Near-Optimal Algorithms for Unique Games

Moses Charikar[*] Konstantin Makarychev [†] Yury Makarychev [‡]

Princeton University

Abstract

Unique games are constraint satisfaction problems that can be viewed as a generalization of Max-Cut to a larger domain size. The Unique Games Conjecture states that it is hard to distinguish between instances of unique games where almost all constraints are satisfiable and those where almost none are satisfiable. It has been shown to imply a number of inapproximability results for fundamental problems that seem difficult to obtain by more standard complexity assumptions. Thus, proving or refuting this conjecture is an important goal. We present significantly improved approximation algorithms for unique games. For instances with domain size k where the optimal solution satisfies $1 - \varepsilon$ fraction of all constraints, our algorithms satisfy roughly $k^{-\varepsilon/(2-\varepsilon)}$ and $1 - O(\sqrt{\varepsilon \log k})$ fraction of all constraints. Our algorithms are based on rounding a natural semidefinite programming relaxation for the problem and their performance almost matches the integrality gap of this relaxation. Our results are near optimal if the Unique Games Conjecture is true, *i.e.* any improvement (beyond low order terms) would refute the conjecture.

1 Introduction

Given a set of linear equations over Z_p with two variables per equation, consider the problem of finding an assignment to variables that satisfies as many equations (constraints) as possible. If there is an assignment to the variables which satisfies all the constraints, it is easy to find such an assignment. On the other hand, if there is an assignment that satisfies almost all constraints (but not all), it seems quite difficult to find a good satisfying assignment. This is the essence of the Unique Games Conjecture of Khot[10].

One distinguishing feature of the above problem on linear equations is that every constraint corresponds to a bijection between the values of the associated variables. For every possible value of one variable, there is a unique value of the second variable that satisfies the constraint. Unique games are systems of constraints – a generalization of linear equations discussed above – that have this uniqueness property (first considered by Feige and Lovász [6]).

[*]moses@cs.princeton.edu Supported by NSF ITR grant CCR-0205594, NSF CAREER award CCR-0237113, MSPA-MCS award 0528414, and an Alfred P. Sloan Fellowship.

[†]kmakaryc@cs.princeton.edu Supported by a Gordon Wu fellowship. Part of this work was done at Microsoft Research.

[‡]ymakaryc@cs.princeton.edu Supported by a Gordon Wu fellowship. Part of this work was done at Microsoft Research.

Definition 1.1 (Unique Game). *A unique game consists of a constraint graph $G = (V, E)$, a set of variables x_u (for all vertices u) and a set of permutations π_{uv} on $[k] = \{1, \dots, k\}$ (for all edges (u, v)). Each permutation π_{uv} defines the constraint $\pi_{uv}(x_u) = x_v$. The goal is to assign a value from the set $[k]$ to each variable x_u so as to maximize the number of satisfied constraints.*

As in the setting of linear equations, instances of unique games where all constraints are satisfiable are easy to handle. Given an instance where $1 - \varepsilon$ fraction of constraints are satisfiable, the Unique Games Conjecture (UGC) of Khot [10] says that it is hard to satisfy even δ fraction of the constraints. More formally, the conjecture is the following.

Conjecture 1 (Unique Games Conjecture [10]). *For any constants $\varepsilon, \delta > 0$, for any $k > k(\varepsilon, \delta)$, it is NP-hard to distinguish between instances of unique games with domain size k where $1 - \varepsilon$ fraction of constraints are satisfiable and those where δ fraction of constraints are satisfiable.*

This conjecture has attracted a lot of recent attention since it has been shown to imply hardness of approximation results for several important problems: MaxCut [11, 15], Min 2CNF Deletion [3, 10], MultiCut and Sparsest Cut [3, 14], Vertex Cover [13], and coloring 3-colorable graphs [5] (based on a variant of the UGC), that seem difficult to obtain by standard complexity assumptions.

Note that a random assignment satisfies a $1/k$ fraction of the constraints in a unique game. Andersson, Engebretsen, and Håstad [2] considered semidefinite program (SDP) based algorithms for systems of linear equations mod p (with two variables per equation) and gave an algorithm that performs (very slightly) better than a random assignment. The first approximation algorithm for general unique games was given by Khot [10], and satisfies $1 - O(k^2 \varepsilon^{1/5} \sqrt{\log(1/\varepsilon)})$ fraction of all constraints if $1 - \varepsilon$ fraction of all constraints is satisfiable. Recently Trevisan [17] developed an algorithm that satisfies $1 - O(\sqrt[3]{\varepsilon \log n})$ fraction of all constraints (this can be improved to $1 - O(\sqrt{\varepsilon \log n})$ [9]), and Gupta and Talwar [9] developed an algorithm that satisfies $1 - O(\varepsilon \log n)$ fraction of all constraints. The result of [9] is based on rounding an LP relaxation for the problem, while previous results use SDP relaxations for unique games.

There are very few results that show hardness of unique games. Feige and Reichman [7] showed that for every positive ε there is c s.t. it is NP-hard to distinguish whether c fraction of all constraints is satisfiable, or only εc fraction is satisfiable.

Our Results. We present two new approximation algorithms for unique games. We state our guarantees for instances where $1 - \varepsilon$ fraction of constraints are satisfiable. The first algorithm satisfies

$$\Omega\left(\min(1, \frac{1}{\sqrt{\varepsilon \log k}}) \cdot (1 - \varepsilon)^2 \cdot \left(\frac{k}{\sqrt{\log k}}\right)^{-\varepsilon/(2-\varepsilon)}\right) \tag{1}$$

fraction of all constraints. The second algorithm satisfies $1 - O(\sqrt{\varepsilon \log k})$ fraction of all constraints and has a better guarantee for $\varepsilon = O(1/\log k)$. We apply the same techniques for d-to-1 games as well.

In order to understand the complexity theoretic implications of our results, it is useful to keep in mind that inapproximability reductions from unique games typically use the "Long Code", which increases the size of the instance by a 2^k factor. Thus, such applications of unique games usually have domain size $k = O(\log n)$. In Figure 1, we summarize known algorithmic guarantees for unique games. In order to compare these different guarantees in the context of hardness applications (*i.e.* $k = O(\log n)$), we compare the range of values of ε (as a function of k) for which each of these algorithms beats the performance of a random assignment.

Algorithm	Guarantee for $OPT = 1 - \varepsilon$	Threshold ε
Khot [10]	$1 - O(k^2 \varepsilon^{1/5} \sqrt{\log(1/\varepsilon)})$	$\tilde{O}(1/k^{10})$
Trevisan [17]	$1 - O(\sqrt[3]{\varepsilon \log n})$	$O(1/k)$
Gupta-Talwar [9]	$1 - O(\varepsilon \log n)$	$O(1/k)$
This paper	$\approx \Omega(k^{-\varepsilon/(2-\varepsilon)})$	const < 1
	$1 - O(\sqrt{\varepsilon \log k})$	$O(1/\log k)$

Figure 1: Summary of results. The guarantee represents the fraction of constraints satisfied for instances where $OPT = 1 - \varepsilon$. The threshold represents the range of values of ε for which the algorithm beats a random assignment (computed for $k = O(\log n)$).

Our results show limitations on the hardness bounds achievable using the UGC and stronger versions of it. Chawla, Krauthgamer, Kumar, Rabani, and Sivakumar [3] proposed a strengthened form of the UGC, conjecturing that it holds for $k = \log n$ and $\varepsilon = \delta = \frac{1}{(\log n)^{\Omega(1)}}$. This was used to obtain an $\Omega(\log \log n)$ hardness for sparsest cut. Our results refute this strengthened conjecture.[1]

The performance of our algorithms is naturally constrained by the integrality gap of the SDP relaxation, *i.e.* the smallest possible value of an integer solution for an instance with SDP solution of value $(1-\varepsilon)|E|$. Khot and Vishnoi [14] constructed a gap instance for the semidefinite relaxation[2] for the Unique Games Problem where the SDP satisfies $(1-\varepsilon)$ fraction of constraints, but the optimal solution can satisfy at most $O(k^{-\varepsilon/9})$ (one may show that their analysis can yield $O(k^{-\varepsilon/4+o(\varepsilon)})$). This shows that our results are almost optimal for the standard semidefinite program.

After we unofficially announced our results, Khot, Kindler, Mossel and O'Donnell [12] showed that a reduction in an earlier version of their paper [11] together with the techniques in the recently proved Majority is Stablest result of Mossel, O'Donnell, and Oleszkiewicz [15] give lower bounds for unique games that almost match the upper bounds we obtain. They establish the following hardness results (in fact, for the special case of linear equations mod p):

Theorem 1.2 ([12], Corollary 13). *The Unique Games Conjecture implies that for every fixed $\varepsilon > 0$, for all $k > k(\varepsilon)$, it is NP-hard to distinguish between instances of unique games with domain size k where at least $1 - \varepsilon$ fraction of constraints are satisfiable and those where $1/k^{\varepsilon/2-\varepsilon}$ fraction of constraints are satisfiable.*

Theorem 1.3 ([12], Corollary 14). *The Unique Games Conjecture implies that for every fixed $\varepsilon > 0$, for all $k > k(\varepsilon)$, it is NP-hard to distinguish between instances of unique games with domain size k where at least $1-\varepsilon$ fraction of constraints are satisfiable and those where $1 - \sqrt{2/\pi}\sqrt{\varepsilon \log k} + o(1)$ fraction of constraints are satisfiable.*

Thus, our bounds are near optimal if the UGC is true – even a slight improvement of the results $1/k^{\varepsilon/(2-\varepsilon)}$ or $1 - O(\sqrt{\varepsilon \log k})$ (beyond low order terms) will disprove the unique games conjecture!

Our algorithms are based on rounding an SDP relaxation for unique games. The goal is to assign a value in $[k]$ to every variable u. The SDP solution gives a collection of vectors $\{u_i\}$ for

[1]An updated version of [3] proposes a different strengthened form of the UGC, which is still plausible given our algorithms. They use a modified analysis to account for the asymmetry in ε and δ to obtain an $\Omega(\sqrt{\log \log n})$ hardness for sparsest cut based on this.

[2]We use a slightly stronger SDP than they used, but their integrality gap construction works for our SDP as well.

every variable u, one for every value $i \in [k]$. Given a constraint π_{uv} on u and v, the vectors u_i and $v_{\pi_{uv}(i)}$ are close. In contrast to the algorithms of Trevisan [17] and Gupta, Talwar [9], our rounding algorithms ignore the constraint graph entirely. We interpret the SDP solution as a probability distribution on assignments of values to variables and the goal of our rounding algorithm is to pick an assignment to variables by sampling from this distribution such that values of variables connected by constraints are strongly correlated. The rough idea is to pick a random vector and examine the projections of this vector on u_i, picking a value i for u for which u_i has a large projection. (In fact, this is exactly the algorithm of Khot [10]). We have to modify this basic idea to obtain our results since the u_i's could have different lengths and other complications arise. Instead of picking one random vector, we pick several Gaussian random vectors. Our first algorithm (suitable for large ε) picks a small set of candidate assignments for each variable and chooses randomly amongst them (independently for every variable). It is interesting to note that such a multiple assignment is often encountered in algorithms implicit in hardness reductions involving label cover. In contrast to previous results, this algorithm has a non-trivial guarantee even for very large ε. As ε approaches 1 (*i.e.* for instances where the optimal solution satisfies only a small fraction of the constraints), the performance guarantee approaches that of a random assignment. Our second algorithm (suitable for small ε) carefully picks a single assignment so that almost all constraints are satisfied. The performance guarantee of this algorithm generalizes that obtained by Goemans and Williamson [8] for $k = 2$. Note that a unique game of domain size $k = 2$ where $1 - \varepsilon$ fraction of constraints is satisfiable is equivalent to an instance of Max-Cut where the optimal solution cuts $1 - \varepsilon$ fraction of all edges. For such instances, the random hyperplane rounding algorithm of [8] gives a solution of value $1 - O(\sqrt{\varepsilon})$, and our guarantee can be viewed as a generalization of this to larger k.

The reader might wonder about the confluence of our bounds and the lower bounds obtained by Khot *et al.* [12]. In fact, both arise from the analysis of the same quantity: *Given two unit vectors with dot product $1 - \varepsilon$, conditioned on the probability that one has projection $\Theta(\sqrt{\log k})$ on a random Gaussian vector, what is the probability that the other has a large projection as well?* This question arises naturally in the analysis of our rounding algorithms. On the other hand, the bounds obtained by Khot *et al.* [12] depend on the noise stability of certain functions. Via the results of [15], this is bounded by the answer to the above question.

In Section 2, we describe the semidefinite relaxation for unique games. In Sections 3 and 4, we present and analyze our approximation algorithms. In Section 5, we apply our results to d-to-1 games. We defer some of the technical details of our analysis to the Appendix.

