Integrality Gaps for Sherali–Adams Relaxations

Moses Charikar[*]
Princeton University

Konstantin Makarychev
IBM T.J. Watson Research Center

Yury Makarychev
Microsoft Research

Abstract

We prove strong lower bounds on integrality gaps of Sherali–Adams relaxations for MAX CUT, Vertex Cover, Sparsest Cut and other problems. Our constructions show gaps for Sherali–Adams relaxations that survive n^δ rounds of lift and project. For MAX CUT and Vertex Cover, these show that even n^δ rounds of Sherali–Adams do not yield a better than $2 - \varepsilon$ approximation.

The main combinatorial challenge in constructing these gap examples is the construction of a fractional solution that is far from an integer solution, but yet admits consistent distributions of local solutions for all small subsets of variables. Satisfying this consistency requirement is one of the major hurdles to constructing Sherali–Adams gap examples. We present a modular *recipe* for achieving this, building on previous work on metrics with a local–global structure. We develop a conceptually simple geometric approach to constructing Sherali–Adams gap examples via constructions of consistent *local* SDP solutions.

This geometric approach is surprisingly versatile. We construct Sherali–Adams gap examples for Unique Games based on our construction for MAX CUT together with a parallel repetition like procedure. This in turn allows us to obtain Sherali–Adams gap examples for any problem that has a Unique Games based hardness result (with some additional conditions on the reduction from Unique Games). Using this, we construct $2 - \varepsilon$ gap examples for Maximum Acyclic Subgraph that rules out any family of linear constraints with support at most n^δ.

1 Introduction

Lift-and-project methods are systematic procedures to construct a sequence of increasing tighter mathematical programming relaxations for 0-1 optimization problems. Such procedures have been proposed by Lovász–Schrijver (LS and LS^+) [15], Sherali–Adams (SA) [16]

and Lasserre [13]. These procedures operate in a sequence of rounds starting from an LP relaxation of an integer program; the strengthened relaxation obtained in r rounds can typically be solved in $n^{O(r)}$ time. Roughly speaking, a feasible solution to the relaxation obtained after r rounds of lift-and-project satisfies the property that for any r variables in the original relaxation, their values can be expressed as a projection of a convex combination of integer solutions.

Recently, there has been significant interest in understanding the strength of lift-and-project procedures for designing better approximation algorithms. The hope that improved approximation algorithms could result from a better understanding of lift-and-project procedures comes from the fact that several families of constraints that have been used to obtain better approximations can in fact be derived from a few rounds of lift-and-project.

Thus far, most results in this realm have been negative, showing integrality gaps for the relaxations obtained by a certain number of rounds by specific families of lift-and-project. A few recent positive results [4, 5] have raised hopes that lift-and-project procedures may indeed yield better approximations for particular problems. We view this study of the strengths and limitations of lift-and-project as an important goal in understanding a potentially powerful algorithmic paradigm; negative examples in this respect can be thought of as lower bounds in a certain restricted model of computation.

In this paper, we consider the Sherali–Adams hierarchy of linear relaxations and show strong lower bounds for several natural problems. This hierarchy is much less understood than the weaker Lovász–Schrijver hierarchy. In fact, we are aware of only two integrality gap results (by de la Vega and Kenyon-Mathieu [6] and Schoenebeck [17] for Lasserre) that hold for more than one round of the Sherali–Adams lift-and-project. Constructing new integrality gap examples for Sherali–Adams is thus an important open question [20, Section 7.3].

In this paper, we prove that the integrality gap for the Sparsest Cut problem is at least $\Omega\left(\sqrt{\frac{\log n}{\log r + \log\log n}}\right)$ after r rounds of the Sherali–Adams lift-and-project. The integrality gap for the MAX CUT and Vertex Cover problems remains $2 - \varepsilon$ even after n^{γ} rounds (for every positive ε and some γ that depends on ε). Our result for MAX CUT improves a result by de la Vega and Kenyon-Mathieu [6] who proved that the integrality gap remains $2 - \varepsilon$ after $\log^c n$ rounds (where $c < 1$ is some fixed constant). Recently several very strong negative results were proved for other hierarchies: Schoenebeck, Trevisan, and Tulsiani [18] proved that the integrality gap for Vertex Cover and MAX CUT is $2 - \varepsilon$ after $\Omega(n)$ rounds of LS (Lovász–Schrijver relaxation of the LP program); Georgiou, Magen, Pitassi, and Tourlakis [7] proved that the integrality gap for Vertex Cover is $2 - \varepsilon$ after $\Omega(\sqrt{\log n/\log\log n})$ rounds of LS_+ (Lovász–Schrijver relaxation of the SDP program). Our results are not directly comparable with these results.

Previous results constructing gap examples for lift-and-project procedures use problem specific constructions. One of our main contributions is a systematic approach to constructing gap examples for the Sherali–Adams hierarchy of relaxations. Roughly speaking, a gap example for Sherali–Adams requires constructing a fractional solution whose objective function value is far from the objective function of any integer solution, yet all small subsets of variables are exactly convex combinations (distributions) of integer solutions. In addition to

2

this, subsets of variables ought to satisfy a strong local consistency property which requires that the distributions of local solutions for subsets S_1 and S_2 should agree on their intersection, i.e. the distribution of integer solutions corresponding to S_1 when projected onto $S_1 \cap S_2$ should be identical to the distribution of integer solutions corresponding to S_2 when projected onto $S_1 \cap S_2$. It is this consistency property that makes it challenging to construct gap examples for Sherali–Adams. Indeed the construction of such gap examples with this local consistency is a challenging combinatorial question.

We build on our previous work on constructing metrics with a local–global structure [3] to present a modular approach to constructing gap examples for Sherali–Adams. These metrics satisfy the property that small subsets embed isometrically into ℓ_2, yet the entire metric is far from ℓ_1 embeddable. We present a conceptually simple geometric approach to constructing locally consistent distributions on integer solutions as follows (see Section 3): Starting with the local–global metric constructions, we obtain a construction of locally consistent SDP solutions for every small subset of variables. (Note that our construction only applies to linear relaxations, yet we find it convenient to produce these gap examples by appealing to the geometry of SDP solutions). We now obtain local distributions on integer solutions by applying SDP rounding algorithms to these local SDP solutions. Our main observation is that the distribution of integer solutions (e.g. cuts) on a subset of points depends only on pairwise distances between these points. This happens because the embedding into ℓ_2 is uniquely defined by the pairwise distances (an analogous statement does not hold for ℓ_1). Note, that the connection between SDPs and integrality gaps for Sherali–Adams hierarchy was not known before. We are hopeful that our techniques can be extended to apply to the stronger Lasserre hierarchy in the future.

This geometric approach turns out to be surprisingly versatile. We first obtain a gap instance for Sparsest Cut using this approach (see Section 4). Then we present gap examples for MAX CUT and Vertex Cover (see Section 5). We use the MAX CUT result together with a parallel-repetition like procedure to obtain gap examples for Unique Games (see Section 6). Using these, we are able to produce gap examples for any optimization problem where there is a reduction from Unique Games to the problem with a certain structure (see Section 7). This allows us to obtain an integrality gap for the Sherali–Adams relaxation of Maximum Acyclic Subgraph. For Maximum Acyclic Subgraph (also known as Linear Ordering), several specific families of constraints with finite support have been studied in the mathematical programming literature (see e.g. [9, 10, 11]). In the past, gap examples have been constructed which are specific to some such family of strengthened constraints. Our result gives a unified construction which shows that no such family of constraints has integrality gap better than $2 - \varepsilon$. In Section 8, we present a gap instance for Multicut. Additionally, in Section 9, we show an integrality gap for Sparsest Cut, which is stronger then the gap we obtain in Section 4 when the number of rounds is small ($r \leq \sqrt{\log n \log \log n}$).

2 Preliminaries

In this section, we define the Sherali–Adams hierarchy. Then we review some results from earlier work [3] where we constructed metric spaces locally embeddable into ℓ_2 that do not (globally) embed into ℓ_1 with small distortion. In the next section, we will present a new method of transforming such metric spaces into integrality gap instances for the Sherali–Adams hierarchy. Later, we show how to apply this method to many combinatorial optimization problems.

