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 $2^{\tilde{O}(\frac{1}{\epsilon^2})}$ for approximating all MAX-rCSP problems. This improves on the best known up to date bound of the form $2^{\tilde{O}(\frac{1}{\epsilon^{2r-2}})}$ 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 $r \geq 2$. 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.

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 $A(i_1, i_2, \dots i_r)$ 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 $S_1 \subseteq V_1, S_2 \subseteq V_2 \dots S_r \subseteq V_r$ we let $A(S_1, S_2, \dots S_r) = \sum_{(i_1,i_2,\dots i_r)\in S_1\times S_2\times\dots S_r} A(i_1, i_2, \dots i_r)$ and then define another norm $||A||_C$ (called the cut norm) :

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

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

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

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

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

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

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

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.

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 $OPT_{\mathcal{S}} - \epsilon' q^r \leq Y \leq OPT_{\mathcal{S}}$ 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 $\epsilon, \epsilon' > 0$,

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

and

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

We have also

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

$$\geq \frac{n^r}{q^r} Y - (\epsilon + \epsilon')\, n^r.$$

Thus, we have

$$|OPT - \frac{n^r}{q^r} Y| \leq (\epsilon + \epsilon')\, n^r$$

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 $V_1, V_2, \dots V_r$, $N = |V_1||V_2|\dots|V_r|$ 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 :*

$$||A - D||_C \leq \epsilon\sqrt{N}||A||_F \qquad (2)$$

$$||A - D||_F \leq ||A||_F \qquad (3)$$

The sum of the squares of the coefficients of the cut arrays is at most $4^r \frac{||A||_F^2}{N}$. (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 $D^{(1)}, D^{(2)}, \dots D^{(t)}$ one by one always maintaining the condition:

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

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

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

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

Let $d_{t+1} = W^{(t)}(S_1^{t+1}; S_2^{t+1}, \dots S_r^{t+1})/(|S_1^{t+1}||S_2^{t+1}|\dots|S_r^{t+1}|)$ be the average of the entries in $S_1 \times S_2 \times \dots S_r$ and let $D^{(t+1)} = CUT(S_1^{t+1}; S_2^{t+1}, \dots S_r^{t+1}, d_{t+1})$. 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 :

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

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

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

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

The proof of the tightness of the upper estimate is included in the full version of this paper.

□

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 $Q_1, Q_2, \dots Q_r$ 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 $Q_1' \subseteq Q_1$, $Q_2' \subseteq Q_2 \dots Q_r' \subseteq Q_r$ such that

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

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

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

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 $Q_1, Q_2, \dots Q_r$; 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 $Q_1, Q_2, \dots Q_r$. 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 $Q_1, Q_2, \dots Q_r$ as random subsets of J^{r-1}, whence (8) holds as if we first pick J' and then $Q_1, Q_2, \dots Q_r$ 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 $R_1 \times R_2 \times \dots R_r$. Suppose $S_1 \subseteq R_1, S_2 \subseteq R_2, \dots S_r \subseteq R_r$ are some fixed subsets. Suppose Q_1 is a random subset of $R_2 \times R_3 \times \dots R_r$ of cardinality p.* [1] *Then, with probability at least $1 - \frac{1}{40(4r)^r}$, we have :*

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

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

Proof. Let $S_2 \times S_3 \dots \times S_r = S$. 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 $X_1 + X_2 + \dots X_p$, where $X_1, X_2, \dots X_p$ is a sample of size p drawn uniformly without replacement from the set of $l = |R_2| \times |R_3| \times \dots |R_r|$ reals - $\{B(z, y))\underline{1}_{y \in S}\}$. For analysis, we also introduce the random variables $Y_1, Y_2, \dots Y_p$ - a sample of size p drawn independently, each uniformly distributed over the same set of reals, but now with replacement. We have

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

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

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

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,

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

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

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

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

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

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),

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

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

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

□

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

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

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

Proof. Let $S_1\subseteq R_1, S_2\subseteq R_2\dots S_r\subseteq R_r$ satisfy $B(S_1,S_2,\dots S_r) = B^+$. 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*

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

Proof. Note that changing any one element of I changes the random variable $G(I,I,\dots I)$ 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:*

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

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

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

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))

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

[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 $Q_1' \subseteq Q_1, Q_2' \subseteq Q_2, \dots Q_r' \subseteq Q_r$. We will apply Lemma (8) to the array G' on V^r obtained by setting

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

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

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

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

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

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

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

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

$$\begin{aligned} \exists Q_1' \subseteq Q_1, \exists Q_2' \subseteq Q_2 \dots G(P(Q_1'), P(Q_2'), \dots P(Q_r')) \geq \\ \frac{|V|^r}{q^r} B^+ - 5\epsilon |V|^{r/2} ||G||_F - 10\epsilon^2 M |V|^r. \end{aligned}$$

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 $F = \{f_1, \dots f_\ell\}$ be a collection of functions where each f_i is a boolean function of exactly r variables picked from $V = \{x_1, \dots x_n\}$. 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, $z = (z_1, z_2, \dots z_r)$, say, we define the r-dimensional array $A^{(z)}$ on V^r by

