Electronic Colloquium on Computational Complexity, Report No. 62 (2006)

Unique k-**SAT** is as Hard as k-**SAT**

Subhas Kumar Ghosh
Honeywell Technology Solutions Laboratory
151/1, Doraisanipalya, Bannerghatta Road,
Bangalore, India, 560076
Email:subhas.kumar@honeywell.com

April 24, 2006

Abstract

In this work we show that Unique k-**SAT** is as Hard as k-**SAT** for every $k \in \mathbb{N}$. This settles a conjecture by Calabro, Impagliazzo, Kabanets and Paturi [CIKP03]. To provide an affirmative answer to this conjecture, we develop a randomness optimal construction of Isolation Lemma(see Valiant and Vazirani [VV85]) for k-**SAT**, using $\Omega(\log s)$ number of random bits, to isolate an element of a family of size $\mathcal{O}(2^s)$. It shown by Chari, Rohatgi and Srinivasan in [CRS93], that $\Omega(\log s)$ number of random bits is the lower-bound of randomness complexity for Isolation Lemma.

1 Introduction

The problem of finding a satisfiable assignment (**SAT**) for a propositional formula in **CNF** (conjunctive normal form) is notably the most important problem in theory of computation. The decision problem for **CNF**-**SAT** was one of the first problem shown to be **NP**-complete[Coo71, Lev73]. This problem is mostly believed to require deterministic algorithm of super polynomial, or even exponential, time complexity. Best known algorithm for **CNF**-**SAT** runs in time $\mathcal{O}\left(2^{n\left(1-\frac{1}{\log m}\right)}\right)$, where m is the number of clauses[Sch03]. A syntactically restricted version of general **CNF**-**SAT** is k-**SAT**, where each clause of a given **CNF** contains at most k literals, for some constant $k \geq 3$. This restriction seem to be of help, and existing algorithms have $\mathcal{O}(2^{\epsilon_k n})$ time complexity for some constant $0 < \epsilon_k < n$ dependent on k. Several work exists on faster algorithms for k-**SAT** (cf. [Dan83], [MS85], [PPSZ98], [PPZ97], [Sch99], [Sch02], [HSSW02], [DGH$^+$02]).

While, **NP**-complete search problems are likely to be intractable, it is essentially under the worst-case complexity measures. Not necessarily every instances of k-**SAT**, for a fixed $k \geq 3$ show worst-case behaviors. This leads to an important question: what structural property makes an instance to show such worst-case behaviors? In this paper we look at certain structural properties of the k-**SAT**, as well as the structural properties of its corresponding set of assignments.

One natural question about the hardness of Satisfiability is following: “Is an instance of **CNF**-**SAT** hard because it has smaller or bigger cardinality of solution space?” Intuition

ISSN 1433-8092

seems to provide contradictory ideas. One might feel that the instances having few satisfiable assignments would be harder. However, algorithms using local search techniques might do worst if there are many unrelated solutions that are equally attractive.

Valiant and Vazirani [VV85] has shown that if uniqueness of solution for problems in **NP** helps one to design a polynomial time procedure to recognize them, then **NP** = **RP**. Formally, If $(\exists Q)\,[\mathbf{USAT}_Q \in \mathbf{P}] \Rightarrow \mathbf{NP} = \mathbf{RP}$. Using hashing techniques, they give a randomized polynomial-time reduction from Formula **SAT** to the instances of Formula **SAT** having unique solution. Implying a probabilistic polynomial time algorithm for Unique Formula **SAT** is unlikely.

However, we can ask a slightly different question: "which is harder? Unique Formula **SAT** or Formula **SAT** having more than one solution?" In other words, Satisfiability for formulas with many satisfying assignments and formulas with fewer satisfying assignments, are they strictly easier from one another? While progress towards this question seems harder. In their work Calabro, Impagliazzo, Kabanets and Paturi[CIKP03] has shown that Unique k-**SAT** is as hard to solve as general k-**SAT**. For each $k \geq 1$, let $s_k = \inf\{\delta \geq 0 : \exists$ a $\mathcal{O}\left(2^{\delta n}\right)$-time randomized algorithm for k-**SAT** $\}$ and, similarly let $\sigma_k = \inf\{\delta \geq 0 : \exists$ a $\mathcal{O}\left(2^{\delta n}\right)$-time randomized algorithm for Unique k-**SAT** $\}$. Denoting $s_\infty = \lim_{k\to\infty} s_k$ and $\sigma_\infty = \lim_{k\to\infty} \sigma_k$, main result of [CIKP03] is following:

Theorem 1.1. $s_\infty = \sigma_\infty$.

While they were able to show that $s_\infty = \sigma_\infty$, they conjecture:

Conjecture 1. $\forall k, s_k = \sigma_k$

Main contribution of this work is an *affirmative* answer to this conjecture.

Before we describe our method, it would be important to understand why method used in [CIKP03] does not help to answer this conjecture. Calabro, Impagliazzo, Kabanets and Paturi gave a modified version of so called *Isolation Lemma* of Valiant and Vazirani[VV85] for k-**CNF**s. We start by describing Isolation Lemma of Valiant and Vazirani[VV85], followed by Isolation Lemma for k-**CNF** in [CIKP03].