Let us first recall the definition of the Sherali–Adams relaxation of the cut polytope (with r rounds of lift-and-project)[1]. For every pair of points i and j, we introduce a variable x_{ij}, which denotes the distance between i and j. We start with the base linear inequalities that define the metric polytope.

$$x_{ij} + x_{jk} - x_{ik} \geq 0$$
$$2 - x_{ij} - x_{jk} - x_{ik} \geq 0$$
$$x_{ij} - x_{ji} = 0$$
$$1 - x_{ij} \geq 0$$
$$x_{ij} \geq 0$$

Now we define the lifted polytope. We introduce a variable y_I for every set $I = \{(i_1, j_1), \ldots, (i_t, j_t)\}$ of size at most $r + 1$. For every two sets I and J such that $|I \cup J| \leq r$, and every base inequality of the form $\mathcal{R}(x) \geq 0$, we write

$$\left[\prod_{(i,j) \in I} x_{ij} \prod_{(i,j) \in J} (1 - x_{ij}) \right] \mathcal{R}(x) \geq 0.$$

We expand the expression as the sum of products of x_{ij}. Then we replace each product $x_{i_1 j_1} \cdot \cdots \cdot x_{i_t j_t}$ with the variable $y_{\{(i_1, j_1), \ldots, (i_t, j_t)\}}$. In other words, we replace each variable x_{ij} in $\mathcal{R}$ with the expression

$$\sum_{B \subset J} (-1)^{|B|} y_{\{(i,j)\} \cup I \cup B}. \tag{1}$$

We obtain a linear inequality on variables y. The set of all such inequalities defines the lifted Sherali–Adams cut polytope. Finally, the Sherali–Adams polytope is the projection of the lifted polytope: a solution x_{ij} lies in the Sherali–Adams polytope if there exists a vector of y_I in the lifted polytope such that $x_{ij} = y_{\{(i,j)\}}$.

We need the following statement, which was implicitly used by de la Vega and Kenyon-Mathieu [6] in their integrality gap construction. For completeness, we give its proof in the appendix.

Lemma 2.1 (de la Vega and Kenyon-Mathieu [6]). *Consider a set X. Suppose that for every subset T of size at most $k = 2r + 3$, we are given a distribution of cuts $\mathcal{D}_T$ on T (i.e.*

---

[1]See [16] and [14] for more background on the Sherali–Adams lift-and-project method.

distribution of subsets of T). Moreover, suppose that for two sets $Q \subset T$, the distributions $\mathcal{D}_Q$ and $\mathcal{D}_T$ agree in the following sense, for every $A \subset Q$

$$\mathcal{D}_Q(\{A\}) = \mathcal{D}_T(\{B : B \cap Q = A\}).$$

Then the vector $(x_{ij})\big|_{i,j \in X}$ defined by

$$x_{ij} = \Pr\left(D_{\{i,j\}} \text{ separates } i \text{ and } j\right) = \mathcal{D}_{\{i,j\}}(\{\{i\}, \{j\}\}).$$

lies in the Sherali–Adams relaxation of the cut polytope (with r rounds of lift-and-project) .

In this paper, we use several results from our paper [3]. We list these results below.

Notation 2.1. *Given a graph G, denote the shortest path metric by $d(\cdot, \cdot)$.*

Theorem 2.2 (Theorem 3.14 part III from [3, ECCC]). *For every n and $k < n$, there exists a bounded degree expander graph $G = (X, E)$ on n vertices such that the metric space X equipped with the distance*

$$\rho(u, v) = \sqrt{1 - (1-\mu)^{d(u,v)} + \mu},$$

where $\mu = \Theta\left(\frac{\log k + \log \log n}{\log n}\right)$ satisfies the following properties.

1. *Every embedding of (X, ρ) into ℓ_1 requires distortion*
   $\Omega\left(\frac{\log n}{\log k + \log \log n}\right)^{1/2}$.

2. *Every subset of X of size k embeds isometrically into ℓ_2.*

Definition 2.3. (IMPLICITLY INTRODUCED IN ARORA, BOLLOBÁS, LOVÁSZ, AND TOURLAKIS [1].) *We say that a graph G is l-path decomposable if every 2-connected subgraph H of G contains a path of length l such that every vertex of the path has degree 2 in H.*

Theorem 2.4 (Theorem 3.3 from [3]). *Suppose $G = (V, E)$ is an l-path decomposable graph. Let $L = \lfloor l/9 \rfloor$; $\mu \in [1/L, 1]$. Then there exists a probabilistic distribution of multicuts of G (or in other words random partition of G in pieces) such that the following properties hold. For every two vertices u and v,*

1. *If $d(u, v) \le L$, then the probability that u and v are separated by the multicut (i.e. lie in different parts) equals $1 - (1 - \mu)^{d(u,v)}$; moreover, if u and v lie in the same part, then the unique shortest path between u and v also lies in that part.*

2. *If $d(u, v) > L$, then the probability that u and v are separated by the multicut is at least $1 - (1 - \mu)^L$.*

3. *Every piece of the multicut partition is a tree.*

Lemma 2.5 (Lemma 3.11 from [3]). *Consider a metric space (Y, ρ) on k points. If for every two distinct points u and v: $|\rho(u, v) - 1| \le \frac{1}{2k}$, then (Y, ρ) is isometrically embeddable into ℓ_2.*

5

3 Sherali–Adams Relaxations and Embedding into ℓ_2

We now show how to transform a metric space locally embeddable into ℓ_2 into a feasible solution of the Sherali–Adams relaxation of the cut polytope via local SDP solutions to the MAX CUT problem. If every subset of $2r + 3$ points is isometrically embeddable in the Euclidean sphere, then the solution constructed satisfies the relaxation obtained from r rounds of Sherali–Adams.

Theorem 3.1. *Let (X, ρ) be a metric space. Assume that every $k = 2r + 3$ points isometrically embed in the Euclidean sphere of radius R. Then the following solution belongs to the Sherali–Adams relaxation of the cut polytope (with r rounds of lift-and-project)*

$$x_{ij} = \frac{1}{\pi} \arccos \left(1 - \frac{\rho(i,j)^2}{2R^2} \right).$$

Proof. We want to construct a distribution of cuts $\mathcal{D}_T$ on every set T of size at most k and then use Lemma 2.1. Embed T in the sphere of radius R centered at the origin. Apply the MAX CUT algorithm of Goemans and Williamson [8]: Choose a random hyperplane passing through the origin. Let one piece of the cut consist of vertices lying on one side of the hyperplane; the other part consist of vertices lying on the other side of the hyperplane.

Note that the distribution of cuts $\mathcal{D}_T$ is completely defined by all pairwise distances $(\rho(u,v))_{u,v \in T}$ since the pairwise distances uniquely determine the set of points on the sphere (up to an isometry of the whole sphere[2]). Moreover, if $Q \subset T$, then the distribution of cuts $\mathcal{D}_T$ restricted to Q depends only on the distances between points in Q. Therefore, Lemma 2.1 applies to the distribution $\mathcal{D}_T$. We conclude that the solution

$$x_{ij} = \Pr \left(D_{\{i,j\}} \text{ separates } i \text{ and } j \right)$$

belongs to the Sherali–Adams relaxation of the cut polytope.

Let us now compute x_{ij}. Since the angle θ_{ij} between embeddings of i and j in the sphere equals (by the law of cosines)

$$\theta_{ij} = \arccos \left(1 - \frac{\rho(i,j)^2}{2R^2} \right).$$

the probability that a random hyperplane separates i and j equals

$$x_{ij} = y_{\{(i,j)\}} = \frac{1}{\pi} \arccos \left(1 - \frac{\rho(i,j)^2}{2R^2} \right).$$

$\square$

---

[2]Here it is important that we embedded points into ℓ_2 rather than ℓ_1.

4   Relaxations for Sparsest Cut

In this section, we prove that the integrality gap for the relaxation of Sparsest Cut (with uniform demands) remains high even after $n^{o(1)}$ rounds of the Sherali–Adams lift-and-project. We consider the following LP relaxation for the Sparsest Cut problem.

$$\min \frac{\sum_{(i,j) \in E} x_{ij}}{\sum_{i,j \in V} x_{ij}}$$

$$\text{subject to: vector of } x_{ij} \text{ lies in the Sherali–Adams relaxation}$$
$$\text{of the cut polytope.}$$

Theorem 4.1. *The integrality gap for the Sparsest Cut relaxation is at least*

$$\Omega \left(\sqrt{\frac{\log n}{\log r + \log \log n}} \right)$$

after r rounds of the Sherali–Adams lift-and-project.

Proof. Consider the metric space (X, ρ) and the underlying expander graph $G = (X, E)$ from Theorem 2.2. We prove that the integrality gap of the Sparsest Cut relaxation for the graph G is $\Omega \left(\sqrt{\log n / (\log r + \log \log n)} \right)$.

Every $k = 2r + 3$ points of X embed into ℓ_2 isometrically. Moreover, we may assume that every subset of k points is isometrically embeddable in the sphere of radius $R = \sqrt{1 + \mu}$. Indeed, we can add a special isolated vertex O to the graph G (G will be still sparse, and the theorem will be still true). Recall that the metric ρ is defined as (for $u \neq v$), $\rho(u, v) = \sqrt{1 - (1 - \mu)^{d(u,v)} + \mu}$. Since the shortest path distance from O to every other vertex equals infinity, $\rho(u, O) = \sqrt{1 + \mu}$, that is, all points $u \in X$ lie on the sphere of radius $R = \sqrt{1 + \mu}$ centered at O.

