Random Sampling and Approximation of MAX-CSP Problems

Noga Alon
Department of Mathematics
Tel Aviv University
Tel Aviv*

W. Fernandez de la Vega
CNRS
Universite Paris Sud
Orsay †

Ravi Kannan
Department of Computer Science
Yale University ‡

Marek Karpinski
Department of Computer Science
University of Bonn §

Abstract

We present a new efficient sampling method for approximating r-dimensional *Maximum Constraint Satisfaction* Problems, MAX-rCSP, on n variables up to an additive error ϵn^r. We prove a new general paradigm in that it suffices, for a given set of constraints, to pick a small uniformly random subset of its variables, and the optimum value of the subsystem induced on these variables gives (after a direct normalization and with high probability) an approximation to the optimum of the whole system up to an additive error of ϵn^r. Our method gives for the first time a polynomial in ϵ^{-1} bound on the sample size necessary to carry out the above approximation. Moreover, this bound is independent in the exponent on the dimension r. The above method gives a completely uniform sampling technique for all the MAX-rCSP problems, and improves the best known sample bounds for the low dimensional problems, like MAX-CUT.

The method of solution depends on a new result on the cut norm of random subarrays, and a new sampling technique for high dimensional linear programs. This method could be also of independent interest.

*Supported in part by a USA-Israeli BSF grant, by the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University. E-mail: noga@math.tau.il

†Supported in part by PROCOPE Project and by IST grant 1496 (RAND-APX). Part of this work was done while visiting Department of Computer Science, Yale University. E-mail: lalo@lri.fr.

‡Supported in part by NSF Grant CCR-9820850. E-mail: kannan@cs.yale.edu.

§Supported in part by DFG grants, Max-Planck Research Prize, and by IST grant 14036 (RAND-APX). Part of this work was done while visiting Department of Computer Science, Yale University. E-mail: marek@cs.uni-bonn.de.

1 Introduction

Suppose r is a fixed integer. In the MAX-rSAT problem, we are given a Conjunctive Normal Form Boolean formula on n variables, with each clause being the OR of precisely r literals. The objective is to maximize the number of clauses satisfied by an assignment to the n variables. The exact problem is NP-hard for $r \geq 2$. This paper has two main results - the first concerns general r, and the second the special case of $r = 2$. The first result is that for any $\epsilon > 0$, there is a positive integer $q \in O(\log(1/\epsilon)/\epsilon^{12})$ such that if we pick at random a subset of q variables (among the n) and solve the "induced" problem on the q variables (maximize the number of clauses satisfied among those containing only those variables and their negations), then the answer multiplied by n^r/q^r is, with high probability, within an additive factor ϵn^r of the optimal answer for the n variable problem. The q needed here will be called the "(vertex) sample complexity" of the problem for obvious reasons.

In fact, we show the same result for all MAX-rCSP problems. (MAX-rCSP problems, also called MAX-rFUNCTION-SAT, are equivalent to MAX-SNP [3]). We note that while, normally, sampling is used to estimate certain specific quantities, here the result actually says that the sample estimates an optimal solution value well. We do not know of any such optimizing results in statistics prior to this work.

The MAX-rSAT and other MAX-rCSP problems all admit fixed factor relative approximation algorithms which run in polynomial time. For some MAX-SNP problems, there have been major breakthroughs in achieving better factors using semi-definite programming and other techniques [9]. More relevant to our paper is the line of work started with the paper of Arora, Karger and Karpinski [3] which introduced the technique of smooth programs, and designed the first polynomial time algorithms for solving MAX-SNP problems (of arity r) to within additive error guarantee ϵn^r, for each fixed $\epsilon > 0$. Frieze and Kannan [7] proved an efficient version of Szémeredi's Regularity Lemma and used it to get a uniform framework to solve all MAX-SNP and some other problems in polynomial time with the same additive error. In [8], they introduced a new way of approximating matrices and more generally r-dimensional arrays, called the "cut-decomposition" and using those, proved a result somewhat similar to the main result here (for each fixed r), but with two important differences - (i) the sample complexity was exponential in $1/\epsilon$ and (ii) their result did not relate the optimal solution value of the whole problem to the optimal solution of the random sub-problems in their original setting; instead it related it to a complicated computational quantity associated with the random sub-problem. We will make central use of cut-decompositions in this paper.

For the special case of $r = 2$, Goldreich, Goldwasser and Ron [10] designed algorithms, where the sample complexity was polynomial in $1/\epsilon$; indeed, by exploiting the special structure of individual problems like the MAX-CUT problem they improved the polynomial dependence. Their results relate the optimal solution value of the whole problem to a complicated function of the random sub-problems like [7] (see also [7], [5] and [2] for higher dimensional cases, or for cases in which our only objective is to decide if we can satisfy almost all constraints). Thus they differ from our new uniform method.

Our second main result is a reduction of the sample complexity for all MAX-2CSP problems to $O(1/\epsilon^4)$. We must remark here that both our main results are derived by general arguments about approximating multi- (and 2-) dimensional arrays by some simple arrays and then using Linear Programming arguments. Unlike previous papers, we do not use problem-specific arguments which dwelve into the special structure of individual problems. The MAX-CUT problem (a special MAX-2CSP problem) has received much attention in this context. Indeed, independently of the papers so far cited, Fernandez de la Vega [6] developed a different algorithm for this problem which within polynomial time, produced a solution with additive error ϵn^2. [10] used the special structure of the problem to derive an algorithm with the best up to now sample complexity $O(1/\epsilon^5)$ (in the sense of (ii) above). Our improved sample complexity argument uses a tightened cut-decomposition argument as well as a better Linear Programming argument.

The global view of our method is the following. We represent MAX-rCSP problems by r-dimensional arrays. In the first stage we use the main result of Section 3 on cut norm of random subarrrays to transfer a cut decomposition of the whole array to a random sample. We use then a cut decomposition of a sample to approximate the value of the objective function. Then, in the second stage, we use linear programs to relate it to the value of the objective function on the whole array by using the main result of Section 4.

For arbitrary dimension r, the sample size for the first stage is $O\left(\frac{1}{\epsilon^6}\right)$, whereas the sample size for the second stage is $O\left(\frac{\log(\frac{1}{\epsilon})}{\epsilon^{12}}\right)$.

We notice, that in order to approximate any problem from MAX-rCSP, it is enough to give a *good absolute approximation* to the optimum of an induced random subsystem. As a consequence, our sample bound above gives, by a direct application of an approximation method of [3], the running times $2^{\tilde{O}\left(\frac{1}{\epsilon^2}\right)}$ for approximating all MAX-rCSP problems. This improves on the best known up to date bound of the form $2^{\tilde{O}\left(\frac{1}{\epsilon^{2r-2}}\right)}$ for the problems of dimension r ([8]).

The paper is organized as follows. Section 2 proves the existence of a Cut decomposition for arrays of dimension $r \geq 2$. This is shown to be essentially optimal in Section 5. Section 3 gives the basic result on the Cut decomposition induced on a random sub-array. In Section 4 we derive an upper bound for the sample size using Linear Programming. In Section 5, we prove a lower bound for the number of Cut arrays in a Cut decomposition. In Section 6, we give an improvement on sample size for all MAX-2CSP problems, including MAX-CUT, improving over the best known upper bounds for these problems. Further we prove also a lower bound on the sample complexity of MAX-CUT.

1.1 Notation

We consider $r-$dimensional arrays, where $r \geq 2$. [The $r = 2$ case gives us matrices.] If $V_1, V_2, \dots V_r$ are (not necessarily distinct) finite sets, an $r-$dimensional array A on $V_1, V_2, \dots V_r$ is a function $A : V_1 \times V_2 \times \dots V_r \longrightarrow \mathbf{R}$. For each $i_1 \in V_1, i_2 \in V_2, \dots i_r \in V_r$, we call $A(i_1, i_2, \ldots i_r)$ an entry of A. We let $||A||_F$ be the square root of the sum of squares of all the entries. [This is sometimes called the Frobenius norm, hence the subscript F.] For any $S_1 \subseteq V_1, S_2 \subseteq V_2 \ldots S_r \subseteq V_r$ we let $A(S_1, S_2, \ldots S_r) = \sum_{(i_1,i_2,\ldots i_r)\in S_1\times S_2\times\ldots S_r} A(i_1, i_2, \ldots i_r)$ and then define another norm $||A||_C$ (called the cut norm) :

$$A^+ = \max_{S_1\subseteq V_1, S_2\subseteq V_2,\ldots S_r\subseteq V_r} A(S_1, S_2, \ldots S_r)$$
$$\text{and } ||A||_C = \max(A^+, (-A)^+).$$

The cut norm was defined and studied by [8].

For any $S_1, S_2, \ldots S_r$, and real value d we define the *Cut Array* $C = CUT(S_1, S_2, \ldots S_r; d)$ by

$$C(i_1, i_2, \ldots i_r) = \begin{cases} d & \text{if } (i_1, i_2, \ldots i_r) \in S_1 \times S_2 \ldots S_r, \\ 0 & \text{otherwise.} \end{cases}$$

The real number d is called the coefficient of the cut array.

We use one other piece of notation : for any $Q \subseteq V_2 \times V_3 \ldots V_r$, we define

$$P(Q) = \{z \in V_1 : A(z, Q) =$$
$$\sum_{(z,i_2,i_3,\ldots i_r):(i_2,i_3,\ldots i_r)\in Q} A(z, i_2, i_3, \ldots i_r) > 0\}.$$

Note that P is with reference to an array A. It will be clear from context which array P is in reference to.

1.2 Main Results

We formulate now the main results of the paper. We denote by MAX-rCSP the class of all r-ary (r-dimensional) *Maximum Constraint Satisfaction Problems* (i.e. the problems defined by the collections of *r-ary boolean functions* $f : \{0,1\}^r \to \{0,1\}$ for r given variables out of the set of n variables with the objective to construct an assignment $s \in \{0,1\}^n$ which maximizes the number of satisfied constraints, cf., e.g., [12]). Given a problem P from MAX-rCSP for a given dimension $r \geq 2$, we call a (randomized) algorithm $\mathcal{A}$ an (absolute) ϵn^r- approximation algorithm for P, if for any instance I of P with n variables, the value $c(\mathcal{A}(I))$ produced by $\mathcal{A}$ on I satisfies, with high probability, $|OPT(I) - c(\mathcal{A}(I))| \leq \epsilon n^r$, where $OPT(I)$ is the value of the optimum. The *sample complexity* of an *r-dimensional ϵn^r-approximation algorithm* (defined for all $\epsilon > 0$) is the number of variables (nodes) in a random sample required by the algorithm as a function of $\frac{1}{\epsilon}$. We are interested in cases in which this complexity is independent of the size of the input size, and is bounded by a function of $\frac{1}{\epsilon}$ only; when this is not the case we say that the the sample complexity is infinite. We call a sample complexity *fully polynomial* if it is $(\frac{1}{\epsilon})^{0(1)}$.

For a fixed dimension r, a problem P from MAX-rCSP is said to have (an *absolute*) *fully polynomial sample complexity* $S = \left(\frac{1}{\epsilon}\right)^{0(1)}$, if for every fixed $\epsilon > 0$, there exists a *constant time ϵn^r-approximation algorithm* for P with a sample complexity S. A class of problems X

will be said to have a sample complexity S if all problems P in X have sample complexity S.

We formulate now our main results.

Theorem 1. *For every dimension r, and every fixed $\epsilon > 0$, MAX-rCSP has a constant time ϵn^r-approximation algorithm with fully polynomial sample complexity $O\left(\frac{\log(\frac{1}{\epsilon})}{\epsilon^{12}}\right)$.*

Theorem 2. *For every fixed $\epsilon > 0$, MAX-2CSP has a constant time ϵn^2-approximation algorithm with a sample complexity $O\left(\frac{1}{\epsilon^4}\right)$.*

The rest of the paper is devoted to the proofs of the above results as well as to the lower bound results on the number of Cut Arrays needed in our cut decompositions, and a lower bound on the sample complexity of MAX-CUT.

1.3 Constant Time Bounds

We show now that the *fully polynomial* sample size bounds of Theorem 1 (and more explicitly of Theorem 8) entail the existence of ϵn^r-approximation algorithms for arbitrary MAX-rCSP problems running, for any fixed $\epsilon > 0$, in time $2^{\tilde{O}\left(\frac{1}{\epsilon^2}\right)}$ and using sample size $O\left(\frac{\log(\frac{1}{\epsilon})}{\epsilon^{12}}\right)$. This improves on the best known so far running time bounds for approximating those problems which were of the form $2^{\tilde{O}\left(\frac{1}{\epsilon^{2r-2}}\right)}$ for r the dimension of a problem [8], and making them asymptotically equal to that of the MAX-CUT. The argument used in the proof of the following theorem is based on a technique of *smooth programs* and the approximation result of Arora, Karger and Karpinski [3]. The crucial point here is the independence of the exponent of $\left(\frac{1}{\epsilon}\right)$ in the running times of smooth programs approximations, on a dimension r.

Theorem 3. *For every fixed dimension r, and every $\epsilon > 0$, MAX-rCSP has ϵn^r-approximation algorithms running in time $2^{\tilde{O}\left(\frac{1}{\epsilon^2}\right)}$ and having sample complexity $O\left(\frac{\log(\frac{1}{\epsilon})}{\epsilon^{12}}\right)$.*

Proof. Let P be a problem on n variables from MAX-rCSP for a given r. We denote by OPT its optimum value. We consider subsystem $\mathcal{S}$ of constraints of P induced by a random sample of its variables of size $q = \Theta\left(\frac{\log(\frac{1}{\epsilon})}{\epsilon^{12}}\right)$. We denote by $\mathrm{OPT}_{\mathcal{S}}$ the optimum value of a subsystem $\mathcal{S}$. We have, by Theorem 8, w.h.p., the following inequality

$$|OPT - \frac{n^r}{q^r} OPT_{\mathcal{S}}| \leq \epsilon n^r. \tag{1}$$

We consider now only a new problem defined by a random subsystem $\mathcal{S}$, and represent it, by using a standard "arithmetization", as a *degree-r Smooth Integer Program*, see for details

[3]. We apply now Theorem 1.10 of [3] to get an $\epsilon' q^r$-approximation algorithm $\mathcal{A}$ for an induced subproblem computing a solution Y which satisfies $OPT_{\mathcal{S}} - \epsilon' q^r \le Y \le OPT_{\mathcal{S}}$ for arbitrary $\epsilon' > 0$. The running time of $\mathcal{A}$ is $q^{O\left(\frac{1}{(\epsilon')^2}\right)} = 2^{\tilde{O}\left(\frac{1}{(\epsilon')^2}\right)}$, with an explicit constant hidden in our O-notation upstairs depending polynomially on a dimension r, see [3].

