Note on strong refutation algorithms for random k-SAT formulas

Hiệp Hàn [1], Yury Person [2], and Mathias Schacht

Institut für Informatik
Humboldt-Universität zu Berlin
Unter den Linden 6, D-10099 Berlin, Germany

Abstract

We present a simple strong refutation algorithm for random k-SAT formulas. Our algorithm applies to random k-SAT formulas on n variables with $\omega(n)n^{(k+1)/2}$ clauses for any $\omega(n) \to \infty$. In contrast to the earlier results of Coja-Oghlan, Goerdt, and Lanka (for $k = 3, 4$) and Coja-Oghlan, Cooper, and Frieze (for $k \geq 5$), which address the same problem for even sparser formulas our algorithm is more elementary.

1 Introduction

The k-SAT problem is among the best studied NP-complete problems. We consider strong refutation algorithms for random k-SAT. Let $X_n = \{x_1, \ldots, x_n\}$ be a set of n propositional variables, let $p = p(n) \in [0, 1]$, and let $\mathcal{F}_k(n, p)$ be the probability space over all k-SAT formulas on X_n, for which each of the $(2n)^k$ possible (ordered) k-clauses will be included independently with probability p. It is well-known that for $p \gg n^{1-k}$ with high probability a random formula $F \in \mathcal{F}_k(n, p)$ is not satisfiable. However, there are no efficient refutation algorithms known. We are interested in deterministic algorithms which w.h.p. reject a k-SAT formula from $\mathcal{F}_k(n, p)$ for $p \gg n^{1-k}$, but which never reject a satisfiable formula.

An algorithm is a strong refutation algorithm if w.h.p. for $F \in \mathcal{F}_k(n, p)$ it approximates $\mathrm{unsat}(F)$ by a factor of $(1 - \varepsilon)$ and never outputs a number

[1] Author is supported by DFG within the RTG "Methods for Discrete Structures".
[2] Author is supported by GIF grant no. I-889-182.6/2005

bigger than unsat(F), where unsat(F) is the minimum number of unsatisfied clauses in F over all possible assignments. A simple averaging argument shows

$$\text{unsat}(F) \leq 2^{-k}|F|, \tag{1}$$

where $|F|$ denotes the number of clauses of F. On the other hand, it follows from Chernoff's inequality that for $p \gg n^{1-k}$ w.h.p. $\text{unsat}(F) \geq (2^{-k} - o(1))|F|$ for $F \in \mathcal{F}_k(n,p)$. From a strong refutation algorithm we demand that it verifies this bound on unsat(F).

Definition 1.1 Let $k \geq 3$, $\varepsilon > 0$, and $p = p(n)$. An algorithm $\mathcal{A}$ is an *ε-strong refutation algorithm for $\mathcal{F}_k(n,p)$* if for a given k-SAT formula F on X_n the algorithm $\mathcal{A}$ outputs an integer $\mathcal{A}(F)$ such that

(i) $\mathcal{A}(F) \leq \text{unsat}(F)$ and

(ii) $\lim_{n\to\infty} \mathbb{P}\big(\mathcal{A}(F) \geq (1-\varepsilon)\,\text{unsat}(F)\big) = 1$ for $F \in \mathcal{F}_k(n,p)$.

Note that for $p \gg n^{1-k}$ the trivial algorithm, which returns $(2^{-k} - \varepsilon)|F|$ for every F, satisfies condition (ii), but fails to fulfill (i).

Refutation and strong refutation algorithms were studied by several researchers and to our knowledge the best strong refutation algorithms for $k = 3, 4$ are due to Coja-Oghlan, Goerdt, and Lanka [2] and for general $k \geq 5$ are due to Coja-Oghlan, Cooper, and Frieze [1] (see also [5,4]). Those authors found ε-strong refutation algorithms for every $\varepsilon > 0$ and $p \gg p_k$, where

$$p_k = \begin{cases} n^{-1.5}(\log n)^6 & \text{if } k = 3, \\ n^{-2} & \text{if } k = 4, \\ n^{-\lfloor k/2 \rfloor} & \text{if } k \geq 5. \end{cases}$$

The algorithms from [2] and [1] relied on tools from linear algebra. We present elementary ε-strong refutation algorithms for every $k \geq 3$ for $p \gg n^{-(k-1)/2}$.