Therefore, by Theorem 3.1, the vector

$$x_{ij} = y_{\{(i,j)\}} = \frac{1}{\pi} \arccos \left(1 - \frac{\rho(i,j)^2}{2R^2} \right)$$

is a feasible solution of the Sherali–Adams relaxation.

Finally, we compute the integrality gap of this solution. Let us estimate the value of x_{ij}. If i and j are adjacent in G, we have $\rho(i, j) = \sqrt{2\mu}$ and $x_{ij} = \Theta(\sqrt{\mu})$. For two typical points i and j, $\rho(i, j) = \Theta(1)$ (since the graph G is an expander with diameter $\Theta(\log n)$) and $x_{ij} = \Theta(1)$. Hence the value of the solution x_{ij} is

$$\Theta(\sqrt{\mu}) = \Theta \left(\sqrt{\frac{\log r + \log \log n}{\log n}} \right).$$

However, since the graph G is an expander the value of the optimal combinatorial solution is $\Theta(1)$. $\qquad \square$

5 Relaxations for MAX CUT and Vertex Cover

In this section, we prove that for every $\varepsilon > 0$ there exists $\gamma > 0$ such that the integrality gap for the MAX CUT and Vertex Cover relaxations is at least $2 - \varepsilon$ after $O(n^\gamma)$ rounds of the Sherali–Adams lift-and-project. Below, we say that the value of MAX CUT is $\alpha \in [0, 1]$, if an α fraction of all edges crosses the cut. Similarly, the value of Vertex Cover is $\alpha \in [0, 1]$, if the vertex cover contains an α fraction of all vertices.

Our integrality gap constructions are similar to that for the Sparsest Cut problem. Essentially, the only difference is that we define a different metric such that the distance between adjacent vertices is very close to 1. Later for the Unique Games problem, we will need this distance to be either close to 0, or 1, depending on the edge. Thus, we introduce weights $\pi_{uv} \in \{-1, +1\}$ for all edges (u, v); the length of the edge (u, v) will be equal to $(1 - (1 - \mu)\pi_{uv})/2$. In the case of MAX CUT and Vertex Cover, all $\pi_{uv} = -1$. Extend the function π_{uv} to all pairs of vertices as follows: if $d(u, v) \le 2L$, let π_{uv} be the product of the weights of all edges along the unique shortest path between u and v; otherwise (if $d(u, v) > 2L$), let $\pi_{uv} = 0$. Particularly, if all weight on edges are equal to -1, then $\pi_{uv} = (-1)^{d(u,v)}$ (when $d(u, v) \le 2L$). We define a new distance on G:

$$\rho_\mu^\pi(u, v) = \frac{1 - (1 - \mu)^{d(u,v)} \pi_{uv}}{2}.$$

Corollary 5.1 implies that ρ_μ^π is a metric (that is, satisfies the triangle inequality). However, we will not use this fact in the paper. The metric ρ_μ^π with $\pi_{uv} = (-1)^{d(u,v)}$ was previously used in the integrality gap constructions by de la Vega and Kenyon-Mathieu [6] and Schoenebeck, Trevisan, and Tulsiani [18]. We denote this metric by ρ_μ^{alt}. We now prove a corollary of Theorem 2.4.

Corollary 5.1. *Suppose $G = (V, E)$ is an l-path decomposable graph. Let $L = \lfloor l/9 \rfloor$; $\mu \in [1/L, 1]$; and f_{uv} be a set of weights (as above). There exists a random mapping $\varphi : V \to \{-1, 1\}$ such that*

1. if $d(u, v) \le L$ then $\mathrm{Pr}\,(\varphi(u) \ne \varphi(v)) = \rho_\mu^\pi(u, v)$;

2. if $d(u, v) \ge L$ then $\frac{1 - (1-\mu)^L}{2} \le \mathrm{Pr}\,(\varphi(u) \ne \varphi(v)) \le \frac{1 + (1-\mu)^L}{2}$.

Proof. Consider the distribution of multicuts from Theorem 2.4. For each piece of the multicut partition (recall that each piece is a tree) pick an arbitrary vertex u and set $\phi(u)$ to be $+1$ or -1 with probability a half. Then propagate values along the edges of the tree so that $\phi(u'') = \pi_{u'u''}\phi(u')$ for adjacent (u', u''). If $d(u, v) \le L$, then the probability that $\varphi(u) \ne \varphi(v)$ equals

$$\mathrm{Pr}\,(u \text{ and } v \text{ are separated by the multicut}) \times \frac{1}{2}$$

$$+ \,\mathrm{Pr}\,(u \text{ and } v \text{ are not separated by the multicut}) \times \frac{1 - \pi_{uv}}{2}.$$

8

If $d(u, v) > L$, then the probability that $\varphi(u) \neq \varphi(v)$ is bounded from below by

$$\Pr\left(u \text{ and } v \text{ are separated by the multicut}\right) \times \frac{1}{2}$$

and from above by

$$\Pr\left(u \text{ and } v \text{ are separated by the multicut}\right) \times \frac{1}{2}$$
$$+ \Pr\left(u \text{ and } v \text{ are not separated by the multicut}\right).$$

Plugging in bounds on the probability that u and v are separated from Theorem 2.4 concludes the proof. $\qquad\square$

Now we prove an analog of Theorem 2.2 for the metric ρ_μ^π.

Theorem 5.2. *Let $G = (V, E)$ be a graph on n vertices with maximum degree Δ; let $k \leq \sqrt{n}$. Suppose that for every t, every subgraph on t vertices is $\Omega(\log(n/t))$-path decomposable (here the constant in Ω-notation may depend on Δ). Then for every subset Y of at most k vertices there exists a mapping ψ_Y from Y to the unit sphere in ℓ_2 such that (for $u \neq v$)*

$$\|\psi(u) - \psi(v)\|_2 = \sqrt{\frac{4\rho_\mu^\pi(u, v) + \mu}{1 + \mu}},$$

where $\mu = c_\Delta(\log k + \log\log n)/\log n$ and c_Δ is a sufficiently large constant. In particular, the map ψ_Y sends adjacent vertices u and v to "almost antipodal" points if $\pi_{uv} = -1$: $\|\psi(u) - \psi(v)\|_2 = \sqrt{\frac{4-\mu}{1+\mu}} = 2 - O(\mu)$; and "very close" points if $\pi_{uv} = 1$: $\|\psi(u) - \psi(v)\|_2 = \sqrt{\frac{3\mu}{1+\mu}} = O(\sqrt{\mu})$.

Proof. The proof goes along the lines of proofs of Theorems 3.7 and 3.13 in [3]. Choose $l = \Theta(\log n)$ so that every subgraph of G on $3 \cdot (\Delta - 1)^l k$ vertices is l-path decomposable. Now consider the ball $B_d(Y, l) = \{x : d(x, Y) \leq l\}$. Since it contains at most $3(\Delta - 1)^l k$ vertices, the graph induced by $B_d(Y, l)$ on G is l-path decomposable. Therefore, there exists a random mapping $\varphi : B_d(Y, l) \to \{-1, 1\}$ satisfying the conditions of Corollary 5.1. Note that for every u and v in Y at distance at most l, the set $B_d(Y, l)$ contains the shortest path between them. Hence all distances appearing in the conditions of Corollary 5.1 do not depend on whether we define them with respect to the shortest path metric in G or with respect to the shortest path metric in $B_d(Y, l)$.

Extend the mapping φ to a new point O by $\varphi(O) = 0$. By Corollary 5.1, for $u, v \in B_d(Y, l)$:

$$\left|\|\varphi(u) - \varphi(v)\|_2^2 - 4\rho_\mu^\pi(u, v)\right| = 4\left|\Pr\left(\phi(u) \neq \phi(v)\right) - \rho_\mu^\pi(u, v)\right|$$
$$\leq 4(1 - \mu)^L \leq 4e^{-\frac{c_\Delta L}{\log n}(\log k + \log\log n)}$$
$$\leq \frac{1}{k \log n} \leq \frac{\mu}{2(k + 1)}$$

9

(when c_Δ is sufficiently large). Define the metric σ on $Y \cup \{O\}$ as follows:

$$\sigma(u,v) = \sqrt{\mu + \underbrace{4\rho_\mu^\pi(u,v) - \|\varphi(u) - \varphi(v)\|_2^2}_{\leq \mu/(2(k+1))}},$$
$$\sigma(u,O) = \sqrt{\mu}.$$

By Lemma 2.5 (c.f. Theorem 3.13 in [3]), the metric σ is Euclidean. Let ν be the direct sum of the isometric embedding of $(Y \cup \{O\}, \sigma)$ into ℓ_2 and the embedding $u \mapsto \varphi(u)$ (the random variable $\varphi(u)$ belongs to an L_2 space, which is isometric to ℓ_2). Then we have

$$\|\nu(u) - \nu(v)\|_2 = \sqrt{4\rho_\mu^\pi(u,v) + \mu}.$$

We may assume that $\nu(O) = 0$. Then $\|\nu(u)\|_2 = \sqrt{1+\mu}$ for $u \in Y$. By scaling ν we get the desired embedding $\psi(u) = \nu(u)/\sqrt{1+\mu}$. $\qquad\square$

We consider the following LP relaxations. The LP relaxation for MAX CUT is

$$\max \sum_{(i,j) \in E} x_{ij}$$
$$\text{subject to: vector of } x_{ij} \text{ lies in the Sherali--Adams relaxation}$$
$$\text{of the cut polytope}$$

The base LP relaxation for Vertex Cover is

$$\min \sum_{i \in V} x_i$$

subject to: (i) $x_i + x_j \geq 1$ for every edge (i,j); (ii) $1 - x_i \geq 0$; (iii) $x_i \geq 0$.