By (1) we have, for all $\epsilon, \epsilon' > 0$,

$$OPT \le \frac{n^r}{q^r}\left(Y + \epsilon' q^r\right) + \epsilon n^r,$$

and

$$OPT \le \frac{n^r}{q^r} Y + \left(\epsilon + \epsilon'\right) n^r.$$

We have also

$$OPT \ge \frac{n^r}{q^r} Y - \epsilon n^r \ge$$

$$\ge \frac{n^r}{q^r} Y - \left(\epsilon + \epsilon'\right) n^r.$$

Thus, we have

$$|OPT - \frac{n^r}{q^r} Y| \le \left(\epsilon + \epsilon'\right) n^r$$

for arbitrary $\epsilon, \epsilon' > 0$.

Therefore an existence of an $\epsilon' q^r$-approximation algorithm computing a solution Y for an induced subproblem which works in time $2^{\tilde{O}\left(\frac{1}{(\epsilon')^2}\right)}$ (cf. [3]) entails, by Theorem 8, an ϵn^r-approximation algorithm for P working in time $2^{\tilde{O}\left(\frac{1}{\epsilon^2}\right)}$ (and using sample size $O\left(\frac{\log(\frac{1}{\epsilon})}{\epsilon^{12}}\right)$) for all $\epsilon > 0$.

□

A similar argument can be applied to Theorem 2, yielding

Theorem 4. *For every $\epsilon > 0$, MAX-2CSP has ϵn^2-approximation algorithms working in time $2^{\tilde{O}\left(\frac{1}{\epsilon^2}\right)}$ and having sample complexity $O\left(\frac{1}{\epsilon^4}\right)$.*

□

2 Existence of Cut Decomposition

In this section, we prove the existence of a certain approximation to any matrix. The approximation will be the sum of a small number of cut-arrays. The sum is taken entry-wise. The proof is elementary and essentially drawn from [8].

Theorem 5. *Suppose A is an array on $V_1, V_2, \ldots V_r$, $N = |V_1||V_2|\ldots|V_r|$ and ϵ is a positive real number. There exist at most $4^r/\epsilon^2$ cut arrays whose sum D approximates A well in the sense :*

$$||A - D||_C \leq \epsilon\sqrt{N}||A||_F \tag{2}$$

$$||A - D||_F \leq ||A||_F \tag{3}$$

The sum of the squares of the coefficients of the

$$\text{cut arrays is at most } 4^r \frac{||A||_F^2}{N}. \tag{4}$$

This upper estimate on the number of cut arrays is tight up to the dependence on the dimension r.

Proof. For an existence argument, we are going to find cut arrays $D^{(1)}, D^{(2)}, \ldots D^{(t)}$ one by one always maintaining the condition:

$$||A - (D^{(1)} + D^{(2)} + \cdots + D^{(t)})||_F^2 \leq (1 - \frac{\epsilon^2 t}{4^r})||A||_F^2. \tag{5}$$

We start with $t = 0$. At a general stage, suppose we already have $D^{(1)}, \ldots D^{(t)}$ satisfying (5). If now $W^{(t)} = A - (D^{(1)} + D^{(2)} + \cdots + D^{(t)})$ satisfies $||W^{(t)}||_C \leq \epsilon\sqrt{N}||A||_F$, then we stop. Otherwise, there exist $S_1, S_2, \ldots S_r$ such that $|W^{(t)}(S_1, S_2, \ldots S_r)| \geq \epsilon\sqrt{N}||A||_F$. If $|S_1| < |V_1|/2$, then since $W^{(t)}(S_1, S_2, \ldots S_r) = W^{(t)}(V_1, S_2, \ldots S_r) - W^{(t)}(V_1 \setminus S_1, S_2, \ldots S_r)$, we have that one of $|W^{(t)}(V_1, S_2, \ldots S_r)|$ or $|W^{(t)}(V_1 \setminus S_1, S_2, \ldots S_r)|$ must be at least $(\epsilon/2)\sqrt{N}||A||_F$. Thus we have that there exist some $S_1 \subseteq V_1$, $|S_1| \geq |V_1|/2$ and $S_2, \ldots S_r$ such that $|W^{(t)}(S_1, S_2, \ldots S_r)| \geq (\epsilon/2)\sqrt{N}||A||_F$. By repeating this with $S_2, S_3, \ldots S_r$, we see that

$$\exists S_1^{t+1}; S_2^{t+1}, \ldots S_r^{t+1} : |S_i^{t+1}| \geq |V_i|/2$$

$$|W^{(t)}(S_1^{t+1}; S_2^{t+1}, \ldots S_r^{t+1})| \geq (\epsilon/2^r)\sqrt{N}||A||_F.$$

Let $d_{t+1} = W^{(t)}(S_1^{t+1}; S_2^{t+1}, \ldots S_r^{t+1})/(|S_1^{t+1}||S_2^{t+1}|\ldots|S_r^{t+1}|)$ be the average of the entries in $S_1 \times S_2 \times \ldots S_r$ and let $D^{(t+1)} = CUT(S_1^{t+1}; S_2^{t+1}, \ldots S_r^{t+1}, d_{t+1})$. Then, noting that subtracting the cut array $D^{(t+1)}$ from $W^{(t)}$ just corresponds to subtracting the average from a set of real numbers, we have :

$$\begin{aligned}
||W^{(t)} - D^{(t+1)}||_F^2 - ||W^{(t)}||_F^2 = \\
\sum_{i_1 \in S_1^{t+1}, i_2 \in S_2^{t+1} \ldots} ((W^{(t)}(i_1, i_2, \ldots i_r) - d_{t+1})^2 \\
-(W^{(t)}(i_1, i_2, \ldots i_r))^2) \qquad (6)\\
= -|S_1^{t+1}||S_2^{t+1}|\ldots|S_r^{t+1}|d_{t+1}^2 = \\
-\frac{W^{(t)}(S_1^{t+1}, S_2^{t+1}, \ldots S_r^{t+1})^2}{|S_1^{t+1}||S_2^{t+1}|\ldots|S_r^{t+1}|} \leq -\frac{\epsilon^2}{2^{2r}}||A||_F^2.
\end{aligned}$$

$$\text{Also, } ||W^{(t)} - D^{(t+1)}||_F^2 - ||W^{(t)}||_F^2 \leq -d_{t+1}^2 N/2^{2r}. \tag{7}$$

We now have (5) satisfied with t one greater. Note that (5) implies that we must stop before t exceeds $2^{2r}/\epsilon^2$. The upper bound on the sum of the d_t^2 follows from adding up the inequalities (7) which yields

$$||A||_F^2 \geq ||A||_F^2 - ||A - (D^{(1)} + D^{(2)} + \ldots D^{(t)})||_F^2 \geq \sum_t d_t^2 N/2^{2r}.$$

The proof of the thigtness of the upper estimate is included in Section 5. □

3 Cut Norm of Random Subarrays

The main purpose of this section is to show that if an array on V^r (where $|V| = n$ is large) has small cut-norm, then so does the array induced by a random subset J of V of cardinality $O(1/\epsilon^6)$.

The outline of the proof is as follows : Suppose G is the array on V^r, and B is the array on J^r. Suppose $Q_1, Q_2, \ldots Q_r$ are random subsets of J^{r-1}, each of cardinality $\Omega(1/\epsilon^2)$. Then, lemma (7) asserts that with high probability, there are subsets $Q_1' \subseteq Q_1$, $Q_2' \subseteq Q_2 \ldots Q_r' \subseteq Q_r$ such that

$$B(P(Q_1'), P(Q_2'), \ldots P(Q_r')) \approx B^+. \tag{8}$$

In other words, we need to consider only $2^{O(1/\epsilon^2)}$ candidate subsets of J to find the $S_1, S_2, \ldots S_r \subseteq J$ approximately maximizing $B(S_1, S_2, \ldots S_r)$ (not all $2^{O(|J|)}$ of them.) Next Lemma (8) shows that if we had already fixed, say $X_1 = P(Q_1'), X_2 = P(Q_2'), \ldots X_r = P(Q_r')$, and then we pick J (independently of X_i), we will have that with high probability

$$G(X_1, X_2, \ldots X_r) \approx \frac{|V|^r}{|J|^r} B(X_1, X_2, \ldots X_r). \tag{9}$$

Multiplying the failure probability with the number of possible subsets of the Q_i (which is $2^{O(1/\epsilon^2)}$), we also get that with high probability, this holds for every subset Q_1' of Q_1, Q_2' of Q_2 etc. If this holds rigorously, we would then clearly be able to infer from (8) and (9) that

$$G^+ \approx \frac{|V|^r}{|J|^r} B^+.$$

A similar inequality also will follow (along the same lines) for $(-G)^+$ and this would finish the proof.

The major problem is that J is not independent of $Q_1, Q_2, \ldots Q_r$; if it were (8) will not hold. To tackle this, we adopt a method of proof reminiscent of the argument of Vapnik and Chervonenkis [15]. We consider a set J' which is J minus all the end points of $r-$ tuples in $Q_1, Q_2, \ldots Q_r$. Noting that $|J| - |J'| \in O(1/\epsilon^2)$, we argue that we get roughly the same probability distributions if we pick, as we described already, J first and then $Q_1, Q_2, \ldots Q_r$ as random subsets of J^{r-1}, whence (8) holds as if we first pick J' and then $Q_1, Q_2, \ldots Q_r$ as random subsets of V^{r-1}, whence we have that (9) holds. Thus, we may actually use both (8) and (9) to get our result.

Lemma 6. *Suppose B is a $r-$dimensional array on $R_1 \times R_2 \times \ldots R_r$. Suppose $S_1 \subseteq R_1, S_2 \subseteq R_2, \ldots S_r \subseteq R_r$ are some fixed subsets. Suppose Q_1 is a random subset of $R_2 \times R_3 \times \ldots R_r$ of cardinality p.* [1] *Then, with probability at least $1 - \frac{1}{40(4r)^r}$, we have :*

$$B(P(Q_1 \cap (S_2 \times S_3 \ldots S_r)), S_2, S_3, \ldots, S_r) \geq$$

$$B(S_1, S_2, \ldots S_r) - \frac{40(4r)^r \sqrt{|R_1||R_2| \ldots |R_r|}}{\sqrt{p}} ||B||_F.$$

Proof. Let $S_2 \times S_3 \ldots \times S_r = S$. We have,

$$B(P(Q_1 \cap S), S) = B(P(S), S) - B(B_1, S) + B(B_2, S), \tag{10}$$

where

$$\begin{aligned} B_1 &= \{z \in R_1 : B(z, S) > 0 \text{ and } B(z, S \cap Q_1) < 0\}, \\ B_2 &= \{z \in R_1 : B(z, S) < 0 \text{ and } B(z, S \cap Q_1) > 0\}, \end{aligned}$$

Consider one fixed $z \in R_1$. Let $X_z = B(z, S \cap Q_1)$. We may write X_z as the sum $X_1 + X_2 + \ldots X_p$, where $X_1, X_2, \ldots X_p$ is a sample of size p drawn uniformly without replacement from the set of $l = |R_2| \times |R_3| \times \ldots |R_r|$ reals - $\{B(z, y))\underline{1}_{y \in S}\}$. For analysis, we also introduce the random variables $Y_1, Y_2, \ldots Y_p$ - a sample of size p drawn independently, each uniformly distributed over the same set of reals, but now with replacement. We have

$$\begin{gathered} E(X_1 + X_2 + \ldots X_p) = \frac{p}{l} B(z, S) \\ \mathrm{Var}(X_1 + X_2 + \ldots X_p) \leq \mathrm{Var}(Y_1 + Y_2 + \ldots Y_p) \leq \\ \frac{p}{l} \sum_{u \in S} B(z, u)^2 \leq \frac{p}{l} \sum_{u \in R_2 \times R_3 \times \ldots R_r} B(z, u)^2, \end{gathered}$$

where the second line is a standard inequality (for example, it follows from Theorem 4 of [11]). Hence, for any $\xi > 0$,

$$\mathbf{Pr}\left(\left| X_z - \frac{p}{l} B(z, S) \right| \geq \xi \right) \leq \frac{p \sum_{u \in R_2 \times R_3 \times \ldots R_r} B(z, u)^2}{l \xi^2} \tag{11}$$

If $z \in B_1$ then $X_z - (p/l)B(z, S) \leq -(p/q)B(z, S)$ and so applying (11) with $\xi = pB(z, S)/l$ we get that for each fixed z,

$$\mathbf{Pr}(z \in B_1) \leq \frac{l \sum_{u \in R_2 \times R_3 \times \ldots R_r} B(z, u)^2}{pB(z, S)^2}.$$

$$\mathbf{E}\left(\sum_{z \in B_1} B(z, S) \right)$$

[1] So, each of the $\binom{|R_2||R_3|\ldots|R_r|}{p}$ subsets is equally likely to be picked to be Q_1.

$$\leq \sum_{\{z\in R_1:\, B(z,S)>0\}} \min\left\{B(z,S), \frac{l\sum_u B(z,u)^2}{pB(z,S)}\right\}$$

$$\leq \sum_{\{z\in R_1:\, B(z,S)>0\}} \sqrt{\frac{l\sum_{u\in R_2\times R_3\ldots R_r} B(z,u)^2}{p}} \tag{12}$$

By an identical argument we obtain

$$\mathbf{E}\left(\sum_{z\in B_2} B(z,S)\right) \geq -\sum_{\{z\in R_1:\, B(z,S)<0\}} \sqrt{\frac{l\sum_u B(u,z)^2}{p}}.$$

Hence, (using the Cauchy-Schwartz inequality),

$$\mathbf{E}(B(P(Q_1\cap S),S)) \geq B(P(S),S) - \sum_{z\in R_1} \sqrt{\frac{l\sum_u B(u,z)^2}{p}}$$

$$\geq B(P(S),S) - \frac{\sqrt{|R_1||R_2|\ldots|R_r|}}{\sqrt{p}}||B||_F.$$

Now, $B(P(S),S) - B(P(S\cap Q_1),S)$ is a nonnegative random variable with expectation at most $\frac{\sqrt{|R_1||R_2|\ldots|R_r|}}{\sqrt{p}}||B||_F$, as argued above. So using Markov inequality, the lemma follows.