Lemma 1.1. Isolation Lemma [VV85]: *There is a randomized polynomialtime algorithm I that, given an n-variable* **CNF** ϕ*, outputs an* $n^{o(1)}$*-variable* **CNF** ϕ'*, such that*

1. *if* ϕ *is unsatisfiable, then so is* ϕ'*, and*
2. *if* ϕ *is satisfiable, then, with probability at least* $1/4$*,* ϕ' *has a unique satisfying assignment.*

[VV85] describes a polynomial time randomized reduction, which takes a Boolean formula ϕ as input and outputs a Boolean formula ϕ' such that with probability at least inverse polynomial in size of ϕ, ϕ' has unique satisfiable assignment, if and only if input ϕ is a satisfiable formula. Output of the reduction is in Unique Formula **SAT**, where Unique Formula **SAT** is the set of all Boolean formulas (not necessarily **CNF**) having exactly one satisfiable assignment. Let ϕ be any **CNF** in n variables $x_1, x_2, \ldots, x_n$. Let $[n] = \{1, 2, \ldots, n\}$ be the coordinates of n-dimensional solution space of ϕ. For any **CNF** ϕ in n variables, we define a witness set of ϕ

$$S \stackrel{\Delta}{=} \{s : (s \in \{0,1\}^n)\,[s \text{ is a satisfiable assignment of } \phi]\}$$

We will view truth assignments to the variables $x_1, x_2, \ldots, x_n$ as n-dimensional $\{0,1\}$ vectors from the vector space $\mathbb{F}_2^n$, where $\mathbb{F}_2 = GF(2)$. For $u, v \in \{0,1\}^n$ we denote by $u \cdot v$ the inner product over $\mathbb{F}_2$ of u and v. Isolation Lemma of [VV85] is based on two important observations.

1. If ϕ is any **CNF** in n variables $x_1, x_2, \ldots, x_n$, and $w_1, \ldots, w_k \in \{0,1\}^n$ then one can construct in linear time a formula ϕ'_k whose solution s' satisfy ϕ and have $s' \cdot w_1 = s' \cdot w_2 = \ldots = s' \cdot w_k = 0$, and from ϕ'_k one can construct a formula ϕ_k with $n^{o(1)}$ new variables such that there is a bijection between solutions of ϕ'_k and ϕ_k defined by equality on the $x_1, x_2, \ldots, x_n$ values.

2. If ϕ is any **CNF** in n variables $x_1, x_2, \ldots, x_n$, and $S \subseteq \{0,1\}^n$ is its solution space, for $w \in \{0,1\}^n$, define $S^{(0)} \triangleq \{s : (s \in S)[w \cdot s = 0]\}$ and $S^{(1)} \triangleq \{s : (s \in S)[w \cdot s = 1]\}$. Then $\mathbf{Pr}\left[\left|S^{(0)}\right| = \left|S^{(1)}\right|\right]$ is very high.

These two observation leads to a **RP** reduction of Formula **SAT** to Unique Formula **SAT**. On input ϕ randomly choose $k \in [n]$ and choose k vectors $w_1, \ldots, w_k \in \{0,1\}^n$ and intersect the solution space S of ϕ. Define formula $\phi'_k \triangleq \phi \wedge \left(x_{i_1} \oplus \ldots \oplus x_{i_j} \oplus 1\right)$, where x_{i_k} are the variables that have value 1 in w_i. If s' satisfies ϕ'_k, and $w_i \cdot s' = 0$, then s' satisfies ϕ. Now define formula $\phi_k = \phi \wedge (y_1 \Leftrightarrow x_{i_1} \oplus x_{i_2}) \wedge (y_2 \Leftrightarrow y_1 \oplus x_{i_3}) \wedge \ldots \wedge \left(y_{j-1} \Leftrightarrow y_{j-2} \oplus x_{i_j}\right) \wedge (y_{j-1} \oplus 1)$. Clearly number of new variables are $n^{o(1)}$, and bijection in solution space of ϕ_k and ϕ'_k is straightforward. Now consider following sets: $H_i = \{v : v \cdot w_i = 0\}$, $S \subseteq \{0,1\}^n$, the solution space of ϕ, and $S_i = S \cap H_1 \cap \ldots \cap H_i$. Since $\mathbf{Pr}\left[\left|S^{(0)}\right| \neq \left|S^{(1)}\right|\right]$ is very high it follows that ϕ_i has roughly $2^{-i}|S|$ solutions, as each H_i roughly halves the solution space S.

However, their proof does not imply Unique k-**SAT** is the worst case of k-**SAT**, or Unique **CNF-SAT** is the worst case of **CNF-SAT**. As shown in [CIKP03], by reduction of [VV85], resulting formula ψ will not be a k-**CNF** even if the input formula ϕ is a k-**CNF**. After applying standard algorithm for converting any Boolean formula into k-**CNF** we will have a formula with $\Omega\left(n^2\right)$ new variables. Thus using their reduction we will convert a satisfiable k-**CNF** formula on n-variables to a uniquely satisfiable k-**CNF** on $\Omega\left(n^2\right)$-variables. If there is an algorithm for Unique k-**SAT** on n variables with running time $\mathcal{O}\left(2^{n^{o(1)}}\right)$, then reduction in [VV85] can be used to create an algorithm for general k-**SAT** with running time $\mathcal{O}\left(2^{n^{o(1)}}\right)$. However, this reduction will not allow us the answer the kind of question we asked above, if working hypothesis is that Unique k-**SAT** requires time $2^{\Omega(n)}$. Because, by their reduction a $\mathcal{O}\left(2^{\sqrt{n}}\right)$ algorithm for Unique k-**SAT** on n variables implies a $\mathcal{O}\left(2^{\Omega(n)}\right)$ algorithm for k-**SAT** on n variables.