Recently, Chlamtac, Makarychev and Makarychev [4] have combined our approach with techniques of metric embeddings. Their approximation algorithm for unique games satisfies $1 - O(\varepsilon\sqrt{\log n \log k})$ fraction of all constraints. This generalizes the result of Agarwal, Charikar, Makarychev, and Makarychev [1] for the Min UnCut Problem (*i.e.* the case $k = 2$). Note that their approximation guarantee is not comparable with ours.

2 Semidefinite Relaxation

First we reduce a unique game to an integer program. For each vertex u we introduce k indicator variables $u_i \in \{0, 1\}$ ($i \in [k]$) for the events $x_u = i$. For every u, the intended solution has $u_i = 1$ for exactly one i. The constraint $\pi_{uv}(x_u) = x_v$ can be restated in the following form:

$$\text{for all } i \quad u_i = v_{\pi_{uv}(i)}.$$

The unique game instance is equivalent to the following integer quadratic program:

$$\text{minimize } \frac{1}{2} \sum_{(u,v)\in E} \left(\sum_{i=1}^{k} |u_i - v_{\pi_{uv}(i)}|^2 \right)$$

$$\begin{aligned} \text{subject to} \quad \forall u \in V \;\; \forall i \in [k] \qquad & u_i \in \{0,1\} \\ \forall u \in V \;\; \forall i,j \in [k], i \neq j \qquad & u_i \cdot u_j = 0 \\ \forall u \in V \qquad & \sum_{i=1}^{k} u_i^2 = 1 \end{aligned}$$

Note that the objective function measures the number of *unsatisfied* constraints. The contribution of $(u,v) \in E$ to the objective function is equal to 0 if the constraint π_{uv} is satisfied, and 1 otherwise. The last two equations say that exactly one u_i is equal to 1.

We now replace each integer variable u_i with a vector variable and get a semidefinite program (SDP):

$$\text{minimize } \frac{1}{2} \sum_{(u,v)\in E} \sum_{i=1}^{k} |u_i - v_{\pi_{uv}(i)}|^2$$

subject to

$$\forall u \in V \; \forall i,j \in [k], i \neq j \qquad \langle u_i, u_j \rangle = 0 \tag{2}$$

$$\forall u \in V \qquad \sum_{i=1}^{k} |u_i|^2 = 1 \tag{3}$$

$$\forall (u,v) \in E \; i,j \in [k] \qquad \langle u_i, v_j \rangle \geq 0 \tag{4}$$

$$\forall (u,v) \in E \; i \in [k] \qquad 0 \leq \langle u_i, v_{\pi_{uv}(i)} \rangle \leq |u_i|^2 \tag{5}$$

The last two constraints are triangle inequality constraints[3] for the squared Euclidean distance: inequality (4) is equivalent to $|u_i - 0|^2 + |v_j - 0|^2 \geq |u_i - v_j|^2$, and inequality (5, right side) is equivalent to $|u_i - v_{\pi_{uv}(i)}|^2 + |u_i - 0|^2 \geq |v_{\pi_{uv}(i)} - 0|^2$. A very important constraint is that for $i \neq j$ the vectors u_i and u_j are orthogonal. This SDP was studied by Khot [10], and by Trevisan [17].

Here is an intuitive interpretation of the vector solution: Think of the elements of the set $[k]$ as states of the vertices. If $u_i = 1$, the vertex is in the state i. In the vector case, each vertex is in a *mixed* state, and the probability that $x_u = i$ is equal to $|u_i|^2$. The inner product $\langle u_i, v_j \rangle$ can be thought of as the joint probability that $x_u = i$ and $x_v = j$. The directions of vectors determine whether two states are correlated or not: If the angle between u_i and v_j is small it is likely that both events "*u is in the state i*" and "*v is in the state j*" occur simultaneously. In some sense later we will treat the lengths and the directions of vectors separately.

3 Rounding Algorithm

We first describe a high level idea for the first algorithm. Pick a random Gaussian vector g (with standard normal independent components). For every vertex u add those vectors u_i whose inner

[3] We will use constraint 4 only in the second algorithm.

product with g are above some threshold τ to the set S_u; we choose the threshold τ in such a way that the set S_u contains only one element in expectation. Then pick a random state from S_u and assign it to the vertex u (if S_u is empty do not assign any states to u). What is the probability that the algorithm satisfies a constraint between vertices u and v? Loosely speaking, this probability is equal to

$$\mathbb{E}\left[\frac{|S_u \cap \pi_{uv}(S_v)|}{|S_u| \cdot |S_v|}\right] \approx \mathbb{E}\left[|S_u \cap \pi_{uv}(S_v)|\right].$$

Assume for a moment that the SDP solution is symmetric: the lengths of all vectors u_i are the same and the squared Euclidean distance between every u_i and $v_{\pi_{uv}(i)}$ is equal to 2ε. (In fact, these constraints can be added to the SDP in the special case of systems of linear equations of the form $x_i - x_j = c_{ij} \pmod p$.) Since we want the expected size of S_u to be 1, we pick threshold τ such that the probability that $\langle g, u_i \rangle \geq \tau$ equals $1/k$. The random variables $\langle g, \sqrt{k} \cdot u_i \rangle$ and $\langle g, \sqrt{k} \cdot v_{\pi_{uv}(i)} \rangle$ are standard normal random variables with covariance $1 - \varepsilon$ (note that we multiplied the inner products by a normalization factor of $\sqrt{k}$). For such random variables if the probability of the event $\langle g, \sqrt{k} \cdot u_i \rangle \geq t \equiv \sqrt{k}\tau$ equals $1/k$, then roughly speaking the probability of the event $\langle g, \sqrt{k} \cdot u_i \rangle \geq t \equiv \sqrt{k} \cdot \tau$ and $\langle g, \sqrt{k} \cdot v_{\pi_{uv}(i)} \rangle \geq t \equiv \sqrt{k} \cdot \tau$ equals $k^{-\varepsilon/2} \cdot 1/k$. Thus the expected size of the intersection of the sets S_u and $\pi_{uv}(S_v)$ is approximately $k^{-\varepsilon/2}$.

Unfortunately this no longer works if the lengths of vectors are different. The main problem is that if, say, u_1 is two times longer than u_2, then $\Pr(u_1 \in S_u)$ is much larger than $\Pr(u_2 \in S_u)$.

One of the possible solutions is to normalize all vectors first. In order to take into account original lengths of vectors we repeat the procedure of adding vectors to the sets S_u many times, but each vector u_i has a chance to be selected in the set S_u only in the first $s_{u,i}$ trials, where $s_{u,i}$ is some integer number proportional to the original squared Euclidean length of u_i.

We now formally present a rounding algorithm for the SDP described in the previous section. In Appendix D, we describe an alternate approach to rounding the SDP.

Theorem 3.1. *There is a polynomial time algorithm that finds an assignment of variables which satisfies*

$$\Omega\left(\min(1, \frac{1}{\sqrt{\varepsilon \log k}}) \cdot (1-\varepsilon)^2 \cdot \left(\frac{k}{\sqrt{\log k}}\right)^{-\varepsilon/(2-\varepsilon)}\right)$$

fraction of all constraints if the optimal solution satisfies $(1-\varepsilon)$ fraction of all constraints.

Rounding Algorithm 1
Input: *A solution of the SDP, with the objective value $\varepsilon \cdot |E|$.*
Output: *An assignment of variables x_u.*

1. Define $\tilde{u}_i = u_i/|u_i|$ if $u_i \neq 0$, 0 otherwise.

   Note that vectors $\tilde{u}_1, \dots, \tilde{u}_k$ are orthogonal unit vectors (except for those vectors that are equal to zero).

2. Pick random independent Gaussian vectors $g_1, \dots, g_k$ with independent components distributed as $\mathcal{N}(0, 1)$.

3. For each vertex u:
   (a) Set $s_{u_i} = \lceil |u_i|^2 \cdot k \rceil$.

(b) For each i project s_{u_i} vectors $g_1, \ldots, g_{s_{u_i}}$ to $\tilde{u}_i$:

$$\xi_{u_i,s} = \langle g_s, \tilde{u}_i \rangle, \ 1 \le s \le s_{u_i}.$$

Note that $\xi_{u_1,1}, \xi_{u_1,2}, \ldots, \xi_{u_1,s_{u_1}}, \ldots, \xi_{u_k,1}, \ldots, \xi_{u_k,s_{u_k}}$ are independent standard normal random variables. (Since u_i and u_j are orthogonal if $i \neq j$, their projections onto a random Gaussian vector are independent). The number of random variables corresponding to each u_i is proportional to $|u_i|^2$.

(c) Fix a threshold t s.t. $\Pr(\xi \ge t) = 1/k$, where $\xi \sim \mathcal{N}(0,1)$ (*i.e.* t is the $(1-1/k)$-quantile of the standard normal distribution; note that $t = \Theta\left(\sqrt{\log k}\right)$).

(d) Pick ξ_{u_i}'s that are larger than the threshold t:

$$S_u = \{(i,s) : \xi_{u_i,s} \ge t\}.$$

(e) Pick at random a pair (i,s) from S_u and assign $x_u = i$.

If the set S_u is empty do not assign any value to the vertex: this means that all the constraints containing the vertex are not satisfied.

We introduce some notation.

Definition 3.2. *Define the distance between two vertices u and v as:*

$$\varepsilon_{uv} = \frac{1}{2} \sum_{i=1}^{k} |u_i - v_{\pi_{uv}(i)}|^2$$

and let

$$\varepsilon_{uv}^i = \frac{1}{2} |\tilde{u}_i - \tilde{v}_{\pi_{uv}(i)}|^2.$$

If u_i and $v_{\pi_{uv}(i)}$ are nonzero vectors and α_i is the angle between them, then $\varepsilon_{uv}^i = 1 - \cos\alpha_i$. For consistency, if one of the vectors is equal to zero we set $\varepsilon_{uv}^i = 1$ and $\alpha_i = \pi/2$.

Lemma 3.3. *For every edge (u,v), state i in $[k]$ and $s \le \min(s_{u_i}, s_{v_{\pi_{uv}(i)}})$ the probability that the algorithm picks (i,s) for the vertex u and $(\pi_{uv}(i), s)$ for v at the step 3.e is*

$$\Omega\left(\min(1, \frac{1}{\sqrt{\varepsilon_{uv}^i \log k}}) \cdot \frac{1}{\sqrt{\log k}} \cdot \left(\frac{\sqrt{\log k}}{k}\right)^{2/(2-\varepsilon_{uv}^i)}\right). \tag{6}$$

Proof. First let us observe that $\xi_{u_i,s}$ and $\xi_{v_{\pi_{uv}(i)},s}$ are standard normal random variables with covariance $\cos\alpha_i = 1 - \varepsilon_{uv}^i$. As we will see later (Lemma B.1) the probability that $\xi_{u_i,s} \ge t$ and $\xi_{v_{\pi_{uv}(i)},s} \ge t$ is equal to (6).

Note that the expected number of elements in S_u is equal to $(s_{u_1} + \ldots + s_{u_k})/k$ which is at most 2. Moreover, as we prove in the Appendix (Lemma B.3), the conditional expected number of elements in S_u given the event $\xi_{u_i,s} \ge t$ and $\xi_{v_{\pi_{uv}(i)},s} \ge t$ is also a constant. Thus by the Markov inequality the following event happens with probability (6): The sets S_u and S_v contain the pairs (i,s) and $(\pi_{uv}(i), s)$ respectively and the sizes of these sets are bounded by a constant. The lemma follows. □

Definition 3.4. *For brevity, denote* $\left(\sqrt{\log k}/k\right)^{2/(2-x)}$ *by* $f_k(x)$.

Remark 3.1. *It is instructive to consider the case when the SDP solution is uniform in the following sense:*

1. *The lengths of all vectors* u_i *are the same and are equal to* $1/\sqrt{k}$.
2. *All* ε^i_{uv} *are equal to* ε.

In this case all s_{u_i} *are equal to 1. And thus the probability that a constraint is satisfied is* k *times the probability (6) which is equal, up to a logarithmic factor, to* $k^{-\varepsilon/(2-\varepsilon)}$. *Multiplying this probability by the number of edges we get that the expected number of satisfied constraints is* $k^{-\varepsilon/(2-\varepsilon)}|E|$.

In the general case however we need to do some extra work to average the probabilities among all states i and edges (u, v).

Recall that we interpret $|u_i|^2$ as the probability that the vertex u is in the state i. Suppose now that the constraint between u and v is satisfied, what is the conditional probability that u is in the state i and v is in the state $\pi_{uv}(i)$? Roughly speaking, it should be equal to $(|u_i|^2 + |v_{\pi_{uv}(i)}|^2)/2$. This motivates the following definition.