For every set of vertices I of size at most $r+1$, we introduce a variable y_I. We require that

1. $y_{\{i\}} = x_i$.

2. For every set of vertices T of size at most $r+3$ there exists a distribution of vertex covers on T such that y_I equals the probability that all vertices from I lie in the vertex cover.

Our second condition implies that the vector of y_I lies in the lifted Sherali--Adams polytope (the proof is similar to the proof of Lemma 2.1).

We will use random Δ-regular graphs in our integrality gap constructions. These graphs were previously used in several integrality gap constructions (e.g. [1], [6], [18]). It is known (see [1], [6], [18]) that for every $\varepsilon > 0$ there exists Δ such that with high probability every cut in a random Δ-regular graph cuts at most $1/2 + \varepsilon$ fraction of all edges and every vertex cover contains at least $1 - \varepsilon$ fraction of all vertices. Arora, Bollobás, Lovász, and Tourlakis [1] proved that with high probability we can remove $o(n)$ edges from a random Δ-regular graph so that every subgraph on k vertices is $\Omega(\log(n/k))$-path decomposable (here, the constant in the Ω-notation depends on Δ).

Theorem 5.3. *I. For every $\varepsilon > 0$ there exists $\gamma > 0$ such that the integrality gap of the relaxation for MAX CUT is $2 - \varepsilon$ after $r = n^\gamma$ rounds of the Sherali–Adams lift-and-project.*

II. There exist at most $1/2 + \varepsilon'$ satisfiable instances of MAX CUT (for every positive constant ε'), such that the LP value after r rounds of Sherali–Adams is at least $1 - \varepsilon$, where

$$\varepsilon = O\left(\sqrt{\frac{\log k + \log \log n}{\log n}}\right).$$

Proof. I. Consider a graph $G = (V, E)$ with maximum degree Δ such that

1. every cut cuts at most $1/2 + \varepsilon/6$ fraction of all edges;

2. every subgraph on k vertices is $\Omega(\log(n/k))$-path decomposable.

Let $\gamma = c\varepsilon^2/c_\Delta$, $r = n^\gamma$, $k = 2(r+3)$, and

$$\mu = c_\Delta \frac{\log k + \log \log n}{\log n} = (1 + o(1))\, c \cdot \varepsilon^2.$$

Applying Theorem 5.2 (with metric ρ_μ^{alt}) and Theorem 3.1, we get that the vector

$$x_{ij} = \frac{1}{\pi} \arccos\left(1 - \frac{4\rho_\mu^{alt}(i,j) + \mu}{2(1+\mu)}\right)$$

is a feasible solution of the Sherali–Adams relaxation.

We now compute the integrality gap of this solution. The distance between every two adjacent vertices i and j is $\rho_\mu^{alt}(i,j) = 1 - \mu/2$, hence

$$x_{ij} = \frac{\arccos(-1 + O(\mu))}{\pi} = 1 - O(\sqrt{\mu}) > 1 - \varepsilon/6,$$

if c is small enough. Therefore, the value of this solution is at least $(1 - \varepsilon/6)|E|$; the value of the optimal combinatorial solution is at most $(1/2 + \varepsilon/6)|E|$. The integrality gap is

$$\frac{1 - \varepsilon/6}{1/2 + \varepsilon/6} > 2 - \varepsilon.$$

II. The same proof works for part (II) with $\varepsilon = \Theta(\sqrt{\mu})$; note that Δ depends only on ε'. $\square$

Theorem 5.4. *For every $\varepsilon > 0$ there exists $\gamma > 0$ such that the integrality gap of the relaxation for Vertex Cover is $2 - \varepsilon$ after $r = n^\gamma$ rounds of the Sherali–Adams lift-and-project.*

Proof. Consider a graph $G = (V, E)$ with maximum degree Δ such that (i) every vertex cover contains at least $1 - \varepsilon/6$ fraction of all vertices; (2) every subgraph on k vertices

is $\Omega(\log(n/k))$-path decomposable. Let $\gamma = c\varepsilon^2/(\Delta^2 c_\Delta)$ (where c is a sufficiently small constant) and $r = n^\gamma$, $k = (\Delta + 1)(r + 2)$, and

$$\mu = c_\Delta \frac{\log k + \log \log n}{\log n} = (1 + o(1))\, c \cdot \varepsilon^2/\Delta^2.$$

For every set T of at most $r + 2$ vertices, let us define a distribution of vertex covers $\mathcal{D}_T$ on T as follows. Denote the set of vertices in T and their neighbors by $N(T)$. Note that $N(T)$ contains at most $k = (\Delta + 1)(r + 2)$ vertices. By Theorem 5.2, there exists an embedding ψ of $N(T)$ in the unit sphere satisfying the following properties.

1. The Euclidean distance between images of adjacent vertices is $2 - O(\mu)$.

2. For every set A of at most k points, its image is completely determined by the shortest path distances between points in A.

Choose a random hyperplane passing through the origin. Choose randomly one of the two half-spaces. Denote it by H. Now we define the vertex cover VC. Let a vertex u be in the vertex cover VC if either $\psi(u) \in H$ or there exists a neighbor v of u such that $\psi(v) \notin H$. Note that VC is a vertex cover of T: for every two adjacent vertices u and v, if $u \notin VC$ then $\psi(v) \in H$ and, hence, $v \in VC$.

Now consider two distributions $\mathcal{D}_{T_1}$ and $\mathcal{D}_{T_2}$. We claim that they agree on $T_1 \cap T_2$. Denote the corresponding embeddings of sets $N(T_1)$ and $N(T_2)$ in the unit sphere by ψ_1 and ψ_2. Observe that the sets $\psi_1(N(T_1 \cap T_2))$ and $\psi_2(N(T_1 \cap T_2))$ are isometric (by property 2). Therefore, we may assume without loss of generality that $\psi_1(u) = \psi_2(u)$ for every $u \in N(T_1 \cap T_2)$. Then $u \in T_1 \cap T_2$ belongs to the vertex cover chosen according to $\mathcal{D}_{T_1}$ if and only if it belongs to the vertex cover chosen according to $\mathcal{D}_{T_2}$ (we choose the same half-space H when we construct both vertex covers).

For every set I of at most $r + 1$ vertices, let y_I be the probability that all vertices in I belong to the vertex cover chosen according to $\mathcal{D}_I$. Let $x_i = y_{\{i\}}$. Since for every set T on at most $r + 2$ vertices, there exists a distribution of vertex covers (namely, $\mathcal{D}_T$) such that y_I is the probability that all vertices in I belong to the vertex cover, the vector of x_i is a feasible solution of our relaxation.

Let us estimate the value of x_i. Vertex i belongs to the vertex cover if either $\psi(i) \in H$, or one of the edges incident to i is not cut by the random hyperplane. The former event happens with probability $1/2$. The latter event happens with probability at most Δ times the probability that one edge is not cut by the random hyperplane. We obtain the following bound:

$$\frac{1}{2} + \Delta \left(1 - \frac{1}{\pi} \arccos \left(1 - \frac{4\rho_\mu^{alt}(i,j) + \mu}{2(1+\mu)}\right)\right) = \frac{1}{2} + \Delta \cdot O(\sqrt{\mu}) \le \frac{1}{2} + \frac{\varepsilon}{6}.$$

Therefore, the value of the solution x is at most $(1/2 + \varepsilon/6)|V|$. The value of the optimal combinatorial solution is at least $(1 - \varepsilon/6)|V|$. Hence the integrality gap is at least $2 - \varepsilon$. $\square$

6 Relaxations for Unique Games

We now present an integrality gap construction for the Unique Games (UG) problem. The problem is as follows: given a graph G, a set of q labels $\mathcal{S}$ and permutations $\pi_{uv} : \mathcal{S} \to \mathcal{S}$ for every edge (u, v), assign a label $\Lambda(u) \in \mathcal{S}$ to every vertex u so as maximize the number of satisfied constraints $\pi_{uv}(\Lambda(u)) = \Lambda(v)$.

