm{null}(B) \subseteq \mathrm{null}(A)$. Then for any $X \in \mathbb{S}_+^C$, the following bound holds:*

$$\sum_j \sigma_j(\Gamma X \Gamma^T) \lambda_j(A, B) \leq \mathrm{Tr}(XA), \tag{9}$$

where σ_j, λ_j are as defined in eqs. (1) and (2), respectively. Moreover for any positive integer r,

$$\sum_{j \geq r} \sigma_j(\Gamma X \Gamma^T) \leq \frac{\mathrm{Tr}(XA)}{\lambda_r(A, B)}. \tag{10}$$

Proof. Unless noted otherwise, we will use $\sigma_j = \sigma_j(\Gamma X \Gamma^T)$ and $\lambda_j = \lambda_j(A, B)$. Given eq. (9), we can derive eq. (10) easily as follows. Since $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_r$:

$$\mathrm{Tr}(XA) \geq \sum_j \sigma_j \lambda_j \geq \sum_{j \geq r} \sigma_j \lambda_j \geq \sum_{j \geq r} \sigma_j \lambda_r = \lambda_r \sum_{j \geq r} \sigma_j.$$

Now we will prove eq. (9). Since $\mathrm{null}(A) \supseteq \mathrm{null}(B)$, $B^\dagger B A = A B^\dagger B = A$. Hence:

$$\mathrm{Tr}(XA) = \mathrm{Tr}(X B^{1/2} B^{\dagger/2} A B^{\dagger/2} B^{1/2}) = \mathrm{Tr}(B^{1/2} X B^{1/2} B^{\dagger/2} A B^{\dagger/2})$$

For q_i's being the eigenvectors of $B^{\dagger/2} A B^{\dagger/2}$, we have $B^{\dagger/2} A B^{\dagger/2} = \sum_j \lambda_j q_{j+r} q_{j+r}^T$ where $r = \mathrm{rank}(B)$. Substituting this into previous equation, we get:

$$= \sum_{j=1}^{n-r} \lambda_j q_{j+r}^T B^{1/2} X B^{1/2} q_{j+r} = \sum_{j=1}^{n-r} \lambda_j q_{j+r}^T B^{1/2} B^\dagger \Gamma^T \Gamma X \Gamma^T \Gamma B^\dagger B^{1/2} q_{j+r}$$

$$= \sum_{j=1}^{n-r} \lambda_j q_{j+r}^T B^{\dagger/2} \Gamma^T Y \underbrace{\Gamma B^{\dagger/2} q_{j+r}}_{\overset{\mathrm{def}}{=} p_j} = \sum_j \lambda_j p_j^T Y p_j \geq \sum_j (\lambda_i){\uparrow}_j \, (p_i^T Y p_i \mid i){\downarrow}_j$$

Assume p_j's were all orthonormal. Then the sequence $(p_j^T Y p_j)_j$ is majorized by $(\sigma_j(Y)) = (\sigma_j(\Gamma X \Gamma^T))_j$. Since $f(x) = \sum_j (\lambda_i){\uparrow}_j \, x{\downarrow}_j$ is Schur-concave, for any y that majorizes x, $f(x) \geq f(y)$:

$$\geq \sum_j (\lambda_i){\uparrow}_j \, \sigma_j = \sum_j \lambda_j \sigma_j.$$

We will finish the proof by proving that p_i's are orthonormal:

$$\langle p_i, p_j \rangle = q_{i'}^T B^{\dagger/2} \Gamma^T \Gamma B^{\dagger/2} q_{j'} = q_{i'}^T B^{\dagger/2} B B^{\dagger/2} q_{j'} = q_{i'}^T B^\dagger B q_{j'} = q_{i'}^T q_{j'} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else.} \end{cases} \qquad \square$$

The following theorem says that given a matrix the sum of whose largest r eigenvalues is large, we can always find a set of $O(r)$ columns that approximates the matrix well (in Frobenius norm). A bound of the form $1 + O(r/r')$ is also given in [BDMI11].

Theorem 5.2. *[GS12b] For any $X \in \mathbb{R}^{R,C}$ with column set C, and positive integers $r \leq r'$, we have*

$$\delta_{r'}(X) \stackrel{\text{def}}{=} \min_{S \in \binom{C}{r'}} \|X_S^\perp X\|_F^2 \leq \frac{r'+1}{r'-r+1} \left(\sum_{i \geq r+1} \sigma_i \right).$$

In particular, for all $\varepsilon \in (0,1)$, $\delta_{r/\varepsilon+r-1} \leq (1+\varepsilon) \sum_{i \geq r+1} \sigma_i$.

We combine Theorems 5.1 and 5.2 in the following corollary.

Corollary 5.3. *Given positive integers r, w and $\varepsilon \in (0,1)$ the following holds. For any $X = [X_S(f)] \in \text{Moment}_{r'}(V, k)$ where*

$$r' = w \left(\frac{r}{\varepsilon} + r - 1 \right),$$

and $A, B \in \mathbb{S}_+^C$ with $\text{null}(B) \subseteq \text{null}(A)$, provided that $B = \Gamma^T \Gamma$ for some Γ of width-w there exists $\mathcal{S}^* \in \binom{V}{r'}$ such that:

$$\text{Tr}(X^T \Pi_{\mathcal{S}^*}^\perp X B) \leq (1+\varepsilon) \frac{\text{Tr}(X^T X A)}{\lambda_{r+1}(A, B)}. \tag{11}$$

For any T, if we replace X and $\mathcal{S}^*$ with $\Pi_T^\perp X$ and $\mathcal{S}^* \cup T$, respectively, then the claim still holds.

Proof. Let $Y \stackrel{\text{def}}{=} X\Gamma^T$. By eq. (10) of Theorem 5.1, we see that

$$\sum_{j \geq r+1} \sigma_j(Y^T Y) \leq \frac{\text{Tr}(X^T X A)}{\lambda_{r+1}(A; B)}.$$

Consequently, for $r'' \stackrel{\text{def}}{=} O(r/\varepsilon)$, we can apply Theorem 5.2 to infer the existence of r'' columns, S, of Y, such that:

$$\|Y_S^\perp Y\|_F^2 \leq (1+\varepsilon) \frac{\text{Tr}(X^T X A)}{\lambda_{r+1}(A; B)}.$$

Since each column of Y is a linear combination of w vectors from X, there exists a set S' of size $|S'| \leq |S| \cdot w = r'' w$ such that each column of $Y_{S'}$ is a linear combination of $X_{S'}$. Finally we let $\mathcal{S}^*$ be the set of all variables which appear in some $f \in S'$. Since $\Pi_{\mathcal{S}^*} X(f) = X(f)$ whenever $f : T \to [k]$ for some $T \subseteq \mathcal{S}^*$, this implies $\Pi_{\mathcal{S}^*} \succeq Y_S^\Pi$. Consequently:

$$\text{Tr}(X^T \Pi_{\mathcal{S}^*}^\perp X B) = \|\Pi_{\mathcal{S}^*}^\perp Y\|_F^2 \leq \|Y_S^\perp Y\|_F^2 \leq (1+\varepsilon) \frac{\text{Tr}(X^T X A)}{\lambda_{r+1}(A; B)}. \qquad \square$$

The following theorem will be our main result in this section.

Theorem 5.4 (Main technical theorem)**.** *Given positive integers r, w and $\varepsilon \in (0,1)$ the following holds. For any $X = [X_S(f)] \in \text{Moment}_{r'}(V, k)$ where*

$$r' \leq \frac{8rw}{\varepsilon^2},$$

and $A, B \in \mathbb{S}_+^C$ with $\text{null}(B) \subseteq \text{null}(A)$, provided that $B = \Gamma^T \Gamma$ for some Γ of width-w, there exists $\mathcal{S}^* \in \binom{V}{r'}$ with:

$$\text{Tr}(X^T \Pi_{\mathcal{S}^*}^\perp X B) + \text{Tr}(X^T \Pi_{\mathcal{S}^*} X A) < (1+\varepsilon) \frac{\text{Tr}(X^T X A)}{\min\{\lambda_{r+1}(A, B), 1\}}. \tag{12}$$

Remark 5.5. We will mainly use Theorem 5.4 with B being a diagonal matrix with non-negative entries on the diagonal. In such case, the diagonal matrix Γ formed by taking the element-wise square root B has width 1 and satisfies $B = \Gamma^T \Gamma$.

Proof. Consider picking the "seed" nodes $\mathcal{S}^*$ in the following iterative way for each $i \in \{1, 2, \ldots\}$. At iteration i, let $S(i)$ be the set of all seeds picked so far (initially $S(0) \leftarrow \emptyset$) and $\mathrm{X}(i+1) \leftarrow \Pi^{\perp}_{S(i)} \mathrm{X}$ (initially $\mathrm{X}(0) \leftarrow \mathrm{X}$). We use Corollary 5.3 on the matrix $\mathrm{X}(i)$ to choose $4rw/\varepsilon$ additional columns, $\Delta S(i+1)$. Finally we set $S(i+1) \leftarrow S(i) \cup \Delta S(i+1)$ and keep iterating until eq. (12) is satisfied. Note that $|S(i)| \leq O\left(\frac{iwr}{\varepsilon}\right)$. Thus it suffices to show that this procedure will stop after $\frac{1}{\varepsilon}$ iterations.

Let $\lambda' \stackrel{\text{def}}{=} \min(\lambda_{r+1}, 1)$ and define

$$\delta_i \stackrel{\text{def}}{=} \mathrm{Tr}(\mathrm{X}(i)^T \mathrm{X}(i)B), \qquad \eta_i \stackrel{\text{def}}{=} \mathrm{Tr}(\mathrm{X}(i)^T \mathrm{X}(i)A), \qquad \eta \stackrel{\text{def}}{=} \mathrm{Tr}(\mathrm{X}^T \mathrm{X}A), \qquad \xi_i \stackrel{\text{def}}{=} \delta_i + \eta - \eta_i.$$

One can easily verify that $\eta = \eta_0 \geq \eta_i \geq \eta_{i+1} \geq 0$, $\delta_i \geq \delta_{i+1}$ and

$$\mathrm{Tr}(\mathrm{X}^T \Pi_{S(i)} \mathrm{X}A) = \eta - \eta_i \implies \mathrm{Tr}(\mathrm{X}^T \Pi^{\perp}_{S(i)} \mathrm{X}B) + \mathrm{Tr}(\mathrm{X}^T \Pi_{S(i)} \mathrm{X}A) = \xi_i.$$

Our goal is to show that, if the procedure did not stop at i^{th} iteration, then:

$$\eta - \eta_i \geq i\varepsilon\eta/2.$$

For $i > \frac{1}{\varepsilon}$, this yield the desired contradiction, meaning that the procedure stopped after that many iterations.