Definition 3.5. *Define a measure* μ_{uv} *on the set* $[k]$*:*

$$\mu_{uv}(T) = \sum_{i\in T} \frac{|u_i|^2 + |v_{\pi_{uv}(i)}|^2}{2}, \text{ where } T \subset [k].$$

Note that $\mu_{uv}([k]) = 1$. *This follows from constraint (3).*

The following lemma shows why this measure is useful.

Lemma 3.6. *For every edge* (u, v) *the following statements hold.*

1. *The average value of* ε^i_{uv} *w.r.t. the measure* μ_{uv} *is less than or equal to* ε_{uv}*:*
$$\sum_{i=1}^{k} \mu_{uv}(i)\varepsilon^i_{uv} \le \varepsilon_{uv}.$$
2. *For every* i,
$$\min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \ge (1-\varepsilon^i_{uv})^2 \mu_{uv}(i)k.$$

Proof. 1. Indeed,

$$\begin{aligned}\sum_{i=1}^{k} \mu_{uv}(i)\cdot\varepsilon^i_{uv} &= \sum_{i=1}^{k} \frac{|u_i|^2 + |v_{\pi_{uv}(i)}|^2 - (|u_i|^2 + |v_{\pi_{uv}(i)}|^2)\cdot\cos\alpha_i}{2}\\ &\le \sum_{i=1}^{k} \frac{|u_i|^2 + |v_{\pi_{uv}(i)}|^2 - 2\cdot|u_i|\cdot|v_{\pi_{uv}(i)}|\cdot\cos\alpha_i}{2}\\ &= \sum_{i=1}^{k} \frac{|u_i - v_{\pi_{uv}(i)}|^2}{2} = \varepsilon_{uv}\end{aligned}$$

Note that here we used the fact that $\langle u_i, v_{\pi_{uv}(i)}\rangle \geq 0$.

2. Without loss of generality assume that $|u_i| \leq |v_{\pi_{uv}(i)}|$, and hence $\min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) = s_{u_i}$. Due to the triangle inequality constraint (5) in the SDP $|v_{\pi_{uv}(i)}|\cos\alpha_i \leq |u_i|$. Thus

$$(1-\varepsilon_{uv}^i)^2\mu_{uv}(i) = \cos^2\alpha_i \cdot \frac{|u_i|^2 + |v_{\pi_{uv}(i)}|^2}{2} \leq |u_i|^2 \leq s_{u_i}/k.$$

□

Lemma 3.7. *For every edge (u,v) the probability that an assignment found by the algorithm satisfies the constraint $\pi_{uv}(x_u) = x_v$ is*

$$\Omega\left(\frac{k}{\sqrt{\log k}} \cdot \min(1, \frac{1}{\sqrt{\varepsilon_{uv}\log k}}) \cdot f_k(\varepsilon_{uv})\right). \tag{7}$$

Proof. Denote the desired probability by P_{uv}. It is equal to the sum of the probabilities obtained in Lemma 3.3 over all i in $[k]$ and $s \leq \min(s_{u_i}, s_{v_{\pi_{uv}(i)}})$. In other words,

$$P_{uv} = \Omega\left(\sum_{i=1}^{k} \min(s_{u_i}, s_{v_{\pi_{uv}(i)}})\frac{1}{\sqrt{\log k}}\min(1, \frac{1}{\sqrt{\varepsilon_{uv}^i \log k}}) f_k(\varepsilon_{uv}^i)\right).$$

Replacing $\min(s_{u_i}, s_{v_{\pi_{uv}(i)}})$ with $(1-\varepsilon_{uv}^i)^2\mu_{uv}(i)\cdot k$ we get

$$P_{uv} = \Omega\left(\frac{k}{\sqrt{\log k}}\sum_{i=1}^{k}\mu_{uv}(i)\min(1, \frac{1}{\sqrt{\varepsilon_{uv}^i\log k}})(1-\varepsilon_{uv}^i)^2 f_k(\varepsilon_{uv}^i)\right).$$

Consider the set $M = \left\{i \in [k] : \varepsilon_{uv}^i \leq 2\varepsilon_{uv}\right\}$. For i in M the term $\sqrt{\varepsilon_{uv}^i \log k}$ is bounded from above by $\sqrt{2\varepsilon_{uv}\log k}$. Thus

$$P_{uv} = \Omega\left(\frac{k}{\sqrt{\log k}}\min(1, \frac{1}{\sqrt{\varepsilon_{uv}\log k}})\sum_{i\in M}\mu_{uv}(i)(1-\varepsilon_{uv}^i)^2 f_k(\varepsilon_{uv}^i)\right).$$

The function $(1-x)^2 f_k(x)$ is convex on $[0,1]$ (see Lemma B.4). The average value of ε_{uv}^i among i in M (w.r.t. the measure μ_{uv}) is at most the average value of ε_{uv}^i among all i, which in turn is less than ε_{uv} according to Lemma 3.6. Finally, by the Markov inequality $\mu_{uv}(M) \geq 1/2$. Thus by Jensen's inequality

$$\begin{aligned} P_{uv} &= \Omega\left(\frac{k}{\sqrt{\log k}}\min(1, \frac{1}{\sqrt{\varepsilon_{uv}\log k}})\mu_{uv}(M)(1-\varepsilon_{uv})^2 \cdot f_k(\varepsilon_{uv})\right) \\ &= \Omega\left(\frac{k}{\sqrt{\log k}}\min(1, \frac{1}{\sqrt{\varepsilon_{uv}\log k}})(1-\varepsilon_{uv})^2 \cdot f_k(\varepsilon_{uv})\right). \end{aligned}$$

This finishes the proof. □

We are now in position to prove the main theorem.

Theorem 3.1. *There is a polynomial time algorithm that finds an assignment of variables which satisfies*

$$\Omega\left(\min(1, \frac{1}{\sqrt{\varepsilon\log k}}) \cdot (1-\varepsilon)^2 \cdot \left(\frac{k}{\sqrt{\log k}}\right)^{-\varepsilon/(2-\varepsilon)}\right)$$

fraction of all constraints if the optimal solution satisfies $(1-\varepsilon)$ fraction of all constraints.

Proof. Let us restrict our attention to a subset of edges $E' = \{(u, v) \in E : \varepsilon_{uv} \leq 2\varepsilon\}$. For (u, v) in E', since $\varepsilon_{uv} \log k \leq 2\varepsilon \log k$, we have

$$P_{uv} = \Omega\left(\frac{k}{\sqrt{\log k}} \min(1, \frac{1}{\sqrt{\varepsilon \log k}})(1 - \varepsilon_{uv})^2 \cdot f_k(\varepsilon_{uv})\right).$$

Summing this probability over all edges (u, v) in E' and using convexity of the function $(1-x)^2 f_k(x)$ we get the statement of the theorem. □

4 Almost Satisfiable Instances

Suppose that ε is $O(1/\log k)$. In the previous section we saw that in this case the algorithm finds an assignment of variables satisfying a constant fraction of constraints. But can we do better? In this section we show how to find an assignment satisfying $1 - O(\sqrt{\varepsilon \log k})$ fraction of constraints.

The main issue we need to take care of is to guarantee that the algorithm always picks only one element in the set S_u (otherwise we loose a constant factor). This can be done by selecting the largest in absolute value $\xi_{u_i,s}$ (at step 3.d). We will also change the way we set s_{u_i}.

Denote by $[x]_r$ the function that rounds x up or down depending on whether the fractional part of x is greater or less than r. Note that if r is a random variable uniformly distributed in the interval $[0, 1]$, then the expected value of $[x]_r$ is equal to x.

Rounding Algorithm 2
Input: *A solution of the SDP, with the objective value $\varepsilon \cdot |E|$.*
Output: *An assignment of variables x_u.*

1. Pick a number r in the interval $[0, 1]$ uniformly at random.

2. Pick random independent Gaussian vectors $g_1, \dots, g_{2k}$ with independent components distributed as $\mathcal{N}(0, 1)$.

3. For each vertex u:

   (a) Set $s_{u_i} = [2k \cdot |u_i|^2]_r$.

   (b) For each i project s_{u_i} vectors $g_1, \dots, g_{s_{u_i}}$ to $\tilde{u}_i$:

   $$\xi_{u_i,s} = \langle g_s, \tilde{u}_i \rangle, \; 1 \leq s \leq s_{u_i}.$$

   (c) Select $\xi_{u_i,s}$ with the largest absolute value, where $i \in [k]$ and $s \leq s_{u_i}$. Assign $x_u = i$.

We first elaborate on the difference between the choice of s_{u_i} in the algorithm above and that in Algorithm 1 presented earlier. Consider a constraint $\pi_{uv}(x_u) = x_v$. Projection $\xi_{u_i,s}$ generated by u_i and $\xi_{v_{\pi_{uv}(i)},s}$ generated by $v_{\pi_{uv}(i)}$ are considered to be *matched*. On the other hand, a projection $\xi_{u_i,s}$ such that the corresponding $\xi_{v_{\pi_{uv}(i)},s}$ does not exist (or vice versa) is considered to be *unmatched*. Unmatched projections arise when $s_{u_i} \neq s_{v_{\pi_{uv}(i)}}$ and the fraction of such projections limits the probability of satisfying the constraint. Recall that in Algorithm 1, we set $s_{u_i} = \lceil |u_i|^2 \cdot k \rceil$. Even if u_i and $v_{\pi_{uv}(i)}$ are infinitesimally close, it may turn out that s_{u_i} and $s_{v_{\pi_{uv}(i)}}$ differ by 1, yielding an unmatched projection. As a result, some constraints that are almost satisfied by the SDP solution (*i.e.* ε_{uv} is close to 0) could be satisfied with low probability (by the first rounding algorithm). In

Algorithm 2, we set $s_{u_i} = [2k \cdot |u_i|^2]_r$. This serves two purposes: Firstly, $\mathbb{E}_r\left[|s_{u_i} - s_{v_{\pi_{uv}(i)}}|\right]$ can be bounded by $2k \cdot |u_i - v_{\pi_{uv}(i)}|^2$, giving a small number of unmatched projections in expectation. Secondly, the number of matched projections is always at least $k/2$. These two properties are established in Lemma 4.3 and ensure that the expected fraction of unmatched projections is small.

Our analysis of Rounding Algorithm 2 is based on the following theorem.

Theorem 4.1. *Let $\xi_1, \cdots, \xi_m$ and $\eta_1, \cdots, \eta_m$ be two sequences of standard normal random variables. Suppose that the random variables in each of the sequences are independent, the covariance of every ξ_i and η_j is nonnegative, and the average covariance of ξ_i and η_i is at least $1-\varepsilon$:*

$$\frac{\mathrm{cov}(\xi_1, \eta_1) + \cdots + \mathrm{cov}(\xi_m, \eta_m)}{m} \geq 1 - \varepsilon.$$

Then the probability that the largest r.v. in absolute value in the first sequence has the same index as the largest r.v. in absolute value in the second sequence is $1 - O(\sqrt{\varepsilon \log m})$.

We informally sketch the proof. See Appendix C for the complete proof. It is instructive to consider the case when $\mathrm{cov}(\xi_i, \eta_i) = 1 - \varepsilon$ for all i. Assume that the first variable ξ_1 is the largest in absolute value among $\xi_1, \ldots, \xi_m$ and its absolute value is a positive number t. Note that the *typical* value of t is approximately $\sqrt{2 \log m - \log\log m}$ (*i.e.* t is the $(1-1/m)$-quantile of $\mathcal{N}(0,1)$). We want to show that η_1 is the largest in absolute value among $\eta_1, \ldots, \eta_m$ with probability $1 - O(\sqrt{\varepsilon \log m})$, or in other words the probability that any (fixed) η_i is larger than η_1 is $O(\sqrt{\varepsilon \log m}/m)$. Let us compute this probability for η_2.

Since $\mathrm{cov}(\eta_1, \xi_1) = 1-\varepsilon$ and $\mathrm{cov}(\xi_2, \eta_2) = 1-\varepsilon$, the random variable η_1 is equal to $(1-\varepsilon)\xi_1 + \zeta_1$; and η_2 is equal to $(1-\varepsilon)\xi_2 + \zeta_2$, where ζ_1 and ζ_2 are normal random variables with variance, roughly speaking, 2ε. We need to estimate the probability of the event

$$\{\eta_2 \geq \eta_1\} = \{(1-\varepsilon)\xi_2 + \zeta_2 \geq (1-\varepsilon)\xi_1 + \zeta_1\} = \{(1-\varepsilon)\xi_2 + \zeta_2 - \zeta_1 \geq (1-\varepsilon)t\}$$

conditional on $\xi_1 = t$ and $\xi_2 \leq t$. For typical t this probability is almost equal to the probability of the event:

$$\{\xi_2 + \zeta \geq t \text{ and } \xi_2 \leq t\} = \{t - \zeta \leq \xi_2 \leq t\} \tag{8}$$

where $\zeta = \zeta_2 - \zeta_1$.