Theorem 1.2 *For every $k \geq 3$, $\varepsilon > 0$, and $o(1) = p(n) \gg n^{-(k-1)/2}$ there is an ε-strong refutation algorithm for $\mathcal{F}_k(n,p)$ with running time $O(n^{k2^{k-1}})$ independent of ε.*

2 Proof of Theorem 1.2

Our work is based on results on quasi-random hypergraphs found in [3]. To every $F \in \mathcal{F}_k(n,p)$ we will associate a k-partite, k-uniform hypergraph H_F in the following way. Let X_n be the variables of F. We denote by $V_n =$

$\{x_1, \bar{x}_1, \ldots x_n, \bar{x}_n\}$ the literals of F and let $V(H_F)$ consist of k copies of V_n, i.e., $V(H_F) = V_n \times [k]$. Moreover, the edges of H_F correspond to the clauses of F, i.e., $\{(v_1, 1), \ldots, (v_k, k)\}$ is an edge of H_F if and only if $v_1 \vee \cdots \vee v_k$ is a clause in F. Clearly, this defines a bijection between all k-SAT formulas on X_n (without multiple occurrences of the same clause) and all k-partite, k-uniform hypergraphs with vertex classes $(V_n \times \{1\})\dot{\cup} \ldots \dot{\cup}(V_n \times \{k\})$. Moreover, it is well-known that $\mathrm{unsat}(F)$ is related to the *discrepancy* of H_F.

Definition 2.1 Suppose $H = (V_1\dot{\cup} \ldots \dot{\cup}V_k, E)$ is a k-partite, k-uniform hypergraph with vertex classes of size N and density $p = |E|/N^k$. For $\varepsilon > 0$ we say that H satisfies $\mathrm{DISC}(\varepsilon)$ if for all subsets $U_1 \subseteq V_1, \ldots, U_k \subseteq V_k$

$$\left| e_H(U_1, \ldots, U_k) - p|U_1| \cdots |U_k| \right| < \varepsilon p N^k.$$

Note that every assignment β of the variables $\{x_1, \ldots, x_n\}$ corresponds to a bipartition of each $V_n \times \{i\}$ into equally large sets of literals $U_i = \{(v, i) \in V_n \times \{i\} \colon \beta(v) = 0\}$ and $(V_n \times \{i\}) \setminus U_i$. Furthermore, the number of clauses not satisfied by β corresponds to the number of edges spanned by $U_1\dot{\cup} \ldots \dot{\cup}U_k$. This observation yields the following.

Fact 2.2 *If* H_F *satisfies* $\mathrm{DISC}(\varepsilon)$ *then* $\mathrm{unsat}(F) \geq (2^{-k} - \varepsilon)|F|$. □

The property DISC cannot be naïvely verified in polynomial time. However, for dense hypergraphs it was shown in [3], that the property DISC is equivalent to an efficiently verifiable property, called DEV, which measures the distribution of the homomorphisms of a certain hypergraph M_k. We will show that the implication "*hypergraphs satisfying* DEV *must satisfy* DISC" still holds for hypergraphs of density $p = o(1)$. Theorem 1.2 then follows from the observation that w.h.p. H_F satisfies DEV if $p \gg n^{-(k-1)/2}$.

The hypergraph M_k is the k-uniform, k-partite hypergraph which arises in the following way. For a k-uniform, k-partite hypergraph A with vertex classes $Y_1, \ldots, Y_k$ and $i \in [k]$ we define $\mathrm{db}_i(A)$, the *doubling of* A *w.r.t.* i, to be the k-uniform hypergraph obtained from A by taking two disjoint copies of A and identifying the vertices of Y_i. For the construction of M_k we will start with a single edge K_k, which can be seen as a k-partite k-uniform hypergraph with partition classes of size 1, and iteratively *double* this w.r.t. $i \in [k]$. More precisely, $M_0 = K_k$ and $M_i = \mathrm{db}_i(M_{i-1})$. Thus, the graph M_2 is the 4-cycle and M_k consists of $k2^{k-1}$ vertices and 2^k edges. Further, let $Y_1^j, \ldots, Y_k^j$ denote the partition classes of M_j and for a k-tuple of vertex sets $\mathcal{V} = (V_1, \ldots, V_k)$ we denote by $\mathrm{Hom}(M_j, \mathcal{V})$ all functions $\varphi\colon \bigcup_{i\in[k]} Y_i^j \to \bigcup_{i\in[k]} V_i$ with $\varphi(Y_i^j) \subseteq V_i$ for all $i \in [k]$. In other words, $\mathrm{Hom}(M_j, \mathcal{V})$ is the set of all (partition respecting) homomorphisms from M_j to the complete k-partite, k-uniform hypergraph on the partition classes $V_1\dot\cup\ldots\dot\cup V_k$. For a k-partite, k-uniform hypergraph H with vertex partition $V_1\dot\cup\ldots\dot\cup V_k$ and density p let $w_H\colon \prod_{i\in[k]} V_i \to [-1,1]$ be the function defined by $w_H(e) = 1-p$ if $e \in E(H)$ and $w_H(e) = -p$ otherwise.

