2007 Conference on Analysis of Algorithms, AofA 07 DMTCS proc. **AH**, 2007, 103–114

Expected number of locally maximal solutions for random Boolean CSPs

Nadia Creignou[1] and Hervé Daudé[2] and Olivier Dubois[3]

[1] *LIF, UMR CNRS 6166 & Université de la Méditerranée, 163 avenue de Luminy, 13 288 Marseille, France*
[2] *LATP, UMR CNRS 6632 & Université de Provence, 39 rue Joliot-Curie, 13 453 Marseille, France*
[3] *LIP6, UMR CNRS 7606 & Université de Paris 6, 4 Place Jussieu, 75 252 Paris, France*

received 14 May 2007, *revised 23rd January 2008*, *accepted tomorrow*.

For a large number of random Boolean constraint satisfaction problems, such as random k-SAT, we study how the number of locally maximal solutions evolves when constraints are added. We give the exponential order of the expected number of these distinguished solutions and prove it depends on the sensitivity of the allowed constraint functions only. As a by-product we provide a general tool for computing an upper bound of the satisfiability threshold for any problem of a large class of random Boolean CSPs.

Keywords: Random structures, Constraint satisfaction problems, Boolean functions, Sensitivity, Satisfiability, Phase transition, Threshold.

Contents

1 Introduction

Constraint satisfaction is recognized as a fundamental problem in computer science, since combinatorial problems from many different application areas (artificial intelligence, databases, automated design, etc.)

can be expressed in a natural way by means of constraints. Informally, an instance of a constraint satisfaction problem (CSP for short) consists of a set of variables, a set of possible values for the variables, and a set of constraints that restrict the combinations of values that certain tuples of variables may take; the question is whether there is an assignment of values to variables that satisfies the given constraints. In the early seventies, many variants of CSPs have been proved to be NP-complete on a problem by problem basis. In contrast Schaefer proposed to study the standard constraint satisfaction problem parameterized by restricting the set of functions $\mathcal{F}$, thus defining the constraint satisfaction problem $\mathsf{CSP}(\mathcal{F})$, so-called *Generalized satisfiability* [Sch78]. He studied the complexity of the associated decision problem and proved that every Boolean $\mathsf{CSP}(\mathcal{F})$ is either NP-complete or solvable in polynomial time, thus revealing the seed of NP-completeness in SAT-like problems. Since then, there has been a growing body of classification results for related problems (see [CKS01] for a survey).

Similarly random instances of Boolean CSPs have early attracted a lot of attention (see [MSL92], [Bay05]). Average case behavior and experiments have provided evidence of the existence of a phase transition for the probability of a random instance of specific CSPs being satisfiable (see [DMSZ01]). While the nature of the phase transition –sharp or coarse- for (Boolean) CSPs is now well-understood [Mol03, CD04b, CD04a], due in particular to a probabilistic adaptation of Schaefer's framework, the location of the threshold is still performed on a problem by problem basis. A long series of works have been devoted to get lower and upper bounds (see respectively, [Ach00, CF90, CR92, FS96] and [FP83, MdlV95, DB97, KKKS98, JSV00]). For some problems these two bounds coincide, namely 2-SAT [CR92, Goe96] and 3-XOR-SAT [DM02].

In this paper we are interested in random symmetric Boolean CSPs, defined in Section 2. Our main contribution is to investigate the evolution of the expected number of distinguished satisfying assignments, so-called *locally maximal solutions* (see [DB97, KKKS98]), when the usual order parameter c ratio of constraints to variables increases. In Section 3 we introduce a new parameter for $\mathsf{CSP}(\mathcal{F})$: the *sensitivity polynomial* of $\mathcal{F}$. This polynomial is based on the sensitivity of Boolean functions, it collects the sensitivity of all points in $f^{-1}(1)$ for every f in $\mathcal{F}$. In Section 4 we express the evolution of the expected number of locally maximal solutions as a function of c. This gives a new picture of the solution-space geometry of random CSPs (see Figure 1). In particular it provides non-trivial upper bounds, c^*, for all these problems, which are better than the ones, $c^\#$, obtained in considering the number of all solutions (see Table 1).

2 Random symmetric Boolean CSPs

A Boolean constraint satisfaction problem consists of a finite set of Boolean variables $\{x_1, \dots, x_n\}$, and a finite set of Boolean constraints. Each such constraint denotes the allowed combinations of values for the variables it affects. A Boolean constraint satisfaction problem asks to determine if a satisfying truth assignment exists. In the symmetric framework a constraint C on the tuple of distinct variables $(x_{i_1}, \dots, x_{i_k})$ is given by a non trivial Boolean function $f : \{0,1\}^k \longrightarrow \{0,1\}$ and a tuple of literals $(x_{i_1}^{\tau_1}, \dots, x_{i_k}^{\tau_k})$, where $\tau_i \in 0,1$, $i = 1, \dots, k$ are called the signs and are such that for any variable x, x^0 denotes the negative literal $\bar{x}$, while x^1 denotes the positive one, x. Such a constraint is denoted by $C = f(x_{i_1}^{\tau_1}, \dots, x_{i_k}^{\tau_k})$, and is referred to as an f-constraint.