Since the variance of the random variable ζ is $O(\varepsilon)$, we can think that $\zeta \approx O(\sqrt{\varepsilon})$. The density of ξ_2 on the interval $[t-\zeta, t]$ is approximately $1/\sqrt{2\pi} e^{-t^2/2} \approx O(\sqrt{\log m}/m)$ (for typical t). Thus probability (8) is equal to $O(\sqrt{\varepsilon \log m}/m)$. This finishes our informal "proof".

Now we are ready to prove the main lemma.

Lemma 4.2. *The probability that the algorithm finds an assignment of variables satisfying the constraint $\pi_{uv}(x_u) = x_v$ is $1 - O(\sqrt{\varepsilon_{uv} \log k})$.*

Proof. If $\varepsilon_{uv} \geq 1/8$ the statement of the lemma follows trivially. So we assume that $\varepsilon_{uv} \leq 1/8$.

Let

$$M = \left\{(i, s) : i \in [k] \text{ and } s \leq \min(s_{u_i}, s_{v_{\pi_{uv}(i)}})\right\};$$

$$M_c = \left\{(i, s) : i \in [k] \text{ and } \min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) < s \leq \max(s_{u_i}, s_{v_{\pi_{uv}(i)}})\right\}.$$

The set M contains those pairs (i, s) for which both $\xi_{u_i,s}$ and $\xi_{v_{\pi_{uv}(i)},s}$ are defined (*i.e.* the *matched* projections); the set M_c contains those pairs for which only one of the variables $\xi_{u_i,s}$ and $\xi_{v_{\pi_{uv}(i)},s}$ is defined (*i.e.* the *unmatched* projections). We will need the following lemmas.

Lemma 4.3. *1. The expected size of M_c is at most $4\varepsilon_{uv}k$:*

$$\mathbb{E}\left[|M_c|\right] \leq 4\varepsilon_{uv}k.$$

2. The set M always contains at least $k/2$ elements: $|M| \geq k/2$.

Proof. 1. First we find the expected value of $|s_{u_i} - s_{v_{\pi_{uv}(i)}}|$ for a fixed i. This value is equal to

$$\mathbb{E}_r\left[\left|[2k \cdot |u_i|^2]_r - [2k \cdot |v_{\pi_{uv}(i)}|^2]_r\right|\right] = 2k \cdot \left||u_i|^2 - |v_{\pi_{uv}(i)}|^2\right|.$$

Now by the triangle inequality constraint (5),

$$2k \cdot \left||u_i|^2 - |v_{\pi_{uv}(i)}|^2\right| \leq 2k \cdot |u_i - v_{\pi_{uv}(i)}|^2.$$

Summing over all i in $[k]$ we finish the proof.

2. Observe that

$$\min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \geq 2k \cdot \min(|u_i|^2, |v_{\pi_{uv}(i)}|^2) - 1$$

and

$$\min(|u_i|^2, |v_{\pi_{uv}(i)}|^2) \geq |u_i|^2 - ||u_i|^2 - |v_{\pi_{uv}(i)}|^2| \geq |u_i|^2 - |u_i - v_{\pi_{uv}(i)}|^2.$$

Summing over all i we get

$$\begin{aligned} |M| &= \sum_{i\in[k]} \min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \geq \sum_{i\in[k]} \left(2k \cdot |u_i|^2 - 2k \cdot |u_i - v_{\pi_{uv}(i)}|^2 - 1\right) \\ &\geq 2k - 4k\varepsilon_{uv} - k \geq k/2. \end{aligned}$$

□

Lemma 4.4. *The following inequality holds:*

$$\mathbb{E}\left[\frac{1}{|M|}\sum_{(i,s)\in M} \varepsilon_{uv}^i\right] \leq 4\varepsilon_{uv}.$$

Proof. Recall that M always contains at least $k/2$ elements. The expected value of $\min(s_{u_i}, s_{v_{\pi_{uv}(i)}})$ is equal to $2k \cdot \min(|u_i|^2, |v_{\pi_{uv}(i)}|^2)$ and is less than or equal to $2k \cdot \mu_{uv}(i)$. Thus we have

$$\begin{aligned} \mathbb{E}_r\left[\frac{1}{|M|}\sum_{(i,s)\in M} \varepsilon_{uv}^i\right] &= \mathbb{E}_r\left[\frac{1}{|M|}\sum_{i=1}^{k} \min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \cdot \varepsilon_{uv}^i\right] \\ &\leq \frac{2}{k}\sum_{i=1}^{k} 2k \cdot \mu_{uv}(i) \cdot \varepsilon_{uv}^i \leq 4\sum_{i=1}^{k} \mu_{uv}(i) \cdot \varepsilon_{uv}^i \leq 4\varepsilon_{uv}. \end{aligned}$$

□

Proof of Lemma 4.2
Applying Theorem 4.1 to the sequences $\xi_{u_i,s}$ $((i,s) \in M)$ and $\xi_{v_{\pi_{uv}(i)},s}$ $((i,s) \in M)$ we get that for given r the probability that the largest in absolute value random variables in the first sequence $\xi_{u_i,s}$ and the second sequence $\xi_{v_{\pi_{uv}(i)},s}$ have the same index (i,s) is

$$1 - O\left(\sqrt{\log|M| \cdot \frac{1}{|M|} \sum_{(i,s)\in M} \varepsilon_{uv}^i}\right).$$

Now by Lemma 4.4, and by the concavity of the function $\sqrt{x}$, we have

$$\mathbb{E}_r\left[1 - O\left(\sqrt{\frac{\log|M|}{|M|} \sum_{(i,s)\in M} \varepsilon_{uv}^i}\right)\right] \geq 1 - O\left(\sqrt{\varepsilon_{uv}\log k}\right).$$

The probability that there is a larger $\xi_{u_i,s}$ or $\xi_{v_{\pi_{uv}(i)},s}$ in M_c is at most

$$\mathbb{E}_r\left[\frac{|M_c|}{|M|}\right] \leq \frac{4\varepsilon_{uv}k}{k/2} = 8\varepsilon_{uv}.$$

Using the union bound we get that the probability of satisfying the constraint $\pi_{uv}(x_u) = x_v$ is equal to

$$1 - O(\sqrt{\varepsilon_{uv}\log k}) - 8\varepsilon_{uv} = 1 - O(\sqrt{\varepsilon_{uv}\log k}).$$

□

Theorem 4.5. *There is a polynomial time algorithm that finds an assignment of variables which satisfies $1 - O(\sqrt{\varepsilon \log k})$ fraction of all constraints if the optimal solution satisfies $(1-\varepsilon)$ fraction of all constraints.*

Proof. Summing the probabilities obtained in Lemma 4.2 over all edges (u,v) and using the concavity of the function $\sqrt{x}$ we get that the expected number of satisfied constraints is $1 - O(\sqrt{\varepsilon \log k})|E|$. □

5 d to 1 Games

In this section we extend our results to d-to-1 games.

Definition 5.1. *We say that $\Pi \subset [k] \times [k]$ is a d-to-1 predicate, if for every i there are at most d different values j such that $(i,j) \in \Pi$, and for every j there is at most one i such that $(i,j) \in \Pi$.*

Definition 5.2 (d to 1 Games). *We are given a directed constraint graph $G = (V,E)$, a set of variables x_u (for all vertices u) and d-to-1 predicates $\Pi_{uv} \subset [k] \times [k]$ for all edges (u,v). Our goal is to assign a value from the set $[k]$ to each variable x_u, so that the maximum number of the constraints $(x_u, x_v) \in \Pi_{uv}$ is satisfied.*

Note that even if all constraints of a d-to-1 game are satisfiable it is hard to find an assignment of variables satisfying all constraints. We will show how to satisfy

$$\Omega\left(\frac{1}{\sqrt{\log k}}\cdot(1-\varepsilon)^4\cdot\left(\frac{k}{\sqrt{\log k}}\right)^{-\frac{\sqrt{d}-1+\varepsilon}{\sqrt{d}+1-\varepsilon}}\right)$$

fraction of all constraints (the multiplicative constant in the Ω notation depends on d). Notice that this value can be obtained by replacing ε in formula (1) with $\varepsilon' = 1-(1-\varepsilon)/\sqrt{d}$ (and changing $(1-\varepsilon)^2$ to $(1-\varepsilon)^4$).

Even though we do not require that for a constraint Π_{uv} each i in $[k]$ belongs to some pair $(i,j)\in\Pi_{uv}$, let us assume that for each i there exists j s.t. $(i,j)\in\Pi_{uv}$; and for each j there exists i s.t. $(i,j)\in\Pi_{uv}$. As we see later this assumption is not important.

In order to write a relaxation for d-to-1 games introduce the following notation:

$$w_{uv}^i = \sum_{j:(i,j)\in\Pi_{uv}} v_j.$$

The SDP is as follows:

$$\text{minimize } \frac{1}{2}\sum_{(u,v)\in E}\left(\sum_{i=1}^{k} |\ u_i - w_{uv}^i\ |^2\right)$$

subject to

$$\forall u\in V\ \forall i,j\in[k], i\neq j \qquad \langle u_i,u_j\rangle = 0 \tag{9}$$

$$\forall u\in V \qquad \sum_{i=1}^{k}|u_i|^2 = 1 \tag{10}$$

$$\forall (u,v)\in V\ i,j\in[k] \qquad \langle u_i,v_j\rangle \geq 0 \tag{11}$$

$$\forall (u,v)\in V\ i\in[k]\ 0\leq\langle u_i,w_{uv}^i\rangle\leq\min(|u_i|^2,|w_{uv}^i|^2) \tag{12}$$

An important observation is that $|w_{uv}^1|^2+\ldots+|w_{uv}^k|^2=1$, here we use the fact that for a fixed edge (u,v) each v_j is a summand in one and only one w_{uv}^i.

We use Algorithm 1 for rounding a vector solution. For analysis we will need to change some notation:

$$\tilde{w}_{uv}^i = \begin{cases} w_{uv}^i/|w_{uv}^i|, & \text{if } w_{uv}^i\neq 0;\\ 0, & \text{otherwise}\end{cases}$$

$$\varepsilon_{uv}^i = \frac{|\tilde{u}_i-\tilde{w}_{uv}^i|^2}{2}$$

$${\varepsilon_{uv}^i}' = 1-\frac{1-\varepsilon_{uv}^i}{\sqrt{d}}$$

$$\mu_{uv}(i) = \frac{|u_i|^2+|w_{uv}^i|^2}{2}$$

The following lemma explains why we get the new dependency on ε.

Lemma 5.3. *For every edge (u,v) and state i there exists j' s.t. $(i,j') \in \Pi_{uv}$ and $|\tilde{u}_i - \tilde{v}_{j'}|^2/2 \leq {\varepsilon^i_{uv}}'$.*

Proof. Let u'_i be the projection of the vector $\tilde{u}_i$ to the linear span of the vectors v_j (where $(i,j) \in \Pi_{uv}$). Let α_i be the angle between $\tilde{u}_i$ and w^i_{uv}; and let β_i be the angle between $\tilde{u}_i$ and u'_i. Clearly, $|u'_i| = \cos\beta_i \geq \cos\alpha_i = 1 - \varepsilon^i_{uv}$. Since all $\tilde{v}_j$ ($(i,j) \in \Pi_{uv}$) are orthogonal unit vectors, there exists $\tilde{v}_{j'}$ s.t. $\langle \tilde{v}_{j'}, u'_i\rangle \geq |u'_i|/\sqrt{d}$. Hence, $\langle \tilde{v}_{j'}, \tilde{u}_i\rangle = \langle \tilde{v}_{j'}, u'_i\rangle \geq (1-\varepsilon^i_{uv})/\sqrt{d}$. □

For every edge (u,v) and state i, find j' as in the previous lemma and define a function[4] $\pi_{uv}(i) = j'$. Then replace every constraint $(x_u, x_v) \in \Pi_{uv}$ with a stronger constraint $\pi_{uv}(x_u) = x_v$. Now we can apply the original analysis of Algorithm 1 to the new problem. In the proof we need to substitute ${\varepsilon^i_{uv}}'$ for ε^i_{uv}, $1-(1-\varepsilon_{uv})/\sqrt{d}$ for ε_{uv}, and $1-(1-\varepsilon)/\sqrt{d}$ for ε. The only missing step is the following lemma.

