Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 70 (2002)

9/8-Approximation Algorithm for Random MAX-3SAT

W. Fernandez de la Vega* Marek Karpinski†

Abstract. We prove that MAX-3SAT can be approximated in polynomial time within a factor 9/8 on random instances.

1 Introduction

Random 3-SAT formulas have been widely studied in the context of structural properties of the general satisfiability problem, cf. [BKPS98, F99, FG01, MF95, KMPS94, FS96, DBM00, A00] and the surveys [DBSZ01] and [GPFW97]. Randomly chosen 3SAT-formulas are empirically difficult for deciding satisfiability and are used often as a benchmark for various testing algorithms.

In this paper we study the problem of approximability (rather than just satisfiability) of random MAX-3SAT. We were originally motivated by a recent paper of Feige [F02] connecting the hardness of approximation of certain combinatorial problems , like MIN-BISECTION, to the problem of efficient approximability of random 3SAT and the problem of refutation of its instances. In particular, we investigate the problem of the possible improvements of the approximation ratio of polynomial algorithms for random MAX-3SAT over Håstad lower bound of 8/7 [H97]. We prove in this paper that there are polynomial time algorithms approximating random

*CNRS, Universite Paris Sud, Orsay. 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. Email: lalo@lri.fr

†Dept. of Computer Science, University of Bonn. Supported in part by DFG, and DAAD 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, and the Isaac Newton Institute for Mathematical Sciences, Cambridge. Email: marek@cs.uni-bonn.de

ISSN 1433-8092

MAX-3SAT (formula by formula) to within a factor 9/8 (a considerable improvement over Håstad's bound).

2 Approximation Algorithms on Random Instances

We consider a standard model of generation of random 3SAT formulas (R3SAT-formulas). Given parameters n for the number of variables and m for the number of of clauses, each clause is generated independently at random by chosing three literals independently and uniformly at random. We denote $\rho = m/n$ and define a parameter $\lambda = \frac{3\rho}{2}$. There are several other models for generating R3SAT-formulas but there are not of significance towards our results.

For a given (generated) R3SAT-formula F, let $m(F)$ denote the maximum number of clauses of F which can be satisfied. For an assignment X, $m_X(F)$ denotes the number of clauses of F satisfied by X.

We call a polynomial time (randomized) algorithm Q an approximation algorithm for the MAX-R3SAT problem with approximation ratio α if for every (generated) formula F, Q outputs an assignment X such that the probability resulting from the input and the inner algorithm's distributions satisfies

$$\Pr\left(\frac{m(F)}{m_X(F)} \leq \alpha\right) \geq 3/4. \tag{1}$$

and

$$\lim_{n\to\infty} \Pr\left(\frac{m(F)}{m_X(F)} \leq \alpha\right) = 1 \tag{2}$$

for any fixed ρ.

We call a polynomial time (randomized) algorithm Q a *value approximation algorithm with approximation ratio* α for the MAX-R3SAT problem if for every (generated) formula F, Q outputs a number $m^*(F)$ such that the probability resulting from the input and the inner algorithm's distributions satisfies

$$\Pr\left(\frac{m(F)}{m^*(F)} \leq \alpha\right) \geq 3/4.$$

and

$$\lim_{n\to\infty} \Pr\left(\frac{m(F)}{m^*(F)} \leq \alpha\right) = 1$$

for any fixed ρ.

3 Main Result

We prove the following main result on the approximability of the MAX-R3SAT problem.

Theorem. *There exists a polynomial time algorithm for approximating MAX-R3SAT to within ratio 9/8.*

An approximation algorithm and a proof of its correctness are given in the next section.

4 A 9/8-Approximation Algorithm for R3SAT

Recall that $\rho = m/n$. In our analysis, we assume that n (and m) are arbitrarily large with ρ fixed. For a formula F, let $m(F)$ denote the maximum number of clauses of F which can be satisfied by a properly chosen assignment of truth values to the variables. We describe an algorithm which, when applied to an F returns a value $m^*(F)$ (together with an assignment X) for which we have that (1) and (2) is satisfied for $\alpha = 9/8$ and any fixed ρ.

Notice that there is no guarantee here as it happens elsewhere that satisfiable formulae are detected with zero error probability.