Lemma 1.2. Isolation Lemma for k-CNFs [CIKP03]: *For every $k \in \mathbb{N}, \epsilon \in \left(0, \frac{1}{4}\right)$, there is a randomized polynomialtime algorithm $I_{k,\epsilon}$ that, given an n-variable k-***CNF** *ϕ, outputs an n-variable k'-***CNF** *ϕ' for $k' = \max\left\{\left\lceil \frac{1}{\epsilon} \ln \frac{2}{\epsilon} \right\rceil, k\right\}$, such that*

1. *if ϕ is unsatisfiable, then so is ϕ', and*

2. *if ϕ is satisfiable, then, with probability at least $\left(32n2^{3nH_2(2\epsilon)}\right)^{-1}$, ϕ' has a unique satisfying assignment.*

where, $H_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$*, is the binary entropy function.*

Isolation Lemma of [CIKP03] results in a randomized reduction from k-**SAT** on n variables to Unique k-**SAT** on n variables with expected time $\mathcal{O}\left(2^{\epsilon_k n}\right)$, where $\epsilon_k \to 0$, as $k \to \infty$. This implies that if general k-**SAT** requires time $2^{\delta n}$ for some k and $\delta > 0$, then for any $\delta' < \delta$, Unique k'-**SAT** requires time $2^{\delta' n}$ for sufficiently large k'.

However, Isolation Lemma of [CIKP03] takes a family of size $\Omega\left(2^s\right)$ to isolate an element of an arbitrary set of size 2^s, by intersecting with randomly chosen sets from the family, and we can only conclude that $s_\infty = \sigma_\infty$. To show that $\forall k, s_k = \sigma_k$, we need a stronger form of reduction which expands the number of variable by polynomial in input size, and has success probability inverse polynomial in input size. In other words, one need to construct polynomial size families, i.e. $\Omega\left(s^{o(1)}\right)$ size families, to isolate an element of an arbitrary set of size 2^s.

1.1 Results

For each $k \geq 1$, let $s_k = \inf\{\delta \geq 0 : \exists$ a $\mathcal{O}\left(2^{\delta n}\right)$-time randomized algorithm for k-**SAT** $\}$ and, similarly let $\sigma_k = \inf\{\delta \geq 0 : \exists$ a $\mathcal{O}\left(2^{\delta n}\right)$-time randomized algorithm for Unique k-**SAT** $\}$. We can state our main result as follows:

Theorem 1.2. $\forall k, s_k = \sigma_k$

Remainder of this paper : In Section-2 we state and prove our main technical result, an Isolation Lemma for k-**CNF**s. Section-3 contains the proof of main result (Theorem-1.2).

2 Isolation Lemma for k-CNFs

Objective of this section will be to prove the following Lemma:

Lemma 2.1. (Isolation Lemma for k-CNFs). *Let q be a prime, and let $2q^2 < n$. For every $k \in \mathbb{N}, \epsilon \in \left[0, \frac{1}{4}\right]$, there is a randomized polynomial time algorithm $I_{k,\epsilon}$ that, given an n-variable k-**CNF** ϕ, outputs an n-variable k'-**CNF** ψ' for $k' = \max\left\{\left\lceil\frac{\ln 2}{\epsilon}\right\rceil, k\right\}$, such that*

1. *if ϕ is unsatisfiable, then so is ψ', and*
2. *if ϕ is satisfiable, then, with probability at least $1/2$, ψ' has a unique satisfying assignment.*

Proof. Proof will have three steps. Let ϕ be any **CNF** in n variables $x_1, x_2, \ldots, x_n$. Let $[n] = \{1, 2, \ldots, n\}$ be the coordinates of n-dimensional solution space of ϕ. For any **CNF** ϕ in n variables, we define a witness set of ϕ

$$S \triangleq \{s : (s \in \{0,1\}^n)\,[s \text{ is a satisfiable assignment of } \phi]\}$$

In the first step we will intersect S, by a family of sets $\mathcal{A} = \{A_1, \ldots, A_m\}$. We will define constraints on the intersection that S will have with $\mathcal{A}$. Let C denote such constraint, if $\psi = \phi \wedge (C = \mathit{true})$ is a formula such that $s \in S$ is a common solution for both ϕ and

ψ, then in first step we will show that all such solutions will be concentrated in a *small* solution space. In second step we will show how to isolate a single assignment from this space. Third step will be to provide a construction of the family $\mathcal{A}$, as required by the first step. With probability at least $1/n$, we can guess an integer $s \in [\log |S|, \log |S| + 1]$, for rest of our argument, let us assume that our guess is correct, and $|S| = 2^s$.

STEP-I: CONCENTRATION

Let $[n] = \{1, 2, \ldots, n\}$, q be a prime, $\mathbb{F}$ be a finite field, and let $L \subseteq \{0, 1, \ldots, q-1\}$ be a set of t integers. For integers r and q we say $r \in L \pmod{q}$ if $r \equiv l \pmod{q}$ for some $l \in L$, we say $r \notin L \pmod{q}$ otherwise. For m to be determined later, let $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ be a family of m sets such that following holds:

1. $(\forall i : 1 \le i \le m)\,[A_i \subseteq [n]]$.
2. $(\forall i : 1 \le i \le m)\,[|A_i| \notin L \pmod{q}]$.
3. $(\forall i, j : 1 \le i \neq j \le m)\,[|A_i \cap A_j| \in L \pmod{q}]$.