In order to lower bound $\eta - \eta_i$, we first recall the bound from Corollary 5.3:

$$\delta_{i+1} \leq (1 + \varepsilon/2)\frac{\eta_i}{\lambda_{r+1}} \leq (1 + \varepsilon/2)\frac{\eta_i}{\lambda'}. \tag{13}$$

Substituting $\xi_{i+1} \geq (1 + \varepsilon)\frac{\eta}{\lambda'}$ in eq. (13):

$$(1 + \varepsilon)\frac{\eta}{\lambda'} \leq \xi_{i+1} = \delta_{i+1} + \eta - \eta_{i+1} \leq (1 + \varepsilon/2)\frac{\eta_i}{\lambda'} + \eta - \eta_{i+1}. \tag{14}$$

After rearranging terms in eq. (14), our proof is complete:

$$\eta - \eta_{i+1} \geq \left(\frac{1 + \varepsilon/2}{\lambda'}\right)(\eta - \eta_i) + \frac{\varepsilon}{2\lambda'}\eta \geq (1 + \varepsilon/2)(\eta - \eta_i) + (\varepsilon/2)\eta$$

$$\vdots$$

$$\geq \frac{\varepsilon i}{2}\eta. \qquad \square$$

6 Algorithms Based on Independent Rounding

In this section, we will present approximation algorithms which are based on independent rounding scheme. Unless stated otherwise, our algorithms will all follow the same outline:

1. For some matrix A, choose seeds $\mathcal{S}^*$ as described in Theorem 5.4 with $B \stackrel{\text{def}}{=} \mathrm{diag}(A)$.

2. Sample $f : \mathcal{S}^* \to [k]$ with probability $\|\mathrm{X}(f)\|^2$.

3. For each $u \in V$, independently sample a label $j \in [k]$ from the distribution $(\|\mathrm{X}_{u|f}(j)\|^2)$.

6.1 Quadratic Integer Programming

Theorem 6.1. *Consider a quadratic integer programming problem with objective function of the form*

$$\mathbf{x}^T A \mathbf{x}$$

for some $A \in \mathbb{S}_+^{V \times [k]}$ subject to various linear constraints. Then by solving $O(r/\varepsilon^2)$-rounds of Lasserre Hierachy relaxation, we can obtain a solution $\mathbf{x}$ such that:

- $\mathbf{x}^T A \mathbf{x} \le (1 + \varepsilon) \frac{\mathsf{OPT}}{\lambda_r(A, \mathrm{diag}(A))}$.

- $\mathbf{x}$ *is obtained by a faithful and independent rounding. In particular, concentration bounds hold for linear constraints.*

Remark 6.2 (Faster Implementation). By using the faster solver from [GS12a], we can implement a matching algorithm for Theorem 6.1 whose running time is $2^{O(r/\varepsilon^3)} n^{O(1/\varepsilon)}$.

Proof. For fixed $\mathcal{S}^*$, the expected value of the rounding is equal to

$$\mathrm{Tr}(\mathrm{X}^T \Pi_{\mathcal{S}^*} \mathrm{X} A) + \mathrm{Tr}(\mathrm{X}^T \Pi_{\mathcal{S}^*}^\perp \mathrm{X} \, \mathrm{diag}(A))$$

by Lemma 4.3. If we choose $\mathcal{S}^*$ using Theorem 5.4 with $B \leftarrow \mathrm{diag}(A)$, then this value is upper bounded by:

$$(1 + \varepsilon) \frac{\mathrm{Tr}(\mathrm{X}^T \mathrm{X} A)}{\lambda_r(A, \mathrm{diag}(A))}. \qquad \square$$

6.2 Maximum Cut and Unique Games

In this section, we obtain our algorithmic result for Unique Games type problems. Let us quickly recall the definition of the Unique Games problem. An instance of Unique Games consists of a graph $G = (V, C)$, a label set $[k]$, and bijection constraints $\pi_{uv} : [k] \to [k]$ for each edge $\{u, v\}$. The goal is to find a labeling $f : V \to [k]$ that minimizes the number of unsatisfied constraints, where $e = \{u, v\}$ is unsatisfied if $\pi_{uv}(f(u)) \ne f(v)$ (we assume the label of the lexicographically smaller vertex u is projected by π_e).

Given graph G and constraints π, we can construct the lift of G, $\widehat{G} = (V \times [k], \widehat{C})$ where:

$$\widehat{C}_{(u,i),(v,j)} \overset{\text{def}}{=} \begin{cases} C_{u,v} & \text{if } \pi_{uv}(i) = j, \\ 0 & \text{else.} \end{cases}$$

We will use $\widehat{L}$ to denote the associated Laplacian matrix of $\widehat{G}$. Now we are ready to express Unique Games as a quadratic programming problem:

$$\min_{\mathbf{x}} \frac{1}{2} \sum_{u < v} C_{uv} \sum_{i \in [k]} (\mathbf{x}_u(i) - \mathbf{x}_v(\pi_e(i)))^2 = \frac{1}{2} \mathbf{x}^T \widehat{L} \mathbf{x},$$

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

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

The corresponding SDP relaxation is given below:

$$\min_x \frac{1}{2} \mathrm{Tr}\left(\mathrm{X}^T \mathrm{X} \widehat{L}\right)$$

$$\text{st} \ \|\mathrm{X}_\emptyset\|^2 = 1, \ \mathrm{X} = [\mathrm{X}_S(f)] \in \mathrm{Moment}_r(V, 2).$$

24

Remark 6.3. Except for the problem of maximum cut, we are unable to apply Theorem 6.1 directly because there is no known way to relate the rth eigenvalue of L to, say, the $\mathrm{poly}(r)^{th}$ eigenvalue of $\widehat{L}$. We instead use the "projection distance" type bound based on column selection (similar to Section 3.5), after constructing an appropriate embedding to relate the problem to the original graph. $\qquad\square$

Remark 6.4. Although we do not explicitly mention in the theorem statements, we can provide similar guarantees in the presence of constraints similar to graph partitioning problems such as

- constraining labels available to each node,
- constraining fraction of labels used among different subsets of nodes.

For example, the guarantee for maximum cut algorithm immediately carries over to maximum bisection with guarantees on partition sizes similar to minimum bisection. $\qquad\square$

6.2.1 Maximum cut

We first start with the simplest problem fitting in the framework for unique games — finding a maximum cut in a graph. The corresponding minimization problem is minimum uncut, where the objective is minimizing the total cost of uncut pairs. The corresponding $\widehat{L}$ is given by:

$$\widehat{L} = \left(\begin{array}{cc} D & -A \\ -A^T & D \end{array} \right) \in \mathbb{S}_+^{V \times [2]}.$$

For convenience, we also present the integral formulation below:

$$\min_{\mathbf{x}} \frac{1}{2} \sum_{u<v} C_{uv} \left[(\mathbf{x}_u(1) - \mathbf{x}_v(0))^2 + (\mathbf{x}_u(0) - \mathbf{x}_v(1))^2 \right],$$
$$\mathrm{st} \ \ \mathbf{x}_u(1) + \mathbf{x}_u(0) = 1 \quad \forall u \in V,$$
$$\mathbf{x} \in \{0,1\}^{V \times [2]}.$$

Theorem 6.5 (Maximum Cut / Minimum Uncut). *Given $G = (V, C)$, for all $\varepsilon \in (0,1)$ and a positive integer r, by solving $O(r/\varepsilon^2)$ rounds of Lasserre Hierachy relaxation, we can find a subset $U \subset V$ such that the total weight of uncut edges by U is at most*

$$\min \left\{ 1 + \frac{2+\varepsilon}{\lambda_{r+1}(\mathcal{L})}, \ \frac{1+\varepsilon}{\min\{2 - \lambda_{n-r-1}(\mathcal{L}), 1\}} \right\} \cdot \mathrm{OPT}_{\mathrm{SDP}}.$$

Remark 6.6 (Faster Implementation). By using the faster solver from [GS12a], we can implement a matching algorithm for Theorem 6.5 whose running time is $2^{O(r/\varepsilon^3)} n^{O(1/\varepsilon)}$.

Proof. The first bound will follow from the more general result for Unique Games (Theorem 6.7 below), so we focus on the second bound claiming an approximation ratio of $(1+\varepsilon)/\min\{2 - \lambda_{n-r-1}, 1\}$. The normalized Laplacian matrix for $\widehat{L}$ is given by

$$\widehat{\mathcal{L}} = \left[\begin{array}{cc} I & -\mathcal{A} \\ -\mathcal{A} & I \end{array} \right],$$

where $\mathcal{A}$ is the normalized adjacency matrix, $\mathcal{A} \overset{\mathrm{def}}{=} D^{-1/2} A D^{-1/2}$.

By direct substitution, it is easy to see that, for every eigenvector q_i of constraint graph G's normalized Laplacian matrix, $\mathcal{L}$, there are two corresponding eigenvectors for $\widehat{\mathcal{L}}$, $\begin{bmatrix} \frac{1}{\sqrt{2}}q_i \\ \frac{1}{\sqrt{2}}q_i \end{bmatrix}$ and

$\begin{bmatrix} \frac{1}{\sqrt{2}}q_i \\ -\frac{1}{\sqrt{2}}q_i \end{bmatrix}$ whose corresponding eigenvalues are given by λ_i and $2 - \lambda_i$ respectively. As a convention, we will refer to these as **even** and **odd** eigenvectors, respectively.

For any node $u \in V$, we can express $X_u(i)$ for $i \in [2]$ as

$$X_u(i) = \|X_u(i)\|^2 X_{\emptyset} + (-1)^i \|X_u(1)\| \|X_u(0)\| Y_u ,$$

where Y_u is a unit vector orthogonal to $X_{\emptyset}$, $\langle X_{\emptyset}, Y_u \rangle = 0$. Note that $X_{\emptyset}^{\perp} X_u(1) = y_u = -X_{\emptyset}^{\perp} X_u(0)$. Hence $X^T X_{\emptyset}^{\perp} X$ has **zero** correlation with even eigenvectors of $\widehat{L}$. Therefore we have the following identity. For any S:

$$\mathrm{Tr}(X^T \Pi_S^{\perp} X \widehat{L}) = \mathrm{Tr}(X^T \Pi_S^{\perp} X(D + A)).$$

We can slightly modify Theorem 5.4 to take into account only the eigenvectors of $\widehat{\mathcal{L}}$ with which $X_{\emptyset}^{\perp} X$ has non-zero correlation. Using the bound from Theorem 5.4, we can then show that total cost of uncut edges is bounded by $(1 + \varepsilon)\frac{\mathrm{OPT_{SDP}}}{\min(\lambda_{r+1}(I+\mathcal{A}),1)}$. The proof is now complete by noting that $\lambda_{r+1}(I + \mathcal{A}) = 2 - \lambda_{n-r-1}(\mathcal{L})$. $\qquad\square$

6.2.2 Unique Games

In this section, we prove our main result for approximating Unique Games. For this problem, we will choose the seed set in a different way, rather than using Theorem 5.4.