We consider separately the case of *"high"* values and the case of *"small"* values of ρ. For values $\rho \geq 102$, the algorithm outputs *random assignment* for every formula. For smaller values of ρ and for each variable the algorithm assigns *greedily* this variable to *true* if the positive literal appears at least as many times as the negative literal. Otherwise the variable is assigned to *false*. We describe now the behavior of the algorithm in detail.

4.1 The Case of "High" Values of ρ

Assume first that $\rho = m/n \geq 102$. (This separation gives near optimal results in our proof method.) In this case we show that $m(F)$ is near to $\frac{7m}{8}$, so that a *random assignment* will give the claimed ratio. Let $\mathrm{val}(A, F)$ be the number of clauses of the random formula F true under the assignment A, and $\mathrm{Bin}(m, p)$ be the binomial random variable with probability p. Then

$$\begin{aligned} \mathbf{E}(\#\{A : \mathrm{val}(A,F) \geq 7m/8(1+\epsilon)\}) &= 2^n \Pr(\mathrm{Bin}(m, 7/8) \geq (7m/8)(1+\epsilon)) \\ &\leq 2^n \exp(-7m\epsilon^2/16), \end{aligned}$$

the last by Hoeffding-Chernoff. This is o(1) for $m/n \geq \frac{16 \log 2}{7\epsilon^2}$. We require that $\epsilon \leq 9/8 - 1 = 0.125$. This is true for $m/n = 101.397 \leq 102$. $\square$

4.2 The Case of "Small" Values of ρ

We consider now the case $\rho = m/n < 102$. We assume for convenience that the clauses of F are ordered. We are going to construct the following greedy algorithm.

For each variable which appears strictly more often in positive than in negative form, we assign it to *true.* and we call the corresponding positive literal "major". We call the corresponding negative literal "minor". Similarly, we assign to *false* every variable which appears stricly more often in negative than in positive form and we call the corresponding negative literal "major". We call the corresponding positive literal "minor". We call neutral all the variables which appears as many times (possibly none) in positive or in negative form and we assign these variables to true. We denote by NEUTRAL the set of literals corresponding to neutral variables. We let MAJOR (resp. MINOR) denote the set of major (respectively minor) literals.

Let $C = (\ell_1, \ell_2, \ell_3)$ be a fixed clause of F (say, the first one). What is the probability that this clause is satisfied in our assignment? We note first that the number of appearances of each fixed literal is asymptotically (as $n \to \infty$) Poisson with parameter $\lambda = \frac{3\rho}{2}$. It will be convenient to introduce two independent random variables P_λ and R_λ and having both this Poisson distribution. Fix attention on ℓ_1. ℓ_1 is true in our assignment either if (i) it is major, which has the probability asymptotic to

$$q_1 = \Pr(P_\lambda \geq R_\lambda),$$

or (ii) it is neutral and *positive*, which has the probability asymptotic to

$$q_2 = (1/2) \Pr(P_\lambda = R_\lambda - 1)$$

Thus, the probability q that ℓ_1 is true satisfies

$$q = q_1 + q_2 \sim \Pr(P_\lambda \geq R_\lambda) + (1/2) \Pr(P_\lambda = R_\lambda - 1)$$

Clearly the dependencies between ℓ_1, ℓ_2, ℓ_3 can be neglected and thus we have that

$$\Pr(\text{C satisfied}) \sim 1 - (1-q)^3$$

This implies of course that the expectation of the total number of clauses satisfied is asymptotic to $m(1-(1-q)^3)$. We proceed now to derive an explicit formula for q. We have that

$$\begin{aligned}\Pr(P_\lambda \geq R_\lambda + 1) &= 1/2 - (1/2)\Pr(P_\lambda = R_\lambda) \\ &= 1/2 - (1/2)\sum_{j=0}^{\infty} e^{-2\lambda}\frac{\lambda^{2j}}{(j!)^2}\end{aligned}$$

and

$$\Pr(P_\lambda = R_\lambda + 1) = \sum_{j=0}^{\infty} e^{-2\lambda}\frac{\lambda^{2j+1}}{j!(j+1)!}$$

Thus

$$q \sim 1/2 + (1/2)\left(\sum_{j=0}^{\infty} e^{-2\lambda}\frac{\lambda^{2j}}{(j!)^2} + \sum_{j=0}^{\infty} e^{-2\lambda}\frac{\lambda^{2j+1}}{j!(j+1)!}\right) \tag{3}$$

We will need now the following Lemma.

