Random generalized satisfiability problems

Nadia CREIGNOU*
LIF
Université de la Méditerranée
Marseille, France

Hervé DAUDÉ †
LATP, UMR CNRS 6632
Université de Provence
Marseille, France

January 31, 2002

Abstract

We develop a coherent model in order to study specific random constraint satisfaction problems. Our framework includes all the generalized satisfiability problems defined by Schaefer. For such problems we study the phase transition. Our model captures problems having a sharp transition as well as problems having a coarse transition. We first give lower and upper bounds for the location of the phase transition. Then we study minimal unsatisfiable formulas. This study leads to the notion of "reasonable" generalized satisfiability problems, for which we conjecture a sharp transition.

1 Introduction

During the past decade there has been an extensive investigation of phase transition phenomena in various NP-complete problems. From the computer scientist's viewpoint one of the goal of this investigation is to bring out the finer structure of NP-complete problems and to explore the relation ship between phase transition phenomena and the average-case performance of algorithms for the problem at hand. Among all the problems for which such a study has been performed, the satisfiability of k-CNF formulas, k-SAT for short, has sparked a lot of interest. Experiments on solving k-CNF random formulas (see for instance [KS94]) have provided evidence of the existence of a satisfiability threshold phenomenon with respect to the ratio, c_k, of the number of clauses to the number of variables of formulas. There are two approaches to address the problem of the existence of a phase transition. The first one is to locate the phase transition, it deals with random graphs and algorithms. For instance for 3-SAT, for an observed sharp threshold of about $c_3 = 4.25$ [SML96] the best lower bound is 3.26 [AS00] and the best upper bound is 4.506 [DBM00], so there still exists a large gap between lower and upper bounds (whereas a

*E-mail: creignou@lim.univ-mrs.fr
†E-mail: daude@gyptis.univ-mrs.fr

sharp threshold is established for 2-SAT ([CR92] and [Goe96]), the critical value being $c_2 = 1$). The second approach consists in describing the nature of the transition (sharp or coarse). In a seminal paper Friedgut [FG99], with an appendix by Bourgain, initiated this approach, he developed a general sharp threshold criterion for monotone subsets of the hypercube. In using this criterion he proved that k-SAT exhibits a sharp threshold for $k \geq 2$ but without specifying its location.

In the last few years there has been a great amount of interest in constraint satisfaction problems (see [SD96] and [AKKKMS98]). In [AKKKMS98] the authors proposed a model for generating random constraint satisfaction instances which exhibit a non-trivial asymptotic behavior. They also suggested to study models where there is some structure, or even no randomness at all, in the choice of forbidden pairs of values for two constrained variables. This is our aim here since we focus our attention on specific random CSP. We propose to study the class of (Boolean) constraint satisfaction problems in which the nature of the constraints is fixed. Informally speaking, each problem in this class is characterized by a finite collection of finitely specified constraint templates, say $\mathcal{F}$.This class of problems, $\mathsf{SAT}(\mathcal{F})$ could be interpreted to be the combinatorial core of complexity and has already been thoroughly tested (see [CKS01] for a survey) as a testbed for abstracting global inferences about the nature of computation. Therefore this class might also provide some hints on the nature of phase transitions.

In Section 2.1 we propose a general framework to make precise the universe set on which probability can be done, $\mathsf{CSP}(\mathcal{F})$. Section 3 deals with the location of the phase transition. Section 4 is devoted to the study of minimal elements of the property $\mathsf{UNSAT}(\mathcal{F})$ for every $\mathcal{F}$. Such a study is motivated by the fact that it emerges from Friedgut's work that such elements are crucial in the analyze of the phase transition.

2 Random Boolean Constraint satisfaction problems

2.1 Boolean Constraint satisfaction problems

Notation: Let f be a function, $f : \{0,1\}^k \longrightarrow \{0,1\}$,where k is a non-negative integer called the arity of f, and $\sigma \in \mathcal{S}_k$. We denote by $f_\sigma : \{0,1\}^k \longrightarrow \{0,1\}$ the function $f_\sigma(a_1, \cdots, a_n) = f(a_{\sigma(1)}, \cdots, a_{\sigma(n)})$. Let $\langle f \rangle$ denote the functions obtained from f by permutation of its variables (i.e. the orbit of f under the action of the group $\mathcal{S}_k$):

$$\langle f \rangle = \{f_\sigma \ / \ \sigma \in \mathcal{S}_k\}.$$

Observe that

$$\#\langle f \rangle = \frac{k!}{\#\{\sigma \in \mathcal{S}_k \ / \ f_\sigma = f\}}.$$

Example 2.1

- *If* $f_{0,3}(x, y, z) = (x \vee y \vee z)$ *, then* $\#\langle f_{0,3} \rangle = 1$.
- *If* $f_{1,3}(x, y, z) = (\neg x \vee y \vee z)$ *, then* $\#\langle f_{1,3} \rangle = 3$.