Theorem 6.7 (Unique Games). *For any instance of Unique Games with label set $[k]$, constraint graph $G = (V, C)$ and constraints π, for all $\varepsilon \in (0,1)$ and a positive integer r, the following holds: By solving $O(r/\varepsilon)$ rounds of Lasserre Hierachy relaxation, we can find a labeling $f : V \to [k]$ such that the total cost of unsatisfied constraints is at most:*

$$\sum_{u<v} C_{uv} \mathbb{1}_{[f(v) \neq \pi_e(f(u))]} \leq \left(1 + \frac{2+\varepsilon}{\lambda_{r+1}(\mathcal{L})}\right) \mathrm{OPT_{SDP}}.$$

Remark 6.8 (Faster Implementation). By using the faster solver from [GS12a], we can implement a matching algorithm for Theorem 6.7 whose running time is $k^{O(r/\varepsilon)} n^{O(1)}$.

Proof. For fixed $\mathcal{S}^*$ (to be chosen later), we can use Lemma 4.3 to express the probability of violating constraint $\pi : [k] \to [k]$ between u and v as:

$$
\begin{aligned}
\frac{1}{2}\mathbb{E}\Big[\sum_{i,j:j=\pi(i)} (x_u(i) - x_v(j))^2\Big] =& \frac{1}{2}\sum_{i,j:j=\pi(i)} \langle \Pi_{\mathcal{S}^*} X_u(i), \Pi_{\mathcal{S}^*} X_v(j)\rangle + \|\Pi_{\bar{\mathcal{S}}^*}^{\perp} X_u(i)\|^2 + \|\Pi_{\bar{\mathcal{S}}^*}^{\perp} X_u(\pi(i))\|^2 \\
=& \frac{1}{2}\sum_{i,j:j=\pi(i)} \|X_u(i) - X_v(j)\|^2 + \langle \Pi_{\bar{\mathcal{S}}^*}^{\perp} X_u(i), \Pi_{\bar{\mathcal{S}}^*}^{\perp} X_v(j)\rangle \\
\leq& \frac{1}{2}\sum_{i,j:j=\pi(i)} \|X_u(i) - X_v(j)\|^2 + \frac{1}{2}\|\Pi_{\bar{\mathcal{S}}^*}^{\perp} X_u(i)\|^2 + \frac{1}{2}\|\Pi_{\bar{\mathcal{S}}^*}^{\perp} X_v(j)\|^2.
\end{aligned}
$$

$$(15)$$

26

Hence the total expected cost of violated constraints is bounded by:

$$\frac{1}{2} \sum_{u < v} C_{uv} \mathbb{E}\Big[\sum_{i,j:j=\pi(i)} (\mathcal{X}_u(i) - \mathcal{X}_v(j))^2 \Big] \leq \frac{1}{2}\left(\eta + \sum_u D_u \sum_i \|\Pi_{\mathcal{S}^*}^\perp \mathrm{X}_u(i)\|^2 \right),$$

where D_u is the degree of node u, and $\eta = \mathrm{Tr}(\mathrm{X}^T \mathrm{X} \widehat{L})$ is twice the SDP value.

Now consider the embedding given in Theorem 6.9 below, $\{\mathrm{X}_u(i)\}_{i \in [k]} \mapsto \mathrm{Y}_u$ with $\mathrm{Y} = [\mathrm{Y}_u]_u$. Then, by Theorem 6.9 (3):

$$\sum_u D_u \sum_i \|\Pi_{\mathcal{S}^*}^\perp \mathrm{X}_u(i)\|^2 \leq \sum_u D_u \|\mathrm{Y}_{\mathcal{S}^*}^\perp \mathrm{Y}_u\|^2 = \mathrm{Tr}(\mathrm{Y}^T \mathrm{Y}_{\mathcal{S}^*}^\perp \mathrm{Y} D).$$

We can use Theorem 5.2 to choose $r' = O(r/\varepsilon)$ columns of $D^{1/2}\mathrm{Y}$, $\mathcal{S}^*$. Now we can mimick the proof of Corollary 5.3 to show that:

$$\mathrm{Tr}(\mathrm{Y}^T \mathrm{Y}_{\mathcal{S}^*}^\perp \mathrm{Y} D) \leq \frac{(1 + \varepsilon/2)\,\mathrm{Tr}(\mathrm{Y}^T \mathrm{Y} L)}{\lambda_r(\mathcal{L})} \leq (2 + \varepsilon) \frac{\mathrm{Tr}(\mathrm{X}^T \mathrm{X} \widehat{L})}{\lambda_r(\mathcal{L})} = \frac{(2 + \varepsilon)\eta}{\lambda_r(\mathcal{L})},$$

where we used Theorem 6.9(2) in the last inequality. Substituting this back into eq. (15), we obtain our final bound on the total expected cost of violated constraints:

$$\frac{1}{2} \sum_{u < v} C_{uv} \mathbb{E}\Big[\sum_{i,j:j=\pi(i)} (\mathcal{X}_u(i) - \mathcal{X}_v(j))^2 \Big] \leq \mathsf{OPT}_{\mathrm{SDP}}\left(1 + \frac{2 + \varepsilon}{\lambda_r(\mathcal{L})} \right). \qquad \square$$

6.2.3 A useful embedding

We now turn to the embedding used in the above proof.

Theorem 6.9. *Given vectors* $(\mathrm{X}_u(i))_{u \in V, i \in [k]}$, *suppose they satisfy the following: For any* $u \in V$, *whenever* $f, g \in [k]^u$ *are two different labelings of* u, $f \neq g$,

$$\langle \mathrm{X}_u(f), \mathrm{X}_u(g) \rangle = 0.$$

Then there exists an embedding $(\mathrm{X}_u(f))_f \mapsto \mathrm{Y}_u$ *for all* $u \in V$ *such that:*

1. *For any* $u \in V$, $\|\mathrm{Y}_u\|^2 = \sum_f \|\mathrm{X}_u(f)\|^2$.

2. *For any* $u, v \in V$ *and any permutation* $\pi \in \mathrm{Sym}([k])$:

$$\sum_{i \in [k]} \|\mathrm{X}_u(i^u) - \mathrm{X}_v(\pi(i)^v)\|^2 \geq \frac{1}{2}\|\mathrm{Y}_u - \mathrm{Y}_v\|^2.$$

3. *For any subset* $S \subseteq V$ *and any* $u \in V$, *if we let* P_S, Q_S *be the projection matrices onto the spans of* $\{\mathrm{X}_s(f)\}_{s \in S, f \in [k]}$ *and* $\{\mathrm{Y}_s\}_{s \in S}$, *respectively, then:*

$$\|Q_S^\perp \mathrm{Y}_u\|^2 \geq \sum_{f \in [k]^u} \|P_S^\perp \mathrm{X}_u(f)\|^2.$$

27

In the rest of this section, we will prove Theorem 6.9. 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 $Y_u \in \mathbb{R}^m \otimes \mathbb{R}^m$ as

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

Observation 6.10. *For vectors $x, y \in \mathbb{R}^m$, $\sum_{i=1}^m \langle x, e_i \rangle \langle y, e_i \rangle = \langle x, y \rangle$.*

The first property of the vectors Y_u follows from this observation easily:

$$\begin{aligned}
\|Y_u\|^2 &= \sum_i \sum_{f,g} \langle \overline{X_u(f)}, e_i \rangle \langle \overline{X_u(g)}, e_i \rangle \langle X_u(f), X_u(g) \rangle \\
&= \sum_{f,g} \langle X_u(f), X_u(g) \rangle \sum_i \langle \overline{X_u(f)}, e_i \rangle \langle \overline{X_u(g)}, e_i \rangle \\
&= \sum_{f,g} \langle X_u(f), X_u(g) \rangle \langle \overline{X_u(f)}, \overline{X_u(g)} \rangle \\
&= \sum_f \|X_u(f)\|^2 .
\end{aligned}$$

We prove the second property in Claim 6.11 and third one in Claim 6.12 below.

Claim 6.11. *For any permutation $\pi \in \mathrm{Sym}([k])$:*

$$\frac{1}{2} \|Y_u - Y_v\|^2 \leq \sum_{f \in [k]} \|X_u(f) - X_v(\pi(f))\|^2$$

Proof. Without loss of generality, we assume π is the identity permutation. We have

$$\begin{aligned}
\frac{1}{2} \|Y_u - Y_v\|^2 &= \frac{\|Y_u\|^2 + \|Y_v\|^2}{2} - \langle Y_u, Y_v \rangle \\
&= \frac{\|Y_u\|^2 + \|Y_v\|^2}{2} - \sum_{f,g} \langle X_u(f), X_v(g) \rangle \sum_i \langle \overline{X_u(f)}, e_i \rangle \langle \overline{X_v(g)}, e_i \rangle \\
&= \sum_f \frac{\|X_u(f)\|^2 + \|X_v(f)\|^2}{2} - \sum_{f,g} \langle \overline{X_u(f)}, \overline{X_v(g)} \rangle^2 \|X_u(f)\| \|X_v(g)\|
\end{aligned}$$

The sum over all pairs is lower bounded by summing only the corresponding pairs:

$$\leq \frac{1}{2} \sum_f \left(\|X_u(f)\|^2 + \|X_v(f)\|^2 - 2\langle X_u(f), X_v(f) \rangle \langle \overline{X_u(f)}, \overline{X_v(f)} \rangle \right)$$

$$= \frac{1}{2} \sum_f \|X_u(f) - X_v(f)\|^2 + \sum_f \langle X_u(f), X_v(f) \rangle \underbrace{\left(1 - \langle \overline{X_u(f)}, \overline{X_v(f)} \rangle \right)}_{\geq 0} \qquad (16)$$

Since the coefficient of $\langle X_u(f), X_v(f) \rangle$ is positive, we can use Cauchy-Schwarz inequality to replace $\langle X_u(f), X_v(f) \rangle$ with $\|X_u(f)\| \cdot \|X_v(f)\|$ in eq. (16) and obtain:

$$\leq \frac{1}{2} \sum_f \|X_u(f) - X_v(f)\|^2 + \sum_f \left(\|X_u(f)\| \cdot \|X_v(f)\| - \langle X_u(f), X_v(f) \rangle \right) \qquad (17)$$

Using inequality $\|X_u(f)\| \cdot \|X_v(f)\| \le \frac{1}{2}\left(\|X_u(f)\|^2 + \|X_v(f)\|^2\right)$ on eq. (17):

$$\le \frac{1}{2}\sum_f \left(\|X_u(f) - X_v(f)\|^2 + \|X_u(f)\|^2 + \|X_v(f)\|^2 - 2\langle X_u(f), X_v(f)\rangle\right)$$

$$= \sum_f \|X_u(f) - X_v(f)\|^2. \qquad \square$$

Claim 6.12.

$$\|Q_S^\perp Y_u\|^2 \ge \sum_f \|P_S^\perp X_u(f)\|^2.$$

Proof. For any $\theta \in \mathbb{R}^S$:

$$\|Y_u - \sum_v \theta_v Y_v\|^2 = \sum_{i=1}^m \left\| \sum_f \langle \overline{X_u(f)}, e_i\rangle X_u(f) - \underbrace{\sum_{v \in S, g} \theta_v \langle \overline{X_v(g)}, e_i\rangle X_v(g)}_{P_S \Theta_i} \right\|^2. \tag{18}$$

Substituting $\alpha_f = P_S^\perp X_u(f)$ and $\beta_f = P_S X_u(f)$ in eq. (18):

$$= \sum_{i=1}^m \left\| \sum_f \langle \overline{X_u(f)}, e_i\rangle (\alpha_f + \beta_f) - P_S \Theta_i \right\|^2$$

$$= \sum_{i=1}^m \left\| \sum_f \langle \overline{X_u(f)}, e_i\rangle \alpha_f \right\|^2 + \left\| \sum_f \langle \overline{X_u(f)}, e_i\rangle \beta_f - P_S \Theta_i \right\|^2$$

$$\ge \sum_{i=1}^m \left\| \sum_f \langle \overline{X_u(f)}, e_i\rangle \alpha_f \right\|^2$$

$$= \sum_{f,g} \langle \alpha_f, \alpha_g\rangle \sum_{i=1}^m \langle \overline{X_u(f)}, e_i\rangle \langle \overline{X_u(g)}, e_i\rangle$$

$$= \sum_{f,g} \langle \alpha_f, \alpha_g\rangle \langle \overline{X_u(f)}, \overline{X_u(g)}\rangle = \sum_f \|\alpha_f\|^2 = \sum_f \|P_S^\perp X_u(f)\|^2. \qquad \square$$

This concludes the proof of Theorem 6.9, therefore also the proof of Theorem 6.7.