Lemma 1. *We have that*

$$q \sim \frac{\mathbf{E}(\max P_\lambda, R_\lambda)}{2\lambda}$$

that is,

$$\Pr(P_\lambda \geq R_\lambda) + (1/2)\Pr(P_\lambda = R_\lambda - 1) = \frac{\mathbf{E}(\max P_\lambda, R_\lambda)}{2\lambda}$$

Proof. Note that we have

$$\frac{\mathbf{E}(\max P_\lambda, R_\lambda)}{2\lambda} = \frac{1}{2\lambda}\left(\sum_{j=1}^{\infty} je^{-2\lambda}\frac{\lambda^{2j}}{(j!)^2} + 2\sum_{j=1}^{\infty} je^{-\lambda}\frac{\lambda^j}{j!}\left(\sum_{k=0}^{j-1} e^{-\lambda}\frac{\lambda^k}{k!}\right)\right).$$

It is a simple matter to check that this expression can be put into the form (3).

□

We take $\rho = 102$ which gives $\lambda = 153$. Then, from (3), we get using computer assisted analysis, $q \sim 0.52279.. \geq 0.52279$ for sufficiently large n. The probability of satisfaction of any fixed clause is thus, for sufficiently large n, at least

$$1 - (1 - 0.52279)^3 = 0.89133...$$

This proves that, for $\lambda = 153$, the expectation of the number of clauses satisfied in our assignment is asymptotic to $0.89133.. < 8/9$. In section

5, we will prove concentration, showing that the approximation ratio 9/8 holds for $\lambda = 153$. In the next section, we prove that q is non-increasing as a function of λ, implying that the proportion of clauses satisfied by our algorithm does not decrease when λ decreases. With the result above this will complete the proof that our algorithm has approximation ratio at least 9/8 for any $\lambda \leq 153$.

4.3 q is Non-increasing

We have to prove that q, given according to Lemma 1 by

$$q = \frac{\mathbf{E}(\max P_\lambda, R_\lambda)}{2\lambda} ,$$

does not increase with λ.

Aside from P_λ and R_λ we introduce additional Poisson random variables P_δ with parameter δ where δ is an arbitrarily small positive real, $P_{\lambda+\delta}$ and $R_{\lambda+\delta}$ both with parameter $\lambda, \delta, \lambda+\delta$ and similarly for $R_\lambda, R_\delta, R_{\lambda+\delta}$. From the fact that the distribution $P_{\lambda+\delta}$ is the convolution of the distributions of P_λ and P_δ it follows that we can wiew the pair $(P_{\lambda+\delta}, R_{\lambda+\delta})$ as the mixture, with coefficients $e^{-2\delta}, 1-e^{-2\delta}$ of the pairs (P_λ, R_λ) and $(P_\lambda, R_{\lambda+\delta})$ or $(P_{\lambda+\delta}, R_\lambda)$. Let $Q(X,Y)$ denote the expectation of $\max(X,Y)$ for two random variables X and Y.

By the above argument, we have that

$$\begin{aligned}
Q(P_{\lambda+\delta}, P_{\lambda+\delta}) &= e^{-2\delta}(Q(P_\lambda, P_\lambda) + 2(1-e^{-\delta})e^{-\delta}Q(P_\lambda+1, P_\lambda) + O(\delta^2) \\
&= (1-2\delta)Q(P_\lambda, P_\lambda) + 2\delta Q(P_\lambda+1, P_\lambda) + O(\delta^2) \\
&= Q(P_\lambda, P_\lambda) + 2\delta\Delta + O(\delta^2),
\end{aligned}$$

where

$$\Delta = Q(P_\lambda+1, R_\lambda) - Q(P_\lambda, R_\lambda).$$

In order to estimate Δ, we simply consider with each pair of values i, j the corresponding pair $i+1, j$ and observe:

- If $i < j$, then the max does not change.

- If $i \geq j$, then the max increases by 1.