In the LP relaxation, we introduce variables $x_u(i)$ and $x_{uv}(ij)$ for every pair of vertices u,v and labels i,j. In the integral solution, $x_u(i) = 1$, if u has label i; and $x_u(i) = 0$, otherwise. Similarly, $x_{uv}(ij) = 1$, if u has label i, and v has label j; and $x_{uv}(ij) = 0$, otherwise. The objective is to maximize the number of satisfied constraints: $\sum_{(u,v) \in E} \sum_{i \in \mathcal{S}} x_{uv}(i\pi_{uv}(i))$. The LP constraints are: for all vertices u, $\sum_{i \in \mathcal{S}} x_u(i) = 1$; for all edges (u, v), $\sum_{j \in \mathcal{S}} x_{uv}(ij) = x_u(i)$. As before, in order to show that the vector of x_u, $x_{uv}(ij)$ belongs to the Sherali–Adams LP relaxation it is enough to prove that for every set of vertices S of size at most $O(r)$ there exists a distribution of integral solutions $\mathcal{D}_S$ such that (i) $x_u(i)$ is the probability that u has label i; and $x_{uv}(ij)$ is the probability that u has label i, and v has label j; (ii) for every two sets $S \subset T$ the distributions $\mathcal{D}_S$ and $\mathcal{D}_T$ coincide on S.

Theorem 6.1. *Fix a number of labels $q = 2^t$, a real $\delta \in (0, 1)$ and let $\Delta = \lceil C(q/\delta)^2 \rceil$ (for a sufficiently large constant C). Then for every positive ε there exists γ depending on ε such that for every sufficiently large n there exists an instance of Unique Games on Δ-regular graph G on n vertices so that (i) The cost of the optimal solution is at most $1/q \cdot (1+\delta)$. (ii) There exists a solution to the LP relaxation obtained after $r = n^\gamma$ rounds of Sherali–Adams of cost $(1 - \varepsilon)$.*

Proof sketch. The underlying graph G is as before a random Δ-regular expander (see Theorem 5.2). The labels are elements of the multiplicative group $\{-1, 1\}^t$. For every edge (u, v) of the graph we pick independently (and uniformly) a random element π_{uv} in $\{-1, 1\}^t$. The constraint between u and v is $\Lambda(u) = \Lambda(v) \cdot \pi_{uv}$.

By the standard probabilistic argument, for sufficiently large C the optimal solution of the Unique Games problem has value at most $(1 + \delta)/q$ with probability close to 1. Indeed, for a fixed assignment of labels, the expected fraction of satisfied constraints (where the expectation is taken only over random choices of π_{uv}) is $1/q$. By the Central Limit Theorem, the probability that the fraction of satisfied constraints deviates from the expectation by more than δ/q is $o(q^{-n})$. Since the total number of fixed assignments is q^n, by the union bound, the Unique Game is at most $(1 + \delta)/q$ satisfiable with probability approaching to 1.

We now need to show that the LP cost of the Unique Game is at least $(1 - \varepsilon)$. Fix s from 0 to $t - 1$ and consider s-th coordinate of every label, and s-th coordinate of every π_{uv}. We get an instance of MAX 2LIN mod 2 — linear equations modulo 2. It is easy to see, that Theorem 5.3 works not only for MAX CUT, but also for MAX 2LIN, we just need to substitute the metric ρ_μ^{alt} by the metric $\rho_\mu^{\pi^s}$ (here π_{uv}^s is the s-th coordinate of π_{uv}). The theorem guarantees that for every set S of size at most $\Omega(r)$ there exists a distribution of cuts such that if $\pi_{uv}^s = 1$, then u and v lie on the same side of the cut; and if $\pi_{uv}^s = -1$, then u and v lie on the opposite sides of the cut with probability $1 - \varepsilon$. Define, $\Lambda^s(u)$ to be -1 or 1 depending on whether u belongs to the cut or not. Then $\Lambda^s(u) = \pi_{uv}^s \cdot \Lambda^s(v)$ with

probability at least $1 - \varepsilon/k$ (for some μ and γ). Thus by the union bound, $\Lambda(u) = \pi_{uv} \cdot \Lambda(v)$ with probability $(1 - \varepsilon)$. This concludes the proof. $\qquad \square$

7   Sherali–Adams Lower Bounds from UG Hardness

In this section, we will show how to convert UGC-hardness results to Sherali–Adams lower bounds. Inapproximability proofs based on the Unique Games Conjecture usually proceed as follows. First they show how to transform every instance U of the Unique Games Problem (UG) to an instance P_U of the given problem $\mathcal{P}$ so that (completeness) if U is $(1 - \varepsilon)$-satisfiable then P_U has a "good" solution; (soundness) if U is not δ-satisfiable then P_U has no "good" solution. To prove the completeness, they show how to convert every solution of U to a solution of P_U. Usually, this conversion is local: the value assigned to every variable/vertex/edge in P_U depends only on labels of a constant number of vertices of U; every "constraint" in P_U is satisfied if a constant number of constraints is satisfied in U.

    We can apply such reduction to the Unique Games instance U presented in the previous section. Since U is only δ-satisfiable, P_U has no "good" solution. However, P_U has "good" local solutions corresponding to $(1 - \varepsilon)$-satisfiable local solutions for U. Using this approach we can convert most known (if not all) UGC-hardness results to Sherali–Adams integrality gaps.

    Now we apply the approach outlined above to a recent result of Guruswami, Manokaran, and Raghavendra [12] that states that it is NP-hard to approximate the Maximum Acyclic Subgraph Problem within a factor of $2 - \varepsilon$ (for every ε) assuming the Unique Games Conjecture.

Theorem 7.1. *For every $\varepsilon > 0$ there exists $\gamma > 0$ s.t. for infinitely many n there exists a weighted directed graph $G = (V_G, E_G)$ on n vertices (the total weight of all edges is 1) with the following properties.*

    1. For every ordering $\phi : V_E \rightarrow [n]$ of V_E, the weight of edges going forward is at most $1/2 + \varepsilon$.

    2. There exists a Sherali–Adams solution of value $1 - \varepsilon$. I.e. there exist local distributions of orderings $\phi_S : S \rightarrow [n]$ on sets $S \subset V_E$ of size at most $k = n^\gamma$ that are locally consistent s.t.
$$\mathbb{E}_{(i,j) \in E_G} \left[\Pr(\phi_{\{i,j\}}(i) < \phi_{\{i,j\}}(j)) \right] \geq 1 - \varepsilon,$$
where edges are sampled with probabilities equal to their weight (i.e. almost all edges go forward).

    First, we briefly describe the reduction of Guruswami, Manokaran, and Raghavendra [12]. They start with a Unique Games instance U on a bipartite graph $(A \cup B, E)$ with label set $\mathcal{S}$. They present a procedure that given U and $\varepsilon > 0$ outputs a directed graph $G_{U,\varepsilon}$ with set of vertices $B \times [m]^{\mathcal{S}}$ (where m is a number, which depends on ε). Two vertices $(b_1, z_1), (b_2, z_2) \in B \times [m]^{\mathcal{S}}$ are connected with an edge if b_1 and b_2 have a common neighbor a in A; the weight of the edge depends on z_1 and z_2.

Theorem 7.2 ([12], Theorem 7.1). *For every $\varepsilon > 0$ there exists $\delta > 0$ s.t. (soundness) if U is not δ-satisfiable then no ordering of G_U has value more than $\frac{1}{2} + \varepsilon$. (completeness) Given a Unique Game labeling $\Lambda : A \cup B \to \mathcal{S}$, let the rank of a vertex (b, z) be $\mathcal{O}(b, z) = z(\Lambda(b)) \in [m]$. The rank defines an ordering: sort vertices w.r.t. their rank, order vertices of the same rank uniformly at random. Let p_e be the probability that the edge $e = ((b_1, z_1), (b_2, z_2))$ goes forward given that the constraints between a and b_1, and a and b_2 are satisfied in U (where a is the common neighbor of b_1 and b_2). Then $\mathbb{E}_e [p_e] \geq 1 - \varepsilon$.*

Proof of Theorem 7.1. Let U be a Unique Games instance from Theorem 6.1 that is at most $\delta/2$-satisfiable with a Sherali–Adams solution of value $1 - \varepsilon/4$. Partition vertices into two sets A and B so that at least half edges go from A to B. We remove edges within A and B and get a bipartite instance of Unique Games.