A truth assignment $I\colon \mathcal{V} = \{x_1, \dots, x_n\} \longrightarrow \{0,1\}$ (which can be seen as an element of $\{0,1\}^n$ and which extends to literals as usual by $I(x^1) = I(x)$ and $I(x^0) = 1 - I(x)$) satisfies such a constraint if $f(I(x_{i_1}^{\tau_1}), \dots, I(x_{i_k}^{\tau_k})) = 1$, which is denoted by $I(C) = 1$.

We are here interested in Boolean constraint satisfaction problem $\mathsf{CSP}(\mathcal{F})$ in which the types of allowed constraints $\mathcal{F}$, i.e., the Boolean functions, are fixed. Thus, throughout the paper $\mathcal{F}$ will denote a finite multi-set of non-trivial Boolean functions of fixed arity k.

Example 2.1 *we get the following well-known problems:*

- 3-SAT $= \mathsf{CSP}(\{f_1\})$ *with* $f_1^{-1}(1) = \{0,1\}^3 \setminus \{000\}$.
- 3-XOR-SAT $= \mathsf{CSP}(\{f_2\})$ *with* $f_2^{-1}(1) = \{001, 010, 100, 111\}$.
- 1-in-3-SAT $= \mathsf{CSP}(\{f_3\})$ *with* $f_3^{-1}(1) = \{001, 010, 100\}$.
- NAE-3-SAT $= \mathsf{CSP}(\{f_4\})$ *with* $f_4^{-1}(1) = \{0,1\}^3 \setminus \{000, 111\}$.

Let L and n be integers, $\mathsf{CSP}_{n,L}(\mathcal{F})$ is a random model for the Boolean constraint satisfaction problem with n variables and L constraints. Each constraint C_i of a random $\mathcal{F}$-formula is chosen as follows:

- Select a tuple of k distinct variables $(x_{i_1}, \ldots, x_{i_k})$ uniformly at random from the set of all k-tuples of variables from $\{x_1, \ldots, x_n\}$.
- Select a sign vector $(\tau_1, \ldots, \tau_k)$ uniformly at random from $\{0,1\}^k$
- Select a function uniformly at random from $\mathcal{F}$.
- Let C_i be the constraint $f(x_{i_1}^{\tau_1}, \ldots, x_{i_k}^{\tau_k})$.

The total number of all possible constraints is thus $|\mathcal{F}| \cdot 2^k \cdot (n)_k$. For any fixed $\mathcal{F}$, we are interested in studying the probability that a random formula in $\mathsf{CSP}_{n,L}(\mathcal{F})$ is satisfiable, we denote this probability by $\mathsf{Pr}_{n,L}(\mathsf{SAT}(\mathcal{F}))$.

3 Locally maximal solutions and Sensitivity

In order to get more insight in the probability that a random formula in $\mathsf{CSP}_{n,L}(\mathcal{F})$ is satisfiable, it is natural to study the expected numbers of assignments that certify the satisfiability of such a formula. Therefore, let $\mathsf{Sol}(\Phi) \subseteq \{0,1\}^n$ denote the set of satisfying assignments, we will be interested in $E_{n,L}(\mathsf{Sol})$. Some other (less numerous) satisfiability certificates are also of interest, namely the *locally maximal satisfying assignments* which are local maxima in the lexicographic ordering of assignments, where the neighborhood of an assignment is the Hamming ball of radius 1. We denote by $\mathsf{MaxSol}(\Phi) \subseteq \{0,1\}^n$ the set of locally maximal assignments of Φ and we will also be interested in $E_{n,L}(\mathsf{MaxSol})$. Formally these assignments can be defined as follows.

Given a truth assignment $I\colon \mathcal{V} = \{x_1, \ldots, x_n\} \longrightarrow \{0,1\}$, for $t = 1, \ldots, n$ let $\hat{I}_t$ denote the truth assignment on $\mathcal{V}$ defined by $\hat{I}_t(x_t) = 1 - I(x_t)$ and $\hat{I}_t(x_s) = I(x_s)$ for $s \neq t$. Given a formula Φ, let I be a satisfying assignment for Φ, i.e., $I(\Phi) = 1$. We say that x_t is a *sensitive variable for I with respect to* Φ if $\hat{I}_t(\Phi) = 0$ and we denote by $\mathcal{P}_s(I, \Phi)$ the set of all such variables:

$$\mathcal{P}_s(I, \Phi) = \{x_t / \hat{I}_t(\Phi) = 0\}.$$

Definition 3.1 *A solution I of a formula Φ is a* locally maximal solution *of Φ if and only if $\mathcal{P}_s(I, \Phi) \supseteq \{x_t / I(x_t) = 0\}$.*