Thus, Δ is just the probability that $P_\lambda \geq Q_\lambda$. By symmetry we have that

$$\Pr(P_\lambda \geq Q_\lambda) = 1/2 + (1/2)\Pr(P_\lambda = Q_\lambda).$$

implying

$$\Delta = \delta/2 + (\delta/2)f(\lambda)$$

where $f(\lambda) = \sum_{k=0}^{\infty} e^{-2\lambda} \frac{\lambda^{2k}}{k!^2}$. In other words the derivative of Q is $1/2 + (1/2)f(\lambda)$. This gives

$$q(\lambda) = 1/2 + \int_0^\lambda \frac{f(\mu)}{2\lambda}$$

whence it follows that q is non-increasing (as we wish to prove) if f is non-increasing. Using again the decomposition above, we have that

$$f(\mu + \delta) = f(\mu) - 2\delta(f(\mu) - g(\mu)) + O(\delta^2)$$

with

$$g(\mu) = \sum_{k=0}^{\infty} e^{-\mu} \frac{\mu^k}{k!} e^{-\mu} \frac{\mu^{k+1}}{(k+1)!}$$

Thus it suffices to prove that $f(\mu) \geq g(\mu), \mu > 0$. We will use the following Lemma.

Lemma 2. *Let $S = (a_0, a_1, ...a_m)$ be a finite sequence of numbers. For any permutation Π of the set $\{0, 1, ...m\}$, let $q(S) = \sum_{j=0}^{m} a_j a_{\Pi(j)}$. The sum S is maximum when Π is the identity.*

Proof. The following simple proof was suggested to us by Yves Verhoeven. We have by the Cauchy-Schwarz inequality that $< S, \Pi S > \leq ||S||.||\Pi S||$, where $<,>$ denotes the scalar product. Also $||\Pi S|| = ||S||$. Therefore $< S, \Pi S > \leq ||S||^2$ which is what we want.

$\square$

We use this Lemma with $a_k = e^{-\mu} \frac{\mu^k}{k!}$. We fix some m, and let $S = (a_0, a_1, ...a_m)$. We define the permutation Π on the set $\{0, 1, ...m\}$ by $\Pi(j) = j + 1$ for $0 \leq j \leq m - 1$ and $\Pi(m) = 0$. Then we have that $q(S)$ tends to $g(\mu)$ and $f(\mu)$ is at least $q(S)$ because of the Lemma. Thus, for any $\epsilon > 0$, we get, choosing m sufficiently large the inequality $g(\mu) \leq f(\mu) + \epsilon$ and this implies of course $g(\mu) \leq f(\mu)$.

5 Proof of Concentration

We now turn to the proof of concentration. We use Chebyshev's inequality. The set of clauses is clearly symmetric. Thus, by a remark of [AS92] (see [AS92], section 4.3) it suffices to prove that for two distinct clauses C and C' we have that

$$\Pr(C \text{ TRUE and } C' \text{ TRUE}) = \Pr(C \text{ TRUE})^2(1 + o(1)) \tag{4}$$

where TRUE means true in the assignment which we define. Let u, v, w, (resp. u', v', w') be the (random) literals in C (resp.C').[For convenience we consider the literals in a clause as being ordered.] For literals $a, b, ...$ let $c\ell(a), c\ell(b)...$, denote the class (MAJOR, MINOR or NEUTRAL) of a, resp. $b...$

Fix a sequence $(c_1, c_2, c_3, c_4, c_5, c_6)$ of classes (i.e., each c_i is either MAJOR, MINOR or NEUTRAL), and write

$$p_1 = \Pr(c\ell(u) = c_1, c\ell(v) = c_2, c\ell(w) = c_3, c\ell(u') = c_4, c\ell(v') = c_5, c\ell(w') = c_6)$$

and

$$p_2 = \Pr(c\ell(u) = c_1, c\ell(v) = c_2, c\ell(w) = c_3).$$
$$\Pr(c\ell(u') = c_1, c\ell(v') = c_2, c\ell(w') = c_3)$$

The proof essentially reduces to the following claim.

Claim. For any two strictly distinct literals u and v, and any two classes c_1 and c_2, we have that

$$\Pr(c\ell(u) = c_1, c\ell(v) = c_2) = \Pr(c\ell(u) = c_1). \Pr(c\ell(v) = c_2)(1 + o(1)) \quad (5)$$

Proof of the claim. Let T_1 denote the (random) number of occurences of v and $\neg v$ conditional on the event $c\ell(v) = c_2$. Then we have that $T_1 \leq \log n$ with probability $1 - O(1/n)$. For a fixed T_1 the class of u is determined by two other random variables T_2 and T_3:

T_2 is the number of clauses which contain u conditional on the fact that already T_1 "places" are used

T_3 is the number of clauses which contain $\neg u$ conditional on the fact that already $T_1 + T_2$ "places" are used.