Lemma 3.6′. *For every edge (u,v) the following statements hold.*

1. *The average value of ε^i_{uv} w.r.t. the measure μ_{uv} is less than or equal to ε_{uv}.*
2. *The average value of ${\varepsilon^i_{uv}}'$ w.r.t. the measure μ_{uv} is less than or equal to $1 - \frac{1-\varepsilon_{uv}}{\sqrt{d}}$.*
3. $\min(s_{u_i}, s_{v_{\pi_{uv}(i)}}) \geq d(1-{\varepsilon^i_{uv}}')^4 \mu_{uv}(i) k$.

Proof. Let α_i be the angle between u_i and w^i_{uv} and let α'_i be the angle between u_i and $v_{\pi_{uv}(i)}$.
1. Indeed,

$$\begin{aligned}\sum_{i=1}^k \mu_{uv}(i)\cdot\varepsilon^i_{uv} &= \sum_{i=1}^k \frac{|u_i|^2 + |w^i_{uv}|^2 - (|u_i|^2 + |w^i_{uv}|^2)\cdot\cos\alpha_i}{2} \\ &\leq \sum_{i=1}^k \frac{|u_i|^2 + |w^i_{uv}|^2 - 2\cdot|u_i|\cdot|w^i_{uv}|\cdot\cos\alpha_i}{2} \\ &= \sum_{i=1}^k \frac{|u_i - w^i_{uv}|^2}{2} = \varepsilon_{uv}.\end{aligned}$$

2. This follows from part 1 and the definition of ${\varepsilon^i_{uv}}'$.
3. Due to the triangle inequality constraint, $|w^i_{uv}|\cos\alpha_i \leq |u_i|$. Thus

$$(1-\varepsilon^i_{uv})^2\mu_{uv}(i) = \cos^2\alpha_i \cdot \frac{|u_i|^2 + |w^i_{uv}|^2}{2} \leq |u_i|^2.$$

Similarly $|v_{\pi_{uv}(i)}|\cos\alpha'_i \leq |u_i|$ and

$$(1-{\varepsilon^i_{uv}}')^2|u_i|^2 \leq \cos^2\alpha'_i\cdot|u_i|^2 \leq |v_{\pi_{uv}(i)}|^2.$$

Combining these two inequalities and noting that $(1-{\varepsilon^i_{uv}}') = (1-\varepsilon^i_{uv})/\sqrt{d}$, we get

$$d(1-{\varepsilon^i_{uv}}')^4\mu_{uv}(i) \leq (1-{\varepsilon^i_{uv}}')^2|u_i|^2 \leq |v_{\pi_{uv}(i)}|^2.$$

The lemma follows. □

[4] The function π_{uv} is not necessarily a permutation.

We now address the issue that for some edges (u, v) and states j there may not necessarily exist i s.t. $(i, j) \in \Pi_{uv}$. We call such j a state of degree 0. The key observation is that in our algorithms we may enforce additional constraints like $x_u = i$ or $x_u \neq i$ by setting $u_i = 1$ or $u_i = 0$ respectfully. Thus we can add extra states and enforce that the vertices are not in these states. Then we add pairs (i, j) where i is a new state, and j is a state of degree 0 (or vice-versa). Alternatively we can rewrite the objective function by adding an extra term:

$$\text{minimize } \frac{1}{2} \sum_{(u,v)\in E} \left(\sum_{i=1}^{k} \left| \ u_i - w_{uv}^i \ \right|^2 + |w_{uv}^0|^2 \right),$$

where w_{uv}^0 is the sum of v_j over j of degree 0.

Acknowledgements

We would like to thank Sanjeev Arora, Uri Feige, Johan Håstad, Muli Safra and Luca Trevisan for valuable discussions and Eden Chlamtac for his comments. The second and third authors thank Microsoft Research for their hospitality.

References

[1] A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. *$O(\sqrt{\log n})$ approximation algorithms for Min UnCut, Min 2CNF Deletion, and directed cut problems.* In Proceedings of the 37th Annual ACM Symposium on Theory of Computing, pp. 573–581, 2005.

[2] G. Andersson, L. Engebretsen, and J. Håstad. *A new way to use semidefinite programming with applications to linear equations mod p.* Journal of Algorithms, Vol. 39, pp. 162–204, 2001.

[3] S. Chawla, R. Krauthgamer, R. Kumar,Y. Rabani, and D. Sivakumar. *On the hardness of approximating multicut and sparsest-cut.* In Proceedings of the 20th IEEE Conference on Computational Complexity, pp. 144–153, 2005.

[4] E. Chlamtac, K. Makarychev and Y. Makarychev. *How to play any Unique Game.* manuscript, February 2006.

[5] I. Dinur, E. Mossel, and O. Regev. *Conditional hardness for approximate coloring.* ECCC Technical Report TR05-039, 2005.

[6] U. Feige and L. Lovász. *Two-prover one round proof systems: Their power and their problems.* In Proceedings of the 24th ACM Symposium on Theory of Computing, pp. 733–741, 1992.

[7] U. Feige and D. Reichman. *On systems of linear equations with two variables per equation.* In Proceedings of the 7th International Workshop on Approximation Algorithms for Combinatorial Optimization, vol. 3122 of Lecture Notes in Computer Science, pp. 117–127, 2004.

[8] M. Goemans , 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] A. Gupta and K. Talwar. *Approximating Unique Games.* In Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms, pp. 99–106, 2006.

[10] S. Khot. *On the power of unique 2-prover 1-round games.* In Proceedings of the 34th ACM Symposium on Theory of Computing, pp. 767–775, 2002.

[11] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. *Optimal inapproximability results for MAX-CUT and other two-variable CSPs?* In Proceedings of the 45th IEEE Symposium on Foundations of Computer Science, pp. 146–154, 2004.

[12] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?* ECCC Report TR05-101, 2005.

[13] S. Khot and O. Regev. *Vertex cover might be hard to approximate to within* $2 - \varepsilon$*.* In Proceedings of the 18th IEEE Conference on Computational Complexity, pp. 379–386, 2003.

[14] S. Khot and N. Vishnoi. *The unique games conjecture, integrality gap for cut problems and the embeddability of negative type metrics into* ℓ_1*.* In Proceedings of the 46th IEEE Symposium on Foundations of Computer Science, pp. 53–62, 2005.

[15] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. *Noise stability of functions with low influences: invariance and optimality.* In Proceedings of the 46th IEEE Symposium on Foundations of Computer Science, pp. 21–40, 2005.

[16] Z. Šidák. *Rectangular Confidence Regions for the Means of Multivariate Normal Distributions.* Journal of the American Statistical Association, vol. 62, no. 318, pp. 626–633, Jun. 1967.

[17] L. Trevisan. *Approximation Algorithms for Unique Games.* In Proceedings of the 46th IEEE Symposium on Foundations of Computer Science, pp. 197–205, 2005.

A Properties of Normal Distribution

For completeness we present some standard results used in the paper.

Denote the probability that a standard normal random variable is bigger than $t \in \mathbb{R}$ by $\tilde{\Phi}(t)$, in other words

$$\tilde{\Phi}(t) \equiv 1 - \Phi_{0,1}(t) = \Phi_{0,1}(-t),$$

where $\Phi_{0,1}$ is the normal distribution function.

Lemma A.1. *1. For every $t > 0$,*

$$\frac{t}{\sqrt{2\pi}(t^2+1)} e^{-\frac{t^2}{2}} < \tilde{\Phi}(t) < \frac{1}{\sqrt{2\pi}t} e^{-\frac{t^2}{2}}$$

2. *There exist positive constants c_1, C_1, c_2, C_2 such that for all $0 < p < 1/3$, $t \geq 0$ and $\rho \geq 1$ the following inequalities hold:*

$$\frac{c_1}{\sqrt{2\pi}(t+1)}e^{-\frac{t^2}{2}} \leq \tilde{\Phi}(t) \leq \frac{C_1}{\sqrt{2\pi}(t+1)}e^{-\frac{t^2}{2}};$$

$$c_2\sqrt{\log(1/p)} \leq \tilde{\Phi}^{-1}(p) \leq C_2\sqrt{\log(1/p)};$$

3. *There exists a positive constant C_3, s.t. for every $0 < \delta \leq 2$ and $t \geq 1/\delta$ the following inequality holds:*

$$\tilde{\Phi}(\delta t + \frac{1}{\delta t}) \geq C_3(t \cdot \tilde{\Phi}(t))^{\delta^2} \cdot t^{-1}.$$

Proof. 1. First notice that

$$\begin{aligned}
\tilde{\Phi}(t) &= \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\left[\left.\frac{-e^{-\frac{x^2}{2}}}{x}\right|_t^\infty - \int_t^\infty \frac{e^{-\frac{x^2}{2}}}{x^2}dx\right] \\
&= \frac{1}{\sqrt{2\pi}t}e^{-\frac{t^2}{2}} - \frac{1}{\sqrt{2\pi}}\int_t^\infty \frac{e^{-\frac{x^2}{2}}}{x^2}dx.
\end{aligned}$$

Thus

$$\tilde{\Phi}(t) < \frac{1}{\sqrt{2\pi}t}e^{-\frac{t^2}{2}}.$$

On the other hand

$$\frac{1}{\sqrt{2\pi}}\int_t^\infty \frac{e^{-\frac{x^2}{2}}}{x^2}dx < \frac{1}{\sqrt{2\pi}t^2}\int_t^\infty e^{-\frac{x^2}{2}}dx = \frac{\tilde{\Phi}(t)}{t^2}.$$

Hence

$$\tilde{\Phi}(t) > \frac{1}{\sqrt{2\pi}t}e^{-\frac{t^2}{2}} - \frac{\tilde{\Phi}(t)}{t^2},$$

and

$$\tilde{\Phi}(t) > \frac{t}{\sqrt{2\pi}(t^2+1)}e^{-\frac{t^2}{2}}.$$

2. This trivially follows from (1).

3. Using (2) we get

$$\begin{aligned}
\tilde{\Phi}(\delta t + \frac{1}{\delta t}) &\geq C \cdot \left(1 + \delta t + \frac{1}{\delta t}\right)^{-1} \cdot e^{-\frac{(\delta t + \frac{1}{\delta t})^2}{2}} \geq C' \cdot (\delta t + 1)^{-1} \cdot e^{-\frac{\delta^2 \cdot t^2}{2}} \\
&\geq C'' \left(\frac{e^{-\frac{t^2}{2}}}{t+1}\right)^{\delta^2} \cdot t^{\delta^2} \cdot t^{-1} \geq C''' \cdot (t \cdot \tilde{\Phi}(t))^{\delta^2} \cdot t^{-1}
\end{aligned}$$

□

We will use the following result of Z. Šidák [16]:

Theorem A.2 (Šidák). *Let $\xi_1, \ldots, \xi_k$ be normal random variables with mean zero, then for any positive $t_1, \ldots, t_k$,*

$$\Pr\left(|\xi_1| \le t_1, |\xi_2| \le t_2, \ldots, |\xi_k| \le t_k\right) \ge \Pr\left(|\xi_1| \le t_1\right) \Pr\left(|\xi_2| \le t_2, \ldots, |\xi_k| \le t_k\right).$$

Note that these random variable do not have to be independent.

Corollary A.3. *Let $\xi_1, \ldots, \xi_k$ be normal random variables with mean zero, then for any positive $t_1, \ldots, t_k$,*

$$\Pr\left(\xi_1 \ge t_1 \;\mid\; |\xi_2| \le t_2, \ldots, |\xi_k| \le t_k\right) \le \Pr\left(\xi_1 \ge t_1\right).$$

Proof. By Theorem A.2,

$$\Pr\left(|\xi_1| \le t_1 \;\mid\; |\xi_2| \le t_2, \ldots, |\xi_k| \le t_k\right) \ge \Pr\left(|\xi_1| \le t_1\right).$$

Thus

$$\begin{aligned} \Pr\left(\xi_1 \ge t_1 \;\mid\; |\xi_2| \le t_2, \ldots, |\xi_k| \le t_k\right) &= \frac{1}{2} - \frac{1}{2}\Pr\left(|\xi_1| \le t_1 \;\mid\; |\xi_2| \le t_2, \ldots, |\xi_k| \le t_k\right) \\ &\le \frac{1}{2} - \frac{1}{2}\Pr\left(|\xi_1| \le t_1\right) = \Pr\left(\xi_1 \ge t_1\right) \end{aligned}$$

□

B Analysis of Algorithm 1

In this section we will prove some technical lemmas we used in the analysis of the first algorithm.

Lemma B.1. *Let ξ and η be correlated standard normal random variables, $0 < \varepsilon < 1$, $t \ge 1$. If $\mathrm{cov}(\xi, \eta) \ge 1 - \varepsilon$, then*

$$\Pr\left(\xi \ge t \text{ and } \eta \ge t\right) \ge C \cdot \min(1, (\sqrt{\varepsilon} t)^{-1}) \cdot t^{-1} \cdot (t \cdot \tilde{\Phi}(t))^{\frac{2}{2-\varepsilon}}. \tag{13}$$

for some positive constant C.

