Super Solutions of Random Instances of Satisfiability

Peng Zhang and Yong Gao[✉]

Department of Computer Science, Irving K. Barber School of Arts and Sciences, University of British Columbia Okanagan, Kelowna V1V 1V7, Canada
yong.gao@ubc.ca

Abstract. We study the probabilistic behaviour of super solutions to random instances of the Boolean Satisfiability (SAT) and Constraint Satisfaction Problems (CSPs). Our analysis focuses on a special type of super solutions, the (1,0)-super solutions. For random k-SAT, we establish the exact threshold of the phase transition of the solution probability for the cases of $k = 2$ and 3, and upper and lower bounds on the threshold of the phase transition for the case of $k \geq 4$. For CSPs, by overcoming difficulties that do not exist in the probabilistic analysis of the standard solution concept, we manage to derive a non-trivial upper bound on the threshold for the probability of having a super solution.

1   Introduction

For many problems arising in uncertain, dynamic, or interactive environments, it is desirable to find solutions that can be modified at a low cost in response to changes of the environment. This requires that a solution has a certain degree of robustness or stability. For example, super solutions have been used to formalize the notion of a robust or stable solution to the Boolean Satisfiability problem and the constraint satisfaction problem [8,10]. An (a, b)-super solution to a CSP instance is a satisfying solution such that, if the values assigned to any set of a variables are no longer available, a new solution can be found by reassigning values to these a variables and at most b other variables.

In general, finding super solutions to SAT and CSPs is NP-complete. One of the fruitful approaches to such hard problems is to understand the typical-case complexity of a hard problem by studying the probabilistic behaviour of random instances [2,9]. By analyzing the threshold phenomena of the solution probability and the correlated easy-hard-easy pattern of the instance hardness of the standard solution concept for SAT and CSPs, much insight has been gained on the effectiveness of the many heuristics widely-used in practice to tackle these problems [6,7,9,11–13].

In this paper, we study the probabilistic behaviour of super solutions to random instances of SAT and CSPs. Our analysis focuses on a special (but highly non-trivial) type of super solutions, the (1,0)-super solutions. We denote the problems of finding (1, 0)-super solution for k-SAT and CSPs by (1, 0)-k-SAT and (1, 0)-CSP respectively.

In Sect. 2, we establish the exact threshold for the probability of having a $(1,0)$-super solutions to random 3-SAT by making use of an observation on the equivalence between a $(1,0)$-k-SAT and a standard satisfying solution of a properly-constructed $(k-1)$-SAT instance. In Sect. 3, we establish upper and lower bounds on the threshold of random $(1,0)$-k-SAT for $k \geq 4$. In Sect. 4, we study possible upper bounds on the threshold of random $(1,0)$-CSPs.

2   Super Solutions for Boolean Satisfiability

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a set of n Boolean variables. A literal is a variable or its negation. A *k-clause* is a disjunction of k different literals and a k-CNF formula is a conjunction of k-clauses. An assignment σ is a mapping $\sigma : X \rightarrow \{1,0\}^n$ and is said to satisfy a k-CNF formula F if each clause of F contains at least one literal that evaluates to true under σ. A satisfying assignment is also called a solution.

2.1   Equivalent Definitions of (1,0)-Super Solutions

As a special case of (a, b)-super solutions, a $(1,0)$-super solution for a k-SAT is a solution such that changing the value assigned to exactly one variable will not violate any clause. Equivalently, a $(1,0)$-super solution is an assignment such that every clause contains at least two literals that evaluate to true under the assignment. Another equivalent condition for a $(1,0)$-super solution is given below and plays a crucial role in our analysis.

Definition 1. *The projection of a clause $C = l_1 \vee \cdots \vee l_k$ is defined to be the conjunction of all $(k-1)$-clauses contained in C, i.e. $\pi(C) = \wedge_{i=1}^{k}(\vee_{j \neq i} l_j)$. We say that C projects onto $\pi(C)$ and call clauses in $\pi(C)$ siblings.*
    The projection of a CNF formula F is defined to be $\pi(F) = \wedge_{C_i \in F} \pi(C_i)$.

The following observation can be proved easily.

Lemma 1. *An assignment (1,0)-satisfies F if and only if it satisfies $\pi(F)$.*

The following theorem complements existing results on the worst-case complexity of super solutions [10].

Theorem 1. $(1,0)$-*k-SAT is in P for $k \leq 3$, and is NP-Complete otherwise.*

Proof. Any instance of $(1,0)$-3-SAT F can be solved by solving the 2-SAT instance of $\pi(F)$, which is in P. For $k \geq 4$, we first prove the NP-completeness of $(1,0)$-4-SAT via a reduction from 3-SAT. Note that, σ satisfies $(l_1 \vee l_2 \vee l_3)$ if and only if it $(1,0)$-satisfies $(l_1 \vee l_2 \vee l_3 \vee 1)$. For any 3-SAT F, we reduce it into a 4-SAT F' as following in three steps. First, create 4 additional variables, $Y = \{y_1, y_2, y_3, y_4\}$ and a 4-SAT F_y of all the possible $\binom{4}{2}$ clauses, where each clause has exactly two negations of variables. In order to $(1,0)$-satisfy F_y, we have $\sigma_y(y_i) = 1, 1 \leq i \leq 4$. Secondly, for each clause c_i in F, add $c_i' = (c_i \vee y_1)$ into F'. Finally, let F' be the conjunction of F' and F_y. Now, σ is a solution of F if and only if it is a $(1,0)$-solution of F'. Thus $(1,0)$-4-SAT is NP-complete. Similar methods can reduce any k-SAT instance to $(1,0)$-$(k+1)$-SAT instance.