□

Lemma 7. *Suppose B is a $r-$dimensional array on $R_1\times R_2\times\ldots R_r$. Let $p\geq 160r^4/\epsilon^2$. Suppose also that Q_i is a random subset of $R_1\times R_2\times\ldots R_{i-1}\times R_{i+1}\ldots R_r$ of cardinality p. Then with probability at least $1-r/(40(4r)^r)$, we have :*

$$\exists Q_1'\subseteq Q_1, \exists Q_2'\subseteq Q_2, \ldots \exists Q_r'\subseteq Q_r,$$

$$B(P(Q_1'),P(Q_2'),\ldots P(Q_r')) \geq B^+ - \epsilon\sqrt{|R_1||R_2|\ldots|R_r|}\;||B||_F.$$

Proof. Let $S_1\subseteq R_1, S_2\subseteq R_2\ldots S_r\subseteq R_r$ satisfy $B(S_1,S_2,\ldots S_r)=B^+$. Applying Lemma (6) r times, we get the current lemma. □

We first need one more simple technical lemma.

Lemma 8. *Suppose G is a r dimensional array on V^r with each entry of absolute value at most M. Let t be a fixed positive integer. Let I be a random subset of V of cardinality t. Then, with probability at least $1-e^{-\epsilon^4 t/8}$ we have*

$$|G(V,V,V,\ldots V) - \frac{|V|^r}{(t)^r}G(I,I,\ldots I)| \leq \epsilon^2 M|V|^r.$$

Proof. Note that changing any one element of I changes the random variable $G(I,I,\ldots I)$ by at most Mt^{r-1}. Thus the lemma follows by standard Martingale inequalities ([4]).

Theorem 9. *Suppose G is a $r-$dimensional array on $V^r = V \times V \times \dots V$ with all entries of absolute value at most M. Let J be a random subset of V of cardinality $q \geq 1000r^7/\epsilon^6$. Let B be the $r-$dimensional array obtained by restricting G to J^r. Then, we have with probability at least 39/40:*

$$||B||_C \leq \frac{q^r}{|V|^r}||G||_C + 10\epsilon^2 M q^r + 5\epsilon q^r \frac{||G||_F}{|V|^{r/2}}.$$

Proof. First we have that $E(||B||_F^2) = \frac{q^r}{|V|^r}||G||_F^2$, so using Markov inequality, we have that with

$$E_1 : ||B||_F \leq 4\frac{q^{r/2}}{|V|^{r/2}}||G||_F \text{ has } \mathbf{Pr}(E_1) \geq 9/10. \tag{13}$$

Let $p = 160r^4/\epsilon^2$. Let $Q_1, Q_2, \dots Q_r$ be r independently, each uniformly randomly picked subsets of J^{r-1}, each of cardinality p. We apply Lemma (7) to B. So, with probability at least 7/8 (using (13))

$$\begin{aligned} \exists Q_1' \subseteq Q_1, \exists Q_2' \subseteq Q_2, \dots \exists Q_r' \subseteq Q_r, G(P(Q_1') \cap J, P(Q_2') \\ \cap J, \dots P(Q_r') \cap J) \geq B^+ - \frac{\epsilon}{3}\frac{q^r}{|V|^{r/2}}||G||_F. \end{aligned} \tag{14}$$

[Here, we mean by $P(Q_1')$ the set $\{z \in V : G(z, Q_1') > 0\}$.] Let J' be obtained from J by removing the at most $r(r-1)p$ end points of the elements of $Q_1 \cup Q_2 \cup \dots Q_r$.

We will make crucial use of the fact that the following two different methods of picking $J, Q_1, Q_2, \dots Q_r$ produce nearly the same joint probability distribution on them :

(i) As above, pick J to be a random subset of V of cardinality q and then pick $Q_1, Q_2, \dots Q_r$ to be independent random subsets of J^{r-1} each of cardinality p. Let $P^{(i)}(J, Q_1, Q_2, \dots Q_r)$ be the probability that we pick $J, Q_1, Q_2, \dots Q_r$ in this experiment. Then, clearly, for each $J, Q_1, Q_2, \dots Q_r$ with $|J| = q, Q_1, Q_2, \dots Q_r \subseteq J^{r-1}, |Q_i| = p$, we have

$$P^{(i)}(J, Q_1, Q_2, \dots Q_r) = \left(\binom{|V|}{q}\binom{q^{r-1}}{p}^r\right)^{-1}.$$

(ii) Now, pick J' to be a random subset of V of cardinality $q - r(r-1)p$. Then pick independently (of J' and of each other) r random subsets $\tilde{Q}_1, \dots \tilde{Q}_r$ of V^{r-1} of cardinality p each. Let $\tilde{J} = J' \cup$ (the set of all end points of elements of $\tilde{Q}_1 \cup \tilde{Q}_2 \dots \tilde{Q}_r$). Let $P^{(ii)}(J', \tilde{Q}_1, \dots \tilde{Q}_r)$ be the probabilities here.

Define E_2 to be the event that all $pr(r-1)$ end points of the elements in $Q_1, Q_2, \dots Q_r$ are all distinct and let E_3 be the event that all the end points of $\tilde{Q}_1, \tilde{Q}_2, \dots \tilde{Q}_r$ are distinct and none of them is in J'. It is easy to see by direct calculation that conditioned on the events E_2, E_3 $P^{(i)}$ and $P^{(ii)}$ are exactly equal. It is also easy to see that

$$P^{(i)}(E_2) = \binom{\binom{q}{r-1}}{p}\binom{\binom{q}{r-1} - p}{p} \dots \binom{\binom{q}{r-1} - (r-1)p}{p} /$$

$$\left[\binom{q^{r-1}}{p}\right]^r \geq 99/100,$$

and $P^{(ii)}(E_3) \geq 99/100$; so we have that the following inequality which we will use shortly :

$$||P^{(i)} - P^{(ii)}||_{\text{TV}} \leq 1/50. \tag{15}$$

Consider one particular collection of subsets $Q_1' \subseteq Q_1, Q_2' \subseteq Q_2, \ldots Q_r' \subseteq Q_r$. We will apply Lemma (8) to the array G' on V^r obtained by setting

$$\begin{aligned} G'(i_1, i_2, \ldots i_r) = G(i_1, i_2, \ldots i_r) \forall (i_1, i_2, \ldots i_r) \in P(Q_1') \\ \times P(Q_2') \times \ldots P(Q_r') \\ G(i_1, i_2, \ldots i_r) = 0 \text{ otherwise} . \end{aligned}$$

Note that $||G'||_F \leq ||G||_F$. Note that we are considering the set-up regarding $P^{(ii)}$; so we may assume that $Q_1, Q_2, \ldots Q_r$ have already been picked. For now, the subsets $Q_1' \subseteq Q_1, Q_2' \subseteq Q_2, \ldots Q_r' \subseteq Q_r$ have been also fixed. Then we pick $J' \subseteq V$ of cardinality $q - r(r-1)p$ independently of $Q_1, Q_2, \ldots Q_r$. Thus applying the lemma, we get the claimed bounds for the probabilities of the events defined below :

$$\begin{aligned} \text{Let } E_8(J', Q_1', Q_2', \ldots Q_r') : \\ \Big| G(P(Q_1'), P(Q_2'), \ldots P(Q_r')) - \frac{|V|^r}{(q - r(r-1)p)^r} \\ G(P(Q_1') \cap J', P(Q_2') \cap J', \ldots, P(Q_r') \cap J') \Big| \\ \leq 10\epsilon^2 M |V|^r \\ \text{Then, } P^{(ii)}(E_8(J', Q_1', Q_2', \ldots Q_r')) \geq 1 - e^{-\epsilon^4 q/16}. \end{aligned}$$

Now using the fact that for a choice of $Q_1, Q_2, \ldots Q_r$, there are $2^{pr} \leq e^{\epsilon^2 q/32}$ choices of $Q_1', Q_2', \ldots Q_r'$, we get :

$$\begin{aligned} E_9(J', Q_1, Q_2, \ldots Q_r) \ : \ \forall Q_1' \subseteq Q_1, \forall Q_2' \subseteq Q_2, \ldots \forall Q_r' \subseteq Q_r \\ E_8(J', Q_1', Q_2', \ldots Q_r') \\ P^{(ii)}(E_9(J', Q_1, Q_2, \ldots Q_r)) \geq 1 - e^{-\epsilon^4 q/32} \geq 99/100. \end{aligned}$$

Noting that $q^r \leq (1 + \epsilon^2)(q - r(r-1)p)^r$ and $|G(P(Q_1') \cap J', P(Q_2') \cap J', \ldots, P(Q_r') \cap J') - G(P(Q_1') \cap J, P(Q_2') \cap J, \ldots, P(Q_r') \cap J)| \leq \epsilon^2 q^r M$, we get (using also (15)) :

$$\begin{aligned} \text{Let } E_{10}(J, Q_1, Q_2, \ldots Q_r) : \\ \forall Q_1' \subseteq Q_1, \forall Q_2' \subseteq Q_2, \ldots \forall Q_r' \subseteq Q_r \\ \Big| G(P(Q_1'), P(Q_2'), \ldots P(Q_r')) - \frac{|V|^r}{(q - r(r-1)p)^r} \\ G(P(Q_1') \cap J, P(Q_2') \cap J, \ldots, P(Q_r') \cap J) \Big| \leq 10\epsilon^2 M |V|^r, \\ P^{(i)}(E_{10}(J, Q_1, Q_2, \ldots Q_r)) \geq 97/100. \end{aligned} \tag{16}$$

Under $E_{10}(J, Q_1, Q_2, \ldots Q_r)$, we have from (14) that

$$\exists Q_1' \subseteq Q_1, \exists Q_2' \subseteq Q_2 \ldots G(P(Q_1'), P(Q_2'), \ldots P(Q_r')) \geq$$

$$\frac{|V|^r}{q^r}B^+ - 5\epsilon|V|^{r/2}||G||_F - 10\epsilon^2 M|V|^r.$$

Thus, we get that with probability at least $79/80$:

$$G^+ \geq \frac{|V|^r}{q^r}B^+ - 10\epsilon^2 M|V|^r - 5\epsilon|V|^{r/2}||G||_F.$$

By an exactly identical argument applied to $-G$, we get also that with probability at least $79/80$,

$$(-G)^+ \geq \frac{|V|^r}{q^r}(-B)^+ - 10\epsilon^2 M|V|^r - 5\epsilon|V|^{r/2}||G||_F.$$

From the last two statements, the Theorem follows.

4 Upper Bound on the Sample Complexity of MAX-rCSP

The purpose of this section is to prove the following theorem.

Theorem 10. *Let r be a fixed integer such that $r \geq 2$. Let $F = \{f_1, ...f_\ell\}$ be a collection of functions where each f_i is a boolean function of exactly r variables picked from $V = \{x_1, ...x_n\}$. Assume that J is a random subset of V of cardinality q where $q = \Omega(\frac{\log(1/\epsilon)}{\epsilon^{12}})$. Let $m^{(V)}$ denote the maximum number of functions in F which can be made true for some assignment of V and $m^{(J)}$ the maximum number of functions in F with all variables in J which can be made true. Then we have that*

$$m^{(V)} \leq m^{(J)}\frac{|V|^r}{q^r} + \epsilon|V|^r \tag{17}$$

$$m^{(V)} \geq m^{(J)}\frac{|V|^r}{q^r} - \epsilon|V|^r \tag{18}$$

with probability at least 2/3.

Note that our Ω hides a factor exponential in r

Proof. For each $0, 1$ sequence z of length r, $z = (z_1, z_2, ...z_r)$, say, we define the r-dimensional array $A^{(z)}$ on V^r by

$$A^{(z)}(i_1, ...i_r) = \text{number of functions in } F \text{ true by setting}$$

$$x_{i_1} = z_1, ...x_{i_r} = z_r$$

Note that the $A^{(z)}$ are not algorithmically constructed. They are used only for the proof. We let $M = \max_{z\in\{0,1\}^r} ||A^{(z)}||_\infty$. We can of course assume $M \leq 2^{2^r}$. □

Let $S : V \to \{0, 1\}$ be any fixed assignment. We will also think of S as the set of true variables under S. Clearly, the number of functions satisfied by S is equal to

$$\sum_{z\in\{0,1\}^r} \sum_{i_1,...i_r:S(i_1)=z_1,...S(i_r)=z_r} A^{(z)}(i_1, ...i_r) \tag{19}$$

Suppose that we have cut decompositions of all the $A^{(z)}$

$$D^{(z)} = A^{(z)} - E^{(z)} = \sum_{t=1}^{s} Cut(S_{t,1}^{(z)}, S_{t,2}^{(z)}, ...S_{t,r}^{(z)}, d_t^{(z)}), 1 \le t \le s,$$

say, with $s = \frac{4^r}{\epsilon^2}$, $||E^{(z)}||_C \le \epsilon M|V|^r$. Using (19), we see that the number of functions which are true in the assignment S and with weights given by the arrays $D^{(z)}$, $z \in \{0,1\}^r$, is equal to $v^*(\nu)$, say, where

$$v^*(\nu) = \sum_{z \in \{0,1\}^r} \sum_{t=1}^{s} d_t^{(z)} \Pi_{i=1}^{r} \nu_{t,i}^{(z)} \tag{20}$$

with $\nu_{t,i}^{(z)} = |S_{t,i}^{(z)} \cap S|$ if $z_i = 1$ and $\nu_{t,i}^{(z)} = |S_{t,i}^{(z)} \cap (V \backslash S)|$ if $z_i = 0$.

