1.0957-APPROXIMATION ALGORITHM FOR RANDOM MAX-3SAT

Wenceslas Fernandez de la Vega[1] and Marek Karpinski[2]

Abstract. We prove that MAX-3SAT can be approximated in polynomial time within a factor 1.0957 on random instances.

Keywords. Random satisfiability, approximate algorithms.

Mathematics Subject Classification. 68W25, 03B70

1. Introduction

Random 3 SAT formulas have been widely studied in the context of structural properties of the general satisfiability problem, cf. [1, 3, 4, 6, 9, 10, 15] and the surveys [5] and [11]. 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 [7] 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 [12]. We prove in this paper that there are polynomial time algorithms approximating random MAX-3SAT (formula by formula) to within a factor 1.0957 (a considerable improvement over Håstad's bound). We note that a previous version of this paper appeared in ECCC [8]. Our approximation ratio was later improved to $\frac{10}{9.5} \sim 1.0526$ by Inderian [14].

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 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, such as fixing beforehand the probability of each possible clause (whence the number of clauses is then a random variable) but they lead to similar 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 Q outputs an assignment X such that the probability resulting from the input and the inner algorithm's distributions satisfies

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

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

$$\Pr \left(\frac{m(F)}{m^*(F)} \leq \alpha \right) \geq 3/4 \tag{2}$$

and

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

for any fixed $\rho > 0$.

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 1.0957.*

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

4. A 1.0957-APPROXIMATION ALGORITHM FOR R3SAT

Recall that $\rho = m/n$. In our analysis, we assume that n (and m) are arbitrarily large with ρ fixed. 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 the theorem is true for $\alpha = 1.0957$ and any fixed $\rho > 0$.

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 $\rho \geq 16.554$, the algorithm outputs the everywhere true 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. Thus the algorithm is deterministic in both cases. We describe now the behavior of this algorithm in detail.

4.1. THE CASE OF "HIGH" VALUES OF ρ

We treat first the case where $\rho = m/n \geq 16.554$. (This separation gives near optimal results in our method of proof.) In this case we shall show that $m(F)$ is near to $\frac{7m}{8}$, so that a *random assignment* will give the claimed ratio. Let $\text{val}(A, F)$ be the number of clauses of the random formula F true under the assignment A. Let $B(n, p)$ denote a binomial random variable with parameters n and p and let $q = 1 - p$. The following inequality is implied immediately by a large deviations bound of Hoeffding (see [13], Th. 1 (2.1), p. 15):

$$\Pr(B(n, p) \geq n(p + t)) \leq \left(\left(\frac{p}{p+t} \right)^{p+t} \left(\frac{q}{q-t} \right)^{q-t} \right)^n.$$

We take $p = 0.875$, $q = 0.125$ and $t = 0.0957$ in the above inequality to get

$$\Pr(\text{val}(A, F) \geq (7m/8)1.0957) \leq 0.95899^m.$$

This gives

$$\mathbf{E}(\#\{A : \text{val}(A, F) \geq (7m/8)1.0957\}) \leq 0.95899^m 2^n.$$

This is $o(1)$ for $\rho \geq 16.554$. Thus for ρ satisfying $\rho \geq 16.554$, using Markov inequality we have that, with probability $1 - o(1)$, there is no assignment satisfying more than $(7m/8)1.0957$ clauses. This clearly gives us the claimed approximation ratio for $\rho \geq 16.554$.

4.2. The case of "small" values of ρ

We consider now the case $\rho = m/n \leq 16.554$. 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. We will make use of the following two propositions.

Proposition 1. Let C be a fixed clause. Given that $\ell \in C$, the distribution of the number of occurrences of ℓ and $\neg \ell$ out of C are both asymptotically Poisson with parameter λ.

Proposition 2. Let k and h be fixed natural integers and assume that $u_1, u_2,...u_k$, $v_1, v_2,...v_h$ are pairwise distinct literals. Assume that for $1 \leq j \leq k$ the literal ℓ_j has a total of n_j occurrences in the clauses with $n_j \in o(n)$. Then the numbers of occurrences of the literals $v_1, v_2, ...v_h$ are asymptotically independent and asymptotically Poisson with parameter λ, as $n \to \infty$.

Let q denote the probability that a fixed literal in a fixed clause is true in the assignment A. We will derive the asymptotic (as $n \to \infty$) value of q by two distinct methods. It will be convenient to introduce two independent random variables P_λ and R_λ and having both this Poisson distribution.