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

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

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

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

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

$$D^{(z)} = A^{(z)} - E^{(z)} = \sum_{t=1}^{s} Cut(S^{(z)}_{t,1}, S^{(z)}_{t,2}, ...S^{(z)}_{t,r}, d^{(z)}_t), 1 \leq t \leq s,$$

say, with $s = \frac{4^r}{\epsilon^2}$, $||E^{(z)}||_C \leq \epsilon M|V|^r$. 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

$$v^*(\nu) = \sum_{z\in\{0,1\}^r} \sum_{t=1}^{s} d^{(z)}_t \Pi_{i=1}^r \nu^{(z)}_{t,i} \qquad (20)$$

with $\nu^{(z)}_{t,i} = |S^{(z)}_{t,i} \cap S|$ if $z_i = 1$ and $\nu^{(z)}_{t,i} = |S^{(z)}_{t,i} \cap (V \setminus S)|$ 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 $\nu^{(z)}_{t,i}$. We say that ν is realizable if there exists $S \subseteq V$ such that

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

and

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

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:

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

with $z_i = 1$

$$\nu^{(z)}_{t,i} - \frac{2\epsilon^3 n}{8^r s} \leq \sum_{j\in S^{(z)}_{t,i}} (1 - x_j) \leq \nu^{(z)}_{t,i} + \frac{2\epsilon^3 n}{8^r s}$$

for all triples (z,t,i) with $z_i = 0$

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

[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 $sr2^{r+1} \leq \frac{\epsilon^3}{8^r s} n$]. 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 $0 \leq x_i \leq 1$. It is easy to see that the combination need not involve both the upper bound and the lower bound on any of the sets $S^{(z)}_{t,i}$. 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,

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

and

$$c^{(\nu)}_0 = \sum_{z\in\{0,1\}^r} \left(\sum_{1\leq t\leq s,\ 1\leq j\leq r} (u^{(z)}_{t,j} \nu^{(z)}_{t,j} + |u^{(z)}_{t,j}| \frac{\epsilon^3 n}{8^r s}) \right) - \sum_{1\leq j\leq r} \sum_{z:z_j=0} \sum_{1\leq t\leq s} u^{(z)}_{t,j}$$

we get that

$$\sum_{i=1}^{n} c^{(\nu)}_i x_i \leq c^{(\nu)}_0 \text{ has no solution } x \text{ with } 0 \leq x_i \leq 1 \qquad (21)$$

$$\text{which is equivalent to } \sum_{i=1}^{n} \text{Min}(c^{(\nu)}_i, 0) > c^{(\nu)}_0. \qquad (22)$$

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 $|c^{(\nu)}_i| \leq \gamma^{(\nu)}$, we have from (22), using the Theorems of Hoeffding [11],

$$\mathbf{Pr}\left(\sum_{i\in J} \text{Min}(c^{(\nu)}_i, 0) \leq \frac{q}{n} c^{(\nu)}_0 - \frac{2\epsilon^3 q}{8^r sn} \gamma^{(\nu)}\right) \leq \exp\left(-\frac{2\epsilon^6 q}{8^{2r} s^2}\right)$$

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

$$\frac{q}{n}\left(\nu^{(z)}_{t,j} - \frac{\epsilon^3 n}{8^r s}\right) \leq \sum_{i\in S^{(z)}_{t,j}\cap J} x_i \leq \frac{q}{n}\left(\nu^{(z)}_{t,j} + \frac{\epsilon^3 n}{8^r s}\right)$$

for all (z,t,j) with $z_j = 1$

$$\frac{q}{n}\left(\nu_{t,j}^{(z)}-\frac{\epsilon^3 n}{8^r s}\right)\leq\sum_{i\in S_{t,j}^{(z)}\cap J}(1-x_i)\leq\frac{q}{n}\left(\nu_{t,j}^{(z)}+\frac{\epsilon^3 n}{8^r s}\right)$$

for all (z,t,j) 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

Pr(LP(J,ν) feasible) $>\alpha$) implies LP(V,ν) 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

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

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 $q\geq 5000r^7/\epsilon^8$. Let B be the r-dimensional array obtained by restricting G to J^r. Then we have, with probability at least $1-1/(4.2^r)$,*

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

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 $1-1/(3.2^r)$,

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

This implies immediately the assertion of the theorem. □

We return now to the proof of Theorem 8.

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

$$||F^{(z)}||_C\leq 16\epsilon 4^rq^rM$$

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

$$|v(\eta)-v^*(\eta)|\leq\sum_{z\in\{0,1\}^r}||F^{(z)}||_C\leq 16\epsilon M8^rq^r. \tag{23}$$

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

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

For each realizable η there is an η', say, belonging to K and for which $||\eta'-\eta||_{\ell_1}\leq\frac{q\epsilon^3}{8^rs}r2^rs\leq\frac{\epsilon^3qr}{4^r}$. 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

$$||\nu'-\frac{|V|}{q}\eta'||_{\ell_1}\leq\frac{\epsilon^3|V|}{8^rs}r2^rs=\frac{\epsilon^3r|V|}{4^r}.$$

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

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

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

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

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.

A refinement of the general method above yields also directly the proof of Theorem 2. The details are given in the final version of this paper.

□

5 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.