In order to estimate $E_{n,L}(\mathsf{Sol})$, one has to compute the number of random formulas satisfied by a given assignment. Thus, a natural parameter defined on any Boolean function $f\colon \{0,1\}^k \longrightarrow \{0,1\}$ naturally emerges: its *weight*, namely $|f^{-1}(1)|$. It turns out that this information can be refined in order to exhibit the parameter that fits exactly the computation of $E_{n,L}(\mathsf{MaxSol})$: the *sensitivity*. A finer description of f can be given in describing how the elements of $f^{-1}(1)$ are distributed on the hypercube $\{0,1\}^k$. Let $v \in \{0,1\}^k$, the number of its neighbors on which the value of f differs is called the *sensitivity* of f at v and is denoted by $s_f(v) = |\{v' \ : \ f(v') \neq f(v), dist(v,v') = 1\}|$. The set $f^{-1}(1)$ can be partitioned into $(k+1)$ sets, each of which consisting in the points whose sensitivity is exactly j, $0 \leq j \leq k$; let $\theta_j(f)$ be the cardinality of each of these sets, that is

$$\theta_j(f) = |\{v : f(v) = 1 \text{ and } s_f(v) = j\}|.$$

Observe that $\displaystyle\sum_{j=0}^{k} \theta_j(f) = |f^{-1}(1)|$. Thus, we define the *sensitivity polynomial* of f as:

$$\mathsf{S}_f(y) = \sum_{r=0}^{k} \theta_r(f) \cdot y^r.$$

This notion is naturally extended to finite multi-sets $\mathcal{F}$ of Boolean functions of same arity k:

$$\mathsf{S}_{\mathcal{F}}(y) = \sum_{r=0}^{k} \theta_r(\mathcal{F}) \cdot y^r, \text{ where } \theta_j(\mathcal{F}) = \sum_{f \in \mathcal{F}} \theta_j(f)$$

4 Expected number of (locally maximal) solutions

4.1 *Solutions versus locally maximal solutions*

We are interested in a comparative study of $E_{n,L}(\mathsf{Sol})$ and $E_{n,L}(\mathsf{MaxSol})$ that are the expected number of solutions and the expected number of locally maximal solutions of a random formula. Specifically, the number L of constraints will be equal to $c \cdot n$ where c is the usual order parameter. Let us notice that our results can be better expressed in introducing a change of variables. For each set $\mathcal{F}$ the usual order parameter will be replaced by y where $c = \gamma_{\mathcal{F}}(y)$, where $\gamma_{\mathcal{F}}$ is defined in the following proposition.

Proposition 4.1 *Let $\mathcal{F}$ be a finite multi-set of Boolean constraint functions of arity k, and $\mathsf{S}_{\mathcal{F}}$ be its sensitivity polynomial. Let $\gamma_{\mathcal{F}}$ be defined on $(1,2]$ by*

$$\gamma_{\mathcal{F}}(y) = \ln\left(\frac{y}{2(y-1)}\right) \cdot \frac{S_{\mathcal{F}}(y)}{S'_{\mathcal{F}}(y)}. \tag{1}$$

The function $\gamma_{\mathcal{F}}$ is continuous and strictly decreasing on $(1,2]$. Since $\gamma_{\mathcal{F}}(2) = 0$ and $\lim\limits_{1+} \gamma_{\mathcal{F}} = +\infty$, this defines a valid change of variables from $(1,2]$ to $[0,+\infty)$.

Proof: Observe that $\ln(\frac{y}{2(y-1)}) < \frac{1}{y(y-1)}$ on $(1,2]$. Moreover, for all polynomial S with positive coefficients and for all $y > 1$, one can verify that $[(S')^2 - SS\text{"} - SS'](y) < 0$. Finally $\gamma'(y)$ is negative since it has the same sign as $\ln(\dfrac{y}{2(y-1)}) \cdot [(S')^2 - SS\text{"})](y) - \dfrac{S(y)S'(y)}{y(y-1)}$. ∎

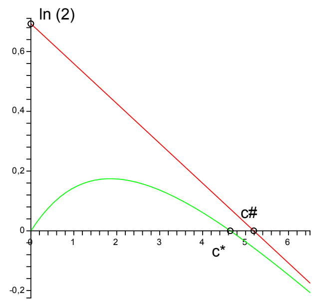

Fig. 1: Evolution of the exponential order of the expected number of solutions (linear curve) versus the exponential order of the expected number of locally maximal solutions.

The following theorem, which is our main result, describes the evolution of the expected number of solutions and locally maximal solutions of random symmetric CSP as a function of the order parameter c. It appears that the entropy, $\frac{1}{n}\log E_{n,cn}(\mathsf{Sol})$ is a linear function of c, while $\frac{1}{n}\log E_{n,cn}(\mathsf{MaxSol})$ is a non-monotone function. In using Maple it is possible to get a synthetic picture of this. Thus, Figure 1 illustrates this theorem for the most famous example, namely $\mathsf{3\text{-}SAT} = \mathsf{CSP}(\{f_1\})$ with $f_1^{-1}(1) = \{0,1\}^3 \setminus \{000\}$.