Definition 2.2 [Restriction of a function] *Let f be a function of arity k. For $1 \leq i \leq k$ and $\varepsilon \in \{0,1\}$ the function denoted by $\Pi_i^{\varepsilon}(f)$ is the restriction of f over $k-1$ variables obtained from f in fixing the value of the i-th variable to ε, that is to say :*

$$\Pi_i^{\varepsilon}(f)(\alpha_1, \ldots, \alpha_{k-1}) = f(\alpha_1, \ldots, \alpha_{i-1}, \varepsilon, \alpha_i, \ldots, \alpha_{k-1}).$$

A truth assignment $\Phi : \{x_1, \ldots, x_n\} \longrightarrow \{0,1\}$ satisfies a constraint generated by a function f and the choice of a tuple $(x_{i_1}, \ldots, x_{i_k})$ if $f(\Phi(x_{i_1}), \cdots, \Phi(x_{i_k})) = 1$. A natural question is how many different constraints can be generated from f over n variables. To answer this question let us consider functions that are nontrivial and which really depend on k variables.

Definition 2.3 [Constraint function] *A function $f : \{0,1\}^k \longrightarrow \{0,1\}$ is said to be a* constraint function of arity k *if $f^{-1}(1) \neq \emptyset$, $f^{-1}(1) \neq \{0,1\}^k$ and f has no fictive variables, that is to say $\Pi_i^1(f) \neq \Pi_i^0(f)$ for every i.*

Let f be a constraint function, $C_n(f)$ denotes the set of all constraints that can be generated from f over n variables. An element of $C_n(f)$ is given by f and a tuple $(x_{i_1}, \ldots, x_{i_k})$ of k distinct variables among n. It can also be given by f, a subset of k elements and a permutation $\sigma \in \mathcal{S}_k$, hence $(f_\sigma, i_1 < i_2 < \cdots < i_k)$ will denote an element of $C_n(f)$.

Proposition 2.4 $\#C_n(f) = \binom{n}{k} \#\langle f \rangle$.

Let $\Phi : \{x_1, \ldots, x_n\} \longrightarrow \{0,1\}$ be a truth assignment and $C \in C_n(f)$ be a constraint, $C = (f_\sigma, i_1 < i_2 < \cdots < i_k)$. We denote by $\Phi(C) = f_\sigma(\Phi(x_{i_1}), \cdots, \Phi(x_{i_k}))$ and $S_n(C) = \{\Phi \text{ such that } \Phi(C) = 1\}$.

Remark: For every C in $C_n(f)$, $\#S_n(C) = 2^{n-k} \# f^{-1}(1)$.

Throughout the paper, $\mathcal{F}(k)$, or simply $\mathcal{F}$ when no confusion can arise, will denote a finite set of constraint functions of arity k over the domain $\{0,1\}$, $\{f_1 \cdots, f_l\}$, such that for every $i, 1 \leq i \leq l$, $f_i \notin \bigcup_{j \neq i} \langle f_j \rangle$. The set $C_n(\mathcal{F})$ is the set of all constraints that can be generated over n variables from any constraint function f_i in $\mathcal{F}$, $C_n(\mathcal{F}) = \bigcup_{i=1}^{l} C_n(f_i)$. There are $N_{\mathcal{F}}$ such constraints,

$$N_{\mathcal{F}} := \#C_n(\mathcal{F}) = \sum_{i=1}^{l} \#C_n(f_i) = \sum_{i=1}^{l} \binom{n}{k} \#\langle f_i \rangle.$$