For $t = 1, 2, ...s,\ i = 1, 2, ...r$ and $z \in \{0,1\}^r$, fix a set ν of values of the $\nu_{t,i}^{(z)}$. We say that ν is realizable if there exists $S \subseteq V$ such that

$$||S_{t,i}^{(z)} \cap S| - \nu_{t,i}^{(z)}| \le \frac{3\epsilon^3}{8^r s} n \text{ for all triples } (z,t,i) \text{ with } z_i = 1, \text{ and}$$

$$||S_{t,i}^{(z)} \cap (V \backslash S)| - \nu_{t,i}^{(z)}| \le \frac{3\epsilon^3}{8^r s} n \text{ for all triples } (z,t,i) \text{ with } z_i = 0.$$

We claim that if ν is not realizable, then the following Linear Program LP(V,ν) which is just a tightening of the above inequalities, is not feasible:

$$\nu_{t,i}^{(z)} - \frac{2\epsilon^3 n}{8^r s} \le \sum_{j \in S_{t,i}^{(z)}} x_j \le \nu_{t,i}^{(z)} + \frac{2\epsilon^3 n}{8^r s} \quad \text{for all triples } (z,t,i)$$

$$\text{with } z_i = 1$$

$$\nu_{t,i}^{(z)} - \frac{2\epsilon^3 n}{8^r s} \le \sum_{j \in S_{t,i}^{(z)}} (1 - x_j) \le \nu_{t,i}^{(z)} + \frac{2\epsilon^3 n}{8^r s}$$

$$\text{for all triples } (z,t,i) \text{ with } z_i = 0$$

$$0 \le x_j \le 1,\ 1 \le j \le n \qquad [LP(V,\nu)]$$

[This is because if LP(V,ν) was feasible, then it would have a basic feasible solution which would have at most $N = sr2^{r+1}$ fractional components; setting the fractional x_i to zero will yield a 0-1 vector realizing ν. We use the obvious fact that for large n, we have that $sr2^{r+1} \le \frac{\epsilon^3}{8^r s} n$]. So, by Linear Programming duality, we see that there exists one inequality obtained as a nonnegative combination of the first N inequalities of LP(V,ν) for which there is no solution x satisfying the bounds $0 \le x_i \le 1$. It is easy to see that the combination need not involve both the upper bound and the lower bound on any of the sets $S_{t,i}^{(z)}$. Thus we get that there are $sr2^r$ real numbers $u^{(z)}_{t,i}, 1 \leq t \leq s,\ 1 \leq i \leq r,\ z \in \{0,1\}^r$ (depending on ν) such that, letting,

$$c_i^{(\nu)} = \sum_{1 \leq j \leq r} \left(\sum_{z:z_j=1} \sum_{t:i \in S^{(z)}_{t,j}} u^{(z)}_{t,j} - \sum_{z:z_j=0} \sum_{t:i \in S^{(z)}_{t,j}} u^{(z)}_{t,j} \right)$$

and

$$c_0^{(\nu)} = \sum_{z \in \{0,1\}^r} \left(\sum_{1 \leq t \leq s,\ 1 \leq j \leq r} (u^{(z)}_{t,j} \nu^{(z)}_{t,j} + |u^{(z)}_{t,j}| \frac{\epsilon^3 n}{8^r s}) \right)$$
$$- \sum_{1 \leq j \leq r} \sum_{z:z_j=0} \sum_{1 \leq t \leq s} u^{(z)}_{t,j}$$

we get that

$$\sum_{i=1}^{n} c_i^{(\nu)} x_i \leq c_0^{(\nu)} \text{ has no solution } x \text{ with } 0 \leq x_i \leq 1 \tag{21}$$

$$\text{which is equivalent to } \sum_{i=1}^{n} \text{Min}(c_i^{(\nu)}, 0) > c_0^{(\nu)}. \tag{22}$$

Let J be a random subset of V of cardinality $q = \Omega\left(\frac{\log(1/\epsilon)}{\epsilon^{12}}\right)$. Let $\gamma^{(\nu)} = \sum_{z \in \{0,1\}^r} \sum_{1 \leq t \leq s} |u^{(z)}_{t,j}|$. Noting that $|c_i^{(\nu)}| \leq \gamma^{(\nu)}$, we have from (22), using the Theorems of Hoeffding [11],

$$\mathbf{Pr}\left(\sum_{i \in J} \text{Min}(c_i^{(\nu)}, 0) \leq \frac{q}{n} c_0^{(\nu)} - \frac{2\epsilon^3 q}{8^r s n} \gamma^{(\nu)} \right) \leq \exp\left(-\frac{2\epsilon^6 q}{8^{2r} s^2} \right)$$

which implies that the following Linear Program $[LP(J,\nu)]$ on the variables $x_i, i \in J$ is unfeasible :

$$\frac{q}{n}\left(\nu^{(z)}_{t,j} - \frac{\epsilon^3 n}{8^r s} \right) \leq \sum_{i \in S^{(z)}_{t,j} \cap J} x_i \leq \frac{q}{n}\left(\nu^{(z)}_{t,j} + \frac{\epsilon^3 n}{8^r s} \right)$$
$$\text{for all } (z,t,j) \text{ with } z_j = 1$$
$$\frac{q}{n}\left(\nu^{(z)}_{t,j} - \frac{\epsilon^3 n}{8^r s} \right) \leq \sum_{i \in S^{(z)}_{t,j} \cap J} (1 - x_i) \leq \frac{q}{n}\left(\nu^{(z)}_{t,j} + \frac{\epsilon^3 n}{8^r s} \right)$$
$$\text{for all } (z,t,j) \text{ with } z_j = 0$$
$$0 \leq x_i \leq 1\ \forall i \in J \qquad [LP(J,\nu)]$$

Let $\alpha = \exp\left(-\frac{2\epsilon^{10} q}{(32)^{2r}} \right)$. To summarize, we have that for any ν,

LP(V,ν) is not feasible implies that LP(J,ν) is not feasible with probability at least $1-\alpha$.

This is of course the same as

$$\mathbf{Pr}(\text{LP}(J,\nu) \text{ feasible}) > \alpha) \text{ implies LP}(V,\nu) \text{ feasible.}$$

This means that, again for any fixed ν, either we are guaranteed the existence of a "good" solution in V, or the probability that LP(J,ν) is feasible is very small. Now, we fix attention on the set K, say, of points with coordinates of the form $\frac{q\epsilon^3}{8^r}\lambda_{t,j}^{(z)}$ where the $\lambda_{t,j}^{(z)}$ are integers. Note that there are at most $\left(\frac{8^r}{\epsilon^3}\right)^{\frac{r8^r}{\epsilon^2}}$ such points. Thus, we can bound above the total probability of having simultaneously LP(J,ν) feasible and LP(V,ν) unfeasible on one point of K by

$$|K|\alpha = \left(\frac{8^r}{\epsilon^3}\right)^{\frac{r8^r}{\epsilon^2}} \exp\left(-\frac{2\epsilon^{10}q}{(32)^{2r}}\right)$$

which is less than 1/3 for $q = \Omega\left(\frac{\log(1/\epsilon)}{\epsilon^{12}}\right)$

For each $z \in \{0,1\}^r$, let $B^{(z)}$ be the matrix induced by $A^{(z)}$ on J^r, and let us write

$$B^{(z)} = F^{(z)} + \sum_{0\le t\le s} Cut(S_{t,1}^{(z)} \cap J, S_{t,2}^{(z)} \cap J, ...S_{t,r}^{(z)} \cap J, d_t^{(z)})$$

say. Then we have that $F^{(z)}$ is the array induced by $E^{(z)}$ on J^r.

The following theorem resembles Theorem 9. However it differs from it in that it does not require a bound for the Frobenius norm (and requires higher sampling size).

Theorem 11. *Suppose G is a r-dimensional array on $V^r = V \times V \times ...V$ with all entries of absolute value at most M. Suppose J is a random subset of V of cardinality $q \ge 5000r^7/\epsilon^8$. Let B be the r-dimensional array obtained by restricting G to J^r. Then we have, with probability at least $1 - 1/(4.2^r)$,*

$$||B||_C \le \frac{q^r}{|V|^r}||G||_C + 5\epsilon^2 q^r M(3 + 4^r/\epsilon).$$

Proof The proof of Theorem 11 mimics the proof of Theorem 9 and we give only a sketch. There are two differences. First we use the trivial upper bound $|V|^{r/2}M(1+4^r/\epsilon)$ for the Frobenius norm of B. Also, we increase the value of p in Lemma 7 by a factor $\Omega(1/\epsilon^2)$ so as to get the assertion of Lemma 2 with ϵ^2 in place of ϵ and with probability at least $1 - 1/(4.2^r)$. We get then that, with probability at least $1 - 1/(3.2^r)$,

$$||B||_C \le \frac{q^r}{|V|^r}||G||_C + 10\epsilon^2 M q^r + 5\epsilon^2 q^r M(1 + 4^r/\epsilon).$$

This implies immediately the assertion of the theorem. □

We return now to the proof of Theorem 8.

Taking $G = F^{(z)}$ gives

$$||F^{(z)}||_C \le 16\epsilon 4^r q^r M$$

simultaneously for all $z \in \{0,1\}^r$ with probability at least 2/3. For $v^*(\eta)$ as already defined (we use η when referring to J, μ, ν when referring to V) and $v(\eta)$ the number of functions with variables in J satisfied by $S^{(\eta)}$ we have

$$|v(\eta) - v^*(\eta)| \leq \sum_{z \in \{0,1\}^r} ||F^{(z)}||_C \leq 16\epsilon M 8^r q^r. \tag{23}$$

Also, since $\max_{z,t} |d_t^{(z)}| \leq 2^r M$,

$$|v^*(\mu) - v^*(\nu)| \leq \frac{8^r M}{\epsilon^2} ||\mu - \nu||_{\ell 1}. \tag{24}$$

For each realizable η there is an η', say, belonging to K and for which $||\eta' - \eta||_{\ell_1} \leq \frac{q\epsilon^3}{8^r s} r 2^r s \leq \frac{\epsilon^3 qr}{4^r}$. We know that, with probability at least $2/3$, there exists simultaneously for all η' in K, a feasible ν' satisfying the inequalities of the Linear Program $[LP(J, \eta)]$ where η is replaced by η', and with

$$||\nu' - \frac{|V|}{q}\eta'||_{\ell_1} \leq \frac{\epsilon^3 |V|}{8^r s} r 2^r s = \frac{\epsilon^3 r |V|}{4^r}.$$

This implies, using (24), $|v^*(\eta') - v^*(\eta)| \leq \frac{8^r M}{\epsilon^2} ||\eta' - (\eta)||_{\ell 1} \leq \epsilon 2^r q^r M$ and, with the above inequality,

$$|v^*(\nu') - \frac{|V|^r}{q^r} v^*(\eta)| \leq \epsilon(r+1)2^r |V|^r M.$$

Now we use (23) twice to get from the above inequality,

$$|v(\nu') - \frac{|V|^r}{q^r} v(\eta)| \leq \epsilon((r+1)2^r + 32.4^r)|V|^r M,$$

which gives, after a rescaling of ϵ, both assertions of the theorem by choosing η such that $v(\eta) = m^{(J)}$.

□

This closes the proof of Theorem 1.

The refinement of the general method to the case of MAX-2CSP and the proof of Theorem 2, as well as a lower bound on sample complexity of MAX-CUT are included in Section 6.

5 Lower Bound on Number of Cut Arrays Needed

In this Section we show that the $c(r)/\epsilon^2$ upper estimate for the number of cut arrays in Theorem 5 is tight (up to the dependence on r), even if we restrict our attention to $\{-1, 1\}$-arrays A, and even if we only require that the sum of the cut arrays D will satisfy (2). Throughout the subsection we assume, whenever this is needed, that ϵ is sufficiently small as a function of r. We also omit all floor and ceiling signs whenever these are not crucial, to simplify the presentation. Note that if we only wish to satisfy (5) in Theorem 5, then its proof implies that $1/\epsilon^2$ cut arrays suffice, as the extra 4^r term appears because of the need to satisfy (4).

The L_1-norm of an array $A : V_1 \times V_2 \cdots \times V_r \mapsto R$ is given by

$$||A||_1 = \sum_{(i_1,i_2,\ldots,i_r)\in V_1\times V_2\cdots\times V_r} |A(i_1,i_2,\ldots,i_r)|.$$

The following lemma supplies a lower bound for the cut-norm of an array in terms of its L_1-norm. The proof is based on the method of [1].

Lemma 12. *Let $A : V_1 \times V_2 \cdots \times V_r \mapsto R$ be an array. Then its cut norm satisfies*

$$||A||_C \geq \frac{||A||_1}{2 \cdot 8^{(r-1)/2} \prod_{j=2}^{r} |V_j|^{1/2}}.$$

The proof (following the ideas of [1]) uses a result of Szarek. Let $c_1, c_2, \ldots, c_n$ be a set of n reals, let $\delta_1, \ldots, \delta_n$ be independent, identically distributed random variables, each distributed uniformly on $\{-1, 1\}$, and define $X = \sum_i \delta_i c_i$.