Definition 2.3 Suppose $H = (V_1\dot\cup\ldots\dot\cup V_k, E)$ is a k-partite, k-uniform hypergraph with vertex classes of size N and density $p = |E|/N^k$. For $\varepsilon > 0$ we say that H satisfies $\mathrm{DEV}(\varepsilon)$ if for $\mathcal{V} = (V_1,\ldots,V_k)$

$$\big|\sum_{\varphi\in\mathrm{Hom}(M_k,\mathcal{V})}\prod_{e\in E(M_k)} w_H(\varphi(e))\big| \le \varepsilon p^{2^k} N^{k2^{k-1}}.$$

Lemma 2.4 *For every $k \ge 3$ and $\varepsilon > 0$ exists n_0 such that for all $N \ge n_0$ the following holds. Suppose $H = (V_1\dot\cup\ldots\dot\cup V_k, E)$ is a k-partite, k-uniform hypergraph with vertex classes of size N. If H satisfies $\mathrm{DEV}(\varepsilon^{2^k})$, then H also satisfies $\mathrm{DISC}(\varepsilon)$.*

The property $\mathrm{DEV}(\delta)$ can be verified in $O(N^{k2^{k-1}})$ time. Lemma 2.4 combined with Fact 2.2 shows that the algorithm $\mathcal{A}$, which for a k-SAT formula F outputs 0 if H_F fails to satisfy $\mathrm{DISC}(\varepsilon^{2^k})$ and outputs $(2^{-k}-\varepsilon)|F|$ otherwise, fulfills part (i) of Definition 1.1. Moreover, the next lemma combined with (1) shows that the algorithm $\mathcal{A}$ also satisfies part (ii) of Definition 1.1 for $F \in \mathcal{F}_k(n,p)$ with $p \gg n^{-(k-1)/2}$.

Lemma 2.5 *For any $k \ge 3$, $\varepsilon > 0$, and $o(1) = p(n) \gg n^{-(k-1)/2}$ we have $\lim_{n\to\infty} \mathbb{P}(H_F$ satisfies $\mathrm{DEV}(\varepsilon)) = 1$ for $F \in \mathcal{F}_k(n,p)$.*

3 Proofs of Lemmas 2.4 and 2.5

Proof of Lemma 2.4 The proof follows the lines of [3, Lemma 13]. Let $H = (V_1\dot\cup\ldots\dot\cup V_k, E)$ be a k-partite, k-uniform hypergraph with vertex classes of size N and density p, which satisfies $\mathrm{DEV}(\varepsilon^{2^k})$. Let $U_1 \subseteq V_1,\ldots,U_k \subseteq V_k$. We show $|e_H(U_1,\ldots,U_k) - p\prod_{i\in[k]}|U_i|| \le \varepsilon p N^k$. Set $\mathcal{U}_i = (V_1,\ldots,V_i,U_{i+1},\ldots,U_k)$ and for $j \in \{0,\ldots,k\}$ let

$$f_H(M_j,\mathcal{U}_j) = \sum_{\varphi\in\mathrm{Hom}(M_j,\mathcal{U}_j)}\prod_{e\in E(M_j)} w_H(\varphi(e)). \tag{2}$$