Such set system with modular intersection has been considered by Frankl and Wilson in [FW81], and Alon in [Alo98], towards explicit construction of Ramsey Graphs. Our proof is based on techniques used in [ABS91], and construction is essentially the construction of [FW81]. With each vector $x \in \{0,1\}^n$ we associate its characteristic vector $x = (x_1, x_2, \ldots, x_n) \in \{0,1\}^n \subseteq \mathbb{F}^n$. For $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$, let $a_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \in \{0,1\}^n \subseteq \mathbb{F}^n$ be the characteristic vector of A_i. Where, $a_{ij} = 1$ if $j \in A_i$ and $a_{ij} = 0$ otherwise. For $x, y \in \mathbb{F}^n$, let $x \cdot y = \sum_{i=1}^{n} x_i \cdot y_i$, denote their standard inner product over $\mathbb{F}^n$. For each $A_i \in \mathcal{A}$ define the following polynomial:

$$f_i(x) \triangleq \prod_{l \in L} (a_i \cdot x - l)$$

Claim 2.1. *Polynomials $f_1, f_2, \ldots, f_m$, are linearly independent.*

Proof. To observe this assume contrary, and let $\sum_{i=1}^{m} \lambda_i f_i(x) = 0$ be a nontrivial linear dependence relation, where $\lambda_i \in \mathbb{F}$. Let i_0 be smallest subscript such that $\lambda_{i_0} \neq 0$. We substitute a_{i_0} for x in the linear relation. Observe that all terms vanishes since $\forall i, j : i \neq j, |A_i \cap A_j| \in L \pmod{q}$, except $f_{i_0}(a_{i_0})$ since $|A_i| \notin L \pmod{q}$ with consequence $\lambda_{i_0} = 0$, a contradiction. □

A polynomial is multilinear if it has degree in each variable at most 1. A monic multilinear polynomial is the product of distinct variables.

Proposition 2.1. *Let $\mathbb{F}^n$ be a field and $\Omega = \{0,1\}^n \subseteq \mathbb{F}^n$. If g is a polynomial of degree $\le t$ in n variables over $\mathbb{F}^n$ then there exists a multilinear polynomial $\hat{g}$ of degree $\le t$ in the same variables such that $\forall x \in \Omega, g(x) = \hat{g}(x)$.*

Proof. Since in the domain Ω, $x_i^2 = x_i$ for each variable, every polynomial is multilinear: expand the polynomial as sum of monomials and for each monomial reduce the exponent of each variable occurring in monomial to 1. □

Claim 2.2. *Polynomials $f_1, f_2, \ldots, f_m$ are basis of a space of dimension $\sum_{i=0}^{t} \binom{n}{i}$. In other words*

$$|\mathcal{A}| \leq \sum_{i=0}^{t} \binom{n}{i}$$

Proof. By proposition-2.1, each polynomial $f_i(x)$ remains valid after replacing f_i by corresponding multilinear polynomial $\hat{f}_i$. Each $\hat{f}_i$ is of degree $\deg \hat{f}_i \leq |L| = t$. That is, each $\hat{f}_i$ is a linear combination of monic multilinear polynomials with $\deg \hat{f}_i \leq t$, and they form a basis of a space of dimension at most $\leq \sum_{i=0}^{t} \binom{n}{i}$. □

We form the following equation with the family of sets $\mathcal{A}$ and solution set S as random constraint:

$$f(x) \triangleq \sum_{A_i \in \mathcal{A}} \lambda_i \prod_{l \in L} (a_i \cdot x - l) = b. \text{ where } \lambda_i, b \in \mathbb{F} \tag{1}$$

by choosing $b \in \mathbb{F}$ and each $\lambda_i \in \mathbb{F}$ independently and uniformly at random. We shall bound the probability that two vectors $x, y \in \{0,1\}^n$ that are reasonably *far* from each other will be separated by this constraint. Let $dist(x, y)$, be the Hamming distance between x, y, i.e. number of coordinates in which x and y differ.

Claim 2.3. *For any two vectors $x, y \in \{0,1\}^n$ with $dist(x, y) \geq r = \lceil \epsilon n \rceil$ we have*

$$\mathbf{Pr}[f(x) = f(y)] \leq \left(\frac{t}{\sqrt{2}q^2}\right)^{2m}$$

Proof. For any two vectors $x, y \in \{0,1\}^n$ we have

$$\mathbf{Pr}[f(x) = f(y)] = \mathbf{Pr}[f(x) - f(y) \equiv 0 \pmod{q}]$$

Now, $f(x) = \sum_{A_i \in \mathcal{A}} \lambda_i f_i(x)$, hence $f(x) - f(y) = \sum_{A_i \in \mathcal{A}} \lambda_i [f_i(x) - f_i(y)]$. Recall, by claim-2.1 all f_i are linearly independent. Thus, when not all $\lambda_i = 0$, if we have $f(x) = f(y)$, following must hold: $\forall i : (f_i(x) = f_i(y)) [\lambda_i \neq 0]$. We have

$$\mathbf{Pr}[f(x) = f(y)] = \prod_{i=1}^{m} \left(\mathbf{Pr}[f_i(x) = f_i(y) | \lambda_i \neq 0] \cdot \mathbf{Pr}[\lambda_i \neq 0]\right)$$