Lemma 1. *We have that*

$$q \sim \frac{\mathbf{E}(\max P_\lambda, R_\lambda)}{2\lambda}. \tag{3}$$

First proof. Let $M = \max(P_\lambda, R_\lambda)$ so that the assertion of the lemma is that $q \sim \frac{\mathbf{E}M}{2\lambda}$. Put $T_\ell = 1$ if the literal ℓ is true in the assignment A. Otherwise $T_\ell = 0$. Let $T = \sum_{\ell \in L} T_\ell$ be the total number of occurences of true literals in F. T concentrates around its expectation. (For a proof, observe that $\mathrm{Var}(T_\ell) = O(1)$, check that $\mathrm{Cov}(T_\ell, T_k) = O(1/n)$ for any pair of literals with distinct underlying variables, T_ℓ and T_k and $\mathrm{Cov}(T_\ell, T_{\neg\ell}) = O(1)$. Then apply Tchebichev's inequality.)
Since the literals within any fixed clause are random within the set of occuring literals, we get that the probability q that a fixed literal in a fixed clause is true

conditionnaly on T satisfies

$$q = \frac{T}{3m}$$
$$\sim \frac{\mathbf{E}M}{3m}$$
$$\sim \frac{\mathbf{E}M}{2\lambda}.$$

The probability that a fixed clause is true satisfies thus

$$\Pr(C \text{ satisfied}) \quad \sim \quad 1 - (1 - q)^3$$
$$\sim \quad 1 - \left(1 - \frac{\mathbf{E}M}{2\lambda}\right)^3.$$

Second proof. 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? Note first that the number of appearances in our formula of each fixed literal is asymptotically (as $n \to \infty$) Poisson with parameter $\lambda = \frac{3\rho}{2}$.

Fix attention on ℓ_1. ℓ_1 is true in our assignment either if (i) it is major, which has probability asymptotic to

$$q_1 = \Pr(P_\lambda \geq R_\lambda), \tag{4}$$

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

$$q_2 = (1/2) \Pr(P_\lambda = R_\lambda - 1). \tag{5}$$

(Note that in (4) and (5) the computation is done for ℓ_1 *given*.) Thus, the probability q that ℓ_1 is true satisfies

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

(We use $\sim$ for asymptotic equivalence, as $n \to \infty$.) We proceed now to derive an explicit formula for q. We have that

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

and

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

Thus

$$q \sim 1/2 + (1/2)\sum_{j=0}^{\infty} \mathrm{e}^{-2\lambda}\frac{\lambda^{2j}}{(j!)^2} + \sum_{j=0}^{\infty} \mathrm{e}^{-2\lambda}\frac{\lambda^{2j+1}}{j!(j+1)!}. \qquad (6)$$

Note that we have

$$\frac{\mathbf{E}(\max P_\lambda, R_\lambda)}{2\lambda} = \frac{1}{2\lambda}\left(\sum_{j=1}^{\infty} j\mathrm{e}^{-2\lambda}\frac{\lambda^{2j}}{(j!)^2} + 2\sum_{j=1}^{\infty} j\mathrm{e}^{-\lambda}\frac{\lambda^j}{j!}\left(\sum_{k=0}^{j-1}\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\right)\right). \qquad (7)$$

Thus, by using (3), we get that the right-hand sides of (7) and (6) are identical.   □

Now we can finish the proof concerning the case $\rho = m/n \leq 16.554$. Fix $\rho = 16.554$ which gives $\lambda = 24.831$. Then, from (6), we get using computer assisted analysis, $q \sim 0.55642$ implying $q \geq 0.55641$ for sufficiently large n. The probability of satisfaction of any fixed clause is thus at least, for sufficiently large n,

$$1 - (1 - 0.55641)^3 = 0.91271.$$

This proves that, for $\rho = 16.554$, the expectation of the number of clauses satisfied in our assignment is asymptotic to $0.91271m = \frac{m}{1.09564}$ and yields that the approximation ratio 1.0957 holds for $\rho = 16.554$. In the next section, we prove that q is non-increasing as a function of λ, implying that the approximation ratio is at least 1.0957 for every $\rho \leq 16.554$. Putting this together with the result of Section 4.1, the proof will be completed by the concentration result of Section 5.