Example 2.5 *Let $\mathcal{F} = \{f_{0,k}, \cdots f_{k,k}\}$, where for $0 \leq i \leq k$, $f_{i,k}(x_1, \cdots, x_k) = (\neg x_1 \vee \cdots \vee \neg x_i \vee x_{i+1} \vee \cdots \vee x_k)$. The set $C_n(\mathcal{F})$ is the set of all k-clauses over n variables, there are $N_{\mathcal{F}} = \#C_n(\mathcal{F}) = \sum_{i=0}^{k} \binom{n}{k} \frac{k!}{i!(k-i)!} = \binom{n}{k} \sum_{i=0}^{k} \binom{k}{i} = 2^k \binom{n}{k}$ such clauses.*

The set of collections of constraints from $C_n(\mathcal{F})$ is denoted by $\mathsf{CSP}_n(\mathcal{F})$ (or simply $\mathsf{CSP}(\mathcal{F})$ when no confusion can arise), $\mathsf{CSP}_n(\mathcal{F}) \subseteq \mathcal{P}(C_n(\mathcal{F}))$.

We will denote by $\mathsf{SAT}_n(\mathcal{F})$ (or simply $\mathsf{SAT}(\mathcal{F})$) the property for such a collection of constraints of being satisfiable and by $\mathsf{UNSAT}(\mathcal{F})$ the property of being unsatisfiable.

Let $s \in \mathsf{CSP}_n(\mathcal{F})$, $s = C_1 \wedge \cdots C_L$, observe that

$$s \in \mathsf{SAT}(\mathcal{F}) \Leftrightarrow \bigcap_{i=1}^{L} S_n(C_i) \neq \emptyset \Leftrightarrow \text{ there exists } \Phi \text{ such that } \prod_{i=1}^{L} \Phi(C_i) = 1.$$

Therefore, in order to get a nontrivial satisfiability property it is necessary that $\bigcap_{C \in C_n(\mathcal{F})} S_n(C) = \emptyset$. For every $\mathcal{F}$ it is not difficult to verify that obtaining a nontrivial satisfiability property $\mathsf{SAT}(\mathcal{F})$ is characterized by the following proposition.

Proposition 2.6 *$\bigcap_{C \in C_n(\mathcal{F})} S_n(C) = \emptyset$ if and only if there exist g_0 and g_1 in $\mathcal{F}$ such that $g_0(\vec{0}) = 0$ and $g_1(\vec{1}) = 0$.*

Therefore, we now restrict our attention to *interesting sets* $\mathcal{F}$, that is sets $\mathcal{F}$ in which there exist g_0 and g_1 in $\mathcal{F}$ such that $g_0(\vec{0}) = 0$ and $g_1(\vec{1}) = 0$. In this way we capture generalized satisfiability problems, first introduced by Schaefer [Sch78], having a non trivial satisfiability property (interesting sets correspond to non-0-valid and non-1-valid sets in Schaefer's terminology).

Example 2.7

k-SAT: *For $0 \leq i \leq k$, let $f_{i,k}$ be defined by $f_{i,k}(x_1, \cdots, x_k) = (\neg x_1 \vee \cdots \vee \neg x_i \vee x_{i+1} \vee \cdots \vee x_k)$. Then, k-SAT $= \mathsf{SAT}(\{f_{0,k}, \cdots, f_{k,k}\})$.*

NAE-k-SAT: *For $0 \leq i \leq k$, let $g_{i,k}$ be the Boolean function defined by $g_{i,k}^{-1}(1) = \{0,1\}^k \setminus \{0^i 1^{k-i}, 1^i 0^{k-i}\}$. Then, NAE-$k$-SAT $= \mathsf{SAT}(\{g_{0,k}, \cdots, g_{k,k}\})$.*

1-in-k-SAT: *For $0 \leq i \leq k$, let $h_{i,k}$ be the Boolean function defined by $h_{i,k}(x_1, \cdots, x_k) = 1$ if and only if $((1 - x_1) + \cdots + (1 - x_i) + x_{i+1} + \cdots + x_k) = 1$. Then, 1-in-$k$-SAT $=$ $\mathsf{SAT}(\{h_{0,k}, \cdots, h_{k,k}\})$.*

k-XOR-SAT: *let $g_0(x_1, \cdots, x_k) = x_1 \oplus \cdots \oplus x_k$ and $g_1(x_1, \cdots, x_k) = x_1 \oplus \cdots \oplus x_k \oplus 1$. Then, k-XOR-SAT $= \mathsf{SAT}(\{g_0, g_1\})$.*