Since each λ_i was chosen uniformly and independently.

$$\mathbf{Pr}[f(x) = f(y)] = \left(\frac{1}{q}\right)^m \prod_{i=1}^{m} \mathbf{Pr}[f_i(x) = f_i(y)]$$

Let us consider the event $f_i(x) = f_i(y)$. Recall, $f_i(w) = \prod_{l \in L} (r - l)$, where $r = a_i \cdot w = \sum_{j=1}^{n} a_{ij} \cdot w_j$. Each f_i is likely to separate two vector $x, y \in \{0,1\}^n$ even when they have very less Hamming distance due to modulo q computation, but if $|x \cap A_i| \in L \pmod{q}$ and

$|y \cap A_i| \in L \pmod{q}$ for some i we may have $f_i(x) = f_i(y)$, and we condition on this, i.e. both $r, r' \notin L \pmod{q}$. We have,

$$\begin{aligned} & \mathbf{Pr}\left[f_i(x) = f_i(y)\right] \\ \leq \; & \mathbf{Pr}\left[f_i(x) = f_i(y) \,|r, r' \notin L \pmod{q}\right] \mathbf{Pr}\left[r, r' \notin L \pmod{q}\right] \\ \leq \; & \left(\frac{t}{q}\right)^2 \mathbf{Pr}\left[f_i(x) = f_i(y) \,|r, r' \notin L \pmod{q}\right] \end{aligned}$$

It must also be that $(x-y)_{A_i} \neq 0$, i.e. over the coordinates in A_i, both vectors are not zero, given that $dist(x,y) \geq r$.

$$\begin{aligned} \leq \; & \left(\frac{t}{q}\right)^2 \mathbf{Pr}\left[f_i(x) = f_i(y) \,|\, (x-y)_{A_i} \neq 0\right] \mathbf{Pr}\left[(x-y)_{A_i} \neq 0\right] \\ \leq \; & \left(\frac{1}{q}\right)\left(\frac{t}{q}\right)^2 \mathbf{Pr}\left[\exists j \in A_i : (x-y)_j \neq 0\right] \\ \leq \; & \left(\frac{1}{q}\right)\left(\frac{t}{q}\right)^2 \left(1 - \frac{\binom{n-r}{|A_i|}}{\binom{n}{|A_i|}}\right) \end{aligned}$$

Note that vectors x, y are in $\{0,1\}^n \subseteq \mathbb{F}^n$, thus above estimate holds.

$$\mathbf{Pr}\left[f_i(x) = f_i(y)\right] \;\leq\; \left(\frac{1}{q}\right)\left(\frac{t}{q}\right)^2 \left(1 - (1-\epsilon)^{|A_i|}\right)$$

With $k' = |A_i| = \frac{\ln 2}{\epsilon}, \forall i$ we have,

$$\begin{aligned} & \mathbf{Pr}\left[f(x) = f(y)\right] \\ \leq \; & \left(\frac{1}{q}\right)^{2m}\left(\frac{t}{q}\right)^{2m} \left(1 - (1-\epsilon)^{\frac{\ln 2}{\epsilon}}\right)^m \\ \leq \; & \left(\frac{t}{q^2}\right)^{2m} \left(1 - e^{-\epsilon\left(\frac{\ln 2}{\epsilon}\right)}\right)^m \\ \leq \; & \left(\frac{t}{q^2}\right)^{2m} 2^{-m} \end{aligned}$$

We have $\mathbf{Pr}\left[f(x) = f(y)\right] \leq \left(\frac{t}{\sqrt{2}q^2}\right)^{2m}$, as required. □

Now we define the set of semi-isolated solutions as

$$si = \{x \in S : (\forall y \in S)\left[dist(x,y) \geq r \rightarrow f(x) \neq f(y)\right]\} \tag{2}$$

For each $x \in S$, we have

$$\begin{aligned} \mathbf{Pr}_f\left[x \in si\right] &= 1 - \mathbf{Pr}_f\left[\exists y \in S\left(dist(x,y) \geq r \wedge f(x) = f(y)\right)\right] \\ &\geq 1 - \sum_{y \in S, dist(x,y) \geq r} \mathbf{Pr}_f\left[f(x) = f(y)\right] \\ &\geq 1 - |S|\left(\frac{t}{\sqrt{2}q^2}\right)^{2m} \end{aligned}$$

Where, the last inequality follows from Claim-2.3. By setting,

$$m = \left(\frac{s}{\log q}\right), \text{ with choice of } t = q-1 \text{ as in Step-III}$$

we obtain:

$$\mathbf{Pr}_f\left[x \in si\right] \geq 1 - \frac{1}{2^s} \tag{3}$$

Alternatively,

$$\begin{aligned} & \sum_{x \in S} \mathbf{Pr}_{f,b}\left[x \in si \wedge f(x) = b\right] \\ = & \sum_{x \in S} \mathbf{Pr}_f\left[x \in si\right] \mathbf{Pr}_b\left[f(x) = b | x \in si\right] \\ \geq & \; |S| \left(1 - \frac{1}{2^s}\right)\left(\frac{1}{q}\right) \end{aligned} \tag{4}$$

where, last inequality follows from equation-3, and by the fact that b was chosen uniformly and independently at random.