From the fact that $T_1, T_2 \leq \log n$ with probability $1 - O(1/n)$, one can deduce (we omit the details) that the conditional distributions of each of T_2 and T_3 lie at ℓ_1 distance $O(\log n/n)$ of their unconditional distributions. This implies easily the claim, with the stronger error coefficient $1 + O(\log n/n$.

□

The claim implies that $p_1 = p_2 \pm O(\log n/n)$ conditionally on u, v, w, u', v', w' being pairwise strictly distinct literals. Now random literals are distinct with probability $1-O(1/n)$ so that we have in fact $p_1 = p_2 \pm O(\log n/n)$ without any conditioning.

This concludes the proof. Actually, we get a bit stronger result than it was needed.

□

6 Value Approximation Algorithms

We notice that if we are interested only in approximating the values of $m^*(F)$ and not in constructing an actual approximating assignment we can use the following value approximation algorithm:

1. Compute $\rho = m/n$,
2. If $\rho < 102$ then output 0.89133 m,
3. If $\rho \geq 102$ then output 8/9 m.

An interesting question arises whether there exist polynomial time *value approximation algorithms* for MAX-R3SAT with much better approximation ratios α than 9/8, or even whether there exist a polynomial time value approximation schemes (VPTAS) approximating MAX-R3SAT within arbitrary approximation ratios $\alpha > 1$.

A possible proof of an existence of a VPTAS for MAX-R3SAT would require though stronger explicit concentration results than the result of this paper.

7 Further Research

An interesting open problem is whether the bound 9/8 for approximation ratio of MAX-R3SAT can be, in fact, improved. Another intriguing question is whether approximation ratio for the measurement of *values* of MAX-R3SAT can be improved considerably over 9/8 (existence of a VPTAS for that problem?).

□

Acknowledgements. We are indebted to Béla Bollobás, Mark Jerrum, Alex Scott, and Yves Verhoeven for interesting remarks and discussions.

References

[A00] D. Achioptas, *Setting Two Variables at a Time Yields a New Lower Bound for Random 3-SAT*, Proc. 32nd ACM STOC, 2000, 28-37

[AS92] N. Alon, J. Spencer, The Probabilistic Method, Wiley 1992

[BKPS98] P. Beame, R. Karp, T. Pitassi, M. Saks, *On the Complexity of Unsatisfiability Proofs for Random k-CNF Formulas*, In Proc. 30th STOC, 1998, 561-571

[DBM00] O. Dubois, Y. Boufkhad and J. Mandler, *Typical 3-SAT Formulae and the Satisfiability Threshold*, Proc. 11th ACM-SIAM SODA (2000), 126-127.

[DBSZ01] O. Dubois, R. Monasson, B. Selman, R. Zecchina, eds, *Phase Transitions in Combinatorial problems*, Theoretical Computer Science, Volume 265, 2001, Numbers 1-2.

[FG01] J. Friedman, A. Goerdt, *Recognizing More Unsatisfiable Random 3SAT Instances Efficiently,*. In Proc. 28th ICALP, 2001, 310-321.

[F02] U. Feige, *Relations between Average Case Complexity and Approximation Complexity*, In Proc. of the 34th ACM STOC, 2002, 534-543

[F99] E. Friedgut, *Necessary and Sufficient Conditions for Sharp Thresholds of Graph Properties and the k-SAT Problem,* Journal of the American mathematical Society 12, 1999, 1017-1054.

[FS96] A. Frieze and S. Suen, *Analysis of Two Simple Heuristics of a Random Instance of k-SAT*, J. of Algorithms 20, 1996, 312-355.

[GPFW97] J. Gu, P. W. Purdom, J. Franco and B. J. Wah, *Algorithms for the Satisfiability (SAT) Problem: A Survey.* In Satisfiability (SAT) Problem, DIMACS, 19-151. American Mathematical Society 1997

[H97] J. Håstad, *Some Optimal Inapproximability Results*, Proc. 29th ACM STOC (1997), 1-10

[KMPS94] A. Kamath, R. Motwani, K. V. Palem, P. G. Spirakis, *Tail Bounds for Occupancy and the Satisfiability Threshold Conjecture*, Proc. 35th IEEE FOCS (1994), 592-603, also in Random Structures and Algorithms 7 (1995), 59-80.

[MF95] A. El Maftouhi, W. Fernandez de la Vega, *On Random 3-SAT*, Combinatorics, Probability and Computing, Volume 4, 1995, 189-195.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'