Proof. Let us represent ξ and η as follows:

$$\xi = \sigma X + \sqrt{1 - \sigma^2} \cdot Y; \quad \eta = \sigma X - \sqrt{1 - \sigma^2} \cdot Y,$$

where

$$\sigma^2 = \mathrm{Var}\left[\frac{\xi + \eta}{2}\right]; \quad X = \frac{\xi + \eta}{2\sigma}; \quad Y = \frac{\xi - \eta}{2\sqrt{1 - \sigma^2}}.$$

Note that X and Y are independent standard normal random variables; and

$$\sigma^2 = \mathrm{Var}\left[\frac{\xi + \eta}{2}\right] = \frac{1}{4}\left[2 + 2\,\mathrm{cov}(\xi, \eta)\,\right] \ge 1 - \frac{\varepsilon}{2}. \tag{14}$$

Notice that $1/2 \le \sigma^2 \le 1$. We now estimate the probability (13) as follows

$$\begin{aligned}\Pr\big(\xi \ge t \text{ and } \eta \ge t\big) &= \Pr\left(\sigma X \ge t + \sqrt{1-\sigma^2}\cdot |Y|\right)\\ &\ge \Pr\left(X \ge \frac{t}{\sigma} + \frac{\sigma}{t}\right)\cdot \Pr\left(|Y| \le \frac{\sigma^2}{\sqrt{1-\sigma^2}\cdot t}\right)\end{aligned}$$

By Lemma A.1 (3) we get

$$\begin{aligned}\Pr\big(\xi \ge t \text{ and } \eta \ge t\big) &\ge C\cdot\left(t^{-1}\cdot (t\tilde{\Phi}(t))^{1/\sigma^2}\right)\cdot \min\left(1, \frac{\sigma^2}{\sqrt{1-\sigma^2}\cdot t}\right)\\ &\ge C'\cdot \min((\sqrt{\varepsilon}\cdot t)^{-1}, 1)\cdot t^{-1}\cdot (t\cdot\tilde{\Phi}(t))^{\frac{2}{2-\varepsilon}}.\end{aligned}$$

□

Corollary B.2. *Let ξ and η be standard normal random variables with covariance greater than or equal to $1-\varepsilon$; let $\tilde{\Phi}(t) = 1/k$. Then*

$$\Pr\left(\xi \ge t \text{ and } \eta \ge t\right) \ge \Omega\left(\min\left(1, \frac{1}{\sqrt{\varepsilon \log k}}\right)\cdot \frac{1}{\sqrt{\log k}}\cdot \left(\frac{k}{\sqrt{\log k}}\right)^{-\frac{2}{2-\varepsilon}}\right).$$

Lemma B.3. *Let ξ, η, ε, k and t be as in Corollary B.2, and let $\xi_1, \dots, \xi_m$ be i.i.d. standard normal random variables and $m \le 2k$, then*

$$\mathbb{E}\left[\sum_{i=1}^{m} I_{\{\xi_i \ge t\}} \mid \xi \ge t \text{ and } \eta \ge t\right] = O(1),$$

where $I_{\{\xi_i \ge t\}}$ is the indicator of the event $\{\xi_i \ge t\}$.

Proof. Let X and Y be as in the proof of Lemma B.1. Put $\alpha_i = \mathrm{cov}(X, \xi_i)$ and express each ξ_i as $\xi_i = \alpha_i X + \sqrt{1-\alpha_i^2}\cdot Z_i$. By Bessel's Inequality $\alpha_1^2 + \cdots + \alpha_m^2 \le 1$ (since random variables ξ_i are orthogonal). We now estimate

$$\begin{aligned}&\Pr(\xi_i \ge t \mid \sigma X \ge t + \sqrt{1-\sigma^2}\cdot |Y|) =\\ &\quad \Pr\left(\xi_i \ge t \mid \sigma X \ge t + \sqrt{1-\sigma^2}\cdot |Y| \text{ and } X \le 4t\right)\cdot \Pr\left(X \le 4t \mid \sigma X \ge t + \sqrt{1-\sigma^2}\cdot |Y|\right)\\ &\quad + \Pr\left(\xi_i \ge t \mid \sigma X \ge t + \sqrt{1-\sigma^2}\cdot |Y| \text{ and } X > 4t\right)\cdot \Pr\left(X > 4t \mid \sigma X \ge t + \sqrt{1-\sigma^2}\cdot |Y|\right)\end{aligned}$$

Notice that

$$\begin{aligned}&\Pr\big(\xi_i \ge t \mid \sigma X \ge t + \sqrt{1-\sigma^2}\cdot |Y| \text{ and } X < 4t\big)\\ &\quad = \Pr\big(\alpha_i X + \sqrt{1-\alpha_i^2}\cdot Z_i \ge t \mid \sigma X \ge t + \sqrt{1-\sigma^2}\cdot |Y| \text{ and } X < 4t\big)\\ &\quad = \int_{t/\sigma}^{4t} \Pr\big(\alpha_i x + \sqrt{1-\alpha_i^2}\cdot Z_i \ge t \mid \sigma x \ge t + \sqrt{1-\sigma^2}\cdot |Y|\big)dF(x)\\ &\quad \le \max_{x\in[t/\sigma, 4t]} \Pr\left(\sqrt{1-\alpha_i^2}\cdot Z_i \ge t - \alpha_i x \mid \sqrt{1-\sigma^2}\cdot |Y| \le \sigma x - t\right)\\ &\quad \overset{\text{by Corollary A.3}}{\le} \max_{x\in[t/\sigma, 4t]} \Pr\left(\sqrt{1-\alpha_i^2}\cdot Z_i \ge t - \alpha_i x\right) \le \Pr\left(Z_i \ge (1-4\alpha_i)t\right).\end{aligned}$$

It suffices to prove that

$$\sum_{i=1}^{m} \Pr\left(Z_i \geq (1-4\alpha_i)t\right) = \sum_{i=1}^{m} \tilde{\Phi}((1-4\alpha_i)t) = O(1).$$

Fix a sufficiently large constant c. The number of α_i that are greater than $1/c$ is $O(1)$. The number of α_i such that $\log^{-1} k \leq \alpha_i \leq 1/c$ is $O(\log^2 k)$ and for them $\tilde{\Phi}((1-4\alpha_i)t) = O(k^{-1/2})$ (since c is a sufficiently large constant). Finally, if $\alpha_i < 1/\log k$, then $\tilde{\Phi}((1-4\alpha_i)t) = O(k^{-1})$. This finishes the proof. □

Lemma B.4. *The function $(1-x)^2 f_k(x)$ is convex on the interval $[0,1]$.*

Proof. Let $m = k/\sqrt{\log k}$. Compute the first and the second derivatives of f_k:

$$\begin{aligned} f_k''(x) &= \left(m^{-\frac{2}{2-x}}\right)'' = -2\log m \cdot \left(\frac{m^{-\frac{2}{2-x}}}{(2-x)^2}\right)' \\ &= 4\log m \cdot \frac{m^{-\frac{2}{2-x}}}{(2-x)^3} \cdot \left(\frac{\log m}{2-x} - 1\right). \end{aligned}$$

Now $\left((1-x)^2 \cdot f_k(x)\right)'' = (1-x)^2 \cdot f_k''(x) - 4(1-x)f_k'(x) + 2f_k(x)$. Observe that $f_k(x)$ is always positive, and $f_k'(x)$ is always negative. Therefore, if $f_k''(x)$ is positive, we are done: $\left((1-x)^2 \cdot f_k(x)\right)'' \geq 0$. Otherwise, we have

$$\begin{aligned} \left((1-x)^2 \cdot f_k(x)\right)'' &= (1-x)^2 \cdot f_k''(x) - 4(1-x)f_k'(x) + 2f_k(x) \geq f_k''(x) + 2f_k(x) \\ &\geq 4\log m \cdot m^{-\frac{2}{2-x}} \left(\frac{\log m}{2} - 1\right) + 2m^{-\frac{2}{2-x}} = 2m^{-\frac{2}{2-x}}(\log m - 1)^2 \geq 0. \end{aligned}$$

□

C Analysis of Algorithm 2

In this section, we present the formal proof of Theorem 4.1. We will follow an informal outline of the proof sketched in the main text. We start with estimating probability (8).

Lemma C.1. *Let ξ and ζ be two independent random normal variables with variance 1 and σ^2 respectively ($0 < \sigma < 1$). Then for every positive t*

$$\Pr\left(\xi \leq t \text{ and } \xi + \zeta \geq t\right) = O(\sigma e^{\frac{(\sigma t+1)^2}{2}} \cdot e^{-\frac{t^2}{2}}).$$

Remark C.1. *In the "typical" case $e^{(\sigma t+1)^2/2}$ is a constant.*

Proof. We have

$$\begin{aligned}
\Pr\Big(\xi \le t \text{ and } \xi + \zeta \ge t\Big) &= \int_0^\infty \Pr\left(\xi \le t \text{ and } \xi + x \ge t\right) dF_\zeta(x) \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty \Pr\left(\xi \le t \text{ and } \xi + x \ge t\right) e^{-\frac{x^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi}}\int_0^\infty \Pr\left(\xi \le t \text{ and } \xi + \sigma\, y \ge t\right) e^{-\frac{y^2}{2}}\,dy \\
&= \frac{1}{\sqrt{2\pi}}\int_0^{t/\sigma} \Pr\left(t-\sigma\, y \le \xi \le t\right) e^{-\frac{y^2}{2}}\,dy + \frac{1}{\sqrt{2\pi}}\int_{t/\sigma}^\infty \Pr\left(t-\sigma\, y \le \xi \le t\right) e^{-\frac{y^2}{2}}\,dy.
\end{aligned}$$

Let us bound the first integral. Since the density of the random variable ξ on the interval $(t-\sigma y, t)$ is at most $\frac{1}{\sqrt{2\pi}}e^{\frac{-(t-\sigma y)^2}{2}}$ and $y \le e^y$, we have

$$\Pr\left(t-\sigma\, y \le \xi \le t\right) \le \sigma y \cdot \frac{1}{\sqrt{2\pi}}e^{\frac{-(t-\sigma y)^2}{2}} \le \frac{\sigma}{\sqrt{2\pi}} \cdot e^{\frac{-t^2}{2}} \cdot e^{(\sigma t+1)y}.$$

Therefore,

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}}\int_0^{t/\sigma} \Pr\left(t-\sigma\, y \le \xi \le t\right) e^{-\frac{y^2}{2}}\,dy \quad &\le \quad \frac{\sigma e^{\frac{-t^2}{2}}}{2\pi}\int_0^{t/\sigma} e^{(\sigma t+1)y} \cdot e^{-\frac{y^2}{2}}\,dy \\
&\le \quad \frac{\sigma e^{\frac{-t^2}{2}}}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{(y-(\sigma t+1))^2}{2}} \cdot e^{\frac{(\sigma t+1)^2}{2}}\,dy \\
&= \quad O\left(\sigma e^{\frac{-t^2}{2}} \cdot e^{\frac{(\sigma t+1)^2}{2}}\right).
\end{aligned}$$

We now upper bound the second integral. If $t \ge 1$, then

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}}\int_{t/\sigma}^\infty \Pr\left(t-\sigma\, y \le \xi \le t\right) e^{-\frac{y^2}{2}}\,dy \quad &\le \quad \frac{1}{\sqrt{2\pi}}\int_{t/\sigma}^\infty e^{-\frac{y^2}{2}}\,dy = \tilde{\Phi}(t/\sigma) = O\left(\frac{e^{-\frac{t^2}{2\sigma^2}}}{t/\sigma+1}\right) \\
&= \quad O\left(\frac{\sigma\, e^{-\frac{t^2}{2}}}{t+\sigma}\right) = O\left(\sigma\, e^{-\frac{t^2}{2}}\right).
\end{aligned}$$

If $t \le 1$, then

$$\frac{1}{\sqrt{2\pi}}\int_{t/\sigma}^\infty \Pr\left(t-\sigma\, y \le \xi \le t\right) e^{-\frac{y^2}{2}}\,dy \quad \le \quad \frac{1}{\sqrt{2\pi}}\int_0^\infty \sigma y \cdot e^{-\frac{y^2}{2}}\,dy = O\left(\sigma\right) = O\left(\sigma\, e^{-\frac{t^2}{2}}\right).$$

The desired inequality follows from the upper bounds on the first and second integrals. □

We need a slight generalization of the lemma.