6.3 Independent Set

Our final algorithmic result is on finding independent sets in a graph. For simplicity, we focus on unweighted graphs though the extension for graphs with non-negative vertex weights is straight-forward. As usual, we denote by $\alpha(G)$ the size of the largest independent set in G. Let $d_{\max}$ denote the maximum degree of a vertex of G.

Theorem 6.13. *Given $0 < \varepsilon < 1$, positive integer r, a graph G with $d_{\max} \ge 3$, by solving $O(r/\varepsilon^2)$ rounds of Lasserre Hierachy relaxation, we can find an independent set $I \subseteq V$ such that*

$$|I| \ge \alpha(G) \cdot \min\left\{ \frac{1}{2d_{\max}}\left(\frac{1}{(1-\varepsilon)\min\{2 - \lambda_{n-r-1}(\mathcal{L}), 1\}} - 1\right)^{-1}, 1 \right\}. \tag{19}$$

Remark 6.14 (Faster Implementation). By using the faster solver from [GS12a], we can implement a matching algorithm for Theorem 6.13 whose running time is $2^{O(r)}n^{O(1)}$.

Remark 6.15. The above bound (19) implies that if λ_{n-r-1}, which is the $(r+1)^{\text{st}}$ largest eigenvalue of the normalized Laplacian $\mathcal{L}$, is very close to 1, then we can find large independent sets in $n^{O(r/\varepsilon^2)}$ time. In particular, if it is at most $1 + \frac{1}{4\mathrm{d}_{\max}}$, then taking $\varepsilon = O(1/\mathrm{d}_{\max})$, we can find an optimal independent set. The best approximation ratio for independent set in terms of $\mathrm{d}_{\max}$ is about $O\left(\frac{\mathrm{d}_{\max}\cdot\log\log \mathrm{d}_{\max}}{\log \mathrm{d}_{\max}}\right)$ [Hal98, Hal02]. The bound in eq. (19) gives a better approximation ratio when $\lambda_{n-r-1} \leq 1 + O\left(\frac{1}{\log \mathrm{d}_{\max}}\right)$. $\qquad\square$

Proof. (of Theorem 6.13) Consider the following integer program for finding largest independent set in G:

$$\max \sum_u \mathbf{x}_u$$
$$\text{st } \mathbf{x}_u\mathbf{x}_v = 0 \qquad \text{for any edge } e = (u,v) \in E\,,$$
$$\mathbf{x} \in \{0,1\}^V.$$

Its SDP relaxation is given by:

$$\max \sum_u \|X_u\|^2$$
$$\text{st } \langle X_u, X_v\rangle = 0 \qquad \text{for any edge } e = (u,v) \in E\,,$$
$$\|X_\emptyset\|^2 = 1, \ X = [X_S(f)] \in \mathrm{Moment}_r(V,2);$$

with value $\sum_u \|X_u\|^2 = \mu$.

We will choose $\mathcal{S}^*$ by applying Theorem 5.4 on matrices $I_V + \mathcal{A}$ and I_V, so that:

$$\mathrm{Tr}(X^T\Pi_{\mathcal{S}^*}^{\perp}XI_V) + \mathrm{Tr}(X^T\Pi_{\mathcal{S}^*}X(I_V + \mathcal{A})) \ \leq\ \frac{\mathrm{Tr}(X^TX(I_V + \mathcal{A}))}{\lambda'}$$
$$=\ \frac{1}{\lambda'}\mathrm{Tr}(X^TXI_V) \quad \text{since} \quad \mathrm{Tr}(X^TX\mathcal{A}) = 0,$$

where $\lambda' = \frac{\min\{\lambda_{r+1}(I+\mathcal{A}),1\}}{1+\varepsilon} = \frac{\min\{2-\lambda_{n-r-1}(\mathcal{L}),1\}}{1+\varepsilon}$.

After sampling f and $x \in \{0,1\}^V$, we convert x into an independent set as follows.

1. For each u, if $x_u = 1$ then let $I \leftarrow I \cup \{u\}$ with probability p_u which we will specify later.

2. After the first step, for each edge $e = \{u,v\}$, if $\{u,v\} \subseteq I$, we choose one end point randomly, say u, and set $I \leftarrow I \setminus \{u\}$.

By construction, final I is an independent set.

Note that for any u, the probability that u will be in the final independent set I is at least:

$$\mathrm{Prob}\big[u \in I\big] \geq \mathbb{E}\big[p_u x_u\big] - \frac{1}{2}\mathbb{E}\Big[\sum_{v\in N(u)} p_u p_v x_u x_v\Big]$$
$$= p_u\|X_u\|^2 - \frac{1}{2}\sum_{v\in N(u)} p_u p_v\langle \Pi_{\mathcal{S}^*}X_u, \Pi_{\mathcal{S}^*}X_v\rangle. \tag{20}$$

By eq. (20), the expected size of the independent set found by the algorithm satisfies

$$\mathbb{E}\big[|I|\big] \geq \sum_u p_u \|\mathrm{X}_u\|^2 - \sum_{\{u,v\}\in E} p_u p_v \langle \Pi_{\mathcal{S}^*}\mathrm{X}_u, \Pi_{\mathcal{S}^*}\mathrm{X}_v\rangle. \tag{21}$$

Note that for every edge $\{u,v\} \in E$,

$$\langle \Pi_{\mathcal{S}^*}x_u, \Pi_{\mathcal{S}^*}x_v\rangle = \sum_{f\in[2]^S} \frac{\langle \mathrm{X}_{\mathcal{S}^*}(f), x_u\rangle\langle \mathrm{X}_{\mathcal{S}^*}(f), \mathrm{X}_v\rangle}{\|\mathrm{X}_{\mathcal{S}^*}(f)\|^2} \geq 0 \,. \tag{22}$$

We now consider two cases.

Case 1: $\langle \Pi_{\mathcal{S}^*}\mathrm{X}_u, \Pi_{\mathcal{S}^*}\mathrm{X}_v\rangle = 0$ for all edges $\{u,v\} \in E$. In this case, we take $p_u = 1$ for all $u \in V$, and by eq. (21), we find an independent set of expected size at least $\mu \geq \alpha(G)$.

Case 2: In this case, we have

$$\sum_{\{u,v\}\in E} \langle \Pi_{\mathcal{S}^*}\mathrm{X}_u, \Pi_{\mathcal{S}^*}\mathrm{X}_v\rangle = \frac{1}{2}\,\mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}A) > 0 \,, \tag{23}$$

where A is the adjacency matrix of G. Let $\mathcal{A} = D^{-1/2}AD^{-1/2}$ be the normalized adjacency matrix, and define

$$\xi \overset{\mathrm{def}}{=} \frac{\mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}\mathcal{A})}{\mathrm{Tr}(\mathrm{X}^T\mathrm{X}I_V)}. \tag{24}$$

By eqs. (22) and (23), we have $\xi > 0$.

We now pick $p_u = \frac{\alpha}{\sqrt{d_u}}$ for all $u \in V$, where we will optimize the choice of α shortly. For this choice, we have

$$
\begin{aligned}
\mathbb{E}\big[|I|\big] &\geq \alpha \sum_u \frac{1}{\sqrt{d_u}}\|\mathrm{X}_u\|^2 - \frac{1}{2}\alpha^2\,\mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}\mathcal{A}) \\
&\geq \frac{\alpha}{\sqrt{\mathrm{d}_{\max}}} \sum_u \|\mathrm{X}_u\|^2 - \frac{1}{2}\alpha^2\,\mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}\mathcal{A}) \\
&= \mu\left(\frac{\alpha}{\sqrt{\mathrm{d}_{\max}}} - \frac{1}{2}\alpha^2 \underbrace{\frac{\mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}\mathcal{A})}{\mathrm{Tr}(\mathrm{X}^T\mathrm{X}I_V)}}_{\xi}\right)
\end{aligned}
$$

This expression is maximized when $\alpha = \frac{1}{\xi\cdot\sqrt{\mathrm{d}_{\max}}}$, for which it becomes:

$$\mathbb{E}\big[|I|\big] \geq \frac{\mu}{2\mathrm{d}_{\max}}\frac{1}{\xi}. \tag{25}$$

We know that:

$$
\begin{aligned}
\frac{\mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}^{\perp}\mathrm{X}) + \mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}(I_v + \mathcal{A}))}{\mathrm{Tr}(\mathrm{X}^T\mathrm{X}I_V)} &= \frac{\mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}^{\perp}\mathrm{X}I_V) + \mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}I_V) + \mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}\mathcal{A})}{\mathrm{Tr}(\mathrm{X}^T\mathrm{X}I_V)} \\
&= \frac{\mathrm{Tr}(\mathrm{X}^T\mathrm{X}I_V) + \mathrm{Tr}(\mathrm{X}^T\Pi_{\mathcal{S}^*}\mathrm{X}\mathcal{A})}{\mathrm{Tr}(\mathrm{X}^T\mathrm{X}I_V)} = 1 + \xi.
\end{aligned}
$$

Thus, for our choice of $\mathcal{S}^*$, we obtain that $\xi \leq \frac{1}{\lambda'} - 1$. Substituting this back into eq. (25), we have

$$\mathbb{E}\big[|I|\big] \geq \frac{\mu}{2\mathrm{d}_{\max}}\frac{1}{1/\lambda' - 1}. \qquad \qquad \square$$

Problem	Denominator	$\lambda_r(G, H)$ **Becomes**
Uniform Sparsest Cut	$\frac{1}{n} \sum_u \mathbf{x}_u(n - \sum_u \mathbf{x}_u)$	$\lambda_r(L)$
Edge Expansion	$\min(\sum_u \mathbf{x}_u, n - \sum_u \mathbf{x}_u)$	$\lambda_r(L)$
Normalized Cut	$\frac{1}{m} \sum_u d_u\mathbf{x}_u(m - \sum_u d_u\mathbf{x}_u)$	$\lambda_r(\mathcal{L})$
Conductance	$\min(\sum_u d_u\mathbf{x}_u, m - \sum_u d_u\mathbf{x}_u)$	$\lambda_r(\mathcal{L})$

Figure 1: Some variations of Sparsest Cut problem with formulations similar to eq. (26). In all cases, Theorem 7.4, Corollaries 7.6 and 7.7 hold. In the above table, d denotes the degrees and m denotes the total capacity.