Theorem 4.2 *Let $\mathcal{F}$ be a finite multi-set of Boolean constraint functions of arity k, and $\mathsf{S}_{\mathcal{F}}$ be its sensitivity polynomial. Let $\gamma_{\mathcal{F}}$, $\chi_{\mathcal{F}}$ and $\Psi_{\mathcal{F}}$ be the functions defined on $(1,2]$ by*

- $\gamma_{\mathcal{F}}(y) = \ln\left(\dfrac{y}{2(y-1)}\right) \cdot \dfrac{S_{\mathcal{F}}(y)}{S'_{\mathcal{F}}(y)}$
- $\chi_{\mathcal{F}}(y) = \gamma_{\mathcal{F}}(y) \cdot \ln(\dfrac{S_{\mathcal{F}}(1)}{2^k \cdot |\mathcal{F}|}) + \ln 2$
- $\Psi_{\mathcal{F}}(y) = (y-1)\ln(y-1) - y\ln(\dfrac{y}{2}) + \gamma_{\mathcal{F}}(y) \cdot \ln\left(\dfrac{S_{\mathcal{F}}(y)}{2^k \cdot |\mathcal{F}|}\right)$

then for any c, and as n tends to infinity, the expected number of solutions and locally maximal solutions of a random $\mathcal{F}$-formula with n variables and $c \cdot n$ constraints satisfy :

$$\frac{1}{n} \log E_{n,cn}(\mathsf{Sol}) \sim \chi_{\mathcal{F}}(\gamma_{\mathcal{F}}^{-1}(c)) \tag{2}$$

$$\frac{1}{n} \log E_{n,cn}(\mathsf{MaxSol}) \sim \Psi_{\mathcal{F}}(\gamma_{\mathcal{F}}^{-1}(c)) \tag{3}$$

Sketch of proof: The classical result that the entropy of any CSP is a linear function of the order parameter c is well-known and the first part of the theorem is an easy consequence of the following combinatorial observation: in our symmetric model, for any truth assignment I

$$\#\{\Phi \in \mathsf{CSP}_{n,L}(\mathcal{F})\ s.t.\ I \in \mathsf{Sol}(\Phi)\} = \Big((n)_k \cdot \sum_{f\in\mathcal{F}} |f^{-1}(1)|\Big)^L \tag{4}$$

Since $\sum_{f\in\mathcal{F}} |f^{-1}(1)| = S_{\mathcal{F}}(1)$, the expected number of solutions is controlled by the parameter $S_{\mathcal{F}}(1)$:

$$E_{n,cn}(\mathsf{Sol}) = 2^n \cdot \left(\frac{S_{\mathcal{F}}(1)}{2^k \cdot |\mathcal{F}|}\right)^{cn} \tag{5}$$

The analysis is more cumbersome for locally maximal solutions. The first step relies on an analogue of (4) for locally maximal solutions. This combinatorial task involves the sensitivity polynomial of the set of Boolean functions $\mathcal{F}$ and is detailed in the last section (see Proposition 5.1). Then we derive the exponential order of the expected number of locally maximal solutions of any CSP. This second analytical task is based on asymptotical results on Stirling numbers [Tem93] (see (7)-(12) in Section 5). ∎

Let us notice that the number of locally maximal solutions is always lower than or equal to the number of solutions. A consequence of the following result on $\chi_{\mathcal{F}}$ and $\Psi_{\mathcal{F}}$ and our main result is that for any CSP, the exponential order of the expected number of locally maximal solutions is always strictly lower than the exponential order of the expected number of solutions.

Proposition 4.3 *For every finite multi-set $\mathcal{F}$ of Boolean constraint functions of fixed arity k, and for any $c > 0$, $\chi_{\mathcal{F}}(\gamma_{\mathcal{F}}^{-1}(c)) > \Psi_{\mathcal{F}}(\gamma_{\mathcal{F}}^{-1}(c))$.*

Proof: Since γ is decreasing, it suffices to prove that for every y in $(1, 2]$, $\Psi(y) < \chi(y)$. Observe that $\Psi(2) = 0$ and $\chi(2) = \ln 2$. Hence, after easy simplifications it comes down to prove that for every $y \in (1, 2)$

$$\frac{S'(y)}{S(y)}\left[1 + \frac{\ln y}{(y-1)\ln(\frac{y}{2(y-1)})}\right] > \frac{\ln(\frac{S(y)}{S(1)})}{(y-1)}.$$

One can use successively that on $(1, 2)$:

$$(y-1)\ln(\frac{y}{2(y-1)}) < \ln 2 \cdot (2-y) \text{ and } \ln 2 \cdot (2-y) + \ln y - \ln 2 > 0.$$

Hence, proving that $\psi(y) < \chi(y)$ comes down to show that

$$\frac{S'(y)}{S(y)} \cdot (y-1) + (y-2) \cdot \ln(\frac{S(y)}{S(1)}) > 0. \tag{6}$$

As we have already noted in the proof of Proposition 4.1, $S'(y)S(y) + S(y)S"(y) - S'^2(y) > 0$. Thus we obtain that the derivative of the left hand side of (6) is positive, this function being equal 0 when $y = 1$ the proof of (6) follows.

▮

4.2 Threshold upper bounds

Let $\mathcal{F}$ be a finite multi-set of Boolean constraint functions of arity k, and $\mathsf{S}_{\mathcal{F}}$ be its sensitivity polynomial. Upper bounds for the satisfiability threshold can be obtained by applying the first moment method to the number of solutions or to the number of locally maximal solutions.