Corollary C.2. *Let ξ and ζ be two independent random normal variables with variance 1 and σ^2 respectively ($0 < \sigma < 1$). Then for every $t \ge 0$ and $0 \le \bar{\varepsilon} < 1$*

$$\Pr\left(\xi + \zeta \ge (1-\bar{\varepsilon})t \quad \middle| \quad |\xi| \le t\right) = O\left(\frac{(\sigma + \bar{\varepsilon}t) \cdot c(\bar{\varepsilon}, \sigma, t) \cdot e^{-t^2/2}}{1 - 2\tilde{\Phi}(t)}\right),$$

where

$$c(\bar{\varepsilon},\sigma,t) = e^{\frac{(\sigma t+1)^2}{2}+\bar{\varepsilon}t^2}.$$

Remark C.2. *As in the previous lemma, in the "typical" case $c(\bar{\varepsilon},\sigma,t)$ is a constant.*

Proof. First note that

$$\begin{aligned}\Pr\left(\xi+\zeta \geq (1-\bar{\varepsilon})t \;\mid\; |\xi|\leq t\right) &\leq \frac{\Pr\left(\xi+\zeta \geq (1-\bar{\varepsilon})t \text{ and } \xi \leq t\right)}{\Pr\left(|\xi|\leq t\right)}\\ &= \frac{\Pr\left(\xi+\zeta \geq (1-\bar{\varepsilon})t \text{ and } \xi \leq t\right)}{1-2\tilde{\Phi}(t)}\end{aligned}$$

Now,

$$\begin{aligned}\Pr\left(\xi+\zeta \geq (1-\bar{\varepsilon})t \text{ and } \xi \leq t\right) &\leq \Pr\left(\xi+\zeta \geq t \text{ and } \xi \leq t\right)\\ &\quad + \Pr\left((1-\bar{\varepsilon})t \leq \xi+\zeta \leq t\right).\end{aligned}$$

By Lemma C.1, the first probability is bounded as follows:

$$\Pr\left(\xi+\zeta \geq t \text{ and } \xi \leq t\right) \leq O\left(\sigma e^{\frac{(\sigma t+1)^2}{2}} \cdot e^{-\frac{t^2}{2}}\right).$$

Since $\mathrm{Var}\left[\xi+\zeta\right] \leq 1+\sigma^2$, the second probability is at most

$$\Pr\left((1-\bar{\varepsilon})t \leq \xi+\zeta \leq t\right) \;\leq\; \bar{\varepsilon}t \cdot e^{-\frac{((1-\bar{\varepsilon})t)^2}{2(1+\sigma^2)}} \leq \bar{\varepsilon}t \cdot e^{\frac{(2\bar{\varepsilon}+\sigma^2)t^2}{2}} \cdot e^{-\frac{t^2}{2}},$$

here we used the following inequality

$$\frac{(1-\bar{\varepsilon})^2t^2}{2(1+\sigma^2)} = \frac{(1-\bar{\varepsilon})^2(1-\sigma^2)t^2}{2(1-\sigma^4)} \geq \frac{(1-2\bar{\varepsilon}-\sigma^2)t^2}{2} \geq \frac{t^2}{2} - \frac{(2\bar{\varepsilon}+\sigma^2)t^2}{2}.$$

The corollary follows. □

In the following lemma we formally define the random variables ζ_1 and ζ_2.

Lemma C.3. *Let ξ_1, ξ_2, η_1 and η_2 be standard normal random variables such that ξ_1 and ξ_2 are independent; η_1 and η_2 are independent; and*

- $\mathrm{cov}(\xi_1,\eta_1) \geq 1-\bar{\varepsilon} \geq 0$ *and* $\mathrm{cov}(\xi_2,\eta_2) \geq 1-\bar{\varepsilon} \geq 0$ *(for some positive $\bar{\varepsilon}$);*
- $\mathrm{cov}(\xi_1,\eta_2) \geq 0$ *and* $\mathrm{cov}(\xi_2,\eta_1) \geq 0$.

Then there exist normal random variables ζ_1 and ζ_2 independent of ξ_1 and ξ_2 with variance at most $2\bar{\varepsilon}$ such that

$$|\eta_1| - |\eta_2| \geq (1-4\bar{\varepsilon})|\xi_1| - (1+3\bar{\varepsilon})|\xi_2| - |\zeta_1| - |\zeta_2|.$$

Proof. Express η_1 as a linear combination of ξ_1, ξ_2, and a normal r.v. ζ_1 independent of ξ_1 and ξ_2:

$$\eta_1 = \alpha_1\xi_1 + \beta_1\xi_2 + \zeta_1,$$

similarly,

$$\eta_2 = \alpha_2\xi_1 + \beta_2\xi_2 + \zeta_2.$$

Note that $\alpha_1 = \mathrm{cov}(\eta_1, \xi_1) \geq 1 - \bar{\varepsilon}$ and $\beta_1 = \mathrm{cov}(\eta_1, \xi_2) \geq 0$. Thus

$$\mathrm{Var}\left[\zeta_1\right] \leq \mathrm{Var}\left[\eta_1\right] - \alpha_1^2 \leq 1 - (1-\bar{\varepsilon})^2 \leq 2\bar{\varepsilon}.$$

Similarly, $\alpha_2 \geq 0$, $\beta_2 \geq 1 - \bar{\varepsilon}$, and $\mathrm{Var}[\zeta_2] \leq 2\bar{\varepsilon}$. Since η_1 and η_2 are independent, we have

$$\alpha_1\alpha_2 + \beta_1\beta_2 + \mathrm{cov}(\zeta_1, \zeta_2) = \mathrm{cov}(\eta_1, \eta_2) = 0.$$

Therefore (note that $\mathrm{cov}(\zeta_1, \zeta_2) \leq 0$; $\alpha_1\alpha_2 \geq 0$; $\beta_1\beta_2 \geq 0$),

$$\alpha_2 = \frac{-\beta_1\beta_2 - \mathrm{cov}(\zeta_1, \zeta_2)}{\alpha_1} \leq \frac{\sqrt{\mathrm{Var}[\zeta_1]\,\mathrm{Var}[\zeta_2]}}{1-\bar{\varepsilon}} \leq \frac{2\bar{\varepsilon}}{1-\bar{\varepsilon}}.$$

Taking into account that $\alpha_2 \leq 1$, we get $\alpha_2 \leq \min(1, \frac{2\bar{\varepsilon}}{1-\bar{\varepsilon}}) \leq 3\bar{\varepsilon}$. Similarly, $\beta_1 \leq 3\bar{\varepsilon}$. Finally, we have

$$\begin{aligned}|\eta_1| - |\eta_2| &\geq (\alpha_1 - \alpha_2)|\xi_1| - (\beta_1 + \beta_2)|\xi_2| - |\zeta_1| - |\zeta_2| \\ &\geq (1 - 4\bar{\varepsilon})|\xi_1| - (1 + 3\bar{\varepsilon})|\xi_2| - |\zeta_1| - |\zeta_2|.\end{aligned}$$

□

In what follows we assume that ξ_1 is the largest r.v. in absolute value among $\xi_1, \ldots, \xi_m$ and its absolute value is t. For convenience we define three events:

$$A_t = \{|\xi_i| \leq t \text{ for all } 3 \leq i \leq m\};$$

$$E_t = A_t \cap \{|\xi_1| = t \text{ and } |\xi_2| \leq t\};$$

$$E = \{|\xi_1| \geq |\xi_i| \text{ for all } i\} = \bigcup_{t \geq 0} E_t.$$

Now we are ready to combine Corollary C.2 and Lemma C.3.

Lemma C.4. *Let $\xi_1, \cdots, \xi_m$ and $\eta_1, \cdots, \eta_m$ be two sequences of standard normal random variables. Suppose that*

1. *the random variables in each of the sequences are independent,*
2. *the covariance of every ξ_i and η_j is nonnegative,*
3. $\mathrm{cov}(\xi_1, \eta_1) \geq 1 - \bar{\varepsilon}$ *and* $\mathrm{cov}(\xi_2, \eta_2) \geq 1 - \bar{\varepsilon}$, *where* $\bar{\varepsilon} \leq 1/7$.

Then

$$\Pr\left(|\eta_1| \leq |\eta_2| \quad | \quad E_t\right) = O\left(\frac{(\sqrt{\bar{\varepsilon}} + \bar{\varepsilon}t) \cdot e^{-t^2/2} \cdot c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t)}{1 - 2\tilde{\Phi}(t)}\right), \tag{15}$$

where $c(\bar{\varepsilon}, \sigma, t)$ is from Corollary C.2.

Proof. By Lemma C.3, we have

$$|\eta_1| - |\eta_2| \geq (1-4\bar{\varepsilon})|\xi_1| - (1+3\bar{\varepsilon})|\xi_2| - |\zeta_1| - |\zeta_2|.$$

Therefore,

$$\begin{aligned}
\Pr\left(|\eta_1| \leq |\eta_2| \;\mid\; E_t\right) &\leq \Pr\left((1+3\bar{\varepsilon})|\xi_2| + |\zeta_1| + |\zeta_2| \geq (1-4\bar{\varepsilon})|\xi_1| \;\mid\; E_t\right) \\
&\leq \Pr\left(|\xi_2| + |\zeta_1| + |\zeta_2| \geq (1-7\bar{\varepsilon})t \;\mid\; E_t\right) \\
&\leq \sum_{s,s_1,s_2\in\{\pm 1\}} \Pr\left(s\xi_2 + s_1\zeta_1 + s_2\zeta_2 \geq (1-7\bar{\varepsilon})t \;\mid\; E_t\right)
\end{aligned}$$

Let us fix signs $s, s_1, s_2 \in \{\pm 1\}$ and denote $\xi = s\xi_2$, $\zeta = s_1\zeta_1 + s_2\zeta_1$, then we need to show that

$$\Pr\left(\xi + \zeta \geq (1-7\bar{\varepsilon})t \;\mid\; E_t\right) = O\left((\sqrt{\bar{\varepsilon}} + \bar{\varepsilon}t) \cdot \frac{e^{-t^2/2} \cdot c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t)}{1 - 2\tilde{\Phi}(t)}\right).$$

Observe that the random variables ξ, ζ and the event A_t are independent of ξ_1, thus

$$\begin{aligned}
&\Pr\left(\xi + \zeta \geq (1-7\bar{\varepsilon})t \;\mid\; E_t\right) \\
&\quad = \Pr\left(\xi + \zeta \geq (1-7\bar{\varepsilon})\,t \;\mid\; A_t \text{ and } |\xi_1| = t \text{ and } |\xi| \leq t\right) \\
&\quad = \Pr\left(\xi + \zeta \geq (1-7\bar{\varepsilon})\,t \;\mid\; A_t \text{ and } |\xi| \leq t\right) \\
&\quad = \Pr\left(\zeta \geq (1-7\bar{\varepsilon})\,t - \xi \;\mid\; A_t \text{ and } |\xi| \leq t\right).
\end{aligned}$$

Since ξ and A_t are independent, for every fixed value of ξ we can apply Corollary A.3. Thus

$$\begin{aligned}
\Pr\left(\zeta \geq (1-7\bar{\varepsilon})t - \xi \;\mid\; A_t \text{ and } |\xi| \leq t\right) &\leq \Pr\left(\zeta \geq (1-7\bar{\varepsilon})t - \xi \;\mid\; |\xi| \leq t\right) \\
&= \Pr\left(\xi + \zeta \geq (1-7\bar{\varepsilon})t \;\mid\; |\xi| \leq t\right).
\end{aligned}$$

Finally, by Corollary C.2 (where $\sigma^2 = \mathrm{Var}\left[\zeta\right] \leq 8\bar{\varepsilon}$),

$$\Pr\left(\xi + \zeta \geq (1-7\bar{\varepsilon})t \;\mid\; |\xi| \leq t\right) \;=\; O\left(\frac{(\sqrt{\bar{\varepsilon}} + \bar{\varepsilon}t) \cdot e^{-t^2/2} \cdot c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t)}{1 - 2\tilde{\Phi}(t)}\right).$$

□

Corollary C.5. *Under assumptions of Lemma C.4,*
1. if $\bar{\varepsilon}t^2 \leq 1$*, then*

$$\Pr\left(|\eta_1| \leq |\eta_2| \;\mid\; E_t\right) = O\left(\sqrt{\bar{\varepsilon}}\, \frac{(t+1) \cdot \tilde{\Phi}(t)}{1 - 2\tilde{\Phi}(t)}\right);$$

2. if $t > 1$*, then*

$$\Pr\left(|\eta_1| \leq |\eta_2| \;\mid\; E_t\right) = O\left(\sqrt{\bar{\varepsilon}}\right).$$

Proof. 1. If $\bar{\varepsilon}t^2 \leq 1$, then $\bar{\varepsilon}t \leq \sqrt{\bar{\varepsilon}}$ and

$$c(7\bar{\varepsilon}, \sqrt{8\bar{\varepsilon}}, t) = e^{\frac{(\sqrt{8\bar{\varepsilon}}t+1)^2}{2} + 7\bar{\varepsilon}t^2} = O(1).$$

Notice that

$$\frac{(\sqrt{\bar\varepsilon}+\bar\varepsilon t)\cdot e^{-t^2/2}}{1-2\tilde\Phi(t)} = O\left(\frac{(\sqrt{\bar\varepsilon}+\bar\varepsilon t)\cdot(t+1)\cdot\tilde\Phi(t)}{1-2\tilde\Phi(t)}\right),$$

since

$$\tilde\Phi(t)=\Theta\left(\frac{e^{-t^2/2}}{t}\right).$$

2. If $\bar\varepsilon > 1/32$ the statement holds trivially. So assume that $\bar\varepsilon \le 1/32$. Then

$$\frac{(\sqrt{8\bar\varepsilon}t+1)^2}{2}+7\bar\varepsilon t^2 \le \frac{3t^2}{8}+O(t).$$

Thus $t\cdot e^{-\frac{t^2}{2}}\cdot c(7\bar\varepsilon,\sqrt{8\bar\varepsilon},t)$ is upper bounded by some absolute constant. Since $t\ge 1$, the denominator $1-2\tilde\Phi(t)$ of the expression (15) is bounded away from 0. □

We now give a bound on the "typical" absolute value of the largest random variable.