2.2 Probabilistic model

Let us first recall some generalities on threshold of monotone properties of the hypercube. Let N be a positive integer. A nonempty subset E of the hypercube $\{0,1\}^N$ is called monotone increasing, or simply monotone, provided if $s \in E$, $s' \in \{0,1\}^N$, $s_i \leq s'_i$ for

$i = 1, \ldots, N$ then $s' \in E$. For $0 \leq p \leq 1$, Let μ_p be the product measure on $\{0,1\}^N$ with weights $1-p$ at 0 and p at 1. Thus, if $|s|$ denotes the size of s

$$\mu_p(\{s\}) = (1-p)^{N-|s|} p^{|s|} \text{ where } |s| = \#\{i = 1, \ldots, N / \ s_i = 1\}.$$

Let $A \subseteq \{0,1\}^*$ be a monotone property, i.e. for every N, $A_N := A \cap \{0,1\}^N$ is monotone. (It is said to be nontrivial if for every N, A_N is nontrivial). Set $f_N(p) = \mu_p(A_N)$. Each $f_N(p)$ defines an increasing one-to-one correspondence from $[0,1]$ onto $[0,1]$ (see [Bol85]). Let us write $p_c(N) = f_N^{-1}(c)$. For $c = 1/2$, $p_{1/2}(N)$ is called the *critical probability* of A. Let $\varepsilon \in]0, 1/2]$, the interval $\tau(N) = [p_\varepsilon(N), p_{1-\varepsilon}(N)]$ is called the *threshold interval*. Let $\delta_\varepsilon(N)$ denote its length, $\delta_\varepsilon(N) = p_{1-\varepsilon}(N) - p_\varepsilon(N)$. One says that A has a *sharp threshold* if for every $\varepsilon \in]0, 1/2]$ the ratio $\frac{\delta_\varepsilon(N)}{p_{1/2}(N)}$ tends to 0 as N tends to infinity, otherwise one says that A has a *coarse threshold*.

Now, consider a set of constraints functions $\mathcal{F}$, $N_\mathcal{F}$ denotes the number of different constraints that can be generated from $\mathcal{F}$ over n variables. $\mathsf{SAT}(\mathcal{F})$ can be considered as a monotone subset of $\{0,1\}^{N_\mathcal{F}}$. Let $S^\mathcal{F}(n,p) \in \{0,1\}^{N_\mathcal{F}}$ denote the random collection of constraints in $C_n(\mathcal{F})$ where each constraint appears independently with probability p. The probability that the random formula $S^\mathcal{F}(n,p)$ is satisfiable is $\mu_p(\mathsf{SAT}(\mathcal{F}))$. We are interested in studying the asymptotic behavior of this probability, when $N_\mathcal{F}$ (or equivalently n) tends to infinity. By abuse of notation we write $p_c(n)$ instead of $p_c(N_\mathcal{F})$.

3 A first estimate of the location of the phase transition

We are interested in studying $\mu_p(\mathsf{SAT}(\mathcal{F}))$. Our goal in this section is specify the scale at which the phase transition occurs for $\mathsf{SAT}(\mathcal{F})$. On the one hand the first moment method gives an upper bound.

Proposition 3.1 *Let $\mathcal{F}$ be an interesting set of constraint functionst. Then $N_\mathcal{F}.p_c = O(n)$, that is to say $p_c(n) = O(n^{1-k})$*

On the other hand it is easy to prove that a large number of constraints is needed to observe the transition as shown by the following estimate.

Lemma 3.2 $\mu_p(\mathsf{UNSAT}(\mathcal{F})) = o(1)$ *as soon as* $N_\mathcal{F} p = o(\sqrt{n})$.

For various satisfiability problems it has been observed that for $k \geq 2$ the transition occurs at $p_c(n) = \Theta(n^{1-k})$ (see [FG99], [AS00] and [CD01]) whereas 1-SAT is a special case for which $p_c(n) = \Theta(n^{-1/2})$. Therefore, it is natural to restrict our attention to functions that do not strongly depend on one component as it is defined as follows.