Observe that $\chi_{\mathcal{F}}$ is strictly increasing from $-\infty$ to $\ln 2$. If $y^{\#}_{\mathcal{F}}$ denotes the unique solution of $\chi_{\mathcal{F}}(y) = 0$ then from (2) or (5) the probability of satisfiability of a random $\mathcal{F}$-formula with n variables and $c \cdot n$ constraints tends to 0 as soon as $c > c^{\#}_{\mathcal{F}}$, where

$$c^{\#}_{\mathcal{F}} = \gamma(y^{\#}_{\mathcal{F}}) = \frac{\ln 2}{\ln(\frac{2^k \cdot |\mathcal{F}|}{S_{\mathcal{F}}(1)})}.$$

In the same way, since $\Psi_{\mathcal{F}}$ is continuous, $\Psi_{\mathcal{F}}(2) = 0$ and $\lim\limits_{y \to 1^+} \Psi_{\mathcal{F}}(y) = -\infty$, let $y^*_{\mathcal{F}} = \inf \Psi^{-1}_{\mathcal{F}}(\{0\})$ be the least solution of $\Psi_{\mathcal{F}}(y) = 0$, then from (3) we get the following upper bound.

Theorem 4.4 *Let $\mathcal{F}$ be a finite multi-set of Boolean constraint functions of arity k, and $\mathsf{S}_{\mathcal{F}}$ be its sensitivity generating function. Let $\gamma_{\mathcal{F}}$ and $\Psi_{\mathcal{F}}$ be the functions defined on $(1, 2]$ by*

- $\gamma_{\mathcal{F}}(y) = \ln\left(\frac{y}{2(y-1)}\right) \cdot \frac{S_{\mathcal{F}}(y)}{S'_{\mathcal{F}}(y)}$
- $\Psi_{\mathcal{F}}(y) = (y-1)\ln(y-1) - y\ln(\frac{y}{2}) + \gamma_{\mathcal{F}}(y) \cdot \ln\left(\frac{S_{\mathcal{F}}(y)}{2^k \cdot |\mathcal{F}|}\right).$

The probability of satisfiability of a random $\mathcal{F}$-formula with n variables and $c \cdot n$ constraints tends to 0 ($\mathsf{Pr}_{n,cn}(\mathsf{SAT}(\mathcal{F})) \longrightarrow 0$) for every $c > c^*_{\mathcal{F}}$, where $c^*_{\mathcal{F}} = \gamma_{\mathcal{F}}(y^*_{\mathcal{F}})$ and $y^*_{\mathcal{F}}$ is the unique number such that for every $y < y^*_{\mathcal{F}}$, $\Psi_{\mathcal{F}}(y) < 0$.

Thus it follows from Proposition 4.3 that for every symmetric Boolean CSP, the threshold upper bound for satisfiability given by the expected number of locally maximal solutions is always better than the one given by the expected number of solutions. The picture given in Figure 1 is indeed generic, for any $\mathcal{F}$ we have :

$$c^*_{\mathcal{F}} < c^{\#}_{\mathcal{F}}.$$

Let us now explain how we can apply our theorem to any specific symmetric CSP in order to get an upper bound for the corresponding threshold. Let us start with the function $h(x, y, z)$ such that $h^{-1}(1) = \{0,1\}^3 \setminus \{000, 110\}$, i.e., $h(x, y, z) = (x \vee y \vee z) \wedge (\bar{x} \vee \bar{y} \vee z)$. We study the CSP generated by h, $\mathsf{CSP}(\{h\})$. According to Schaefer's theorem this problem is NP-complete, and according to Creignou

Name	$f^{-1}(1)$	S_f	$c^{\#}$	c^*
3-SAT	$\{0,1\}^3 \setminus \{000\}$	$4+3y$	5.19089	4.64248
NAE-3-SAT	$\{0,1\}^3 \setminus \{000,111\}$	$6y$	2.40942	2.19573
3-XOR-SAT	$\{000,110,011,101\}$	$4y^3$	1.	0.95662
1-in-3-SAT	$\{000,110,101\}$	$3y^3$	0.706695	0.61493

Tab. 1: Upper bounds for well-known problems

and Daudé's result it has a sharp threshold [CD04b, Theorem 4.2]. The sensitivity generating function of h is $S(y) = 2 + 2y + 2y^2$. We have to solve the equation

$$(y-1)\ln(y-1) - y\ln(\frac{y}{2}) + \gamma(y)\cdot\ln\left(\frac{S(y)}{2^3}\right) = 0$$

MAPLE provides the following numerical estimates: $y^* = 1.060952$, $\gamma(y^*) = 2.20850$, and hence for every $c > 2.20850$ we have $\mathsf{Pr}_{n,cn}(\mathsf{SAT}(\{h\})) \longrightarrow 0$.

In applying Theorem 4.2 to the well-known problems mentioned in Example 2.1, the upper bounds shown in Table 1 are obtained. Better upper bounds have already been obtained for some of these problems with specific techniques. This is the case for 3-SAT, an upper bound has been computed at 4.506 in [DBM00], or for 3-XOR-SAT, the exact threshold has even been computed in [DM02]. However the technique used for each of these problem has not been generalized to any class of problems. Our Theorem 4.2 is a general tool giving for any problem being a symmetric Boolean CSPs, an upper bound of the satisfiability threshold better than the one derived straightforwardly from the expected number of solutions.