On the other hand, given f, and b, number of semi-isolated solutions $x \in S$ such that $f(x) = b$ is at most $2^{nH_2(\epsilon)}$, since, every pair of such solutions must be Hamming distance less than r apart, where, $H_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$, is the binary entropy function. Implying:

$$\sum_{x \in S} \mathbf{Pr}_{f,b}\left[x \in si \wedge f(x) = b\right] \leq 2^{nH_2(\epsilon)} \mathbf{Pr}_{f,b}\left[(\exists x \in S)\left[x \in si \wedge f(x) = b\right]\right] \tag{5}$$

Combining inequality-4 and inequality-5, we obtain:

$$\begin{aligned} \mathbf{Pr}_{f,b}\left[(\exists x \in S)\left[x \in si \wedge f(x) = b\right]\right] & \geq |S|\left(1 - \frac{1}{2^s}\right)\left(\frac{1}{q}\right) 2^{-nH_2(\epsilon)} \\ & \geq q^{-1} 2^{s - nH_2(\epsilon)} \end{aligned} \tag{6}$$

This completes STEP-I of the lemma.

STEP-II: ISOLATION

Let ϕ be any **CNF** in n variables $x_1, x_2, \ldots, x_n$. Suppose that formula $\psi = \phi \wedge (f(x) = b)$ is satisfiable, and solutions are in a Hamming ball of radius less than r, and diameter $d \leq 2\lfloor \epsilon n \rfloor$. Where, $f(x) = b$ is a random constraint as defined in equation-1. Consider assignments u and v to disjoint set of variables of ϕ. Let uv denote union of variables in u and v. Let uv and uv' denote two distinct solutions of $\phi \wedge (f(x) = b)$ that are farthest apart. Let C be the set of variables on which uv and uv' differ. Surely, $|C| \leq d$. Fix $D \supseteq C$ such that $|D| = d$. Let β_D be an assignment to the variables in D such that $\forall i \in C, x_i = v_i$ and $\forall i \in D \setminus C, x_i = u_i$. Our reduction algorithm chooses a random set D' with $|D'| = d$, and a random assignment $\alpha_{D'}$ to the variables in D'. Let $\psi' = \phi \wedge (f(x) = b) \wedge (x_{D'} = \alpha_{D'})$. Now observe that by the arguments above, if $D' = D$ and $\alpha_{D'} = \beta_D$, then uv is the unique solution to ψ', as any other solution wv to ψ' would also be a solution to $\psi = \phi \wedge (f(x) = b)$, but $dist(wv, uv') > dist(uv, uv')$. Finally, probability of guessing D and β_D is at least:

$$\frac{2^{-d}}{\binom{n}{d}} \geq 2^{-n(2\epsilon + H_2(2\epsilon))} \tag{7}$$

Combining equation-7 with Step-I (equation-6), and noting that with probability at least $1/n$ we can guess an integer $s \in [\log |S|, \log |S| + 1]$, for $\psi' = \phi \wedge (f(x) = b) \wedge (x_{D'} = \alpha_{D'})$ we have:

$$\begin{aligned}\mathbf{Pr}\left[\psi' \text{ has unique solution}\right] &\geq 2^{-n(2\epsilon + H_2(2\epsilon))} (qn)^{-1} 2^{s - nH_2(\epsilon)} \\ &\geq (qn)^{-1} 2^{s - 2nH_2(2\epsilon)}\end{aligned}$$

With this we conclude STEP-II of the lemma.

STEP-III: CONSTRUCTION

In this step we will show how to construct such families following ideas of [FW81]. Let $n > 8$, and $n > 2q^2$. Consider set of all subsets of $[n]$ of cardinality $q^2 - 1$. We will consider only those subsets for which following hold:

$$|A \cap B| \not\equiv -1 \pmod q, \text{ and }, A, B \subseteq [n], \text{ and }, |A|, |B| = q^2 - 1$$

Imagine a graph $G(n, q)$, on these subsets with vertex set as set of all subsets of $[n]$ of cardinality $q^2 - 1$, and edge between any two subset is included iff for $i \neq j, |A_i \cap A_j| \not\equiv -1 \pmod q$, holds.

Claim 2.4. *A clique on graph $G(n, q)$ satisfies the required properties of the family of sets $\mathcal{A} = \{A_1, \dots, A_r\}$, with $L = \{0, 1, \dots, q-2\}$ and $|L| = q - 1$.*

Proof. If $|A_i \cap A_j| \not\equiv -1 \pmod q$ for all $i \neq j$, then $|A_i \cap A_j| \in L \pmod q$, when $L = \{0, 1, \dots, q-2\}$. Also, $|A_i| \notin L \pmod q, \forall i$ since $\forall i, |A_i| = q^2 - 1 \notin L \pmod q$. □

Claim 2.5. *Let q be a prime, and $n > 2q^2$. Graph $G(n, q)$ has $\binom{n}{q^2-1}$ vertices and no clique of size more than $2\binom{n}{q-1}$.*

Proof. By claim-2.2

$$|\mathcal{A}| \leq \sum_{i=0}^{t} \binom{n}{i}$$

Now note that $|L| = t = q - 1 \leq n/4$, as $2q^2 < n$. With $t \leq n/4$, we have that

$$|\mathcal{A}| \leq \sum_{i=0}^{q-1} \binom{n}{i} \leq \binom{n}{q-1} \left[1 + \sum_{i=0}^{q-2} \frac{\binom{n}{i}}{\binom{n}{q-1}}\right] \leq 2\binom{n}{q-1}$$

□

Claim-2.4, and Claim-2.5 together shows the existence of the family of sets with desired properties. By [FW81], with $n = q^3$, we can have a graph $G(n, q)$ on $q^{\mathcal{O}(q^2)}$ vertices having clique size no more than $q^{\mathcal{O}(q)}$, choosing $n = q^3$.