Lemma 13. (Szarek [14]) *In the above notation,*

$$E(|X|) \geq 2^{-1/2}(c_1^2 + \ldots + c_n^2)^{1/2} \quad (\geq \frac{|c_1| + \ldots + |c_n|}{\sqrt{2n}}.)$$

Corollary 14. *Let $c_1, \ldots, c_n$ be reals, and let S be a random subset of $\{1, 2, \ldots, n\}$ taken uniformly among all 2^n subsets. Let Y be the random variable $Y = \sum_{i\in S} c_i$. Then*

$$E(|Y|) = \frac{\sum_{S\subset\{1,\ldots,n\}} |\sum_{i\in S} c_i|}{2^n} \geq \frac{\sum_i |c_i|}{\sqrt{8n}}$$

Proof. For every vector $\delta = (\delta_1, \ldots, \delta_n) \in \{-1, 1\}^n$ define $S_\delta = \{i : \delta_i = 1\}$ and $S'_\delta = \{i : \delta_i = -1\}$. Then, by the triangle inequality

$$|\sum_{i\in S_\delta} c_i| + |\sum_{i\in S'_\delta} c_i| \geq |\sum_i \delta_i c_i|.$$

As δ ranges over all 2^n members of $\{-1, 1\}^n$, S_δ, as well as S'_δ range over all 2^n subsets of $\{1, 2, \ldots, n\}$ implying that $2E(|Y|) \geq E(|X|)$, where X is as above. The result now follows from Lemma 13. $\Box$

Proof of Lemma 12. We prove, by induction on t, that for every $0 \leq t \leq r$ there are subsets $S_{r-t+1} \subset V_{r-t+1} \ldots S_r \subset V_r$ such that

$$\sum_{i_1\in V_1} \cdots \sum_{i_{r-t}\in V_{r-t}} | \sum_{i_{r-t+1}\in S_{r-t+1}} \cdots \sum_{i_r\in S_r} A(i_1, i_2, \ldots, i_r)| \geq \frac{||A||_1}{8^{t/2} \prod_{j=r-t+1}^{r} |V_j|^{1/2}}. \tag{25}$$

For $t = 0$ there is nothing to prove. Assuming the assertion holds for $t-1 < r$, we prove it for t. For each $(r-t)$-tuple $i_1, i_2, \ldots, i_{r-t}$ and each $i \in V_{r-t+1}$ define

$$c_i = c_i(i_1, i_2, \ldots, i_{r-t})$$
$$= \sum_{i_{r-t+2} \in S_{r-t+2}} \cdots \sum_{i_r \in S_r} A(i_1, i_2, \ldots, i_{t-r}, i, i_{t-r+2}, \ldots, i_r),$$

and apply Corollary 14 with $n = |V_{r-t+1}|$. Summing the resulting inequalities for all $(i_1, \ldots, i_{r-t}) \in V_1 \times \cdots \times V_{r-t}$ we conclude that the average (over $S_{r-t+1} \subset V_{r-t+1}$) of the sum

$$\sum_{i_1 \in V_1} \cdots \sum_{i_{r-t} \in V_{r-t}} |\sum_{i_{r-t+1} \in S_{r-t+1}} \cdots \sum_{i_r \in S_r} A(i_1, i_2, \ldots, i_r)|$$

is at least

$$\frac{1}{\sqrt{8|V_{r-t+1}|}} \frac{||A||_1}{8^{(t-1)/2} \prod_{j=r-t+2}^{r} |V_j|^{1/2}} = \frac{||A||_1}{8^{t/2} \prod_{j=r-t+1}^{r} |V_j|^{1/2}}.$$

Therefore, there is a set $S_{r-t+1} \subset V_{r-t+1}$ for which (25) holds, showing that it indeed holds for all $t \leq r$.

In particular, for $t = r-1$ there are sets $S_2 \subset V_2, \ldots, S_r \subset V_r$ such that

$$\sum_{i_1 \in V_1} |\sum_{i_2 \in S_2} \cdots \sum_{i_r \in V_r} A(i_1, i_2, \ldots, i_r)| \geq \frac{||A||_1}{8^{(r-1)/2} \prod_{j=2}^{r} |V_j|^{1/2}}. \tag{26}$$

Fixing such sets S_i, either the contribution of the positive terms $\sum_{i_2 \in S_2} \cdots \sum_{i_r \in V_r} A(i_1, i_2, \ldots, i_r)$ gives at least half of (26), or the contribution of the absolute values of the negative terms gives at least half the sum. In each case we can define S_1 as the set of those $i_1 \in V_1$ that correspond to those contributing terms and conclude that

$$||A||_C \geq |\sum_{i_1 \in S_1} \cdots \sum_{i_r \in S_r} A(i_1, \ldots, i_r)|$$
$$\geq \frac{||A||_1}{2 \cdot 8^{(r-1)/2} \prod_{j=2}^{r} |V_j|^{1/2}}.$$

This completes the proof. □

From now on we restrict our attention in this subsection to arrays $A : V_1 \times V_2 \times \cdots \times V_r \mapsto \{-1, 1\}$ where $|V_i| = n$ for all i. We need the following simple fact.

Lemma 15. *There exists a family $\mathcal{F}$ of r-dimensional arrays, each mapping $V_1 \times V_2 \times \cdots \times V_r$, where $|V_i| = n$ for each i, into $\{-1, 1\}$ such that $|\mathcal{F}| \geq 2^{n^r/2}$ and for each two distinct members $A, B \in \mathcal{F}$, $||A - B||_1 > \frac{n^r}{5}$.*

Proof. Let $H(x) = -x \log_2 x - (1-x) \log_2(1-x)$ be the binary entropy function. By the Gilbert-Varshamov bound (see, e.g., [13]), for every (large) m there are at least $2^{(1-H(1/10))m}$ ($> 2^{m/2}$) vectors of length m over $\{-1, 1\}$, where the Hamming distance between each pair exceeds $m/10$. Taking $m = n^r$ and viewing these vectors as arrays mapping $V_1 \times \cdots \times V_r$ to $\{-1,1\}$, the desired result follows, as the difference between any two distinct arrays in the collection will have more than $n^r/10$ nonzero entries, each of which is either 2 or -2. $\Box$

We can now prove the main result of this subsection.

Theorem 16. *For every fixed dimension $r \geq 2$ there exists some $c(r) > 0$ so that for every $\epsilon > 0$ there are $n, N = n^r$ and an r-dimensional array $A : V_1 \times \cdots \times V_r \mapsto \{-1,1\}$, where $|V_i| = n$ for all i, such that for every array D which is the sum of less than $c(r)/\epsilon^2$ cut arrays,*

$$||A - D||_C > \epsilon n^r \ (\ = \epsilon\sqrt{N}||A||_F)$$

Proof. We prove the theorem for all ϵ which is sufficiently small as a function of r, and with $c(r) = \frac{1}{4r\cdot 40^2 \cdot 8^{r-1}}$. Clearly this implies the result for all ϵ (with a possibly smaller $c = c(r)$). Define

$$n = \frac{1}{8\cdot(40)^{2/(r-1)}\epsilon^{2/(r-1)}},$$

and note that $N = n^r < 1/(2\epsilon^4)$. By Lemma 15 there is a family $\mathcal{F}$ of $2^{n^r/2}$ arrays $A : V_1 \times V_2 \times \cdots \times V_r \mapsto \{-1,1\}$ such that for every two distinct members $A, B \in \mathcal{F}$, $||A - B||_1 > N/5$. By Lemma 12 this implies that for every such A, B,

$$||A - B||_C \geq \frac{||A - B||_1}{2\cdot 8^{(r-1)/2} n^{(r-1)/2}} > \frac{n^{(r+1)/2}}{10\cdot 8^{(r-1)/2}} = 4\epsilon n^r, \tag{27}$$

where the last equality follows from the definition of n.

Therefore, $\mathcal{F}$ is a large set of arrays, so that the cut-distance between any pair of them is large. To complete the proof we show that at least one member of $\mathcal{F}$ cannot be approximated well (in the cut metric) by a sum of a small number of cut arrays. To do so, suppose that for each member A of $\mathcal{F}$ there is an array D which is a sum of at most t cut arrays, such that $||A - D||_C \leq \epsilon n^r$. Call a cut-array ϵ-nice if it is an array of the form $CUT(S_1, S_2, \ldots, S_r; d)$ where d is an integral multiple of ϵ/t. An obvious rounding procedure implies that for each member of $\mathcal{F}$ there is an array D which is the sum of at most t ϵ-nice cut arrays, such that $||A - D||_C < 2\epsilon n^r$.

We next prove an upper bound for the total possible number of such arrays D. Note, first, that as $n^r < 1/(2\epsilon^4)$, the absolute value of no entry of such a D can exceed $1 + 1/\epsilon^3 < 2/\epsilon^3$ (since otherwise the cut-norm of $A - D$ would exceed $2\epsilon n^r$ simply by considering a single entry). As each entry of D is also an integral multiple of ϵ/t it follows that there are at most $4t/\epsilon^4$ possibilities for each such entry. There are at most 2^{nr} possibilities for choosing the sets $S_1, \ldots, S_r$ in each cut array $CUT(S_1, \ldots, S_r; d)$, and as D is the sum of t such arrays there are at most 2^{nrt} possibilities for choosing the defining sets of all of them. Once these are chosen, we have to choose the densities d of these arrays. Each of those is an integral multiple of ϵ/t, but the trouble is that its absolute value may be large (as there may be cancellations between them, while forming D). It is thus better to bound the number of possibilities of all these densities as follows. Let $d_1, \ldots, d_t$ be the densities. Since we have already chosen all sets S_i in all the cut arrays whose sum is D, we can express each entry of

D as a sum of a subset of the densities d_i. At most t of the characteristic vectors of these subsets span all the characteristic vectors of all other subsets we have, and thus if we are given the values of D in these entries, we can solve for all other entries of D. There are at most n^{rt} ways to choose t entries of D, and then there are at most $(4t/\epsilon)^t$ possibilities for the values of D in these entries (as each entry is an integral multiple of ϵ/t whose absolute value does not exceed $2/\epsilon^3$.) Therefore, the total number of possible arrays D is at most

$$n^{rt}(\frac{4t}{\epsilon})^t 2^{nrt}.$$

Each member of $\mathcal{F}$ is within cut-distance smaller than $2\epsilon n^r$ from at least one of these arrays D, and the cut-distance between any two distinct members of $\mathcal{F}$ exceeds $4\epsilon n^r$, by (27). It thus follows that the number of arrays D is at least as large as $\mathcal{F}$, implying that

$$\log|\mathcal{F}| = \frac{n^r}{2} \le rt\log n + t\log(4t/\epsilon) + nrt < 2trn,$$

where here we used the fact that n is much bigger than $\log n + \log(4t/\epsilon)$. The last inequality implies that

$$t \ge \frac{n^{r-1}}{4r} = \frac{1}{4r \cdot 40^2 \cdot 8^{r-1}\epsilon^2},$$

completing the proof. □

6 MAX-CUT and MAX-2CSP

6.1 Cut norm of random submatrices

In this section, we will prove that the cut decomposition obtained for arrays in the last section also holds with high probability for a submatrix of A "induced" by a random subset J of $\{1, 2, \ldots n\}$ of size $\Omega(1/\epsilon^4)$. [A "random" subset of size q is obtained by taking q independent samples with replacement.] More precisely,

Theorem 17. *Suppose J is a random subset of $\{1, 2, \ldots n\}$ of cardinality $q \ge 10^8/\epsilon^4$ and let B be the $q \times q$ submatrix of W in the rows and columns in J. Then, with probability at least 9/10, we have*

$$||B||_C \le 41\epsilon q^2.$$

The use of this theorem for the MAX-CUT problem will become clear later. But we motivate its use with the solution of another problem :

Definition 1 *The* **maximum submatrix sum** *of A (denoted $MSS(A)$) is*

$$\max_{S,T\subseteq\{1,2,\ldots n\}} A(S,T).$$

Corollary 18. *If A satisfies $|A_{ij}| \leq 1$, $q \geq 10^8/\epsilon^4$ is a positive integer, J is a random subset of $\{1, 2, \ldots n\}$ of cardinality q, and F is the $q \times q$ submatrix of A in the J rows and columns, then with probability at least 9/10, we have*

$$|MSS(A) - \frac{n^2}{q^2} MSS(F)| \leq 4\epsilon n^2.$$

We will actually prove a more general result (Theorem 19) from which Theorem 17 will follow.

Theorem 19. *Suppose G is a $n \times n$ matrix with rows and columns indexed by $\{1, 2, \ldots n\}$. Suppose $|G_{ij}| \leq M$ for all i, j. Suppose ϵ is a real number satisfying*

$$\frac{||G||_F}{\sqrt{2}nM} \leq \epsilon \leq 10 \qquad ; \qquad \frac{10^3}{n^{1/8}} \leq \epsilon.$$

Let q be any positive integer which is at least $10^8/\epsilon^4$. Then for a random subset J of $\{1, 2, \ldots n\}$, with $|J| = q$, the submatrix B of G consisting of its rows and columns in J satisfies both the following inequalities with probability at least 9/10 :

$$MSS(B) \leq 4\epsilon^2 M q^2 + MSS(G)\frac{q^2}{n^2} + 8\epsilon q^2 \frac{||G||_F}{n}.$$

Remark 1. It is not difficult to see from the above theorem that in fact if A satisfies $|A_{ij}| \leq 1$, and F is the submatrix of A induced on J, then

$$|MSS(A) - \frac{n^2}{q^2} MSS(F)| = O(\epsilon n^2).$$

[What remains to be proved is that $MSS(F)$ is not too small - this part is easy - if S, T realize the MSS of A, then w.h.p., $S \cap J$, $T \cap J$ give a fairly high submatrix sum too, just by standard sampling arguments which we omit here.]

Thus, at least for the MSS problem, these methods say that a random subset of size $\Omega(1/\epsilon^4)$ does the job. The proof for the maximum cut is more difficult, primarily because, whereas in the MSS problem, the choice of the set of rows and columns are both open (to us) and not interlinked, in the maximum cut problem, they have to be complementary sets. These difficulties are tackled in the next section via Linear Programming techniques. **Proof.** First we have that $E(||B||_F^2) = \frac{q^2}{n^2}||G||_F^2$, so using Markov inequality, we have that

$$E_1 : ||B||_F \leq 4q\frac{||G||_F}{n} \text{ has } \mathbf{Pr}(E_1) \geq 9/10. \tag{28}$$

Let $p = 1600/\epsilon^2$. We pick random subsets U, Q of J, each of cardinality p, independently of each other. [There are auxillary random sets which we use for the proof.] We let J' denote $J \setminus (U \cup Q)$. We will make crucial use of the fact that the following two different methods of picking J', U, Q produce nearly the same joint probability distribution on them :

(i) As above, pick J to be a random subset of V of cardinality q, and then pick Q, U to be independent random subsets of J of cardinality p each. Let $P^{(i)}(J, Q, U)$ be the probability that we pick J, Q, U in this experiment. Then, clearly, for each J, Q, U with $|J| = q, Q, U \subseteq J, |U| = |Q| = p$, we have $P^{(i)}(J, Q, U) = 1/\binom{n}{q}\binom{q}{p}\binom{q}{p}$.