Now we apply the reduction from Theorem 7.2 to U and get a graph $G_{U,\varepsilon/2}$. The number of vertices in $G_{U,\varepsilon/2}$ is linear in the number of vertices of U; denote it by n. Since U is at most δ-satisfiable, no ordering of G_U has value more than $\frac{1}{2} + \varepsilon$. Now for every set of vertices S of G_U, consider the set T of corresponding vertices in B and their neighbors: $T = \{b : (b, z) \in S\} \cup \{a : (a, b) \in E, (b, z) \in S\}$. Sample a local labeling Λ_T of vertices in T. The labeling Λ_T defines a rank for every vertex in S. We randomly map vertices in S of rank $r \in [m]$ to distinct positions in $\{(r-1)n/m+1, \ldots, rn/m\}$ and obtain an ordering ϕ_S. Since local labelings Λ_T are consistent, local orderings ϕ_S are also consistent. Since Λ_T satisfies each Unique Games constraint with probability at least $1 - \varepsilon/4$, for every edge $e = ((b_1, z_1), (b_2, z_2))$ within S both constraints between a and b_1, and a and b_2 are satisfied with probability at least $1 - \varepsilon/2$ (here a is the common neighbor of b_1 and b_2). Therefore, the edge e goes forward with probability at least $(1 - \varepsilon/2)p_e$. Thus

$$\mathbb{E}_{e=(i,j)} \left[\Pr(\phi_{\{i,j\}}(i) < \phi_{\{i,j\}}(j)) \right] \geq (1 - \varepsilon/2)\mathbb{E}_e [p_e]$$
$$\geq (1 - \varepsilon/2)(1 - \varepsilon/2) \geq 1 - \varepsilon.$$

$\square$

8 Relaxations for Multicut

We now present a Sherali–Adams integrality gap for the Min Multicut problem. In this problem, we are given a graph and a set of source terminal pairs (s_i, t_i). Our goal is to find a multicut separating all source terminal pairs that minimizes the number of cut edges. As before for every two vertices u and v we introduce a variable x_{uv}. The objective is to minimize $\sum_{(u,v) \in E} x_{uv}$, given that the vector x_{uv} lies in the Sherali–Adams relaxation (with r rounds) of the cut polytope and subject to an extra constraint $x_{s_i t_i} = 1$ for every i (i.e. s_i and t_i are always separated).

Theorem 8.1. *The integrality gap for the Min Multicut problem is*

$$\Omega \left(\sqrt{\frac{\log n}{\log r + \log \log n}} \right)$$

15

after r rounds of Sherali–Adams relaxation.

The proof is based on a black box reduction from MAX CUT. Consider a MAX CUT instance G. Construct a new graph G': For every original vertex v introduce two new vertices v_{-1} and v_1. If (u, v) are connected by an edge in G, we connect u_1 with v_{-1} and u_{-1} with v_1 in G'. The source terminal pairs are (u_{-1}, u_1) i.e. the multicut should separate every u_{-1} from u_1. Notice, that if the cost of a multicut is δ, then there exists a cut in G of cost $1 - 2\delta$: indeed, if we remove all edges cut by the multicut from G, then the remaining graph does not contain odd cycles and thus is bipartite (formally, we remove an edge (u, v) if (u_1, v_{-1}) or (u_{-1}, v_1) is cut by the multicut). Thus if the cost of MAX CUT in G is less than $2/3$, then the cost of Multicut is at least $1/6$. On the other hand, any feasible LP solution x_{ij} for MAX CUT corresponds to a feasible LP solution $\tilde{x}_{u_a v_b}$ for Multicut:

$$\tilde{x}_{u_{-1}v_1} = \tilde{x}_{u_1v_{-1}} = 1 - x_{uv}; \quad \tilde{x}_{u_1v_1} = \tilde{x}_{u_{-1}v_{-1}} = x_{uv}.$$

Note that in the standard MAX CUT relaxation we can introduce variables x_{uu} and add a constraint $x_{uu} = 0$, since u always lies in the same part where u, itself, lies. This implies

$$\tilde{x}_{u_{-1}u_1} = x_{uu} = 1.$$

If the cost of the optimal LP solution for MAX CUT is $1 - \varepsilon$, then the cost of the optimal LP solution for Multicut is at most ε. Hence, the integrality gap is $\Omega(1/\varepsilon)$. By Theorem 5.3, after r rounds of Sherali–Adams, $\varepsilon = O\left(\sqrt{\frac{\log r + \log \log n}{\log n}}\right)$.

Thus the integrality gap is $\Omega\left(\sqrt{\frac{\log n}{\log r + \log \log n}}\right)$ (the hidden constant in Ω notation depends on Δ which we fix to be a small constant).

9 Integrality Gap for Sparsest Cut for Small Number of Rounds

In this section, we will prove the following theorem. We borrow many ideas from [6].

Theorem 9.1. *The integrality gap for the Sparsest Cut relaxation is at least $\Omega\left(\frac{\log n}{r + \log \log n}\right)$ after r rounds of the Sherali–Adams lift-and-project.*

Definition 9.2. *Denote the graph obtained from a graph H by randomly and independently removing each of its edges with probability μ by H_μ. Denote the set of removed ("cut") edges by C_μ.*

We will need the following lemma. The proof is similar to the proof of Theorem 3.3 from [3].

Lemma 9.3. *Consider an l-path decomposable graph H. The probability that two vertices u and v lie in the same connected component of H_μ is between $(1 - \mu)^{d(u,v)}$ and $(1 - \mu)^{d(u,v)} + (1 - \mu)^L m/L$, where m is the number of vertices in H and $L = \lfloor l/3 \rfloor$.*

Proof. Denote the probability that two vertices x and y lie in the same component of H_μ by p_{xy}. Our goal is to prove

$$(1 - \mu)^{d(u,v)} \le p_{uv} \le (1 - \mu)^{d(u,v)} + (1 - \mu)^L m/L. \tag{2}$$

The lower bound on p_{uv} in (2) is obvious, since with probability at least $(1 - \mu)^d$ no edge is cut on the shortest path between u and v.

Now we prove the upper bound by induction. Without loss of generality we assume that H is connected (otherwise it suffices to prove the lemma for each connected component).

Since H is l-path decomposable, one of the following cases should hold:

1. H is an edge.

2. H is a one vertex union of l-path decomposable subgraphs.

3. H is 2-connected. It is the union of an l-path decomposable subgraph and a path of length l. The subgraph and the path do not have common vertices except for the endpoints of the path.

Therefore, we can iteratively decompose H into small subgraphs. We stop when each remaining subgraph is an edge. Denote by $D(H)$ the number of times we decompose a subgraph according to item (3). We will prove that

$$p_{uv} \le (1 - \mu)^{d(u,v)} + (1 - \mu)^L D(H).$$

Since $D(H) \le m/(l - 2) \le m/L$ this bound implies the desired upper bound on p_{uv}.

Let us verify the base case: if H is an edge, then $p_{uv} = 1 - \mu$ (for $u \ne v$). Now we prove the induction step. Consider two possibilities.

1. H is a one vertex union of two subgraphs H_1 and H_2. If u and v lie in the same subgraph then the bound holds by the induction hypothesis. Now suppose that u is in H_1 and v is in H_2. Denote the cut point by c. We have,

$$\begin{aligned}
p_{uv} = p_{uc} p_{vc} &\le \min\left((1 - \mu)^{d(u,c)} + (1 - \mu)^L D(H_1), 1\right) \\
&\quad \times \min\left((1 - \mu)^{d(v,c)} + (1 - \mu)^L D(H_2), 1\right) \\
&\le (1 - \mu)^{d(u,c)+d(v,c)} + (1 - \mu)^L D(H_1) + (1 - \mu)^L D(H_2) \\
&= (1 - \mu)^{d(u,v)} + (1 - \mu)^L D(H).
\end{aligned}$$

2. H is the union of a subgraph H_0 and a path P (H_0 and P intersect only at the endpoints of P; the length of P is at least l). Divide P into 3 subpaths of length at least L each. Note that one of these subpaths does not contain neither u nor v as an internal point. Denote this subpath by P_1. Denote the union of the other two subpaths and H_0 by H_1. Applying the induction hypothesis to H_1, we get

$$\Pr\left(u \text{ and } v \text{ are not separated by the cut in } H_1\right) \le (1 - \mu)^{d(u,v)} + (1 - \mu)^L D(H_1).$$

The probability that that P_1 is not cut is at most $(1 - \mu)^L$. By the union bound,

$$p_{uv} \leq (1 - \mu)^{d(u,v)} + (1 - \mu)^L D(H_1) + (1 - \mu)^L = (1 - \mu)^{d(u,v)} + (1 - \mu)^L D(H).$$

$\square$

Definition 9.4. *Consider a graph H of girth at least $2L + 1$. We say that two vertices u and v are connected only by a non-canonical path if:*

1. *u and v are connected by a path in H_μ; and*

2. *no path of length less than or equal to L connects u and v in H_μ.*

In other words, the second condition means that either the distance in H between u and v is greater than L or the shortest path between them is cut by C_μ.