7 Algorithms Based on Threshold Based Rounding

In this section, we will present approximation algorithms which use the threshold based rounding scheme. All problems share the following structure: Given a capacity graph, $G = (V, C)$, and a *connected* demand graph, $H = (V, D)$, find a non-empty $U \subset V$ which minimizes the total capacity cut subject to various constraints involving the total demand cut.

Recall that, for L_H and Γ_H being the Laplacian and edge-node incidence matrices of graph H, respectively, we have $L_H = \Gamma_H^T \Gamma_H$. Since H is connected, null space of L_H consists only of constant vector. Hence $\mathrm{null}(L_H) \subseteq \mathrm{null}(L_G)$. Also the width of Γ_H is 2. Therefore matrices L_G and L_H satisfy all the requirements in Corollary 5.3, which we use for seed selection. The following is the randomized rounding algorithm we will use in this section:

1. Choose seeds $\mathcal{S}^*$ as described in Corollary 5.3 where $A \overset{\mathrm{def}}{=} L_G$, $B \overset{\mathrm{def}}{=} L_H$ and $\Gamma \overset{\mathrm{def}}{=} \Gamma_H$.

2. Sample $f : \mathcal{S}^* \to [k]$ with probability $\|\mathrm{X}(f)\|^2$.

3. Choose a threshold $\tau \in (0, 1]$ uniformly at random.

4. For each $u \in V$, set $\mathbf{x}_u \leftarrow 1$ if $\|\mathrm{X}_{u|f}\|^2 \geq \tau$ and $\mathbf{x}_u \leftarrow 0$ otherwise.

In all our analysis, we will prove that the procedure will output correct answer with positive probability. But we can easily derandomize the above procedure by enumerating over all:

(i) $f : \mathcal{S}^* \to \{0, 1\}$ in step (2),

(ii) $\tau \in \{\|\mathrm{X}_{u|f}\|^2 \mid u \in V\}$ in step (3).

After these changes, the running time of rounding is $2^{O(r)} \mathrm{poly}(n)$, which does not affect the total running time.

7.1 Sparsest Cut and Variations

In the problem of Non-Uniform Sparsest Cut, our goal is to minimize the ratio of total capacity cut versus total demand cut:

$$\Phi(G, H) \overset{\mathrm{def}}{=} \min_{\mathbf{x} \in \{0,1\}^V : L_H \mathbf{x} \neq 0} \frac{\mathbf{x}^T L_G \mathbf{x}}{\mathbf{x}^T L_H \mathbf{x}}. \tag{26}$$

This includes problems such as: Uniform Sparsest Cut, Normalized Cut, etc... (see also Figure 1). The following is a relaxation of eq. (26):

$$\min \quad \frac{\sum_{u<v} C_{u,v} \|X_u - X_v\|^2}{\sum_{u<v} D_{u,v} \|X_u - X_v\|^2} = \frac{\text{Tr}(X^T X L_G)}{\text{Tr}(X^T X L_H)} \tag{27}$$

$$\text{st} \quad \sum_{u<v} D_{u,v} \|X_u - X_v\|^2 > 0,$$

$$\|X_\emptyset\|^2 = 1, \ X = [X_S(f)] \in \text{Moment}_r(V, 2).$$

We will refer to the optimum of eq. (27) as Φ_{SDP}. Even though eq. (27) is not convex (it is quasi-convex), there is an equivalent convex formulation.

Lemma 7.1. *The following SDP is equivalent to eq. (27):*

$$\min \quad \text{Tr}(W^T W L_G) \tag{28}$$

$$\text{st} \quad \text{Tr}(W^T W L_H) = 1,$$

$$\|W_\emptyset\|^2 > 0, \ W = [W_S(f)] \in \text{Moment}_r(V, 2);$$

Remark 7.2. The constraint $\|W_\emptyset\|^2 > 0$ in eq. (28) is redundant, but we included it for the sake of clarity.

(of Lemma 7.1). Given a feasible solution X for formulation (27), let $W = [W_S(f)]$ be $W \leftarrow \frac{1}{\sqrt{\text{Tr}(X^T X L_H)}} X$. It is easy to see that $\sum_{u<v} D_{u,v} \|W_u - W_v\|^2 = 1$ and objective values are equal. Finally $\|W_\emptyset\|^2 = \frac{1}{\sum_{u<v} D_{u,v} \|X_u - X_v\|^2} > 0$ since $0 < \sum_{u<v} D_{u,v} \|X_u - X_v\|^2$.

For the other direction of equivalence, suppose W is a feasible solution of eq. (28). For each $T \in \binom{V}{\leq r'}, f \in \{0,1\}^T$, let $X_T(f) \leftarrow \frac{1}{\|W_\emptyset\|} W_T(f)$. It is easy to see that the objective values are equal. The rest of the proof for X being a feasible solution of eq. (27) follows similarly to the previous direction. $\square$

Remark 7.3. Our rounding procedure is scale invariant, thus eq. (28) is sufficient for rounding purposes. But we chose to first present eq. (27) as it is more intuitive.

Theorem 7.4. *Given capacity graph $G = (V, C)$ and a connected demand graph $H = (V, D)$, positive r and ε; one can find $x \in \{0,1\}^V$ such that*

$$\frac{x^T L_G x}{x^T L_H x} \leq \Phi_{SDP} \left(1 - (1+\varepsilon) \frac{\Phi_{SDP}}{\lambda_r(G, H)}\right)^{-1}$$

provided that $\lambda_r(G, H) \geq (1+\varepsilon)\Phi_{SDP}$, by rounding $O(r/\varepsilon)$-rounds of Lasserre Hierachy relaxation.

Remark 7.5 (Faster Implementation). By using the faster solver from [GS12a], we can implement a matching algorithm for Theorem 7.4 whose running time is $2^{O(r/\varepsilon)} n^{O(1)}$.

Proof. For fixed f, the probability of separating u and v is at most $\|X_{u|f} - X_{v|f}\|^2$. Taking expectation over f, we see that:

$$\mathbb{E}_{f,\tau}\left[(x_u - x_v)^2\right] \leq \mathbb{E}_f\left[\|X_{u|f} - X_{v|f}\|^2\right] = \|X_u - X_v\|^2.$$

By Claim 4.7, we can lower bound the probability of separating u and v by:

$$\mathbb{E}_{f,\tau}\left[(x_u - x_v)^2\right] \geq \|X_u - X_v\|^2 - \|\Pi_{\tilde{S}^*}^\perp(X_u - X_v)\|^2.$$

Consequently,

$$\frac{\mathbb{E}\left[x^T L_G x\right]}{\mathbb{E}\left[x^T L_H x\right]} \leq \frac{\text{Tr}(X^T X L_G)}{\text{Tr}(X^T X L_H) - \text{Tr}(X^T \Pi_{\tilde{S}^*}^\perp X L_H)}.$$

Since we choose $\mathcal{S}^*$ as in Corollary 5.3, we can use eq. (11) to bound $\mathrm{Tr}(X^T \Pi_{\mathcal{S}^*}^{\perp} X L_H)$:

$$\mathrm{Tr}(X^T \Pi_{\mathcal{S}^*}^{\perp} X L_H) \leq (1+\varepsilon) \frac{\mathrm{Tr}(X^T X L_G)}{\lambda_r} = \frac{\mathrm{Tr}(X^T X L_H)}{\lambda_r} \Phi_{SDP}.$$

Substituting this back into the previous upper bound yields:

$$\frac{\mathbb{E}_{f,\tau}[x^T L_G x]}{\mathbb{E}[x^T L_H x]} \leq \frac{\mathrm{Tr}(X^T X L_G)}{\mathrm{Tr}(X^T X L_H)(1 - (1+\varepsilon)\frac{\Phi_{SDP}}{\lambda_r})} = \frac{\Phi_{SDP}}{1 - (1+\varepsilon)\frac{\Phi_{SDP}}{\lambda_r}}.$$

This implies the existence of f and τ for which the claim holds. $\qquad\square$

Even though Theorem 7.4 is not guaranteed to output a good solution when $\lambda_r \leq \Phi_{SDP}$, we can still obtain an unconditional bound on the integrality gap:

Corollary 7.6. *The integrality gap after $O(r)$ rounds is bounded by:*

$$\frac{\Phi}{\Phi_{SDP}} \leq O\Big[\max\big(\Phi/\lambda_r, 1\big)\Big].$$

Proof. Let $\varepsilon \leftarrow \frac{1}{2}$. If $\frac{\lambda}{\Phi_{SDP}} \geq \frac{1+\varepsilon}{1-\varepsilon} = 3$; then the algorithm will output a subset whose sparsity is at most $2\Phi_{SDP}$, which implies $\frac{\Phi}{\Phi_{SDP}} \leq 2$. Otherwise, $\frac{\lambda}{3} < \Phi_{SDP}$. Taking the reciprocal of both sides and multiplying by Φ, we see that $\frac{\Phi}{\Phi_{SDP}} < \frac{3\Phi}{\lambda}$. $\qquad\square$

Finally we highlight the two interesting regimes of Theorem 7.4 in Corollary 7.7.

Corollary 7.7. *Given capacity graph $G = (V, C)$ and a connected demand graph $H = (V, D)$, positive r and δ; one can find a subset whose sparsity is at most:*

- *(Near Optimal) $\Phi(1 + \delta)$ if $\Phi < \frac{1}{2}\delta \cdot \lambda_r$;*
- *(Constant Factor) $\frac{\Phi}{\delta}$ if $\Phi < (1 - 2\delta)\lambda_r$.*

by rounding $O(r/\delta)$-rounds of Lasserre Hierachy relaxation.

7.2 Minimizing Capacity Cut Under Packing Constraints

In this section, we will consider the problem of finding a subset of nodes with respect to some packing constraint while minimizing the total capacity cut. This includes problems such as balanced separator, minimum bisection. Formally, given a graph G with Laplacian matrix L, a non-negative diagonal matrix $D = \mathrm{diag}(d) \in \mathbb{S}_+^V$ (think of $d \in \mathbb{R}_+^V$ as node weights with $\sum_u d_u = m$) and positive real μ; we consider the following problem:

$$\begin{aligned} \min \quad & \mathbf{x}^T L \mathbf{x} \\ \text{st} \quad & \sum_u d_u \mathbf{x}_u = \mathbf{x}^T D \mathbf{x} = \mu, \\ & \mathbf{x} \in \{0,1\}^V. \end{aligned} \tag{29}$$

Our main result is the following.