On the other hand we can identify the set $[n]$ as elements of $\mathbb{F}^3$, with subsets of $[n]$ represented as $\langle a, b, f_i(a, b)\rangle : a, b \in \mathbb{F}, f_i \in \mathbb{F}[x, y]$. Then choosing a set of linearly independent polynomials of degree at most t we can construct the desired family $\mathcal{A}$.

Now we combine all three steps of the lemma to conclude as claimed. We have shown that the formula ψ' is satisfiable iff ϕ is, moreover,

$$\mathbf{Pr}\left[\psi' \text{ has unique solution}\right] \geq (qn)^{-1} 2^{s - 2nH_2(2\epsilon)}$$

Recall, we have $m \log q = s$, and we have gussed s. This implies that when we have $\epsilon \in [0, 1/4]$, then with $m = 2n/\log q + \log n/\log q - 1/\log q + 1 = o\left(n/\log q\right)$ and choosing q such that $q^3 = n$, above probability can be bounded by $1/2$. To be specific:

$$\begin{aligned} \mathbf{Pr}\left[\psi' \text{ has unique solution}\right] &\geq (qn)^{-1}\, 2^{m \log q - 2nH_2(2\epsilon)} \\ &\geq (qn)^{-1}\, 2^{m \log q - n} \\ &\geq \frac{1}{2} \end{aligned}$$

We have also shown that such families are explicitly constructable, i.e. there exists an algorithm that given n, m and q, outputs $A_i \in \mathcal{A}$ in time $T = \text{poly}\left(\log m, q\right)$. It is interesting to note that, this construction generalizes the method of [VV85]. In fact observe that in the previous step, choosing $q = 2, L = \{0\}$ makes the construction equivalent to [VV85], and $m = o(n)$ ensures that $\mathbf{Pr}_{f,b}\left[\left(\exists x \in \mathcal{F}\right)\left[x \in si \wedge f\left(x\right) = b\right]\right] \geq 1/2$ (compare this with Theorem-4 and Theorem-5 in [VV85]). □

3 Application of Isolation Lemma

In this section we prove the Main Theorem (Theorem-1.2).

Lemma 3.1. $s_k \leq \sigma_k + \mathcal{O}\left(1/n\right)$.

Proof. Let $\epsilon \leq \ln 2/k$. For every $\gamma > 0$ and every k, we have a randomized algorithm $U_{k,\gamma}$ for Unique k-**SAT** that runs in time $\mathcal{O}\left(2^{(\sigma_k+\gamma)n}\right)$. Now given any k-**CNF** ϕ we apply Isolation Lemma - 2.1 to obtain a ψ' using our randomized polynomial time reduction algorithm $I_{k,\epsilon}$ with success probability $p \geq 1/2$, that if ϕ is satisfiable, then, with probability at least $1/2$, ψ' has a unique satisfying assignment. Running $I_{k,\epsilon}$, $\mathcal{O}\left(p^{-1}\right)$ times with same input ϕ, and then running algorithm $U_{k,\gamma}$ on the obtained formula ψ' on each run, we obtain a randomized algorithm for k-**SAT** that has running time $\mathcal{O}\left(2 \cdot 2^{(\sigma_k+\gamma)n}\right) \leq \mathcal{O}\left(2^{(\sigma_k+\gamma+\mathcal{O}(1/n))n}\right)$. As $\gamma \to 0$, lemma follows. □

Observe that our Main Theorem-1.2, $\forall k \in \mathbb{N}, s_k = \sigma_k$, follows directly from the proof of the Lemma-3.1, by taking sufficiently large n, as $1/n \to 0$ when $n \to \infty$.

4 Comments

One important aspect of Isolation Lemma is its randomness complexity, i.e. the number of random bits required to isolate a vector. Isolation Lemma of [VV85] uses quadratically many numbers of random bits. In [BCGL89] it was observed that one can use $universal_2$ ([CW79]) families of Hash functions in the reduction of [VV85], and using such family $\mathcal{H} = \{h_k \in \mathcal{H}_k\}, h_k : \{0,1\}^n \to \{0,1\}^k \; (1 \leq k \leq q-1)$, can reduce the number of random bits to $2q\left(n\right)$. Construction of [CRS93] uses $\mathcal{O}\left(\log z + \log n\right)$ random bits to isolate a family of size z over ground set of cardinality n. Naik, Regan, and Sivakumar [NRS95] developed a reduction scheme that uses a quasilinear number of bits to achieve constant success probability. In [CRS93] it was also shown that any randomized scheme requires $\Omega\left(\log z\right)$ random bits to isolate a family of size z. In [CRS93], consideration was, if it matters that the size of the family is known. Isolation Lemma presented here for k-**CNF**s uses $\Omega(s)$ random bits to isolate a family of size 2^s when s is known, as $m = s/\log q$, $b \in \mathbb{F}$, and all $\lambda_i \in \mathbb{F}$, we have $\Omega(m \log q) = \Omega(s)$ randomness complexity. Also, from the randomness requirement it is very clear that it does matter to know s. When s is unknown, we require, $\Omega(n)$ random bits. Recently Klivans and Spielman [KS01] in their work on identity testing of multivariate polynomials considered the use of Isolation Lemma. To provide a randomized reduction from non-zero multivariate polynomials to non-zero univariate polynomials they extend the result in [CRS93] to set systems of linear form over a base set, with comparable randomness complexity that of [CRS93].