5 Proof of Theorem 4.2

In this section we give more details on the estimate of the exponential number of locally maximal solutions. The first step is the following combinatorial result.

Proposition 5.1 *Let $\mathcal{F}$ be a finite multi-set of Boolean constraint functions of arity k and let $\sum_{r=0}^{k}\theta_r(\mathcal{F})\cdot y^r$ be its sensitivity polynomial. The expected number of locally maximal solutions for a random $\mathcal{F}$-formula with n variables and L constraints is*

$$\begin{aligned} E_{n,L}(\mathsf{MaxSol}) &= \sum_{i=0}^{n} \sum_{\substack{L_0+L_1+\cdots+L_k=L \\ L1+2L_2+\cdots+kL_k\geq i}} \frac{\binom{n}{i}\cdot 2^i}{(|\mathcal{F}|\cdot 2^k\cdot (n)_k)^L} \\ &\cdot \binom{L}{L_0\dots L_k}\cdot \mathcal{T}(L_1+2L_2+\cdots+kL_k, i) \\ &\cdot \prod_{r=0}^{k}\Big[\theta_r(\mathcal{F})\cdot (n-r)_{(k-r)}\Big]^{L_r} \end{aligned}$$

where $\mathcal{T}(L_1+2L_2+\cdots+kL_k, i)$ denotes the number of applications φ from $\{1,\cdots,L_1+2L_2+\cdots+kL_k\}$ onto $\{1,\ldots,i\}$, with the additional requirement that for every $j\in\{0,\cdots,k\}$ and for each $0\leq i\leq$

$j(L_j - 1)$ *with* $i \equiv 0(j)$, $\#\{\varphi(L_1 + 2L_2 + \cdots + (j-1)L_{j-1} + i + 1), \ldots, \varphi(L_1 + 2L_2 + \cdots + (j-1)L_{j-1} + i + j)\} = j$.

Proof:

$$E_{n,L}(\mathsf{MaxSol}) = \sum_{(I,\mathcal{P})} \frac{\mathcal{N}(I,\mathcal{P})}{(|\mathcal{F}| \cdot 2^k \cdot (n)_k)^L},$$

where $\mathcal{N}(I,\mathcal{P})$ is the number of $\mathcal{F}$-formula Φ with n variables and L constraints such that $I \in \mathsf{Sol}(\Phi)$ and $\mathcal{P}_s(I,\Phi) = \mathcal{P}$. Our first analysis will be to estimate $\mathcal{N}(I,\mathcal{P})$ for fixed $(I,\mathcal{P})$.

Given a constraint $C = f(x_{i_1}^{\tau_1}, \ldots, x_{i_k}^{\tau_k})$ satisfied by I, we say that the position j is a *sensitive position* in C with respect to I if $f(I(x_{i_1}^{\tau_1}), \ldots, 1 - I(x_{i_j}^{\tau_j}), \ldots, I(x_{i_k}^{\tau_k})) = 0$. A constraint C is said to be l-sensitive with respect to I if $|\mathcal{P}_s(I,C)| = l$. Due to the symmetry of our model the number of such constraints is $\theta_l(f) \cdot (n)_k$. In order to count the number of formulas $\Phi = (C_1, \ldots, C_L)$ in $\mathsf{CSP}_{n,L}(\mathcal{F})$ such that $I \in Sol(\Phi)$ and $\mathcal{P}_s(I,\Phi) = \mathcal{P}$, we partition the constraints of Φ in $k+1$ classes, according to their degree of sensitivity. For $j = 0, \ldots, k$, let L_j be the number of j-sensitive constraints in Φ, $L_0 + \ldots + L_k = L$.

The number of f-formulas of the form

$$\Phi = (C_1^0, \ldots, C_{L_0}^0, C_{L_1}^1, \ldots, C_{L_1}^1, \ldots, C_1^k, \ldots, C_{L_k}^k),$$

such that $I \in Sol(\Phi)$ and $\mathcal{P}_s(I,\Phi) = \mathcal{P}$, with $|\mathcal{P}| = i$, is:

$$\mathcal{T}(L_1 + 2 \cdot L_2 + \cdots k \cdot L_k, i) \prod_{r=0}^{k} (\theta_r(f) \cdot (n-r)_{(k-r)})^{L_r},$$

where $\mathcal{T}(L_1 + 2L_2 + \cdots + kL_k, i)$ corresponds to the number of ways of choosing the $L_1 + 2 \cdot L_2 + \cdots k \cdot L_k$ variables from $\mathcal{P}$ in order to fill the sensitive positions so that all variables from $\mathcal{P}$ occur, with the additional requirement that variables occurring in the same constraint are pairwise distinct.