(ii) Now, pick J' to be a random subset of V of cardinality $q-2p$. Then pick independently (of J' and of each other) two random subsets $\tilde{Q}, \tilde{U}$ of V of cardinality p each. Let $\tilde{J} = J' \cup \tilde{Q} \cup \tilde{U}$. Let $P^{(ii)}(\tilde{J}, \tilde{Q}, \tilde{U})$ be the probabilities here.

It is easy to see by direct calculation that conditioned on the events $E_2 : Q \cap U = \emptyset$ and $E_3 : (\tilde{Q} \cap \tilde{U} = \emptyset) \wedge (J' \cap (\tilde{Q} \cup \tilde{U}) = \emptyset)$, $P^{(i)}$ and $P^{(ii)}$ are exactly equal. It is also easy to see that $P^{(i)}(E_2) = \frac{(q-p)!(q-p)!}{q!(q-2p)!} \geq 99/100$ and $P^{(ii)}(E_3) \geq 99/100$; so we have that

$$||P^{(i)} - P^{(ii)}||_{\mathrm{TV}} \leq 1/50. \tag{29}$$

We will use the above facts later. Now, for any subset Y of the rows of G, we define $P(Y)$ to be the set of (indices of) columns of G whose sum in the Y rows is nonnegative

$$P(Y) = \{j \in V : \sum_{i \in Y} G_{ij} \geq 0\}.$$

Similarly, for subset X of the columns of G, we define $P(X)$ as the set of rows which have nonnegative sum in the X columns. Suppose $S_0, T_0 \subseteq J$ satisfy

$$G(S_0, T_0) = \max_{S,T \subseteq J} G(S, T).$$

So, we may assume that $T_0 = P(S_0) \cap J$.

Lemma 20. *The probability of the event $E(Q, U)$ defined below is at least 19/20 :*

$$E(Q, U) : \exists Q' \subseteq Q, U' \subseteq U : G(P(Q') \cap J, P(U') \cap J) \geq G(S_0, T_0) - 8\epsilon q^2 \frac{||G||_F}{n}. \tag{30}$$

Proof. Let $Z = P(S_0) \cap J$ and $Z' = P(U \cap S_0) \cap J$. We have,

$$G(S_0, Z') = G(S_0, Z) - G(S_0, B_1) + G(S_0, B_2), \tag{31}$$

where

$$\begin{aligned} B_1 &= \{z \in J : G(S_0, z) > 0 \text{ and } G(U \cap S_0, z) < 0\}, \\ B_2 &= \{z \in J : G(S_0, z) < 0 \text{ and } G(U \cap S_0, z) > 0\}. \end{aligned}$$

Consider one fixed z. Let $X_z = G(U \cap S_0, z)$. We may write X_z as the sum $X_1 + X_2 + \ldots X_p$, where $X_1, X_2, \ldots X_p$ is a sample of size p drawn uniformly without replacement from the set of q reals - $\{G(y, z)\underline{1}_{y \in S_0}\}$. For analysis, we also introduce the random variables $Y_1, Y_2, \ldots Y_p$ - a sample of size p drawn independently, each uniformly distributed over the same set of reals, but now with replacement. We have

$$E(X_1 + X_2 + \ldots X_p) = \frac{p}{q} G(S_0, z),$$

$$\mathrm{Var}(X_1 + X_2 + \ldots X_p) \leq Var(Y_1 + Y_2 + \ldots Y_p) \leq \frac{p}{q} \sum_{u \in S_0} G(u, z)^2 \leq \frac{p}{q} \sum_{u \in J} G(u, z)^2,$$

where the second line is a standard inequality (for example, it follows from Theorem 4 of [11]). Hence, for any $\xi > 0$,

$$\mathbf{Pr}\left(\left|X_z - \frac{p}{q}G(S_0, z)\right| \geq \xi\right) \leq \frac{p\sum_{u\in J} G(u, z)^2}{q\xi^2} \tag{32}$$

If $z \in B_1$ then $X_z - (p/q)G(S_0, z) \leq -(p/q)G(S_0, z)$ and so applying (11) with $\xi = pG(S_0, z)/q$ we get that for each fixed z,

$$\mathbf{Pr}(z \in B_1) \leq \frac{q\sum_{u\in J} G(u, z)^2}{pG(S_0, z)^2}.$$

$$\begin{aligned}
\mathbf{E}\left(\sum_{z\in B_1} G(S_0, z)\right) &\leq \sum_{\{z\in J:\, G(S_0,z)>0\}} \min\left\{G(S_0, z), \frac{q\sum_{u\in J} G(u, z)^2}{pG(S_0, z)}\right\} \\
&\leq \sum_{\{z\in J:\, G(S_0,z)>0\}} \sqrt{\frac{q\sum_{u\in J} G(u, z)^2}{p}} \qquad (33)
\end{aligned}$$

By an identical argument we obtain

$$\mathbf{E}\left(\sum_{z\in B_2} G(S_0, z)\right) \geq -\sum_{\{z\in J:\, G(S_0,z)<0\}} \sqrt{\frac{q\sum_{u\in J} G(u, z)^2}{p}}.$$

Hence, (using the Cauchy-Schwartz inequality),

$$\mathbf{E}(G(S_0, Z')) \geq G(S_0, Z) - \sum_{z\in J}\sqrt{\frac{q\sum_{u\in J} G(u, z)^2}{p}} \geq G(S_0, Z) - \frac{q}{\sqrt{p}}||B||_F.$$

Now, $G(S_0, Z) - G(S_0, Z')$ is a nonnegative random variable with expectation at most $q||B||_F/\sqrt{p}$, as argued above. So using Markov inequality, we have

$$\mathbf{Pr}\left(G(S_0, P(S_0)\cap J) - G(S_0, P(U\cap S_0)\cap J) \geq \epsilon q||B||_F\right) \leq \frac{1}{40}. \tag{34}$$

Now, by definition of P, we have

$$G(P(P(U\cap S_0)\cap J)\cap J, P(U\cap S_0)\cap J) \geq G(S_0, P(U\cap S_0)\cap J). \tag{35}$$

By an exactly symmetric argument now applied to the random choice of columns Q, we get analogously to (34), that with probability at least 39/40, the following holds :

$$\begin{aligned}
G(P(P(U\cap S_0)\cap J)\cap J, P(U\cap S_0)\cap J) &- G(P(P(U\cap S_0)\cap J\cap Q)\cap J, P(U\cap S_0)\cap J) \\
&\leq \epsilon q||B||_F. \qquad (36)
\end{aligned}$$

By applying (34),(35) and (36) and using (29), the lemma follows with $Q' = P(U\cap S_0)\cap Q$ and $P' = U\cap S_0$.

□

The lemma below is a particular "large-deviations" result. While the proof is standard, it differs from the usual ones in its hypothesis which upper bound each real as well as the sum of squares. [We note that if we did not have the upper bound on the sum of squares, the upper bound one usually gets on the probability in the lemma depends on γ^2 rather than γ.]

Lemma 21. *Suppose $a_1, a_2, \ldots a_r$ are any reals with $|a_i| \leq M$ for all i and $\sum_{i=1}^r a_i^2 \leq Nr$. Let $X_1, X_2, \ldots X_q$ be a sample of size q picked by sampling uniformly without replacement from the set $\{a_1, a_2, \ldots a_r\}$. Then, for any real $\gamma \in [\frac{2N}{M^2}\ ,\ 100]$, we have :*

$$\mathbf{Pr}\left(\left|\sum_{t=1}^q X_t - \frac{q}{r}\sum_i a_i\right| \geq \gamma M q\right) \leq 2e^{-\gamma q/4}.$$

Proof. Let λ be a positive real to be chosen later. Let $\bar{a} = \frac{1}{r}\sum_{i=1}^r a_i$ and $b_i = a_i - \bar{a}$ and let $Y_1, Y_2, \ldots Y_q$ be a sample of size q drawn with replacement from the same set of reals - $\{a_1, a_2, \ldots a_r\}$. (To be used just in the proof.) Let $\Lambda = \gamma M q$. □

$$\mathbf{Pr}\left(\sum_{t=1}^q X_t \geq q\bar{a} + \Lambda\right) \leq E\left(e^{\lambda \sum_t X_t}\right) e^{-\lambda q \bar{a}} e^{-\lambda\Lambda} \leq E\left(e^{\lambda \sum_t Y_t}\right) e^{-\lambda q \bar{a}} e^{-\lambda\Lambda}$$

the last since e^x is a convex function - from Theorem 4 of [11]

$$\leq \left(E\left(e^{\lambda(Y_1 - \bar{a})}\right)\right)^q e^{-\lambda\Lambda}. = \frac{1}{r^q}\left(\sum_{i=1}^r e^{\lambda b_i}\right)^q e^{-\lambda\Lambda}.$$

The b_i satisfy the constraints $\sum_i b_i^2 \leq \sum_i a_i^2 \leq Nr$ and $|b_i| \leq 2M$. The maximum of the last expression subject to these two constraints is attained when $r_0 = \text{Min}(r, \frac{Nr}{4M^2})$ of the b_i 's are $2M$ each and the rest are zero. Thus, we have, by choosing $\lambda = 1/(4M)$ in the above,

$$\mathbf{Pr}\left(\sum_{t=1}^q X_t \geq q\bar{a} + \Lambda\right) \leq \frac{1}{r^q}[r_0 e^{1/2} + r - r_0]^q e^{-\Lambda/(4M)} \leq (1 + \frac{N}{4M^2})^q e^{-\Lambda/(4M)} \leq e^{-\gamma q/4},$$

using $(1 + (N/4M^2) \leq e^{N/4M^2}$. This bounds the probability of $\sum X_t$ being too large. To bound the probability of this sum being too negative, we just use the same argument with the set of a_i replaced by the set of $-a_i$. This then yields the lemma. □

Now we go back to the proof of Theorem 19. Let $q' = (q - 2p)/2$. Let L be a random set of q' pairs (i, j) picked by sampling without replacement from the set of n^2 pairs in $V \times V$. Under the assumption that $n \geq 100q'^2$, we have that with probability at least 49/50, no two pairs in L share an endpoint. We assume this from now on.

For the moment, fix attention on one particular $Q' \subseteq Q$ and one particular $U' \subseteq U$. Let $X = P(Q')$ and $Y = P(U')$. By Lemma 21, the event E_{10} defined below has the claimed probability bound (by putting $r = n^2, q = q'; N = ||G||_F^2/n^2$ in that lemma)

$$E_{10}(L) : \left|G(X,Y) - \frac{n^2}{q'}\sum_{(i,j)\in L\cap X\times Y} G_{i,j}\right| \leq \epsilon^2 M n^2 \text{ has } \mathbf{Pr}(E_{10}) \geq 1 - 2e^{-\epsilon^2 q'/4}. \quad (37)$$

Now imagine having picked already Q, U and having fixed this $Q' \subseteq Q$ and $U' \subseteq U$. Now we pick J', L independently of Q, U. We will say that L belongs to J' if the at most $2q'$ end points of pairs in L all belong to J'. It is easy to check by direct counting that

$$E_{10}(L) \text{ holds for all } L \text{ belonging to } J' \text{ implies} \tag{38}$$

$$E_{11}(J') \text{ holds, where } E_{11}(J') : \left| G(X,Y) - \frac{n^2}{q'^2} G(J' \cap X, J' \cap Y) \right| \leq 2\epsilon^2 M n^2. \tag{39}$$

It is clear that each L belongs to exactly one J'. So, we have :

$$|\{J' : \neg E_{11}(J')\}| \leq |\{L : \neg E_{10}(L)\}| \leq \mathbf{Pr}(\neg E_{10}) \binom{n^2}{q'} \leq 2e^{-q'\epsilon^2/4} \binom{n^2}{q'}.$$

The total number of J' 's is $\binom{n}{2q'}$ which under the assumption that $n \geq 100 q'^2$, is at least $1/2$ of $\binom{n^2}{q'}$. Thus, we have that

$$\mathbf{Pr}(E_{11}) \geq 1 - 4e^{-q'\epsilon^2/4}.$$

Let now

$$E_{12}(J') : \forall Q' \subseteq Q, \forall U' \subseteq U : |G(P(Q'), P(U')) - \frac{n^2}{q'^2} G(J' \cap P(Q'), J' \cap P(U'))| \leq 2\epsilon^2 n^2.$$

Then, clearly, the event $\neg E_{12}$ is the union of at most 2^{2p} events $\neg(E_{11}$ - one for each $Q' \subseteq Q; U' \subseteq U$. So, by the above, $\mathbf{Pr}(E_{12}) \geq 1 - 2^p \mathbf{Pr}(\neg E_{11}$ for one fixed Q', U'). Thus we get

$$\mathbf{Pr}(E_{12}) \geq 19/20.$$

But under $E_{12}(J')$, we have

$$\forall Q', \forall U', G(J' \cap P(Q'), J' \cap P(U')) \leq 2\epsilon^2 M q'^2 + ||G||_C q^2/n^2.$$

which implies that (since $|G_{i,j}| \leq M$),

$$\forall Q', \forall U', G(J \cap P(Q'), J \cap P(U')) \leq 4\epsilon^2 M q^2 + ||G||_C \frac{q^2}{n^2}.$$

This along with Lemma 20 implies that $G(S_0, T_0) \leq 4\epsilon^2 M q^2 + ||G||_C \frac{q^2}{n^2} + 8\sqrt{\gamma} q^2 \frac{||G||_F}{n}$. with probability at least 9/10 proving the theorem.