Definition 3.3 *A Boolean function* $f : \{0,1\}^k \longrightarrow \{0,1\}$ strongly depends on one component *if:*

$\exists \varepsilon \in \{0,1\}, \exists i \in \{1, \ldots, k\}, \forall (a_1, \ldots, a_k) \in \{0,1\}^k, f(a_1, \ldots, a_k) = 1 \Rightarrow a_i = \varepsilon$

Example 3.4 *The function* $f : \{0,1\}^k \longrightarrow \{0,1\}$ *where* $f^{-1}(1) = \{(0,0,0),(0,1,0),(0,1,1)\}$ *strongly depends on its first component.*

We are then able to give a better lower bound.

Proposition 3.5 *Suppose that* $\mathcal{F}$ *does not contain any constraint function that strongly depends on one component. Then* $N_{\mathcal{F}}.p_c = \Omega(n)$*, that is to say* $p_c(n) = \Omega(n^{1-k})$

Thus, we get the following:

Theorem 3.6 *Suppose that* $\mathcal{F}$ *does not contain any constraint function that strongly depends on one component. Then,* $p_c = \Theta(n^{1-k})$

4 On minimal elements

Our aim is to provide a bound on the number of variables occurring in a minimal unsatisfiable collection of constraints. Let us define what are minimal elements.

Definition 4.1 [minimal unsatisfiable collection] *A collection of constraints* $s \in \mathsf{CSP}(\mathcal{F})$ *is said to be* minimal *for* $\mathsf{UNSAT}(\mathcal{F})$ *if* $s \in \mathsf{UNSAT}(\mathcal{F})$ *and for all* $s' \leq s$, $s' \neq s$, $s' \in \mathsf{SAT}(\mathcal{F})$.

Let us now introduce some definitions.

Definition 4.2 *A Boolean constraint function* $f : \{0,1\}^k \longrightarrow \{0,1\}$ strongly depends on a 2XOR-relation *if :*
$\exists i,j \in \{1,\ldots,k\}, \forall (a_1,\ldots,a_k) \in \{0,1\}^k, f(a_1,\ldots,a_k) = 1 \Rightarrow a_i \oplus a_j = 1.$

Example 4.3 *The function* $g : \{0,1\}^k \longrightarrow \{0,1\}$ *where* $g^{-1}(1) = \{(0,1,0),(0,1,1),(1,0,1)\}$ *strongly depends on a* 2XOR*-relation since* $g(x,y,z) = 1 \Rightarrow x \oplus y = 1$.

Definition 4.4 *A Boolean constraint function* f *is said to be* reasonable *if* f *neither strongly depends on one component nor strongly depends on a* 2XOR*-relation. A set of constraint functions* $\mathcal{F}$ *is said to be* reasonable *if every function in* $\mathcal{F}$ *is reasonable.*

We get the following theorem

Theorem 4.5 *Let* $\mathcal{F}$ *be an interesting set of constraint functions. The set* $\mathcal{F}$ *is reasonable if and only if for every* s *minimal for* $\mathsf{UNSAT}(\mathcal{F})$ *we have*

$$|Var(s)| \leq (k-1)|s| - 1.$$

Proof: Given in the appendix. ∎

This result show that the $\mathsf{UNSAT}(\mathcal{F})$ property cannot be approximated by the property of having certain small unsatisfiable formulas as a subformula. More precisely we have the following result.

Proposition 4.6 *Let $\mathcal{F}$ be an interesting set of constraint functions. If $\mathcal{F}$ is reasonable then*

$(C1)$ $\forall K \in \mathbf{N}, \forall c \in (0,1)$,

$$\mu_{p_c(N)}(s \geq s' \text{ where } s' \in \mathsf{UNSAT}(\mathcal{F}) \text{ and } |s'| \leq K) \xrightarrow[N \to +\infty]{} 0.$$

It is worth pointing out that in his analyze Friedgut suggested that condition $(C1)$ is the one which really reveals the nature of the threshold and that it should be sufficient to prove the sharpness of the threshold for non artificial properties of the hypercube. Therefore, for Boolean constraint satisfaction problems $\mathsf{SAT}(\mathcal{F})$ which are natural problems one may raise the following conjecture.

Conjecture 4.7 *Let $\mathcal{F}$ be an interesting and reasonable set of constraint functions. Then, $\mathsf{SAT}(\mathcal{F})$ exhibits a sharp threshold*

Up to now satisfiability properties for which sharpness of the phase transition has been proved : k-SAT [FG99] for $k \geq 2$, k-XOR-SAT [CD01], NAE-k-SAT and 1-in-k-SAT [ACIM01] for $k \geq 3$ confirm our conjecture since each of these problems deals with interesting and reasonable sets (see Example 2.7).