Corollary 9.5. *Let H be an l-path decomposable graph of girth at least $2L + 1$. Then the probability that two given vertices are connected only by a non-canonical path is at most $(1 - \mu)^L(m/L + 1)$, where m is the number of vertices in H.*

Proof. Denote these vertices by u and v. If $d(u, v) \geq L$ then the probability that u and v are connected with any path in h_μ is at most

$$(1 - \mu)^{d(u,v)} + (1 - \mu)^L m/L \leq (1 - \mu)^L(m/L + 1).$$

If $d(u, v) \leq L$ then the desired probability equals

$\Pr\big(u$ and v are connected with a path$\big) - \Pr\big($the shortest path between u and v is not cut$\big)$
$$\leq (1 - \mu)^{d(u,v)} + (1 - \mu)^L m/L - (1 - \mu)^{d(u,v)} < (1 - \mu)^L(m/L + 1).$$

$\square$

Definition 9.6. *Consider a graph G of girth at least $2L + 1$ and a set of vertices T. Define the subgraph $\mathrm{clos}(T)$, the closure of T, as follows:*

- *$\mathrm{clos}(T)$ contains all vertices of T;*

- *for every two vertices $u, v \in T$ at distance at most L, $\mathrm{clos}(T)$ contains the shortest path between them.*

Note that $|\mathrm{clos}(T)| \leq L|T|^2$.

Proof of Theorem 9.1. Let $G = (V, E)$ be a bounded degree expander on n vertices of girth at least l such that every its subgraph on $k = 2r + 3$ vertices is l-path decomposable, where $l = \Theta(\log n)$. The existence of such graph was proved by Arora, Bollobás, Lovász, and Tourlakis [1] (the graph is a random 3-regular graph with $o(n)$ edges removed).

18

Let $L = \lceil l/3 \rceil$. Choose $\mu = \Theta(\frac{k + \log \log n}{L}) = \Theta(\frac{k + \log \log n}{\log n})$ so that

$$2^{2k} k^4 (1 - \mu)^L \leq \mu. \tag{3}$$

For every subset $T \subset V$ of vertices, we are going to define a random cut of T; then we will use Lemma 2.1. However, we will proceed in two steps: first we will define distributions of random cuts that will only approximately satisfy conditions of Lemma 2.1, and then we will slightly change these distributions.

We cut every edge of G independently with probability μ and obtain the graph G_μ. Consider cut edges that lie in $\mathrm{clos}(T)$. They partition the vertices of $\mathrm{clos}(T)$ into connected components. Let $\mathcal{P}_T$ be the restriction of this partition to T. Choose every piece of $\mathcal{P}_T$ with probability $1/2$. Let S_T be the union of chosen pieces.

Lemma 9.7. *Given a set of vertices T of size at most k and $Q \subset T$, consider the probabilistic partitions $\mathcal{P}_T$ and $\mathcal{P}_Q$. Denote the restriction of $\mathcal{P}_T$ to Q by $\mathcal{P}_T\big|_Q$. Then $\left| \Pr\left(\mathcal{P}_T\big|_Q \neq \mathcal{P}_Q \right) \right| \leq k^4 (1 - \mu)^L$.*

Proof. If $\mathcal{P}_T\big|_Q \neq \mathcal{P}_Q$ then there exist vertices $u, v \in Q$ that belong to one piece in one partition and to different pieces in the other partition. Now if u and v belong to the same piece in $\mathcal{P}_Q$ then they must belong to the same piece in $\mathcal{P}_T\big|_Q$ since $\mathrm{clos}(T)$ is a subgraph of $\mathrm{clos}(Q)$. On the other hand, if u and v belong to the same piece of $\mathcal{P}_T\big|_Q$ but they belong to different pieces in $\mathcal{P}_Q$ then they are connected only by a non-canonical path in $\mathrm{clos}(T)$. The probability of this event is at most

$$(1 - \mu)^L \left(\frac{|\mathrm{clos}(T)|}{L} + 1 \right) \leq (1 - \mu)^L \left(\frac{k^2 L}{L} + 1 \right) = (k^2 + 1)(1 - \mu)^L.$$

The total number of pairs (u, v) is at most $\binom{k}{2}$, so by the union bound (for $k \geq 2$)

$$\Pr\left(\mathcal{P}_T\big|_Q \neq \mathcal{P}_Q \right) \leq \frac{k(k+1)(k^2+1)}{2}(1 - \mu)^L \leq k^4 (1 - \mu)^L.$$

$\square$

Corollary 9.8. *Let T be a set of size at most k and $Q \subset T$. Then $TV(S_T \cap Q, S_Q) \leq k^4 (1 - \mu)^L$, where TV is the total variation distance.*

Our goal now is to construct random cuts $\hat{S}_T$ (which are perturbations of cuts S_T) so that $\hat{S}_T \cap Q$ and $\hat{S}_Q$ have exactly the same distribution. Note that by the inclusion–exclusion formula, for every set T and $A \subset T$, we have $\Pr(S_T = A) = \sum_{B \subset T \setminus A}(-1)^{|B|} \Pr(A \cup B \subset S_T)$. We use this identity to define a signed measure $\mathcal{E}_T$ on subsets (cuts) of T. Let $q_T = \Pr(S_T = T)$ and $\mathcal{E}_T(\{A\}) = \sum_{B \subset T \setminus A}(-1)^{|B|} q_{A \cup B}$. The measure of sets that contain a fixed set $M \subset T$

is

$$\mathcal{E}_T(\{A : M \subset A \subset T\}) = \sum_{A:M\subset A\subset T} \sum_{B\subset T\setminus A} (-1)^{|B|} q_{A\cup B}$$

$$= \sum_{C\subset T\setminus M} \left(\sum_{B\subset C}(-1)^{|B|}\right) q_{C\cup M} = q_{\varnothing\cup M} = q_M.$$

where $C = (A \cup B) \setminus M$. In particular,

$$\mathcal{E}_T(2^T) = \mathcal{E}_T(\{A : \varnothing \subset A \subset T\}) = q_\varnothing = 1,$$

where 2^T is the set of all subsets of T. By Corollary 9.8,

$$|q_{A\cup B} - \Pr(A \cup B \subset S_T)| = |\Pr(A \cup B = S_{A\cup B}) - \Pr(A \cup B = S_T \cap (A \cup B))| \le k^4(1-\mu)^L.$$

Therefore, we have (we use bound (3))

$$|\mathcal{E}_T(\{A\}) - \Pr(S_T = A)| \le \sum_{B\subset T\setminus A} k^4(1 - \mu)^L \le \frac{\mu}{2^k}.$$

Define the random set (cut) $\hat{S}_T$ as follows. For $A \subset T$, let

$$\Pr\left(\hat{S}_T = A\right) = \frac{\mu}{2^{|T|}} + (1 - \mu)\mathcal{E}_T(\{A\}).$$

We need to verify that this definition is valid. Indeed, all the probabilities are non-negative

$$\Pr\left(\hat{S}_T = A\right) \ge \frac{\mu}{2^{|T|}} + (1 - \mu)\left(\Pr(S_T = A) - \frac{\mu}{2^k}\right)$$

$$\ge \frac{\mu}{2^{|T|}} - (1 - \mu)\frac{\mu}{2^k} \ge 0.$$

For a given set $M \subset T$ the probability that $\hat{S}_T$ contains M equals

$$\sum_{A:M\subset A\subset T}\left(\frac{\mu}{2^{|T|}} + (1 - \mu)\mathcal{E}_T(\{A\})\right) = \frac{\mu}{2^{|M|}} + (1 - \mu)q_M.$$

In particular, the total probability of all elementary events $\{\hat{S}_T = A\}$ equals $\frac{\mu}{2^{|\varnothing|}} + (1-\mu)q_\varnothing = 1$. Finally,

$$\Pr\left(\hat{S}_T \cap Q = A\right) = \sum_{B\subset Q\setminus A} (-1)^{|B|} \Pr\left(A \cup B \subset \hat{S}_T \cap Q\right)$$

$$= \sum_{B\subset Q\setminus A} (-1)^{|B|} \Pr\left(A \cup B \subset \hat{S}_T\right)$$

$$= \sum_{B\subset Q\setminus A} (-1)^{|B|} \left(\frac{\mu}{2^{|A\cup B|}} + (1 - \mu)q_{A\cup B}\right) = \Pr\left(\hat{S}_Q = A\right).$$

Therefore, by Lemma 2.1, the solution defined by

$$x_{ij} = \Pr\left(\hat{S}_{\{i,j\}} = \{i\} \text{ or } \hat{S}_{\{i,j\}} = \{j\}\right)$$

belongs to the Sherali–Adams relaxation of the cut polytope. Let us compute the value of this solution. Note that

$$\Pr\left(S_{\{i,j\}} = \{i\}\right) = \frac{\mu}{4} + (1 - \mu)\mathcal{E}_{\{i,j\}}(\{i\})$$
$$= \frac{\mu}{4} + (1 - \mu)(q_{\{i\}} - q_{\{i,j\}}) = \frac{\mu}{4} + (1 - \mu)\left(\frac{1}{2} - q_{\{i,j\}}\right).$$