□

Proof of Theorem 17: We apply Theorem 19 to W as follows : we take $G = W$. We see that by Theorem (5), $|G_{ij}| \leq \sqrt{s}(2||A||_F)/n \leq 8/\epsilon$. We may thus take $M = 8/\epsilon$ in Theorem 19. Then we get that $\max_{S,T \subseteq J} B(S,T) \leq 41\epsilon q^2$ holds with probability at least 9/10. Applying the same argument to $-W$ also, we see that $\max_{S,T \subseteq J} -B(S,T) \leq 41\epsilon q^2$ holds with probability at least 9/10. So, Theorem 17 follows.

□

6.2 Linear Programming Formulation for the Max-Cut Problem

It is easy to see that the value of any cut $(S, \bar{S})$ in G, is determined to within ϵn^2 by just the quanities $|S_t \cap S|, t = 1, 2, \ldots s$ and $|T_t \cap \bar{S}|, t = 1, 2, \ldots, s$, namely :

$$D(S, \bar{S}) = \sum_{t=1}^{s} d_t |S_t \cap S||T_t \cap \bar{S}| \qquad \text{and}$$

$$|A(S, \bar{S}) - D(S, \bar{S})| \leq \epsilon n^2.$$

Noting that $\sum_t |d_t| \leq 2\sqrt{s}$, it is possible to show that if we approximate each $|S_t \cap S|$ and $|T_t \cap \bar{S}|$ to within an error of plus or minus $\epsilon n/s$, then we can get the value of $D(S, \bar{S})$ to an error of plus or minus ϵn^2. This then leads us to an algorithm : enumerate all $O(s/\epsilon)^{2s}$ candidate sets of values for $\{|S_t \cap S|, |T_t \cap \bar{S}| : t = 1, 2, \ldots s\}$, check which ones are realizable and then take the one that attains the best value of $\sum_{t=1}^{s} d_t |S_t \cap S||T_t \cap \bar{S}|$. This was the strategy adopted in [8].

Here we will give a finer (self-contained) analysis wherein we essentially enumerate the values of $|S_t \cap S|$, $|T_t \cap S|$ to within plus or minus ϵn (rather than $\epsilon n/s$.)

We represent a set S by its characteristic vector - $x \in \{0, 1\}^n$. Then it is easy to see that the value of the cut $(S, \bar{S})$ is $\sum_{i,j} x_i(1 - x_j)A_{ij}$. Also, $|S \cap S_t| = \sum_{i \in S_t} x_i$, and $|\bar{S} \cap T_t| = \sum_{i \in T_t}(1 - x_i)$. For any $n-$ vector x, we define

$$a(x) = \left(\sum_{i \in S_1} x_i, \sum_{i \in S_2} x_i, \ldots \sum_{i \in S_s} x_i, \sum_{i \in T_1}(1 - x_i), \sum_{i \in T_2}(1 - x_i), \right.$$
$$\left. \ldots \sum_{i \in T_s}(1 - x_i) \right).$$

Let $\mathcal{A} = \{a \in \mathbf{R}^{2s} : \frac{a_i}{\epsilon n} \in \{0, 1, \ldots \lfloor (1/\epsilon) \rfloor\} \text{ for } i = 1, 2, \ldots 2s\}.$

For each vector a, define $P(a) = \{b \in \mathbf{R}^{2s} : |b_i - a_i| \leq 2\epsilon n\}$. We say that $a \in \mathcal{A}$ is realizable if there is a 0-1 vector x with $|a(x) - a|_\infty \leq 2\epsilon n$; in that case, we will say that x **realizes** a. If a is not realizable, then the following Linear Program - LP(a) is unfeasible :

$$[LP(a)] a_t - 2\epsilon n + 4s \leq \sum_{i \in S_t} x_i \leq a_t + 2\epsilon n - 4s \tag{40}$$

$$a_{t+s} - 2\epsilon n + 4s \leq \sum_{i \in T_t}(1 - x_i) \leq a_{t+s} + 2\epsilon n - 4s \text{ for } t =$$

$$\text{for } t = 1, 2, \ldots s, \quad 0 \leq x_i \leq 1$$

[This is because if LP(a) was feasible, then it would have a basic feasible solution which would have at most $4s$ fractional components; setting the fractional x_i to zero will yield a 0-1 vector weakly realizing a.] So, by Linear Programming duality, we see that there exists one inequality obtained as a nonnegative combination of the first $4s$ inequalities of (LP(a)) for which there is no solution x satisfying the bounds $0 \leq x_i \leq 1$. It is easy to see that the combination need not involve both the upper bound and the lower bound on any one S_t or T_t. So, we get that there are $2s$ real numbers $u_1^{(a)}, u_2^{(a)}, \ldots u_{2s}^{(a)}$ such that

$$\text{Letting } c_i^{(a)} = \sum_{t: i \in S_t} u_t^{(a)} - \sum_{t: i \in T_t} u_{t+s}^{(a)} \qquad \text{and}$$

$$c_0^{(a)} = \sum_{t=1}^{2s} u_t^{(a)} a_t + \sum_{t=1}^{2s} |u_t^{(a)}|(2\epsilon n - 4s) - \sum_{t=s+1}^{2s} u_t^{(a)} \tag{41}$$

$$\sum_{i=1}^{n} c_i^{(a)} x_i \leq c_0^{(a)} \text{ has no soultion } x \text{ with } 0 \leq x_i \leq 1 \tag{42}$$

$$\text{which is equivalent to } \sum_{i=1}^{n} \operatorname{Min}(c_i^{(a)}, 0) > c_0^{(a)}. \tag{43}$$

Now, for the rest of this paper, J is a random subset of V of cardinality q, which satisfies :

$$q \geq 10^8 \log(2/\epsilon)/\epsilon^4 \tag{44}$$

Noting that $|c_i^{(a)}| \leq \sum_{t=1}^{2s} |u_t^{(a)}|$, we have from (43), using the Theorems of Hoeffding [11],

$$\mathbf{Pr}\left(\sum_{i \in J} \operatorname{Min}(c_i^{(a)}, 0) \leq \frac{q}{n} c_0^{(a)} - \frac{\epsilon q}{2} [\sum_{t=1}^{2s} |u_t^{(a)}|]\right)$$

$$\leq \left(\frac{2}{\epsilon}\right)^{-2s} /100,$$

which implies that the following Linear Program on the variables $x_i, i \in J$ is infeasible :

$$\frac{q}{n}(a_t - \frac{5}{4}\epsilon n) \leq \sum_{i \in S_t \cap J} x_i \leq \frac{q}{n}(a_t + \frac{5}{4}\epsilon n)$$

$$\frac{q}{n}(a_{t+s} - \frac{5}{4}\epsilon n) \leq \sum_{i \in T_t \cap J} (1 - x_i) \leq \frac{q}{n}(a_{t+s} + \frac{5}{4}\epsilon n) \text{ for } t =$$

$$1, 2, \ldots s \qquad [LP(J, a)] 0 \leq x_i \leq 1 \forall i \in J.$$

So we have the following fact which will be used later :

$$\mathbf{Pr}\,(\exists \text{an unrealizable } a \text{ with } LP(J, a) \text{ feasible })$$

$$\leq \left(\frac{2}{\epsilon}\right)^{2s} \left(\frac{2}{\epsilon}\right)^{-2s} /100 \leq \frac{1}{100}. \tag{45}$$

Now, we will deal with the realizable a and show that for these, the max-cut in J is not much greater than the max-cut in G.

Lemma 22. *For any a, define :*

$$c_0(a) = \sum_{i \notin E(a)} \sum_{t: i \in T_t} d_t a_t \qquad c_i(a) = \sum_{t: i \in S_t} d_t a_{t+s} - \sum_{t: i \in T_t} d_t a_t \text{ for } i1 =, 2, \ldots n. \tag{46}$$

Then for any realizable a and any $x \in \mathbf{R}^n$ with $0 \le x_i \le 1$, we have

$$\sum_{i=1}^{n} c_i(a)^2 \le 2500 n^3 \tag{47}$$

$$\left| \sum_{i=1}^{n} c_i(a) x_i + \sum_{t=1}^{s} d_t a_t a_{t+s} + c_0(a) - \sum_{i,j} x_i (1 - x_j) A_{ij} \right| \le \frac{8}{\epsilon} |a - a(x)|_\infty^2 + \epsilon n^2, \tag{48}$$

$$\text{and, w.h.p.,} \quad \left| \sum_{i \in J} \frac{q}{n} c_i(a) x_i + \frac{q}{n} c_0(a) + \frac{q^2}{n^2} \sum_{t=1}^{s} d_t a_t a_{t+s} - \sum_{i,j \in J} x_i (1 - x_j) A_{ij} \right| \le \frac{8q^2}{\epsilon n^2} |a - a(x)|_\infty^2 + 20 \epsilon q^2. \tag{49}$$

$$\tag{50}$$

Proof. Fix attention on a particular a, x. Let $\Delta = a(x) - a$.

$$\begin{aligned}
D(x, \bar{x}) &= \sum_{t=1}^{s} d_t |x \cap S_t| |\bar{x} \cap T_t| \\
&= \sum_{t=1}^{s} d_t (a_t + \Delta_t)(a_{t+s} + \Delta_{t+s}) \\
&= \sum_{t=1}^{s} d_t a_t a_{t+s} + \sum_{t=1}^{s} d_t a_t \Delta_{t+s} + \sum_{t=1}^{s} d_t \Delta_t a_{t+s} + \delta_1, \\
&\qquad \text{where} \quad |\delta_1| \le \sum_t |d_t| |a - a(x)|_i nfty^2 \\
&= \sum_{t=1}^{s} d_t (a_t + \Delta_t) a_{t+s} + \sum_{t=1}^{s} d_t a_t (\sum_{i \in T_t} (1 - x_i) - a_{t+s}) + \delta_1 \\
&= \sum_{i=1}^{n} \left(x_i \sum_{t: i \in S_t} d_t a_{t+s} \right) + \sum_{i=1}^{n} \left((1 - x_i) \sum_{t: i \in T_t} d_t a_t \right) - \sum_{t=1}^{s} d_t a_t a_{t+s} + \delta_1.
\end{aligned} \tag{51}$$

Then, we see that from (51),

$$A(x,\bar{x}) = D(x,\bar{x}) + \delta_2 = \sum_{i=1}^{n} c_i(a)x_i + c_0(a)$$

$$+\sum_{t=1}^{s} d_t a_t a_{t+s} + \delta_1 + \delta_2,$$

where $\delta_2 \le \epsilon n^2$. This proves the second statement of the lemma. The third statement follows by exactly analogous argument, noting that the cut decomposition of A also holds for A restricted to J as proved in the last section.

The upper bound on the sum of all the $c_i(a)^2$ is proved as follows : since a is realizable, there is some $x \in \{0,1\}^n$ such that $|a - a(x)|_\infty \le 2\epsilon n$. Let l be any natural number between 1 and n and consider the l largest $c_i(a)$ among $i \notin x$. Adding these i to x (and dropping them from $\bar{x}$) changes the value of the cut by at most ln. But by part (ii) of the lemma, the value of the cut changes by at least the sum of these $c_i(a)$ minus $20\epsilon n^2$; so we have that the sum of these $c_i(a)$ is at most $ln + 20\epsilon n^2$. Similarly, the sum of the largest l of the $c_i(a)$ among $i \in x$ is at most $ln + 20\epsilon n^2$. Thus the sum of the largest l $c_i(a)$ overall is at most $2ln + 40\epsilon n^2$. Similarly, the sum of the smallest l of the $c_i(a)$ is at least $-2ln - 40\epsilon n^2$. For the moment, renumber the i such that $|c_1(a)| \ge |c_2(a)| \ldots |c_n(a)|$. Then, by the above, we have (the upper bound on individual $|c_l(a)|$ follows from their definition)

$$\sum_{i=1}^{l} |c_i(a)| \le 2ln + 40\epsilon n^2; \quad |c_l(a)| \le n \sum_t |d_t| \le \frac{8n}{\epsilon}$$

$$\text{for } l = 1, 2, \ldots n.$$

It is easy to see that under these constraints, the maximum value of $c_l(a)^2$ is attained when the first $\frac{40\epsilon n}{[(8/\epsilon)-2]}$ of the $|c_l(a)|$ are $8n/\epsilon$ each and the rest of the $|c_l(a)|$ are $2n$ thus proving part 1 of the lemma.