Theorem 7.8. *Given a problem of the form eq. (29) and positive integer r, if $\delta > 0$ satisfies*

$$\lambda_r(L, D) \geq \frac{1+\epsilon}{\delta} \frac{\eta}{\mu^2/m},$$

then, by rounding $O(r/\varepsilon)$ rounds of Lasserre Hierarchy relaxation, one can find $x \in \{0,1\}^V$ such that:

(i) $\left| \varkappa^T D \varkappa - \mu \right| \leq \delta^{1/3} \mu$,

(ii) $\varkappa^T L \varkappa \leq \dfrac{\eta}{1 - \delta^{1/3}}$.

Here η is the optimum value of relaxation.

Remark 7.9 (Faster Implementation)**.** By using the faster solver from [GS12a], we can implement a matching algorithm for Theorem 7.8 whose running time is $2^{O(r/\varepsilon)} n^{O(1)}$.

Proof. Since we choose $\mathcal{S}^*$ as in Corollary 5.3, we can use eq. (11) to bound $\mathrm{Tr}(\mathrm{X}^T \Pi_{\mathcal{S}^*}^{\perp} \mathrm{X} D)$:

$$\mathrm{Tr}(\mathrm{X}^T \Pi_{\mathcal{S}^*}^{\perp} \mathrm{X} D) \leq (1 + \varepsilon) \frac{\mathrm{Tr}(\mathrm{X}^T \mathrm{X} L)}{\lambda_r} \leq \frac{\delta \mu^2}{m}.$$

Recall that $\mathrm{Tr}(\mathrm{X}^T \Pi_{\mathcal{S}^*}^{\perp} \mathrm{X} D) = \mathbb{E}_f \big[\sum_u d_u \| \mathrm{X}_{\emptyset|f}^{\perp} \mathrm{X}_{u|f} \|^2 \big]$. Hence there exists $f : S \to \{0, 1\}$ for which:

$$\frac{\delta \mu^2}{m} \geq \sum_u d_u \| \mathrm{X}_{\emptyset|f}^{\perp} \mathrm{X}_{u|f} \|^2. \tag{30}$$

We define $y_u \overset{\text{def}}{=} \| \mathrm{X}_{u|f} \|^2$ so that $\| \mathrm{X}_{\emptyset|f}^{\perp} \mathrm{X}_{u|f} \|^2 = \| \mathrm{X}_{u|f} \|^2 - \| \mathrm{X}_{u|f} \|^4 = y_u(1 - y_u)$ and $\delta \frac{\mu^2}{m} \geq \sum_u d_u y_u(1 - y_u)$.

Claim 7.10. $\mathrm{Var}_\tau \big[\sum_u d_u \varkappa_u \big] \leq \delta \mu^2$. *In particular,*

$$\mathrm{Prob}_\tau \big[\big| \sum_u d_u \varkappa_u - \mu \big| > \delta^{1/3} \mu \big] < \delta^{1/3}.$$

Proof. We can express $\mathrm{Var}_\tau \big[\sum_u d_u \varkappa_u \big]$ as:

$$\mathrm{Var}_\tau \Big[\sum_u d_u \varkappa_u \Big] = \sum_{u,v} d_u d_v \mathbb{E} \big[\varkappa_u \varkappa_v \big] - \Big[\sum_u d_u \mathbb{E} \big[\varkappa_u \big] \Big]^2 = \sum_{u,v} d_u d_v \min(y_u, y_v) - \mu^2$$

$$= \frac{1}{2} \sum_{u,v} d_u d_v (y_u + y_v) - \frac{1}{2} \sum_{u,v} d_u d_v |y_u - y_v| - \mu^2$$

$$\leq m \sum d_u y_u - \frac{1}{2} \sum_{u,v} d_u d_v (y_u - y_v)^2 - \mu^2$$

$$= m \sum_u d_u y_u(1 - y_u) + \frac{1}{2} \sum_{u,v} d_u d_v (2 y_u y_v) - \mu^2$$

$$= m \sum_u d_u y_u(1 - y_u) \leq \delta \mu^2.$$

The second bound follows from the first one via Chebyshev's inequality. $\qquad \square$

Given Claim 7.10, the rest of the proof is easy. For random τ, with probability $> 1 - \delta^{1/3}$, the corresponding $\varkappa$ satisfies the first condition as proven in Claim 7.10. Since the expected cut cost is bounded by η in Lemma 4.6, we can use Markov's inequality to show that $\varkappa^T L \varkappa$ will be at most $(1 - \delta^{1/3})^{-1} \eta$ with probability $\leq 1 - \delta^{1/3}$. Hence with non-negative probability, both properties are satisfied simultaneously. $\qquad \square \qquad\qquad \square$

Acknowledgments

References

[ABS10]   Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for Unique Games and related problems. In *FOCS*, pages 563–572, 2010. 3, 5, 6, 7

[AG11]   Sanjeev Arora and Rong Ge. New tools for graph coloring. In *APPROX-RANDOM*, pages 1–12, 2011. 3, 8, 41, 43

[AKK⁺08] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K. Vishnoi. Unique games on expanding constraint graphs are easy. In *STOC*, pages 21–28, 2008. 5, 6, 7, 14

[AL08]   Reid Andersen and Kevin J. Lang. An algorithm for improving graph partitions. In *SODA*, pages 651–660, 2008. 5

[AMS11]   Christoph Ambühl, Monaldo Mastrolilli, and Ola Svensson. Inapproximability results for maximum edge biclique, minimum linear arrangement, and sparsest cut. *SIAM J. Comput.*, 40(2):567–596, 2011. 2

[ARV09]   Sanjeev Arora, Satish Rao, and Umesh V. Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2), 2009. 7

[BBH⁺12] Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *STOC*, pages 307–326, 2012. 2, 6

[BDMI11] Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail. Near optimal column-based matrix reconstruction. In *FOCS*, pages 305–314, 2011. 6, 13, 16, 21

[BRS11]   Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS*, pages 472–481, 2011. 3, 7, 8, 39

[CDK12]   Eden Chlamtac, Michael Dinitz, and Robert Krauthgamer. Everywhere-sparse spanners via dense subgraphs. In *FOCS*, pages 758–767, 2012. 13

[Chl07]   Eden Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *FOCS*, pages 691–701, 2007. 7

[CS08]   Eden Chlamtac and Gyanit Singh. Improved approximation guarantees through higher levels of SDP hierarchies. In *APPROX-RANDOM*, pages 49–62, 2008. 3, 7

[CT11]   Eden Chlamtac and Madhur Tulsiani. Convex relaxations and integrality gaps. In *Handbook on Semidefinite, Cone and Polynomial Optimization*. Springer, 2011. 2, 7

[DMN13]   Anindya De, Elchanan Mossel, and Joe Neeman. Majority is stablest: discrete and sos. In *STOC*, pages 477–486, 2013. 6

[DR10]   Amit Deshpande and Luis Rademacher. Efficient volume sampling for row/column subset selection. In *FOCS*, pages 329–338, 2010. 6

[DV06]   Amit Deshpande and Santosh Vempala. Adaptive sampling and fast low-rank matrix approximation. In *APPROX-RANDOM*, pages 292–303, 2006. 6

[Fei02]   Uriel Feige. Relations between average case complexity and approximation complexity. In *STOC*, pages 534–543, 2002. 2

[GS11]   Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with PSD objectives. In *FOCS*, pages 482–491, 2011. 3, 8

[GS12a]   Venkatesan Guruswami and Ali Kemal Sinop. Faster SDP hierarchy solvers for local rounding algorithms. In *FOCS*, pages 197–206, 2012. 5, 8, 24, 25, 26, 30, 33, 35, 41, 42

[GS12b]   Venkatesan Guruswami and Ali Kemal Sinop. Optimal column-based low-rank matrix reconstruction. In *SODA*, pages 1207–1214, 2012. 6, 13, 16, 22

[GS13]   Venkatesan Guruswami and Ali Kemal Sinop. Approximating non-uniform sparsest cut via generalized spectra. In *SODA*, pages 295–305, 2013. 3, 8

[Hal98]   Magnús M. Halldórsson. Approximations of independent sets in graphs. In *APPROX*, pages 1–13, 1998. 30

[Hal02]   Eran Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM J. Comput.*, 31(5):1608–1623, 2002. 30

[Hås01]   Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. 4

[HJ90]   Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, 1990. 21

[Kho06]   Subhash Khot. Ruling out PTAS for Graph Min-Bisection, Dense k-Subgraph, and Bipartite Clique. *SIAM J. Comput.*, 36(4):1025–1071, 2006. 2

[KLPT11]   Jonathan A. Kelner, James R. Lee, Gregory N. Price, and Shang-Hua Teng. Metric uniformization and spectral bounds for graphs. *Geometric and Functional Analysis*, 21(5):1117–1143, 2011. 3

[KMN11]   Anna R. Karlin, Claire Mathieu, and C. Thach Nguyen. Integrality gaps of linear and semi-definite programming relaxations for knapsack. In *IPCO*, volume 6655, pages 301–314. Springer Berlin Heidelberg, 2011. 3, 7, 10

[Kol10]   Alexandra Kolla. Spectral algorithms for unique games. In *CCC*, pages 122–130, 2010. 6

[KPS10]   Subhash Khot, Preyas Popat, and Rishi Saket. Approximate Lasserre integrality gap for unique games. In *APPROX-RANDOM*, pages 298–311, 2010. 7

[KS09]   Subhash Khot and Rishi Saket. SDP integrality gaps with local ℓ_1-embeddability. In *FOCS*, pages 565–574, 2009. 2

[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. In *FOCS*, pages 53–62, 2005. 6

[Las02]   Jean B. Lasserre. An explicit equivalent positive semidefinite program for nonlinear 0-1 programs. *SIAM J. Optimization*, 12(3):756–769, 2002. 2, 6, 10, 11

[Lau03]   Monique Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, 28(3):470–496, 2003. 2

[LS91]   László Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. Optimization*, 1:166–190, 1991. 2

[MM10]   Konstantin Makarychev and Yury Makarychev. How to play unique games on expanders. In *WAOA*, pages 190–200, 2010. 5

[Oli10]   Roberto Oliveira. The spectrum of random k-lifts of large graphs (with possibly large k). *Journal of Combinatorics*, 1(3-4):285–306, 2010. 6

[OZ13]   Ryan O'Donnell and Yuan Zhou. Approximability and proof complexity. In *SODA*, pages 1537–1556, 2013. 6

[PP93]   James K. Park and Cynthia A. Phillips. Finding minimum-quotient cuts in planar graphs. In *STOC*, pages 766–775, 1993. 3

[Rag08]   Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *STOC*, pages 245–254, 2008. 7

[RS09]   Prasad Raghavendra and David Steurer. Integrality gaps for strong SDP relaxations of Unique Games. In *FOCS*, pages 575–585, 2009. 2

[RS10]   Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *STOC*, pages 755–764, 2010. 2

[RST12]   Prasad Raghavendra, David Steurer, and Madhur Tulsiani. Reductions between expansion problems. In *CCC*, 2012. 2

[RT12]   Prasad Raghavendra and Ning Tan. Approximating CSPs with global cardinality constraints using SDP hierarchies. In *SODA*, 2012. 3

[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 Mathematics*, 3:411–430, 1990. 2

[Sch08]   Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *FOCS*, pages 593–602, 2008. 7

[TSSW00]   Luca Trevisan, Gregory B. Sorkin, Madhu Sudan, and David P. Williamson. Gadgets, approximation, and linear programming. *SIAM J. Comput.*, 29(6):2074–2097, 2000. 4

[Tul09]   Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *STOC*, pages 303–312, 2009. 7

A Analysis of Other Rounding Algorithms Using Column Selection Framework

A.1 2-CSPs

Given a 2-CSP problem on variables $[n]$ and labels $[k]$, let G be its constraint graph. For convenience, we assume G is regular; however all our bounds still hold when G is non-regular. We use A to denote G's normalized adjacency matrix and λ_i to denote the ith smallest eigenvalue of G's normalized Laplacian matrix. Finally we will use $uv \sim G$ to denote sampling a constraint with probability proportional to the weight of constraint between u and v.