Now, if $d(i,j) \leq L$ then $\text{clos}(\{i,j\})$ is the path between i and j. Therefore, $q_{\{i,j\}}$ equals

$$\frac{\Pr\left(\mathcal{P}_{\{i,j\}} \text{ sep. } i,j\right)}{4} + \frac{\Pr\left(\mathcal{P}_{\{i,j\}} \text{ doesn't sep. } i,j\right)}{2} = \frac{1 - (1 - \mu)^{d(i,j)}}{4} + \frac{(1 - \mu)^{d(i,j)}}{2}$$
$$= \frac{1 + (1 - \mu)^{d(i,j)}}{4}.$$

If $d(i,j) \geq L$ then i and j are disconnected in $\text{clos}(\{i,j\})$. Therefore, $q_{\{i,j\}} = 1/4$. We have

$$\Pr\left(S_{\{i,j\}=\{i\}}\right) = \begin{cases} \frac{\mu}{4} + (1 - \mu)\frac{1 - (1-\mu)^{d(i,j)}}{4}, & \text{if } d(i,j) \leq L; \\ \frac{1}{4}, & \text{if } d(i,j) > L. \end{cases}$$

Finally,

$$x_{ij} = \begin{cases} \frac{\mu}{2} + (1 - \mu)\frac{1 - (1-\mu)^{d(i,j)}}{2}, & \text{if } d(i,j) \leq L; \\ \frac{1}{2}, & \text{if } d(i,j) > L. \end{cases}$$

The distance between adjacent vertices in G is $\mu(2 - \mu)/2 < \mu$. The distance between two vertices at distance $\Theta(\log n)$ is $\Theta(1)$. Therefore, the cost of the solution x_{ij} is $O(\mu) = O\left(\frac{k + \log\log n}{\log n}\right)$. Since the graph G is an expander, the cost of combinatorial solution is $\Theta(1)$. We proved that the integrality gap is $\Omega\left(\frac{\log n}{k + \log\log n}\right)$. $\qquad\square$

References

[1] S. Arora, B. Bollobás, L. Lovász, and I. Tourlakis. *Proving Integrality Gaps without Knowing the Linear Program.* Theory of Computing, vol. 2, pp. 19–51, 2006.

[2] S. Arora, L. Lovász, I. Newman, Y. Rabani, Y. Rabinovich, and S. Vempala. *Local versus global properties of metric spaces.* Proceedings of the 17th Symposium on Discrete Algorithms, pp. 41–50, 2006.

[3] M. Charikar, K. Makarychev, and Y. Makarychev. *Local Global Tradeoffs in Metric Embeddings.* ECCC Technical Report. Proceedings of the 48th Symposium on Foundations of Computer Science, pp. 713–723, 2007.

[4] Eden Chlamtac. *Approximation Algorithms Using Hierarchies of Semidefinite Programming Relaxations.* Proceedings of the 47th Symposium on Foundations of Computer Science, pp. 691-701, 2006.

[5] E. Chlamtac and G. Singh. *Improved Approximation Guarantees through Higher Levels of SDP Hierarchies.* Proceedings of the 11th International Workshop, APPROX 2008, pp. 49–62.

[6] W. F. de la Vega and C. Kenyon-Mathieu. *Linear Programming Relaxations of Maxcut.* Proceedings of the 18th Symposium on Discrete Algorithms, pp. 53–61, 2007.

[7] K. Georgiou, A. Magen, T. Pitassi, and I. Tourlakis. *Integrality gaps of $2 - o(1)$ for vertex cover SDPs in the Lovász–Schrijver hierarchy.* Proceedings of the 48th Symposium on Foundations of Computer Science, pp. 702–712, 2007.

[8] M. Goemans and D. Williamson. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming.* Journal of the ACM, vol. 42, no. 6, pp. 1115–1145, Nov. 1995.

[9] M. Grötschel, M. Jünger, and G. Reinelt, *A cutting plane algorithm for the linear ordering problem.* Operations Research 32 (1984), pp. 1195-1220.

[10] M. Grötschel, M. Jünger, and G. Reinelt, *On the acyclic subgraph polytope.* Mathematical Programming, 32 (1985), pp. 28–42.

[11] M. Grötschel, M. Jünger, and G. Reinelt, *Facets of the linear ordering polytope.* Mathematical Programming, 32 (1985), pp. 43–60.

[12] V. Guruswami, R. Manokaran, and P. Raghavendra. *Beating the Random Ordering is Hard: Inapproximability of Maximum Acyclic Subgraph.* Proceedings of the 49th Symposium on Foundations of Computer Science, pp. 573–582, 2008.

[13] J. B. Lasserre. *Global Optimization with Polynomials and the Problem of Moments.* SIAM J. Optim. vol. 11, no. 3, pp. 796–817, 2001.

[14] M. Laurent. *A comparison of the Sherali–Adams, Lovász–Schrijver and Lasserre relaxations for $0 - 1$ programming.* Mathematics of Operations Research, vol. 28, pp. 470–496, 2003.

[15] L. Lovász and A. Schrijver. *Cones of matrices and set-functions and 0-1 optimization.* SIAM J. Optim., vol. 1, pp. 166–190, 1991.

[16] H. D. Sherali and W. P. Adams. *A Hierarchy of Relaxations between the Continuous and Convex Hull Representations for Zero-One Programming Problems.* SIAM J. Discrete Math.,, vol. 3, pp. 411–430, 1990.

[17] G. Schoenebeck. *Linear Level Lasserre Lower Bounds for Certain k-CSPs.* Proceedings of the 49th Symposium on Foundations of Computer Science, pp. 593–602, 2008.

[18] G. Schoenebeck, L. Trevisan, and M. Tulsiani. *Tight Integrality Gaps for Lovász–Schrijver LP Relaxations of Vertex Cover and Max Cut.* Proceedings of the 39th Symposium on Theory of Computing, pp. 302–310, 2007.

[19] G. Schoenebeck, L. Trevisan, and M. Tulsiani. *A Linear Round Lower Bound for Lovász–Schrijver SDP Relaxations of Vertex Cover.* Proceedings of the 22nd Conference on Computational Complexity, pp. 205–216, 2007.

[20] I. Tourlakis. *New lower bounds for Approximation Algorithms in the Lovász–Schrijver Hierarchy.* Ph.D. Thesis, Princeton University, 2006.

A   Proof of Lemma 2.1

First we define the lifted solution. Consider a set $I = \{(i_1, j_1), \ldots, (i_t, j_t)\}$ of size at most $r + 1$. Denote

$$Q_I = \{i_1, j_1, i_2, j_2, \ldots, i_t, j_t\}.$$

Let y_I be the probability that a cut chosen according to $\mathcal{D}_{Q_I}$ separates every i_s from j_s. Note that x_{ij} is indeed the projection of $y_{\{(i,j)\}}$: $x_{ij} = y_{\{(i,j)\}}$.

We verify now that the vector of y_I lies in the lifted Sherali–Adams polytope. Let I and J be two sets of pairs of points such that $|I \cup J| \leq r$; let $\mathcal{R}(x) \geq 0$ be one of the base inequalities. Consider the set of points T appearing in I, J and $\mathcal{R}$. Let $\tilde{x}_{ij}$ be the conditional probability w.r.t. $\mathcal{D}_T$ that points i and j are separated given that every pair of points in the set I is separated and every pair of points in the set J is not separated. According to (1) we replace each variable x_{ij} in $\mathcal{R}$ with

$$
\begin{aligned}
x_{ij} \rightsquigarrow \sum_{B \subset J} (-1)^{|B|} y_{\{(i,j)\} \cup I \cup B} &= \sum_{B \subset J} (-1)^{|B|} \times \\
&\quad \Pr\big(\mathcal{D}_{Q_{\{(i,j)\} \cup I \cup B}} \text{ separates every pair in } \{(i,j)\} \cup I \cup B\big) \\
&= \sum_{B \subset J} (-1)^{|B|} \Pr(\mathcal{D}_T \text{ separates every pair in } \{(i,j)\} \cup I \cup B) \\
&= \Pr\big(\mathcal{D}_T \text{ separates } i \text{ and } j, \text{ separates every pair in } I, \\
&\qquad \text{does not separate any pair in } J\big) \\
&= \tilde{x}_{ij} \times \Pr_{\mathcal{D}_T}\big(\text{every pair in } I \text{ is separated};\\
&\qquad \text{every pair in } J \text{ is not separated}\big).
\end{aligned}
$$

Now it is easy to see that the constraint on the lifted polytope obtained from $\mathcal{R}$ is

$$
\mathcal{R}(\tilde{x}) \times \Pr_{\mathcal{D}_T}\big(\text{every pair of points in } I \text{ is separated and}
$$

$$
\text{every pair of points in } J \text{ is not separated}\big) \geq 0.
$$

Note that by the definition the vector $(\tilde{x}_{ij})_{i,j \in T}$ equals a convex combination of cut metrics and, therefore, lies in the cut polytope. Hence the inequality above holds.