4.3. LIM q IS NON-INCREASING

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

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

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 $\lambda + \delta$ 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 the pair $(P_{\lambda+\delta}, R_{\lambda+\delta})$ is the mixture, with coefficients $\mathrm{e}^{-2\delta}, 1 - \mathrm{e}^{-2\delta}$ 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. The above mixture argument gives,

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

where

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

In order to estimate Δ, we simply consider aside 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

$$\Delta = 1/2 + (1/2)f(\lambda)$$

where $f(\lambda) = \sum_{k=0}^\infty e^{-2\lambda} \frac{\lambda^{2k}}{k!^2}$. In other words the derivative of Q is $1 + f(\lambda)$. 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 $q(S)$ 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 $\langle S, \Pi S \rangle \leq ||S||.||\Pi S||$, where $\langle, \rangle$ denotes the scalar product. Also $||\Pi S|| = ||S||$. Therefore $\langle S, \Pi S \rangle \leq ||S||^2$ 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 needed for the case of small values of ρ, $\rho \leq 16.554$.

The set of clauses is clearly symmetric. Thus, by a remark of [2] (see [2], Sect. 4.3) it suffices to prove that for two distinct clauses C and C' we have that

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

where FALSE means false in the assignment which we defined. Thus we have that $\Pr(C \text{ FALSE}) = (1 - q)^3$ with q defined by (6). Let us say that a literal ℓ in a clause is *good* if it does not satisfy the clause (that is, ℓ either belongs to MINOR or is negative and belongs to NEUTRAL). Let u, v, w, resp. u', v', w' be the (random) literals in C (resp. C'). We have to check that

$$\Pr(u', v', w' \text{ are good}|u, v, w \text{ are good}) = (1 - q)^3(1 + o(1)). \qquad (8)$$

With probability at least $1 - O(1/n)$ it is the case that all the literals u, v, w or their complements are disjoint from u', v', w' and their complements. We can thus apply Proposition 2 which implies that the conditional distributions of the occurences of each of u', v', w' and their complements are asymptotically Poisson and pairwise independent and we can derive (8) just in the same way as we derived it's unconditional analogue.

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 $\rho \leq 16.554$ then output 0.91271 m;
3. if $\rho > 16.554$ then output 7/8 m.

An interesting question arises whether there exist *value approximation algorithms* for MAX-R3SAT with still better approximation ratios than that of [14], 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 existence of a VPTAS for MAX-R3SAT would require though stronger concentration results than the result of this paper.

7. Further research

We already mentioned the open problem of improving substancially the approximation ratio. Another intriguing question is whether approximation ratio for the

measurement of *values* of MAX-R3SAT can be considerably improved (existence of a VPTAS for that problem?).

REFERENCES

[1] D. Achioptas, Setting two variables at a time yields a new lower bound for random 3-SAT, in *Proc. 32th STOC* (2000) 28–37.

[2] N. Alon and J. Spencer, *The Probabilistic Method*. Wiley (1992).

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

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

[5] O. Dubois, R. Monasson, B. Selman and R. Zecchina, eds, Phase Transitions in Combinatorial problems, *Theor. Comput. Sci.* **265** (2001) Nos. 1–2.

[6] J. Friedman, A. Goerdt, Recognizing more unsatisfiable random 3SAT instances efficiently, in *Proc. 28th ICALP* (2001) 310-321.

[7] U. Feige, Relations between Average Case Complexity and Approximation Complexity, in *Proc. 34th ACM STOC* (2002) 534–543

[8] W. Fernandez de la Vega and Marek Karpinski, *9/8 Approximation Algorithm for Random MAX-3SAT*, ECCC tracts, TR02-070, Dec. 13 (2002). Revision accepted Jan. 10 (2003)

[9] E. Friedgut, Necessary and sufficient conditions for sharp thresholds of graph properties and the k-SAT problem. *J. Amer. Math. Soc.* **12** (1999) 1017–1054.

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

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

[12] J. Håstad, Some optimal innasproximability results, in *Proc. 29th ACM STOC* (1997) 1–10.

[13] W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** (1964) 13–30.

[14] Y. Interian, Approximation Algorithm for Random MAX-kSAT, in International Conference on the Theory and Applications of Satisfiability testing (2004).

[15] A. El Maftouhi and W. Fernandez de la Vega, In Random 3-SAT. *Combin. Probab. Comput.* **4** (1995) 189–195.