Note that by definition $f_H(M_0,\mathcal{U}_0) = e_H(U_1,\ldots,U_k) - p\prod_{i\in[k]}|U_i|$ and, since H satisfies $\mathrm{DEV}(\varepsilon^{2^k})$, $f_H(M_k,\mathcal{U}_k) \le \varepsilon^{2^k} p^{2^k} N^{k2^{k-1}}$. On the other hand, we can rewrite (2) in the following way. For an arbitrary ordering $\boldsymbol{y} = (y_1,\ldots,y_{2^j})$ of the vertices in the j-th vertex class $Y_{j+1}(M_j)$ of M_j, we fix the image of $\boldsymbol{y}$ to be $\boldsymbol{v} = (v_1,\ldots,v_{2^j}) \in U_{j+1}^{2^j}$, i.e. map y_i to v_i for all $i \in [2^j]$, and extend this choice to a homomorphism $\varphi \in \mathrm{Hom}(M_j,\mathcal{U}_j)$. Consequently,

$$f_H(M_j,\mathcal{U}_j) = \sum_{\boldsymbol{v}\in U_{j+1}^{2j}} \sum_{\substack{\varphi\in\mathrm{Hom}(M_j,\mathcal{U}_j)\\ \varphi(\boldsymbol{y})=\boldsymbol{v}}} \prod_{e\in E(M_j)} w_H(\varphi(e)). \tag{3}$$

Recall, that $M_{j+1} = \mathrm{db}_{j+1}(M_j)$ arises from M_j by fixing the $(j+1)$-st vertex class $Y_{j+1}(M_j)$ of M_j and "doubling" all the edges together with the remaining vertices. Thus, applying the Cauchy-Schwarz inequality to $f_H(M_j,\mathcal{U}_j)$ (to the form stated in (3)), we obtain $f_H(M_j,\mathcal{U}_j)^2 \le |U_{j+1}|^{2^j} f_H(M_{j+1},\mathcal{U}_{j+1})$ for every $j\in\{0,\dots,k-1\}$. Applying this inductively for $j=0,\dots,k-1$ we obtain

$$\left|f_H(M_0,\mathcal{U}_0)\right|^{2^k} \le \prod_{i\in[k]} |U_i|^{2^{k-1}} \left|f_H(M_k,\mathcal{U}_k)\right| \le \varepsilon^{2^k} p^{2^k} N^{k2^k}.$$

Consequently, $\left|e(U_1,\dots,U_k) - p\prod_{i\in[k]}|U_i|\right| = |f_H(M_0,\mathcal{U}_0)| \le \varepsilon p N^k$. □

Proof of Lemma 2.5 For $k\ge 3$ and $\varepsilon>0$ let $o(1) = p \gg n^{-(k-1)/2}$. Set $\delta = \varepsilon/(12\cdot 2^{2^k})$ and let $\mathcal{M}$ be the set of all spanning subgraphs of M_k. Let $\mathcal{B}$ be the set of all labeled k-uniform hypergraphs B on $v_B < k2^{k-1}$ vertices such that there is a surjective homomorphism from M_k to B. For a k-partite hypergraph A let X_A be the random variable denoting the number of labeled partition respecting copies of A in H_F with $F\in\mathcal{F}_k(n,p)$.

Claim 3.1 *With high probability we have*

(a) $X_A = (1\pm\delta)\mathbb{E}X_A$ for all $A\in\mathcal{M}$ and (b) $\sum_{B\in\mathcal{B}} X_B < \delta X_{M_k}$.

Proof (sketch) For part (a) we note that, since every vertex of M_k is contained in precisely two edges, the hypergraph M_k is balanced, i.e., $e_{M_k}/v_{M_k} = 2^k/k2^{k-1} = 2/k \ge e_A/v_A$ for all (not necessarily spanning) subhypergraphs $A\subseteq M_k$. Moreover, it is easy to check that for the p considered here, we have $\mathbb{E}X_A \ge \mathbb{E}X_{M_k} \to \infty$ for every $A\in\mathcal{M}$. Hence, part (a) follows easily from Chebyshev's inequality applied in a similar way as, e.g., in [6, Theorem 3.4].

Due to part (a), it suffices to show that w.h.p. $X_B \le \delta p^{2^k}(2n)^{k2^{k-1}}/(2|\mathcal{B}|)$ for every $B\in\mathcal{B}$ to conclude assertion (b). Let $B\in\mathcal{B}$ and set $q = 2^k - e_B$ and $r = k2^{k-1} - v_B$. Hence, $p^{2^k}(2n)^{k2^{k-1}} = (1-o(1))(p(2n)^{r/q})^q\mathbb{E}X_B$ and below we will show that $r\ge (k-1)q/2$, which due to our choice of p yields that $\mathbb{E}X_B = o(p^{2^k}(2n)^{k2^{k-1}})$ and assertion (b) follows from Markov's inequality.