Embedding Consider the embedding used in Lemma 5.3 of [BRS11] which is used to convert k vectors $\mathrm{X}_u(i)$ into a single vector. Given a partial assignment $f \in [k]^S$ and $u \in [n]$ with $(\mathrm{X}_{u|f}(i))_{i \in [k]} \subset \mathbb{R}^{[m]}$, we define $\mathrm{X}_u(f)$ as the following vector.

$$\mathrm{X}_u(f) \stackrel{\text{def}}{=} \frac{1}{\sqrt{k}} \sum_j \frac{(\mathrm{X}_{\emptyset|f}^\perp \mathrm{X}_{u|f}(j))^{\otimes 2}}{\|\mathrm{X}_{\emptyset|f}^\perp \mathrm{X}_{u|f}(j)\|} \tag{31}$$

Seed Selection and Rounding We will give only an overview of the seed selection procedure. At i^{th} level, we choose a seed set of size $O(r/\varepsilon)$, S_i, from the matrix $\mathrm{X}(f_i) = [\mathrm{X}_u(f_i)]_{u \in [n]}$ where X_u's are defined in eq. (31). After choosing seed set S_i, we sample an assignment $g_i \in [k]^{S_i}$ (conditioned on f_{i-1}) that satisfies

$$\delta_{f_{i-1}, g_i} \leq \mathbb{E}_{g \sim \|\mathrm{X}_{S|f_{i-1}}(g)\|^2} \left[\delta_{f_{i-1}, g}\right]$$

where δ_f is defined in eq. (33) and set $f_i \leftarrow f_{i-1} \circ g_i$. We repeat the seed selection procedure as long as $\epsilon_{f_i} > \varepsilon$ where ϵ_f is as defined in eq. (32).

The rounding procedure remains the same — independent labeling for each CSP variable from the respective conditional distributions. Formally, for each variable $u \in [n]$, we choose a label $i \in [k]$ with probability $\|\mathrm{X}_{u|f_i}(i)\|^2$ independently at random. In Theorem A.5, we will show that $\ell = O\left(\frac{k^2}{\varepsilon^2}\right)$, i.e. seed selection will terminate after choosing at most ℓ sets.

Analysis Let us begin by defining the quantity

$$\epsilon_f \stackrel{\text{def}}{=} \mathbb{E}_{uv \sim G} \sum_{(i,j) \in [k]^2} \left| \mathbb{E}\left[\mathcal{X}_{uv|f}(ij)\right] - \mathbb{E}\left[\mathcal{X}_{u|f}(i)\right] \mathbb{E}\left[\mathcal{X}_{v|f}(j)\right]\right|$$

$$= \mathbb{E}_{uv \sim G} \sum_{i,j} \left|\text{Cov}\left[\mathcal{X}_{u|f}(i), \mathcal{X}_{v|f}(j)\right]\right| = \mathbb{E}_{uv \sim G} \sum_{i,j} \left|\langle \mathrm{X}_{\emptyset|f}^\perp \mathrm{X}_{u|f}(i), \mathrm{X}_{\emptyset|f}^\perp \mathrm{X}_{u|f}(j)\rangle\right|. \tag{32}$$

As shown in [BRS11], the above gives an upper bound on the expected extra fraction of unsatisfied constraints in the rounded solution compared to the SDP optimum (when performing rounding after conditioning on assignment f). Therefore, when $\varepsilon_f \leq \varepsilon$, we get an additive ε-error approximation. Our goal is prove (which we will do in Theorem A.5) that for $\ell \leq \tilde{O}(k/\varepsilon)$, we must have $\varepsilon_{f_\ell} \leq \varepsilon$.

If we define the quantity δ_f measuring the expected total variances of each $\mathcal{X}_{u|f}(i)$ as

$$\delta_f \stackrel{\text{def}}{=} \mathbb{E}_u \sum_{i \in [k]} \text{Var}\left[\mathcal{X}_{u|f}(i)\right] = \mathbb{E}_u \sum_{i \in [k]} \|\mathrm{X}_{\emptyset|f}^\perp \mathrm{X}_{u|f}(i)\|^2, \tag{33}$$

then it is easy to see that $\epsilon_f \leq k\delta_f$ by Cauchy-Schwarz.
We will first relate eq. (32) to the inner products of the embedded vectors $X_u(f)$.

Claim A.1. $\mathbb{E}_{uv \sim G} \left[\langle X_u(f), X_v(f) \rangle \right] \geq \left(\frac{\epsilon_f}{k} \right)^2.$

Proof. We have

$$k \langle X_u(f), X_v(f) \rangle = \sum_{ij} \frac{\langle X_{\emptyset|f}^{\perp} X_{u|f}(i), X_{\emptyset|f}^{\perp} X_{v|f}(j) \rangle^2}{\|X_{\emptyset|f}^{\perp} X_{u|f}(i)\| \|X_{\emptyset|f}^{\perp} X_{v|f}(j)\|} \geq \frac{\left(\sum_{ij} \left| \langle X_{\emptyset|f}^{\perp} X_{u|f}(i), X_{\emptyset|f}^{\perp} X_{v|f}(j) \rangle \right| \right)^2}{\sum_{ij} \|X_{\emptyset|f}^{\perp} X_{u|f}(i)\| \|X_{\emptyset|f}^{\perp} X_{v|f}(j)\|} \quad (34)$$

where the second step uses Cauchy Schwarz. Since

$$\sum_i \|X_{\emptyset|f}^{\perp} X_{u|f}(i)\| \leq \sqrt{k} \left(\sum_i \|X_{\emptyset|f}^{\perp} X_{u|f}(i)\|^2 \right)^{1/2} \leq \sqrt{k} \left(\sum_i \|X_{u|f}(i)\|^2 \right)^{1/2} = \sqrt{k}$$

the expected value of the above lower bound (34) for $uv \sim G$ is at least ε_f^2 / k. $\qquad \square$

We now upper bound the lengths of the embedded vectors.

Claim A.2. $\|X_u(f)\|^2 \leq \sum_{i \in [k]} \left\| X_{\emptyset|f}^{\perp} X_{u|f}(i) \right\|^2.$ *In particular, $\mathbb{E}_u \|X_u(f)\|^2 \leq \delta_f$.*

Proof. $\frac{1}{k} \|X_u(f)\|^2 = \mathbb{E}_{ij} \frac{\langle X_{\emptyset|f}^{\perp} X_{u|f}(i), X_{\emptyset|f}^{\perp} X_{u|f}(j) \rangle^2}{\|X_{\emptyset|f}^{\perp} X_{u|f}(i)\| \|X_{\emptyset|f}^{\perp} X_{u|f}(j)\|} \leq \left(\mathbb{E}_i \left\| X_{\emptyset|f}^{\perp} X_{u|f}(i) \right\| \right)^2 \leq \mathbb{E}_i \left\| X_{\emptyset|f}^{\perp} X_{u|f}(i) \right\|^2.$ $\qquad \square$

Now for fixed f we will upper bound the expected value of $\delta_{f,g}$ over $g \sim \|X_{S|f}(g)\|^2$ in terms of the projection distance of the embedded vectors from the subspace spanned by $X_v(f)$ for $v \in S$. (Below, $X_S(f)^{\perp}$ denotes the projection onto the orthogonal complement of $\mathrm{span}\{X_v(f) \mid v \in S\}$.)

Claim A.3. $\mathbb{E}_{g \sim \|X_{S|f}(g)\|^2} \left[\delta_{f,g} \right] \leq \mathbb{E}_{u \sim G} \left[\left\| X_S^{\perp}(f) X_u(f) \right\|^2 \right].$

Proof. We know that $\mathbb{E}_g \left[\left\| X_{\emptyset|g,f}^{\perp} X_{u|g,f}(i) \right\|^2 \right] = \left\| \Pi_{S|f}^{\perp} X_{u|f}(i) \right\|^2.$ Since $X_{\emptyset|f}$ is in the span of $\Pi_{S|f}$ $\Pi_{S|f}^{\perp} X_{u|f}(i) = \Pi_{S|f}^{\perp} X_{\emptyset|f}^{\perp} X_{u|f}(i)$. Similarly for any $v \in S$ and $j \in [k]$, the vector $X_{\emptyset|f}^{\perp} X_{v|f}(j)$ is in the span of $\Pi_{S|f}$. By using the same arguments from Claim 6.12, namely the embedding used here preserves linearity, we obtain $\sum_i \left\| \Pi_{S|f}^{\perp} X_{\emptyset|f}^{\perp} X_{u|f}(i) \right\|^2 \leq \left\| X_S^{\perp}(f) X_u(f) \right\|^2.$ Taking expectation over u completes the proof. $\qquad \square$

Using the above, we can prove the main claim about the seed selection procedure, namely that, assuming λ_r is close enough to 1, the expected variance δ_f can be reduced by a geometric factor by conditioning on the assignment to a further $O(r/\varepsilon)$ nodes.