Moreover, it is worth pointing out that 1-SAT and 2-XOR-SAT, which are the core of non-reasonable sets, both exhibit a coarse threshold.

5 Conclusion

Our study provides some hints for the study of phase transition. Up to now satisfiability problems have been studied on a problem-by-problem basis, thus blinding us from seeing the "essence" of thresholds phenomena. Our model suggests that 2-XOR-SAT is the core of the coarseness property for the phase transition of generalized satisfiability problems. Of course our framework can immediately be extended to finite domains (thus capturing 3-colorability for instance). It would be interesting to embark on an investigation of phase transition for such random constraint satisfaction problems.

References

[ACIM01] D. Achlioptas, A. Chtcherba, G. Istrate and C. Moore. The phase transition in NAESAT and 1-in-k-SAT. *Proceedings of the 11th ACM-SIAM Symposium on Discrete Algorithms*, SODA'2001, pages 721–722, 2001.

[AKKKMS98] D. Achlioptas, L.M. Kirousis, E. Kranakis, D. Krizanc, M. Molloy and Y.C. Stamatiou. Random constraint satisfaction: a more accurate picture. *Random structures and algorithms*, 12:253–269, 1998.

[AS00] D. Achlioptas and G.B. Sorkin. Optimal myopic algorithms for random 3-SAT. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 590–600, 2000.

[Bol85] B. Bollobás. *Random graphs*. Academic Press, 1985.

[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*, pages 620–627, IEEE, 1992.

[CD99] N. Creignou and H. Daudé. Satisfiability threshold for random XOR-CNF formulæ. *Discrete Applied Mathematics*, 96-97(1-3): 41–53, 1999

[CD01] N. Creignou and H. Daudé. Coarse and sharp thresholds for random k-XOR-CNF satisfiability. *Submitted*, 2001.

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

[DBM00] O. Dubois, Y. Boufkhad and J. Mandler. Typical random 3-SAT formulae and the satisfiability threshold. *Proceedings of the 11th ACM-SIAM Symposium on Discrete Algorithms*, SODA'2000, pages 124–126, 2000.

[FG99] E. Friedgut and an appendix by J. Bourgain. Sharp thresholds of graph properties, and the k-sat problem. *Journal of the A.M.S.*, 12(4):1017–1054, 1999.

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

[KS94] S. Kirkpatrick and B. Selman. Critical behavior in the satisfiability of random Boolean expressions. *Science*, 264:1297–1301,1994.

[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.

[SML96] B. Selman, D. Mitchell and H. Levesque. Generating hard satisfiability problems *Artificial intelligence* , 81(1-2):17–29, 1996.

[SD96] B.M. Smith and M.E. Dyer. Locating the phase transition in binary constraint satisfaction problems. *Artificial intelligence journal* , 81(1-2):155–181, 1996.

Appendix

Proof of Theorem 4.5:
Let s be a minimal unsatisfiable formula, $s \in \mathsf{UNSAT}(\mathcal{F})$ with $|s| \geq 2$. Suppose that we construct s in adding constraints one at a time (at time t we have t constraints). Since s is minimal each new constraint can be chosen in such a way that it has at least one variable in common with the previously selected constraints. Thus, at each time we have at most $(k-1)$ new variables in addition. Therefore $|Var(s)| \leq (k-1)|s|+1$. Our goal is now to characterize sets $\mathcal{F}$ such that one can get $|Var(s)| \geq (k-1)|s|$ for some s. At each time of our process we will say that a constraint is *free* at time t if it has $(k-1)$ variables appearing only once in the formula formed at time t. Variables occurring only once in free constraints at time t are also said to be *free* at time t. The other variables are called *dependent*. To get $|Var(s)| = (k-1)|s|+1$ it is clear that one has to add a free constraint at each time. Similarly, to get $|Var(s)| = (k-1)|s|$ one as to add a free constraint at each time of the process but at time t_0 where one has to add a constraint with $(k-2)$ new variables. Let F_t be the number of free constraints at time t.

Claim 5.1 *For* $2 < t < t_0$*,* $F_t \geq F_{t-1} \geq 2$.