This analysis can be extended to a finite set of constraint functions $\mathcal{F}$ in replacing the $\theta_r(f)$ by $\theta_r(\mathcal{F}) = \sum_{f \in \mathcal{F}} \theta_r(f)$. Thus we obtain:

$$\begin{aligned}\mathcal{N}(I,\mathcal{P}) \quad = \quad & \sum_{\substack{L_0 + L_1 + \cdots + L_k = L \\ L1 + 2L_2 + \cdots + kL_k \geq i}} \binom{L}{L_0 \ldots L_k} \cdot \mathcal{T}(L_1 + 2L_2 + \cdots + kL_k, i) \\ \cdot \quad & \prod_{r=0}^{k} \Big[\theta_r(\mathcal{F}) \cdot (n-r)_{(k-r)}\Big]^{L_r}\end{aligned}$$

In taking into account the number of pairs $(I,\mathcal{P})$ (there are $\binom{n}{i} \cdot 2^i$ such pairs with $|\mathcal{P}| = i$) and all the different possible distributions of the number of constraints according to their sensibility, each requiring $L_0 + L_1 + \cdots + L_k = L$ and $L_1 + 2L_2 + \cdots + kL_k \geq i$, we get the expression of $E_{n,L}(\mathsf{MaxSol})$ as stated in the proposition. ▮

The asymptotical behavior of the number of applications from $\{1, \cdots, a\}$ onto $\{1, \cdots, b\}$ is well-studied and based on Stirling numbers of the second kind. It follows from the above result that the exponential order of $E_{n,L}(\mathsf{MaxSol})$ is governed by exponential order of Stirling numbers of the second kind and of multinomial coefficients.

Observe that in Proposition 5.1 only indices r for which $\theta_r \neq 0$ occur. Suppose that $\theta_{i_0} \neq 0, \ldots, \theta_{i_s} \neq 0$ and for all $j \notin \{i_0, \ldots, i_s\}$, $\theta_j = 0$. In setting $\mathbf{L_{i_j}} = \beta_{\mathbf{i_j}} \cdot \mathbf{L}$ and $\mathbf{i} = \alpha \cdot \mathbf{n}$, in considering that $\beta_{i_0} = 1 - \beta_{i_1} - \ldots - \beta_{i_s}$ and in using asymptotic estimates of Stirling numbers (see [Tem93]), we obtain that when $L = cn$ and n tends to infinity:

$$\frac{1}{n} \log E_{n,cn}(\mathsf{MaxSol}) \sim \max_{\mathcal{D}_c} \Phi_c, \tag{7}$$

where Φ_c is a function of $(s+1)$ variables on a domain $\mathcal{D}_c$ given by:

$$\begin{aligned}
\Phi_c(\alpha, \beta_{i_1}, \ldots, \beta_{i_s}) &= \alpha \ln(e^{x_0} - 1) + cB \ln \frac{cB}{x_0 e} \\
&+ c \sum_{j=0}^{s} \beta_{i_j} \ln \frac{\theta_{i_j}(\mathcal{F})}{\beta_{i_j}} \\
&- \alpha \ln \alpha - (1-\alpha)\ln(1-\alpha) + \alpha \ln 2 \\
&- c \ln(2^k |\mathcal{F}|),
\end{aligned}$$

with

$$\mathcal{D}_c = \left\{ (\alpha, \overrightarrow{\beta}),\ 0 \leq \alpha \leq 1;\ 0 \leq \beta_{i_j} \leq 1,\ \sum_{j=1}^{s} \beta_{i_j} \leq 1;\ c \cdot B \geq \alpha \right\},$$

where $\mathbf{B} = \sum_{\mathbf{j=0}}^{\mathbf{s}} \mathbf{i_j} \cdot \beta_{\mathbf{i_j}}$, and $\mathbf{x_0}$ denotes the implicit solution of the equation

$$1 - e^{-x_0} = \frac{\alpha}{cB} \cdot x_0. \tag{8}$$

It appears that the global maximum of Φ_c on $\mathcal{D}_c$ is located in the interior of $\mathcal{D}_c$, and thus is a point at which all the partial derivatives are 0. In using the change of variables introduced in Proposition 4.1 we show that there is a unique stationary point which is completely determined by c. Indeed, annihilating the partial derivates one gets

$$2(1-\alpha)(e^{x_0} - 1) = \alpha \tag{9}$$

and

$$\frac{(1 - \beta_1 - \cdots - \beta_k) \cdot \theta_j(\mathcal{F}) \cdot c^j \cdot (\beta_1 + 2\beta_2 + \ldots + k\beta_k)^j}{x_0^j \cdot \theta_0(\mathcal{F}) \cdot \beta_j} = 1 \tag{10}$$

Plugging these equations in (8) we get α and the β_j's respectively as a function of x_0

$$\alpha = \frac{2(e^{x_0} - 1)}{2e^{x_0} - 1)},$$

$$\beta_j = \frac{\theta_j(\mathcal{F}) \cdot (1 - e^{x_0}/2)^{-j}}{\sum_{j=0}^{k} \theta_j(\mathcal{F}) \cdot (1 - e^{x_0}/2)^{-j}} \text{ for all } j.$$

Therefore, in setting $y = \frac{2e^{x_0}}{2e^{x_0}-1}$, we get α and the β_j's as a function of y:

$$\alpha = 2 - y, \beta_j = \frac{\theta_j \cdot y^j}{\sum_{j=0}^{k} \theta_j \cdot y^j}.$$