□

Let v be the value of the max-cut in the graph G. For each $a \in \mathcal{A}$, define a linear program

$$[LP(a)]\text{Max} \sum_i c_i(a)x_i + c_0(a)$$

$$a_t - 2\epsilon n + 4s \le \sum_{i \in S_t} x_i \le a_t + 2\epsilon n + 4s$$

$$a_{t+s} - 2\epsilon n + 4s \le$$

$$\sum_{i \in T_t} (1 - x_i) \le a_{t+s} + 2\epsilon n - 4s \text{ for } t = 1, 2, \ldots s \tag{52}$$

$$0 \le x_i \le 1.$$

The maximum value of this LP (if feasible) is at most $v + 1000\epsilon n^2 - \sum_t d_t a_t a_{t+s}$. Then by Linear Programming duality, there exist real numbers $u_1^{(a)}, u_1^{(a)}, \ldots u_{2s}^{(a)}$ such that with $c_i^{(a)}$

defined as in (41), we have

$$\sum_i (c_i^{(a)} - c_i(a))^- > -c_0(a) + c_0^{(a)} - v - 1000\epsilon n^2 + \sum_t d_t a_t a_{t+s}.$$

I.e., with $R = \{i : c_i^{(a)} - c_i(a) < 0\}$,

$$\sum_{i \in R} c_i^{(a)} - c_i(a) > -c_0(a) + c_0^{(a)} - v - 1000\epsilon n^2 + \sum_t d_t a_t a_{t+s}.$$

Each $c_i^{(a)}$ is at most $\sum_t |u_t|$ in absolute value; so using the standard Hoeffding inequalities, we get that with probability at least $1 - (2/\epsilon)^{-2s}/200$:

$$\sum_{i \in R \cap J} c_i^{(a)} > \frac{q}{n} \sum_{i \in R} c_i^{(a)} - \frac{3}{4}\epsilon q \sum_t |u_t^{(a)}|. \tag{53}$$

Using part (i) of lemma (22), we see that $\sum_i |c_i(a))|^2 \leq 2500n^3$. Also, we have $|c_i(a)| \leq \frac{8}{\epsilon} n$. So applying lemma (21), with $N = 2500n^2$, $M = \frac{8}{\epsilon}n$, we get that with probability at least $1 - (2/\epsilon)^{-2s}/200$:

$$\sum_{i \in R \cap J} -c_i(a) > \frac{q}{n} \sum_i (-c_i(a)) - 200\epsilon q n. \tag{54}$$

Adding (53) and (54), we get with probability at least $1 - (2/\epsilon)^{-2s}/100$:

$$\sum_{i \in R \cap J} (c_i^{(a)} - c_i(a)) > \frac{q}{n}[-c_0(a) + c_0^{(a)} - v - 1000\epsilon n^2] - \epsilon q(200n + \frac{3}{4}\sum_t |u_t^{(a)}|) + \frac{q}{n}\sum_t d_t a_t a_{t+s}.$$

The last inequality implies that the optimum value of the Linear Program below is upper bounded by $\frac{qv}{n} + 1005\epsilon q n + \frac{q}{n}\sum_t d_t a_t a_{t+s}$:

$$\text{Max} \sum_{i \in J} c_i(a) x_i + c_0(a)$$

$$\frac{q}{n}(a_t - \frac{5}{4}\epsilon n) \leq \sum_{i \in S_t \cap J} x_i \leq \frac{q}{n}(a_t + \frac{5}{4}\epsilon n)$$

$$\frac{q}{n}(a_{t+s} - \frac{5}{4}\epsilon n) \leq \sum_{i \in T_t \cap J} (1 - x_i) \leq \frac{q}{n}(a_{t+s} + \frac{5}{4}\epsilon n)$$

$$\text{for } t = 1, 2, \ldots s 0 \leq x_i \leq 1 \forall i \in J.$$

For every x feasible to the above LP, we have using (49),

$$\left| \sum_{i,j \in J} A_{ij} x_i (1 - x_j) - \frac{q}{n}\left(\sum_{i \in J} c_i(a) x_i + c_0(a)\right) + \frac{q^2}{n^2}\sum_t d_t a_t a_{t+s} \right| \leq 4000\epsilon q^2$$

$$\text{Thus, } \sum_{i,j \in J} A_{ij} x_i (1 - x_j) \leq \frac{q^2 v}{n^2} + 7000\epsilon q^2. \tag{55}$$

Thus with probability at least 19/20, we have that for every a, for which the Linear Program LP(J,a) is feasible, the optimal solution to it is at most $\frac{q^2 v}{n^2} + 7000\epsilon q^2$. Further, from (45), we have that for every a which is not realizable, $LP(J,a)$ is infeasible (with probability); so under this, the maximum cut in J is clearly at most the maximum over all feasible $LP(J,a)$ of the maximum value of $A_{ij}x_ix_j$, where the $x_i, i \in J$ satisfy the constraints of $LP(J,a)$ [at most because, we have relaxed the integrality constraints.] Thus we get our result on MAX-CUT. $\square$

6.3 A Lower Bound for the Sample Complexity of MAX-CUT

We formulate our sample lower bound for MAX-CUT in a general black-box model of computation. In this model any algorithm within the black-box can sample a graph according to uniform or biased but fixed distribution depending only on the number n of vertices and output an approximate value of the maximum cut. The sample complexity of such an algorithm is the number of sampled vertices. We denote it by $S(\mathrm{MAX-CUT})$. It depends of course on the required accuracy ϵ.

Theorem 23.

$$S(\mathrm{MAX-CUT}) = \Omega(1/\epsilon^2)$$

Proof. Theorem 23 will be deduced from the next theorem.

Theorem 24. *Let ϵ be any sufficiently small positive real. Suppose G is an undirected graph on n vertices and c is a sufficiently small positive number. Suppose $s = c/\epsilon^2$ is an integer and H is the induced graph on a random subset of s vertices of G. Then,*

$$\mathbf{Pr}(|\max-\mathrm{cut}(G) - \frac{n^2}{s^2}\max-\mathrm{cut}(H)| \geq \epsilon n^2) > 1/5$$

for any sufficiently small ϵ and infinitely many G.

Proof. We consider the sequence of graphs $(G_n)_{n=1,2,..}$ where G_n is the complete bipartite graph with color classes C and C', $|C| = n$, $|C'| = 2n$. (Thus G_n has $3n$ vertices.)

Let Q denote the number of vertices in the sample that belong to C. Clearly, $\max-\mathrm{cut}(H) = Q(s-Q)$. Q has a Binomial distribution $\mathrm{B}(m,p)$ with parameters $m = s = c/\epsilon^2$ and $p = 1/3$, and with variance $mp(1-p) = \frac{2c}{9\epsilon^2}$. By the Central Limit Theorem, if we set

$$\delta = \max_t \left| \mathbf{Pr}\left(\frac{3\epsilon}{\sqrt{2c}}(Q - \frac{c}{3\epsilon^2}) \leq t\right) - \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds \right|$$

we have that δ tends to 0 as ϵ tends to 0. Fix $t = t_o$ such that $\int_{-\infty}^{-t_o} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds = 1/4$. Then, by the preceding assertion we have that

$$\mathbf{Pr}\left(\frac{3\epsilon}{\sqrt{2c}}(Q - \frac{c}{3\epsilon^2}) \leq -t_o\right) > 1/5$$

for any sufficiently small ϵ. Let us rewrite this as follows

$$\mathbf{Pr}\left(Q - \frac{c}{3\epsilon^2} \leq -\frac{t_o\sqrt{2c}}{3\epsilon}\right) > 1/5$$

We have thus, with probability at least 1/5,

$$\begin{aligned}
\max - \mathrm{cut}(H) &\leq (\frac{c}{3\epsilon^2} - \frac{t_o\sqrt{2c}}{3\epsilon})(\frac{2c}{3\epsilon^2} + \frac{t_o\sqrt{2c}}{3\epsilon}) \\
&\leq (\frac{2c^2}{9\epsilon^4} - \frac{t_o\sqrt{2}c^{3/2}}{3\epsilon^3}) \\
\max - \mathrm{cut}(G) & \\
-\frac{n^2}{s^2}\max - \mathrm{cut}(H) &\geq \epsilon n^2 \frac{t_o\sqrt{2}}{3\sqrt{c}} \\
&\geq 9\epsilon n^2
\end{aligned}$$

if $c \leq \frac{2t_o^2}{729}$.

Thus it follows that, with probability $\Omega(1)$, any algorithm that samples only c/ϵ^2 many vertices, will not be able to distinguish between the graphs G_n on $3n$ vertices, and the graphs B_n, for B_n complete bipartite graphs on $3n$ vertices with color classes of sizes $n - \epsilon n$ and $2n + \epsilon n$. We note that

$$|\max - \mathrm{cut}(G_n) \ - \ \max - \mathrm{cut}(B_n)| = (\epsilon - \epsilon^2)n^2,$$

and therefore, any $(\epsilon/3)n^2$-approximation algorithm for MAX-CUT should distinguish between the graphs G_n, and B_n, a contradiction. □

6.4 An Improvement for MAX-2CSP

An instance of MAX-2CSP on a set V of logical variables is a set of binary logical constraints on these variables, which are called the "constraints", and the aim is to find an assignment which makes the number of satisfied constraints maximum.

We can assume that the instance I is, in fact, an instance of MAX-2DNF. We only need for this to replace each constraint by an equivalent set of conjunctions. We let $A^{(1)}$ denote the $n \times n$ matrix with V as set of rows and set of columns associated with the sets of constrains in I of the form $x \wedge y$:

$$A^{(1)}_{i,j} = 1 \text{ if } x_i \wedge x_j \in I, \ 0 \text{ otherwise}$$

Similarly, we let $A^{(2)}$, (resp. $A^{(3)}$) denote the matrices associated with the constrains of the form $x \wedge \neg y$, (resp.$\neg x \wedge \neg y$.) The main idea is to use separate cut decompositions $(S_t^{(i)}, S_{s+t}^{(i)}, d_t^{(i)}), 1 \leq t \leq s, i = 1, 2, 3$, for $A^{(1)}$, $A^{(2)}$ and $A^{(3)}$. We can apply Theorem 17 for $i = 1, 2, 3$, with the result that the cut decompositions induced on a a random subset J of $\{1, 2, ..n\}$ of cardinality $q = \Omega(1/\epsilon^4)$ by the decompositions $(S_t^{(i)}, S_{s+t}^{(i)}, d_t^{(i)}), 1 \leq t \leq s$, of

$A^{(i)}, i = 1, 2, 3$, satisfy simultaneously to the cut norm bounds $||B^{(i)}||_C \leq 41\epsilon q^2, i = 1, 2, 3$, where $B^{(i)}$ is the matrix induced on J by the difference $W^{(i)}$ between $A^{(i)}$ and its cut decomposition. Then, if a is any assignment to the variables in J, the number of constrains of $A^{(1)}$ satisfied by this assignment is of the form $A^{(1)}(a^{-1}(1), a^{-1}(1))$ and we approximate this by $(A^{(1)} - W^{(1)})(a^{-1}(1), a^{-1}(1))$. This involves an error at most $||B^{(1)}||_C$ by the definition of the cut norm. A similar observation applies for $i = 2$ and $i = 3$. Thus, exactly as for MAX-CUT, we can use on the sample the cut decomposition induced by the cut decomposition of the whole set.

The function to be maximized is

$$f(x) =$$

$$\sum_{i,j \in \{1,..,n\}} \left(x_i x_j A^{(1)}_{ij} + x_i(1 - x_j)A^{(2)}_{ij} + (1 - x_i)(1 - x_j)A^{(3)}_{ij} \right),$$

$$x_i \in [0, 1], 1 \leq i \leq n.$$

Similarly as for MAX-CUT, we can approximate this function in terms of the weighted sizes of the intersections of the vector x with the sets of rows and the sets of columns defining the cut decompositions: If for each vector $x \in [0, 1]^n$ we define the vector $a = a(x) \in \mathbf{R}^{6s}$ by:

$$\begin{aligned}
a_j &= \sum_{i \in S^{(1)}_j} x_i, \quad a_{j+s} = \sum_{i \in S^{(1)}_{j+s}} x_i \\
a_{j+2s} &= \sum_{i \in S^{(2)}_j} x_i, \quad a_{j+2s} = \sum_{i \in S^{(2)}_{j+s}} x_i \\
a_{j+4s} &= \sum_{i \in S^{(3)}_j} x_i, \quad a_{j+5s} = \sum_{i \in S^{(3)}_{j+s}} x_i.
\end{aligned}$$

then we have that, $f(x)$ is well approximated by the expression

$$\sum_{t=1}^{s} d^{(1)}_t a_t a_{t+s} + \sum_{t=1}^{s} d^{(2)}_t a_{t+2s}(|S^{(2)}_{t+s}| - a_{t+3s}) +$$
$$\sum_{t=1}^{s} d^{(3)}_t (|S^{(3)}_t| - a_{t+4s})(|S^{(3)}_{t+s}| - a_{t+5s}) \tag{56}$$

As for MAX-CUT, we can approximate this expression in the vicinity of a feasible point by a linear function of x. We can thus extend to this case the linear programming arguments used for the MAX-CUT problem. □

7 Acknowledgements

We thank Alan Frieze, Oded Goldreich and Claire Kenyon for helpful comments.

References

[1] N. Alon, B. Doerr, T. Luczak and T. Schoen, On the discrepancy of combinatorial rectangles, to appear.

[2] N. Alon and A. Shapira, Testing satisfiability, Proc. of the 13^{th} Annual ACM-SIAM SODA, ACM Press (2002), to appear.

[3] S.Arora, D.Karger and M.Karpinski, *Polynomial time approximation schemes for dense instances of NP-hard problems*, Proceedings of the 27th Annual ACM Symposium on Theory of Computing (1995) 284-293; the full paper appeared in J. Comp. System. Sci. **58** (1999), 193-210.

[4] K. Azuma, Weighted Sum of Certain Dependent random Variables, Tôhoku Math.J. 3 (1967), 357-367.

[5] A. Czumaj, C. Sohler, *Testing Hypergraph Coloring*, Proc. 28th ICALP'01, LNCS Vol. 2076, Springer, 2001, 493-505.

[6] W.Fernandez-de-la-Vega, *MAX-CUT has a Randomized Approximation Scheme in Dense Graphs*, Random Structures and Algorithms 8 (1996) 187-199.

[7] A.M.Frieze and R.Kannan, *The Regularity Lemma and approximation schemes for dense problems*, Proceedings of the 37th Annual IEEE Symposium on Foundations of Computing, (1996) 12-20.

[8] A. M. Frieze and R. Kannan *Quick Approximation to matrices and applications*, Combinatorica **19** (2) (1999) pp 175-200.

[9] M.X.Goemans and D.P.Williamson, *.878-Approximation algorithms for MAX CUT and MAX 2SAT*, Proceedings of the 26th Annual ACM Symposium on Theory of Computing, ACM Press (1994), 422-431.

[10] O.Goldreich, S.Goldwasser and D.Ron, *Property testing and its connection to learning and approximation*, Proc. 37th IEEE FOCS 96, 339-348; the full paper appeared in J. ACM **45** (1998), 653-750.

[11] W.Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association 58 (1963) 13-30.

[12] S. Khanna, M. Sudan and D. Williamson, *A complete classification of the approximability of maximization problems derived from boolean constraint satisfaction*, Proc. 29th ACM STOC (1997), 11-20.

[13] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North Holland, Amsterdam, 1977.

[14] S. J. Szarek, On the best constants in the Khinchin Inequality, Studia Math. 58 (1976), 197-208.

[15] V.N. Vapnik and A.Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and its Applications*, 16(2) (1971), 264–280.