Lemma C.6. *The following inequality holds:*

$$\Pr\left(|\xi_1|\ge 2\sqrt{\log m} \quad | \quad E\right)\le \frac{1}{m}.$$

Proof. Note that the probability of the event E is $1/m$, since all random variables $\xi_1,\dots,\xi_m$ are equally likely to be the largest in absolute value. Thus we have

$$\Pr\left(|\xi_1|\ge 2\sqrt{\log m} \quad | \quad E\right)\le \frac{\Pr\left(|\xi_1|\ge 2\sqrt{\log m}\right)}{\Pr(E)} \le \frac{1}{m^2}\Big/\frac{1}{m}=\frac{1}{m}.$$

□

Lemma C.7. *Let $\xi_1,\cdots,\xi_m$ and $\eta_1,\cdots,\eta_m$ be two sequences of standard normal random variables as in Theorem 4.1. Assume that $\mathrm{cov}(\xi_1,\eta_1)\ge 1-\bar\varepsilon$ and $\mathrm{cov}(\xi_2,\eta_2)\ge 1-\bar\varepsilon$, where $\bar\varepsilon < \min(1/(4\log m),1/7)$. Then*

$$\Pr\left(|\eta_1|\le|\eta_2| \quad | \quad E\right)=O\left(\frac{\sqrt{\bar\varepsilon}\log m}{m}\right).$$

Proof. Write the desired probability as follows:

$$\begin{aligned}\Pr\left(|\eta_1|\le|\eta_2| \quad | \quad E\right) &= \Pr\left(|\eta_1|\le|\eta_2| \text{ and } |\xi_1|\le 2\sqrt{\log m} \quad | \quad E\right)\\ &\quad + \Pr\left(|\eta_1|\le|\eta_2| \text{ and } |\xi_1|\ge 2\sqrt{\log m} \quad | \quad E\right)\end{aligned}$$

First consider the case $|\xi_1|\le 2\sqrt{\log m}$. Denote by $dF_{|\xi_1|}$ the density of $|\xi_1|$ conditional on E. Then

$$\begin{aligned}\Pr\left(|\eta_1|\le|\eta_2| \text{ and } |\xi_1|\le 2\sqrt{\log m} \quad | \quad E\right) &= \int_0^{2\sqrt{\log m}} \Pr\left(|\eta_1|\le|\eta_2| \quad | \quad E \text{ and } |\xi_1|=t\right) dF_{|\xi_1|}(t)\\ &= \int_0^{2\sqrt{\log m}} \Pr\left(|\eta_1|\le|\eta_2| \quad | \quad E_t\right) dF_{|\xi_1|}(t)\end{aligned}$$

Now by Corollary C.5,

$$\int_0^{2\sqrt{\log m}} \Pr\left(|\eta_1| \le |\eta_2| \quad | \quad E_t\right) dF_{|\xi_1|}(t) = \int_0^{2\sqrt{\log m}} O\left(\frac{2\sqrt{\bar\varepsilon \log m}\ \tilde\Phi(t)}{1-2\tilde\Phi(t)}\right) dF_{|\xi_1|}(t).$$

Let us change the variable to $x = 1-2\tilde\Phi(t)$. What is the probability density function of $1-2\tilde\Phi(|\xi_1|)$ given E? For each i the r.v. $1-2\tilde\Phi(|\xi_i|)$ is uniformly distributed on the interval $[0,1]$. Now $|\xi_i| > |\xi_j|$ if and only if $1-2\tilde\Phi(|\xi_i|) > 1-2\tilde\Phi(|\xi_j|)$, therefore $1-2\tilde\Phi(|\xi_1|)$ is distributed as the maximum of m independent random variables on $[0,1]$ given E. Its density function is $(x^m)' = mx^{m-1}$ (for $x \in [0,1]$). We have

$$\begin{aligned}\int_0^{2\sqrt{\log m}} \frac{2\sqrt{\bar\varepsilon \log m}\ \tilde\Phi(t)}{1-2\tilde\Phi(t)} dF_{|\xi_1|}(t) &\le \int_0^{\infty} \frac{2\sqrt{\bar\varepsilon \log m}\ \tilde\Phi(t)}{1-2\tilde\Phi(t)} dF_{|\xi_1|}(t)\\ &= \int_0^1 \frac{2\sqrt{\bar\varepsilon \log m}\cdot(1-x)/2}{x}\cdot mx^{m-1}dx = m\sqrt{\bar\varepsilon\log m}\int_0^1 (1-x)x^{m-2}dx\\ &= m\sqrt{\bar\varepsilon \log m}\left(\frac{1}{m-1}-\frac{1}{m}\right) = \frac{\sqrt{\bar\varepsilon\log m}}{m-1}.\end{aligned}$$

Now consider the case $|\xi_1| \ge 2\sqrt{\log m}$, by Corollary C.5,

$$\Pr\left(|\eta_1| \le |\eta_2| \quad | \quad E \text{ and } |\xi_1| \ge 2\sqrt{\log m}\right) = O\left(\sqrt{\bar\varepsilon}\right).$$

By Lemma C.6,

$$\Pr\left(|\xi_1| \ge 2\sqrt{\log m} \quad | \quad E\right) \le \frac{1}{m}.$$

This concludes the proof. □

Now we will prove a lemma, which differs from Theorem 4.1 only by one additional condition (4).

Lemma C.8. *Let $\xi_1, \cdots, \xi_m$ and $\eta_1, \cdots, \eta_m$ be two sequences of standard normal random variables. Let $\varepsilon^i = \operatorname{cov}(\xi_i, \eta_i)$. Suppose that*

1. *the random variables in each of the sequences are independent,*
2. *the covariance of every ξ_i and η_j is nonnegative,*
3. $\frac{1}{m}\sum_{i=1}^m \varepsilon^i = \varepsilon$,
4. $\varepsilon^i \le \min(1/(4\log m), 1/7)$.

Then the probability that the largest r.v. in absolute value in the first sequence has the same index as the largest r.v. in absolute value in the second sequence is $1 - O(\sqrt{\varepsilon \log m})$.

Proof. By Lemma C.7,

$$\Pr\left(|\eta_1| \le |\eta_2| \quad | \quad |\xi_1| \ge \max_{j\ge 2}|\xi_j|\right) = O\left(\frac{\sqrt{\log m}}{m}\sqrt{\max\left(\varepsilon^1, \varepsilon^2\right)}\right).$$

Applying the union bound, we get

$$\Pr(|\eta_1| \le \max_{i\ge 2}|\eta_j| \quad | \quad |\xi_1| \ge \max_{j\ge 2}|\xi_j|) = O\left(\frac{\sqrt{\log m}}{m}\sum_{i=2}^{m}\sqrt{\max\left(\varepsilon^1,\varepsilon^i\right)}\right)$$

$$= O\left(\frac{\sqrt{\log m}}{m}\cdot\left(m\sqrt{\varepsilon^1}+\sum_{i=1}^{m}\sqrt{\varepsilon^i}\right)\right) \overset{\text{by Jensen's inequality}}{\le} O\left(\sqrt{\log m}(\sqrt{\varepsilon^1}+\sqrt{\varepsilon})\right).$$

Since the probability that $|\xi_i| = \max_j |\xi_j|$ equals $1/m$ for each i, the probability that the largest r.v. in absolute value among ξ_i, and the largest r.v. in absolute value among η_j have different indexes is at most

$$O\left(\frac{1}{m}\sum_{i=1}^{m}\sqrt{\log m}\cdot(\sqrt{\varepsilon^i}+\sqrt{\varepsilon})\right) \le O\left(\sqrt{\log m}\cdot(\sqrt{\varepsilon}+\sqrt{\varepsilon})\right) = O\left(\sqrt{\varepsilon\log m}\right).$$

□

Proof of Theorem 4.1. Denote $\varepsilon^i = 1 - \mathrm{cov}(\xi_i,\eta_i)$. Then $(\varepsilon^1+\cdots+\varepsilon^m) \le m\varepsilon$. We may assume that $\varepsilon < \min(1/(4\log m), 1/7)$ — otherwise, the theorem follows trivially.

Consider the set $I = \{i : \varepsilon_i < \min(1/(4\log m), 1/7)\}$. Since $\varepsilon < \min(1/(4\log m), 1/7)$, the set I is not empty. Applying Lemma C.8 to random variables $\{\xi_i\}_{i\in I}$ and $\{\eta_i\}_{i\in I}$, we conclude that the the largest r.v. in absolute value among $\{\xi_i\}_{i\in I}$ has the same index as the largest r.v. in absolute value among $\{\xi_i\}_{i\in I}$ with probability

$$1 - O\left(\sqrt{\log|I|\cdot\frac{1}{|I|}\sum_{i\in I}\varepsilon^i}\right) = 1 - O\left(\sqrt{\varepsilon\log m}\right).$$

Since each ξ_i is the largest r.v. among $\xi_1,\ldots,\xi_m$ in absolute value with probability $1/m$, the probability that the largest r.v. among $\xi_1,\ldots,\xi_m$ does not belong to $\{\xi_i\}_{i\in I}$ is $\frac{m-|I|}{m}$. Similarly, the probability that the largest r.v. among $\eta_1,\ldots,\eta_m$ does not belong to $\{\eta_i\}_{i\in I}$ is $(m-|I|)/m$. Therefore, by the union bound, the probability that the largest r.v. in absolute value among ξ_i, and the largest r.v. in absolute value among η_j have different indexes is at most

$$1 - O(\sqrt{\varepsilon\log m}) - 2\,\frac{m-|I|}{m}. \tag{16}$$

We now upper bound the last term.

$$2\,\frac{m-|I|}{m} \overset{\text{by the Markov inequality}}{\le} 2\,\frac{\varepsilon}{\min(1/(4\log m), 1/7)}$$

$$\le 2\,(4\log m + 7)\varepsilon = O(\varepsilon\log m) = O(\sqrt{\varepsilon\log m}).$$

(Here we use that $\varepsilon\log m < 1$.)

Plugging this bound into (16) we get that the desired probability is $1 - O(\sqrt{\varepsilon\log m})$. This finishes the proof. □

D An Alternate Approach

We would like to present an alternative version of Algorithm 1. It demonstrates another approach for taking into account the lengths of vectors u_i: we can choose a separate threshold t_{u_i} for each u_i. This algorithm achieves the same approximation ratio. The analysis uses similar ideas and we omit it here.

Input: *A solution of the SDP, with the objective value* $\varepsilon \cdot |E|$.

Output: *An assignment of variables* x_u.

1. Define $\tilde{u}_i = u_i/|u_i|$ if $u_i \neq 0$, 0 otherwise.

2. Pick a random Gaussian vector g with independent components distributed as $\mathcal{N}(0, 1)$.

3. For each vertex u:

   (a) For each i project the vector g to $\tilde{u}_i$:

   $$\xi_{u_i} = \langle g, \tilde{u}_i \rangle.$$

   (b) Fix a threshold t_{u_i} s.t. $\Pr(\xi_{u_i} \geq t_{u_i}) = |u_i|^2$ (*i.e.* t_{u_i} is the $(1 - |u_i|^2)$-quantile of the standard normal distribution). Note that t_{u_i} depends on u_i.

   (c) Pick ξ_{u_i}'s that are larger than the threshold t_{u_i}:

   $$S_u = \{i : \xi_{u_i} \geq t_{u_i}\}.$$

   (d) Pick at random a state i from S_u and assign $x_u = i$.