Let $\varphi\colon M_k \to B$ be a surjective homomorphism. For $e\in E(B)$ let $\{f_1,\dots,f_m\} = \varphi^{-1}(e) \subseteq E(M_k)$. Fix f_1 and call f_i, $i\ne 1$, a *lost edge* and any vertex $v\in f_i\setminus f_1$ a *lost vertex*. There are q lost edges and every lost edge contains at least $(k-1)$ lost vertices (f_i and f_1 intersect in at most one vertex, since M_k is a linear hypergraph). On the other hand, the number of lost vertices is at most r and every lost vertex is contained in at most two (lost) edges. Thus, by double counting we have $q(k-1) \leq 2r$. $\square$

We deduce Lemma 2.5 from Claim 3.1. Let $\mathrm{Inj}(M_k, \mathcal{V}) \subseteq \mathrm{Hom}(M_k, \mathcal{V})$ be the set of all injective mappings $\varphi \in \mathrm{Hom}(M_k, \mathcal{V})$. Thus, every $\varphi \in \mathrm{Inj}(M_k, \mathcal{V})$ corresponds to an $\tilde{A} \subseteq H$ which is a labeled copy of some $A \in \mathcal{M}$ in H, whereas any $\varphi \in \mathrm{Hom}(M_k, \mathcal{V}) \setminus \mathrm{Inj}(M_k, \mathcal{V})$ corresponds to a $\tilde{B} \subset H$ which is labeled copy of a hypergraph $B \in \mathcal{B}$. Let $\hat{X}_A$ be the number of induced copies of A. Since $p = o(1)$ we have w.h.p. $\hat{X}_A = (1-o(1))X_A$ and $(1-p)^{k2^{k-1}} \geq 1-\delta$. Since w.h.p. $e(H_F)/(2n)^k = (1+o(1))p$, part (a) of Claim 3.1 yields w.h.p.

$$\sum_{\varphi \in \mathrm{Inj}(M_k,\mathcal{V})} \prod_{e \in E(M_k)} w_H(\varphi(e)) = (1-o(1)) \sum_{A \in \mathcal{M}} (1-p)^{e_A}(-p)^{2^k - e_A} \hat{X}_A$$
$$= p^{2^k} \sum_{A \in \mathcal{M}} (1 \pm 3\delta)(-1)^{2^k - e_A}(2n)^{k2^{k-1}} \leq 6\delta 2^{2^k} p^{2^k}(2n)^{k2^{k-1}} \leq \tfrac{\varepsilon}{2} p^{2^k}(2n)^{k2^{k-1}}.$$

Moreover, due to parts (a) and (b) of the Claim 3.1 w.h.p. we can bound

$$\sum_{\varphi \in \mathrm{Hom}(M_k,\mathcal{V}) \setminus \mathrm{Inj}(M_k,\mathcal{V})} \prod_{e \in E(M_k)} w_H(\varphi(e)) \leq \sum_{B \in \mathcal{B}} X_B \leq \delta X_{M_k} \leq \tfrac{\varepsilon}{2} p^{2^k}(2n)^{k2^{k-1}}.$$

Thus for $F \in \mathcal{F}_k(p,n)$ the hypergraph H_F satisfies w.h.p. $\mathrm{DEV}(\varepsilon)$. $\square$

References

[1] Coja-Oghlan, A., C. Cooper and A. Frieze, *An efficient regularity concept for sparse graphs and matrices*, in: C. Mathieu, editor, *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms* (2009), pp. 207–216.

[2] Coja-Oghlan, A., A. Goerdt and A. Lanka, *Strong refutation heuristics for random k-SAT*, Combin. Probab. Comput. **16** (2007), pp. 5–28.

[3] Conlon, D., H. Hàn, Y. Person and M. Schacht, *Weak quasi-randomness for uniform hypergraphs*, submitted.

[4] Feige, U., J. H. Kim and E. Ofek, *Witnesses for non-satisfiability of dense random 3CNF formulas*, in: *Proceedings of 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, 2006, pp. 497–508.

[5] Feige, U. and E. Ofek, *Easily refutable subformulas of large random 3CNF formulas*, Theory Comput. **3** (2007), pp. 25–43.

[6] Janson, S., T. Łuczak and A. Ruciński, “Random graphs,” Wiley-Interscience, New York, 2000, xii+333 pp.