Proof: If the dependent variable of the constraint added at time t is chosen among the variables free at time $t-1$ then $F_t = F_{t-1}$, otherwise $F_t = F_{t-1}+1$. ▮

At time t_0, for a new constraint C_0 one chooses $(k-2)$ new variables, and u and v two other variables already occurring in the formula.

- If u and v were both free at time $t-1$ and belonged to two different free constraints, then $F_{t_0} = F_{t_0-1} - 2$.
- If u and v were both free at time $t-1$ but belonged to the same free constraint, then $F_{t_0} = F_{t_0-1} - 1$.
- If u was free at time $t-1$ but v was not (or the converse), then $F_{t_0} = F_{t_0-1} - 1$.
- If neither u nor v was free at time $t-1$, then $F_{t_0} = F_{t_0-1}$.

Suppose that $t_0 < |s|$. As in the claim we have $F_{t+1} = F_t$ or $F_t + 1$ for $t > t_0$, thus $F_{|s|} \geq F_{t_0}$. Moreover if $F_{t_0} = 0$ at time t_0 then $F_{t_0+1} = 1$ for there is no free constraint at time t_0. Therefore, if $t_0 < |s|$, $F_{|s|} \geq 1$. Thus, if $|Var(s)| \geq (k-1)|s|$ and if $F_{|s|} = 0$ (i.e. s does not contain any free constraint when the process comes at end) then at any time $t < |s|$ one has to choose only one variable which is free in one of the two constraints that are free at time $t-1$, that is to say $F_t = 2$. Therefore we have proved the following lemma.

Lemma 5.2 *A minimal unsatisfiable formula $s \in \mathsf{UNSAT}(\mathcal{F})$ which verifies $|Var(s)| \geq (k-1)|s|$ either has a free constraint (at time $|s|$) or can be described by a cycle of length $|s|$ of the form:*

$$s = (f_1(x_1, \ldots, x_2), f_2(x_2, \ldots, x_3), \ldots, f_{|s|}(x_{|s|}, \ldots, x_1)),$$

(the variables not specified are variables occurring only once)

We are now in a position to prove the theorem. First let us prove that if $\mathcal{F}$ is reasonable then $|Var(s)| \leq (k-1)|s| - 1$. By contradiction, suppose that $|Var(s)| \geq (k-1)|s|$. Observe that s cannot contain any free constraint since $\mathcal{F}$ does not contain any function that strongly depends on one component and since s is minimal. Therefore, following the above lemma s is described by a cycle form $s = (f_1(x_1, \ldots, x_2), f_2(x_2, \ldots, x_3), \ldots, f_{|s|}(x_{|s|}, \ldots, x_1))$. But by assumption $\mathcal{F}$ is reasonable, therefore every f_i does not depend on a 2XOR-relation. Hence there exists $(a_1, \ldots a_k) \in \{0,1\}^k$ such that $f_i(a_1, \ldots, a_k) = 1$ and $a_1 = a_k$. Thus we get a satisfying assignment for s, a contradiction.

Conversely, let us prove that the property of the theorem is not satisfied for any interesting set of constraint functions $\mathcal{F}$ which is not reasonable. Indeed, let $f : \{0,1\}^k \longrightarrow \{0,1\}$ be a non reasonable constraint function. If f strongly depends on one component, then suppose w.l.o.g. that $f(x_1, \ldots, x_k) = 1 \Rightarrow x_1 = 0$. Since $\mathcal{F}$ is interesting there is a constraint function g in $\mathcal{F}$ such that $g(0, \ldots, 0) = 0$. Then the following $\mathcal{F}$-collection of constraints furnishes a counter-example:

$$f(x_1, y_2^1, \ldots, y_k^1) \wedge f(x_2, y_2^2, \ldots, y_k^2) \wedge \cdots f(x_k, y_2^k, \ldots, y_k^k) \wedge g(x_1, \ldots, x_k).$$

If f strongly depends on a 2XOR-relation, then suppose w.l.o.g. that $f(x_1, \ldots, x_k) = 1 \Rightarrow x_1 \oplus x_2 = 1$. Then the following $\mathcal{F}$-collection of constraints furnishes a counter-example:

$$f(x_1, x_2, y_3^1, \ldots, y_k^1) \wedge f(x_2, x_3, y_3^2, \ldots, y_k^2) \wedge f(x_3, x_1, y_3^3, \ldots, y_k^3).$$

▮