References

[ABS91] N. Alon, L. Babai, and H. Suzuki. Multilinear polynomials and frankl-ray-chaudhuri-wilson type intersection theorems. *J. Comb. Theory Ser. A*, 58(2):165–180, 1991.

[Alo98] Noga Alon. The shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998.

[BCGL89] S. BenDavid, B. Chor, O. Goldreich, and M. Luby. On the theory of averagecase complexity. In *STOC '89: In Proc. 21st Annual ACM Symposium on the Theory of Computing*, pages 204–216, New York, NY, USA, 1989. ACM Press.

[CIKP03] Chris Calabro, Russell Impagliazzo, Valentine Kabanets, and Ramamohan Paturi. The complexity of unique k-**SAT**: An isolation lemma for k-**CNF**s. In *18th IEEE Annual Conference on Computational Complexity*, pages 135–141, 2003.

[Coo71] Stephen A. Cook. The complexity of theorem-proving procedures. In *STOC '71: Proceedings of the third annual ACM symposium on Theory of computing*, pages 151–158, New York, NY, USA, 1971. ACM Press.

[CRS93] S. Chari, P. Rohatgi, and A. Srinivasan. Randomnessoptimal unique element isolation, with applications to perfect matching and related problems. In *STOC '93: In Proc. 25th Annual ACM Symposium on the Theory of Computing*, pages 458–467, New York, NY, USA, 1993. ACM Press.

[CW79] J. Carter and M. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18:143–154, 1979.

[Dan83] E. Dantsin. Two propositional proof systems based on the splitting method. *Zapiski Nauchnykh Seminarov LOMI, 105:24-44, 1981. (in Russian),English translation in Journal of Soviet Mathematics*, 22(3):1293–1305, 1983.

[DGH$^+$02] Evgeny Dantsin, Andreas Goerdt, Edward A. Hirsch, Ravi Kannan, Jon Kleinberg, Christos Papadimitriou, Prabhakar Raghavan, and Uwe Schöning. A deterministic $(2 - 2/(k+1))^n$ algorithm for k-**SAT** based on local search. *Theor. Comput. Sci.*, 289(1):69–83, 2002.

[FW81] Peter Frankl and Richard Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1:357–368, 1981.

[HSSW02] T. Hofmeister, U. Schöning, R. Schuler, and O. Watanabe. A probabilistic 3−**SAT** algorithm further improved. In *In Proceedings of the Nineteenth Annual Symposium on Theoretical Aspects of Computer Science*, pages 192–202, 2002.

[KS01] Adam R. Klivans and Daniel Spielman. Randomness efficient identity testing of multivariate polynomials. In *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 216–223, New York, NY, USA, 2001. ACM Press.

[Lev73] L. Levin. Universal'nyie perebornyie zadachi (universal search problems: in russian). *Problemy Peredachi Informatsii, English translation in [Tra84]*, 9(3):265–266, 1973.

[MS85] B. Monien and E. Speckenmeyer. Solving satisfiability in less than 2^n steps. *Discrete Applied Mathematics*, 10:287–295, 1985.

[NRS95] A. Naik, K. Regan, and D. Sivakumar. On quasilinear-time complexity theory. *Theoretical Computer Science*, 148(2):325–349, 1995.

[PPSZ98] R. Paturi, P. Pudlák, M. E. Saks, and F. Zane. An improved exponential-time algorithm for k-sat. In *FOCS '98: Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 628, Washington, DC, USA, 1998. IEEE Computer Society.

[PPZ97] R. Paturi, P. Pudlák, and F. Zane. Satisfiability coding lemma. In *FOCS '97: Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS '97)*, page 566, Washington, DC, USA, 1997. IEEE Computer Society.

[Sch99] Uwe Schöning. A probabilistic algorithm for k-**SAT** and constraint satisfaction problems. In *FOCS '99: Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, page 410, Washington, DC, USA, 1999. IEEE Computer Society.

[Sch02] Uwe Schöning. A probabilistic algorithm for k-**SAT** and constraint satisfaction problems. *Algorithmica*, 32:615–623, 2002.

[Sch03] R. Schuler. An algorithm for the satisfiability problem of formulas in conjunctive normal form. Technical report, Technical Report, Universit at Ulm, 2003.

[Tra84] B. A. Trakhtenbrot. A survey of russian approaches to perebor (brute-force search) algorithms. *Annals of the History of Computing*, 6(4):384–400, 1984.

[VV85] L G Valiant and V V Vazirani. **NP** is as easy as detecting unique solutions. In *STOC '85: Proceedings of the seventeenth annual ACM symposium on Theory of computing*, pages 458–463, New York, NY, USA, 1985. ACM Press.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/