Lemma A.4. *Given $f \in [k]^{S_0}$, positive real $\varepsilon > 0$ and positive integer r with $\lambda_{r+1} \geq 1 - \frac{\varepsilon^2}{2k^2}$, if $\epsilon_f \geq \varepsilon$ then there exists a set of $O(rk^2/\varepsilon^2)$-columns of $X(f)$, S and $g \in [k]^S$ such that $X_{S|f}(g) \neq 0$ and:*

$$\delta_{f,g} \leq \delta_f - \Omega \left(\frac{\varepsilon^2}{k^2} \right). \quad (35)$$

Proof. Let $\rho \overset{\text{def}}{=} \varepsilon/k$, and $\mu \overset{\text{def}}{=} \mathbb{E}_u\|X_u\|^2$ where for notational convenience we suppress the dependence on f and denote $X_u(f)$ by X_u. Observe that

$$\frac{1}{n}\operatorname{Tr}\left[X^T X A\right] = \mathbb{E}_{uv \sim G}\langle X_u, X_v\rangle \geq (\varepsilon_f/k)^2 \geq \rho^2$$

by Claim A.1. This implies $\frac{1}{n}\operatorname{Tr}\left[X^T X L\right] \leq \mathbb{E}_u\|X_u\|^2 - \rho^2 = \mu - \rho^2$. From Corollary 5.3, we know that volume sampling $O(r/\rho^2)$ columns from X yields a set S for which:

$$\mathbb{E}_u\|X_S^{\perp}X_u\|^2 \leq \left(1 + O(\rho^2)\right)\frac{(1/n)\operatorname{Tr}\left[X^T X L\right]}{1 - \max(1-\lambda_{r+1}, 0)} \leq \left(1 + O(\rho^2)\right)\frac{\mu - \rho^2}{1 - \frac{\rho^2}{2}}$$

Since $\rho \leq 1$, we have $(1 - \rho^2/2)^{-1} \leq \left(1 + \frac{3}{4}\rho^2\right)$:

$$\begin{aligned}
&\leq \left(1 + O(\rho^2)\right)(\mu - \rho^2)\left(1 + \frac{3}{4}\rho^2\right) \leq \left(1 + O(\rho^2)\right)\left(\mu - \frac{\rho^2}{4}\right) \\
&\leq \mu - \Omega\left(\rho^2\right) \\
&\leq \delta_f - \Omega\left(\rho^2\right) \qquad \text{(by Claim A.2)}.
\end{aligned}$$

By Claim A.3, $\mathbb{E}_g\left[\delta_{f,g}\right] \leq \mathbb{E}_u\|X_S^{\perp}X_u\|^2$, which means there exists g for which $\delta_{f,g} \leq \delta_f - \Omega\left(\frac{\varepsilon^2}{k^2}\right)$. $\square$

We put together everything in the following theorem.

Theorem A.5. *For $\ell = O\left(\frac{k^2}{\varepsilon^2}\right)$, seed selection procedure will output a partial assignment f_ℓ with $\epsilon_{f_\ell} \leq \varepsilon$.*

Remark A.6 (Faster Implementation)**.** By using the faster solver from [GS12a], we can implement a matching algorithm for Theorem A.5 whose running time is $2^{\text{poly}(k/\varepsilon)r}n^{\text{poly}(k/\varepsilon)}$.

Proof. Suppose $\epsilon_{f_i} > \varepsilon$ for all $i \leq \ell$. Then by Lemma A.4, for each $i \leq \ell$:

$$0 \leq \delta_{f_i} \leq \delta_{f_0} - i\Omega\left(\frac{\varepsilon^2}{k^2}\right) \leq 1 - \Omega\left(\frac{i\varepsilon^2}{k^2}\right) \implies \delta_{f_\ell} < 0,$$

which is a contradiction. $\square$

A.2 Partial Coloring of 3-Colorable Graphs

The rounding algorithm is given below:

1. Let $X_u \leftarrow \sum_{i=1}^{3} e_i \otimes X_{\emptyset}^{\perp}X_u(i)$ (same as in [AG11]).

2. Use Theorem 5.2 to choose S, an r'-subset of vectors from $(X_u)_{u \in [n]}$.

3. Sample $f \sim \|X_S(f)\|^2$.

4. For each $u \in V$, let $\mathbf{x}_u \leftarrow j$ if $\|X_{u|f}(j)\|^2 > \frac{1}{2}$ for some $j \in \{1, 2, 3\}$. If no such j exists, let $\mathbf{x}_u \leftarrow 0$.

5. Output the partial coloring, $\mathbf{x} \in \{0, 1, 2, 3\}^V$.

Theorem A.7. *Given a 3-colorable d-regular graph G on n nodes, positive real $1 > \varepsilon > 0$ and positive integer r, suppose its rth largest eigenvalue of normalized Laplacian matrix, λ_{n-r}, satisfies*

$$\lambda_{n-r} \leq \frac{4 - \delta}{3}$$

for some positive real $\delta > 0$. Then, for the choice of $r' = O(r/\delta\varepsilon)$, by solving r' rounds of Lasserre Hierachy relaxation, we can find a partial coloring which colors at least $(1 - \varepsilon)\frac{\delta}{2+\delta}n$ nodes.

Remark A.8 (Faster Implementation). By using the faster solver from [GS12a], we can implement a matching algorithm for Theorem A.7 whose running time is $2^{O(r)}n^{O(1)}$.

We have the following as an immediate corollary of Theorem A.7:

Corollary A.9. *Given a 3-colorable d-regular graph G, for any positive integer r with $\lambda_{n-r} \leq \frac{10}{9} - \Omega(1)$, we can find a partial coloring on $\frac{n}{4}$ nodes and an independent set of size at least $\frac{n}{12}$ in time $n^{O(r)}$.*

Before we begin the proof of Theorem A.7, we will state some simple claims. As the method applies for k-colorable graphs with different parameters, below for clarity we first use k for the number of colors, and then later set $k = 3$.

Claim A.10. *For any edge (u, v) of G,*

$$\frac{1}{2}\|X_u + X_v\|^2 \leq 1 - \frac{2}{k}.$$

In particular, if we use A to denote the normalized adjacency matrix of G, then:

$$\mathrm{Tr}\left[X^T X(I + A)\right] \leq n\left(1 - \frac{2}{k}\right).$$

Proof.

$$\frac{1}{2}\|X_u + X_v\|^2 = \frac{1}{2}\sum_{i\in[k]}\left(\|X_\emptyset^\perp X_u(i)\|^2 + \|X_\emptyset^\perp X_v(i)\|^2 + 2\langle X_\emptyset^\perp X_u(i), X_\emptyset^\perp X_v(i)\rangle\right)$$

$$= \frac{1}{2}\sum_{i\in[k]}\Bigg(\|X_u(i)\|^2 - \|X_u(i)\|^4 + \|X_v(i)\|^2 - \|X_v(i)\|^4$$

$$+ 2\langle X_u(i), X_v(i)\rangle - 2\langle X_\emptyset, X_u(i)\rangle\langle X_\emptyset, X_v(i)\rangle\|X_\emptyset\|^2\Bigg)$$

Using $\langle x_u(i), x_v(i)\rangle = 0$, we can rewrite this as:

$$= 1 - \frac{1}{2}\sum_{i\in[k]}\left(\|X_u(i)\|^4 + \|X_v(i)\|^4 + 2\|X_u(i)\|^2\|X_v(i)\|^2\right)$$

$$= 1 - \frac{1}{2}\sum_{i\in[k]}\left(\|X_u(i)\|^2 + \|X_v(i)\|^2\right)^2.$$

At this point, observe that $\sum_{i\in[k]}\left(\|X_u(i)\|^2 + \|X_v(i)\|^2\right)^2$ is a convex function on $\|X_u(i)\|^2$ and $\|X_v(j)\|^2$'s. Since $\sum_i \|X_u(i)\|^2 = \sum_j \|X_v(j)\|^2 = 1$, it is minimized when $\|X_u(i)\|^2 = \|X_v(j)\|^2 = \frac{1}{k}$. Substituting this into the above expression, we see that:

$$\frac{1}{2}\|X_u + X_v\|^2 \leq 1 - \frac{k}{2}\left(\frac{2}{k}\right)^2 = 1 - \frac{2}{k}.$$

For the final part, observe that:

$$\text{Tr}\left[X^T X(I+A)\right] = \frac{1}{d} \sum_{\{u,v\} \in E(G)} \|X_u + X_v\|^2 \leq \frac{2|E(G)|}{d}\left(1 - \frac{2}{k}\right) = n\left(1 - \frac{2}{k}\right). \qquad \square$$

Claim A.11. *Given a graph G and positive integer r, for λ_r being the rth smallest eigenvalue of corresponding normalized graph Laplacian matrix, the following holds:*

$$\sum_{j \geq r} \sigma_j(X^T X) \leq n\frac{1 - 2/k}{2 - \lambda_r}.$$

Proof. Follows from using the upper bound from Claim A.10 on inequality:

$$\sum_{j \geq r} \sigma_j(X^T X) \leq \frac{1}{\lambda_r} \text{Tr}\left[X^T X(I+A)\right] . \qquad \square$$

Claim A.12. *Assume u is uncolored. Then:*

$$\sum_i \|X_{\emptyset|f}^\perp X_{u|f(i)}\|^2 \geq \frac{1}{2}.$$

Proof. Note that $\|X_{\emptyset|f}^\perp X_{u|f}(i)\|^2 = \|X_{u|f}(i)\|^2(1-\|X_{u|f}(i)\|^2)$. If u is uncolored, then $1-\|X_{u|f}(i)\|^2 \geq \frac{1}{2}$ for all $i \in [k]$[8], in which case we have:

$$\sum_i \|X_{\emptyset|f}^\perp X_{u|f}(i)\|^2 \geq \frac{1}{2}\sum_i \|X_{u|f}(i)\|^2 = \frac{1}{2}. \qquad \square$$

For a subset S of vertices of G, we denote by $X_S^\perp$ the projection operator onto the orthogonal complement of $\text{span}\{X_u \mid u \in S\}$.

Lemma A.13. *If we sample $f : S \to [3]$ with probability $\|X_S(f)\|^2$:*

$$\mathbb{E}_f\left[\sum_i \|X_{\emptyset|f}^\perp X_{u|f}(i)\|^2\right] \leq \|X_S^\perp X_u\|^2.$$

Proof. Follows from Claim 4.2 and Lemma 4.4. $\qquad \square$

Proof of Theorem A.7. Let $\delta' = \frac{1}{2}\delta$ and $\varepsilon' = \varepsilon\delta'$. By Theorem 5.2, we know that there exists $r' = O(r/\varepsilon')$ columns, S, such that:

$$\sum_u \|X_S^\perp X_u\|^2 \leq (1+\varepsilon)\sum_{j \geq r} \sigma_j(X^T X) \leq n(1+\varepsilon')\frac{1 - 2/k}{2 - \lambda_r}.$$

Using Markov inequality, the fraction of uncolored nodes is bounded by:

$$\leq 2n(1+\varepsilon)\frac{1 - 2/k}{2 - \lambda_r} = \frac{2(1+\varepsilon')}{3(2 - \lambda_r)}n \quad \text{(for } k = 3\text{)}.$$

For $\lambda_r \leq \frac{4}{3} - \frac{2}{3}\delta'$, this expression becomes $\frac{1+\varepsilon'}{1+\delta'}n$, which implies

$$\mathbb{E}\left[\text{fraction of colored nodes}\right] \geq 1 - \frac{1 + \varepsilon'}{1 + \delta'} = \frac{\delta' - \varepsilon'}{1 + \delta'} = \frac{\delta'}{1 + \delta'}(1 - \varepsilon) = \frac{\delta/2}{1 + \delta/2}(1 - \varepsilon).$$

To prove that the coloring output is legal, notice that for any pair of adjacent nodes $(u, v) \in E(G)$, both $\|X_{u|f}(i)\|^2$ and $\|X_{v|f}(i)\|^2$ cannot be larger than $1/2$ both at the same time. $\qquad \square$

---

[8]This follows from the threshold rounding algorithm used in [AG11] for coloring, which colors u with color i if $\|X_{u|f}(i)\|^2 > 1/2$.