Finally, from (8), c can also be expressed as a function of y:

$$c = \ln\left(\frac{y}{2(y-1)}\right) \cdot \frac{S(y)}{S'(y)} = \gamma(y). \tag{11}$$

In plugging in Φ_c the above coordinates of the stationary point, elementary calculations reveal that the maximum value of Φ_c is completely determined by c and can be expressed by the function Ψ occurring in our theorem:

$$\max_{\mathcal{D}_c} \Phi_c = \Phi_c(2 - \gamma^{-1}(c), \theta_{i_1}(\mathcal{F}), \theta_{i_2}(\mathcal{F}) \cdot \gamma^{-1}(c), \ldots, \theta_{i_s}(\mathcal{F}) \cdot (\gamma^{-1}(c))^{s-1}) = \Psi(\gamma^{-1}(c)).$$

Thus, from (7) we get:

$$\frac{1}{n} \log E_{n,cn}(\mathsf{MaxSol}) \sim \Psi(\gamma^{-1}(c)) \tag{12}$$

References

[Ach00] D. Achlioptas. Setting 2 variables at a time yields a new lower bound for random 3-sat. In *Proceedings of the 32nd ACM Symposium on Theory of Computing (STOC 2000)*, pages 28–37, 2000.

[Bay05] D. Bayley. Phase transitions for generalized sat problems: upper bounds and experiments. Manuscript, 2005.

[CD04a] N. Creignou and H. Daudé. Coarse and sharp transitions for random generalized satisfiability problems. In *Proceedings of the third colloquium on mathematics and computer science, Vienna (Austria)*, pages 507–516. Birkhäuser, 2004.

[CD04b] N. Creignou and H. Daudé. Combinatorial sharpness criterion and phase transition classification for random csps. *Information and Computation*, 190(2):220–238, 2004.

[CF90] M.T. Chao and J. Franco. Probabilistic analysis of a generalization of the unit-clause selection heuristics for the k-satisfiability problem. *Information Science*, 51(3):289–314, 1990.

[CKS01] N. Creignou, S. Khanna, and M. Sudan. *Complexity classifications of Boolean constraint satisfaction problems*, volume 7. SIAM Monographs on discrete mathematics and applications, 2001.

[CR92] V. Chvátal and B. Reed. Mick gets some (the odds are on his side). In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science (FOCS'92)*, pages 620–627, 1992.

[DB97] O. Dubois and Y. Boufkhad. A general upper bound for the satisfiability threshold of random r-sat formulae. *Journal of Algorithms*, 24(2):395–420, 1997.

[DBM00] O. Dubois, Y. Boufkhad, and J. Mandler. Typical random 3-sat formulae and the satisfiability threshold. In *Proceedings of the 11th ACM-SIAM Symposium on Discrete Algorithms (SODA 2000)*, pages 721–722, 2000.

[DM02] O. Dubois and J. Mandler. The 3-xor-sat threshold. In *Proceedings 43rd Symposium on Foundations of Computer Science (FOCS 2002), Vancouver (British Columbia, Canada)*, pages 769–778, November 2002.

[DMSZ01] O. Dubois, R. Monasson, B. Selman, and R. Zecchina, editors. *Phase transitions for combinatorial problems*, volume 265 of *Theoretical Computer Science*, 2001.

[FP83] J. Franco and M. Paull. Probabilistic analysis of the davis-putnam procedure for solving the satisfiability problem. *Discrete Applied Mathematiques*, 5:77–87, 1983.

[FS96] A.M. Frieze and S. Suen. Analysis of two simple heuristics on a random instance of the k-sat problem. *Journal of Algorithms*, 20(2):312–355, 1996.

[Goe96] A. Goerdt. A threshold for unsatisfiability. *Journal of of Computer and System Sciences*, 53(3):469–486, 1996.

[JSV00] S. Janson, Y.C. Stamatiou, and M. Vamvakari. Bounding the unsatisfiability threshold of random 3-sat. *Random Structures and Algorithms*, 17(2):79–102, 2000. Erratum , Random Structures and Algorithms, 18(1):99–100, 2001.

[KKKS98] L.M. Kirousis, E. Kranakis, D. Krizanc, and Y.C. Stamatiou. Approximating the unsatisfiability threshold of random formulas. *Random Structures and Algorithms*, 12(3):253–269, 1998.

[MdlV95] A. El Maftouhi and W. Fernandez de la Vega. On random 3-sat. *Combinatorics, Probability and Computing*, 4(3):189–195, 1995.

[Mol03] M. Molloy. Models for random constraint satisfaction problems. *SIAM Journal on Computing*, 32(4):935–949, 2003.

[MSL92] D. Mitchell, B. Selman, and H.J. Levesque. Hard and easy distributions of sat problems. In *Proceedings 10th National Conference on Artificial Intelligence, San Jose (CA, USA)*, pages 459–465. AAAI, 1992.

[Sch78] T.J. Schaefer. The complexity of satisfiability problems. In *Proceedings 10th STOC, San Diego (CA, USA)*, pages 216–226. Association for Computing Machinery, 1978.

[Tem93] N.M. Temme. Asymptotic estimates of stirling numbers. *Stud. appl. Math.*, 89:223–243, 1993.