2.2   Random Models of k-SAT

We denote by $F_k(n, m)$ the standard random model for k-CNF formulas on n variables, where the m clauses are selected uniformly at random without replacement from the set of all possible $2^k \binom{n}{k}$ k-clauses. We say that a sequence of events $\mathcal{E}_n$ occurs *with high probability* (w.h.p.) if $\lim_{n \to \infty} \mathbb{P}[\mathcal{E}_n] = 1$. As sometimes it is hard to directly analyze $F_k(n, m)$ due to the dependence created by selecting the clauses without replacement, we consider two related models. The first model selects from all $2^k \binom{n}{k}$ proper clauses with replacement. The second model selects each literal uniformly and independently with replacement. Both models may result in improper formula and the second model may have improper clauses. As long as k is fixed, the number of improper clauses and repeated clauses is $o(n)$. Therefore, with-high-probability properties of (1,0)-satisfiability hold in all these three models simultaneously. For notation convenience, we denote all three models by $F_k(n, m)$. When $k \leq 3$, we use the first model. When $k \geq 4$, we use the second model. We also assume that k is fixed.

Due to Lemma 1, the probability for $F_k(n, m)$ to be (1,0)-satisfiable equals the probability for its projection $\pi(F)$ to be satisfiable. This, however, does not imply that the probability for a random $F_k(n, m)$ to be $(1, 0)$-satisfiable equals the probability for $F_{k-1}(n, km)$ to be satisfiable. The following result on the exact threshold of the solution probability of (1,0)-2-SAT is not hard to establish.

Theorem 2. $F_2(n, m)$ *is (1,0)-satisfiable w.h.p. when* $m = o(\sqrt{n})$ *and is (1,0)-unsatisfiable w.h.p. when* $m = \omega(\sqrt{n})$.

Proof. We say that two clauses are conflicting if some literal in one clause is the negation of some literal in the other clause. Note that a 2-CNF formula F is (1,0)-satisfiable if and only if no conflicting clauses exists. Let $F = C_1 \wedge \cdots \wedge C_m$ and $X_{i,j}$ be the indicator variable that C_i conflicts with C_j. Then, $\mathbb{E}[X_{i,j}] = \frac{2(2(n-1)-1)+1}{2^2 \binom{n}{2}} = \frac{4n-5}{2n(n-1)}$. Denote by $X = \sum_{(i,j)} X_{i,j}$ the number of conflicting pairs in F.

$$\mathbb{E}[X] = \binom{m}{2} \mathbb{E}[X_{i,j}] = \frac{m^2}{n}(1 - o(1))$$

When $m = o(\sqrt{n})$, $\lim_{n \to \infty} \mathbb{P}[X > 0] \leq \lim_{n \to \infty} \mathbb{E}[X] = 0$.

Let $t = \binom{m}{2}$ and $p = \mathbb{E}[X_{i,j}]$, then $\mathbb{E}[X] = tp$. Note that, X^2 is composed of t^2 items of $X_{i,j} X_{i',j'}$. Group these items according to $h = |\{i, j, i', j'\}|$. We see that $\mathbb{E}[X_{i,j} X_{i',j'}]$ equals p when $h = 2$, and equals p^2 otherwise. Thus, $\mathbb{E}[X^2] = tp + (t^2 - t)p^2$. When $m = \omega(\sqrt{n})$,

$$\lim_{n \to \infty} \mathbb{P}[X > 0] \geq \lim_{n \to \infty} \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} = \lim_{n \to \infty} \frac{tp}{tp + 1 - p} = 1,$$

where the first inequality is due to the Cauchy-Schwarz inequality.

2.3   An Exact Threshold for the Solution Probability of (1,0)-3-SAT

We use the equivalence (Lemma 1) for a $(1,0)$-super solution to study the threshold for the solution probability of random $(1,0)$-3-SAT. We upper bound (resp. lower bound) the probability for F to be $(1,0)$-unsatisfiable by the probability of some necessary (resp. sufficient) condition on the satisfiability of its projection $\pi(F)$ (a 2-CNF formula). The conditions were proposed in [4]. It is important to note that while $\pi(F)$ is a 2-CNF formula obtained from a random 3-CNF formula $F_3(n, m)$, $\pi(F)$ itself is not distributed as the random 2-CNF formula $F_2(n, m)$. This is the major obstacle we have to deal with in our analysis.

Theorem 3. $F_3(n, rn)$ *is (1,0)-satisfiable w.h.p. if* $r < 1/3$ *and is (1,0)-unsatisfiable w.h.p. if* $r > 1/3$.

The proof of the above result is presented in two lemmas. In the proof, we use F to denote a random formula $F_3(n, rn)$, $m = rn$, and write $N = 2^3 \binom{n}{3}$. A *bicycle* ([4]) of length $s \geq 2$, is a conjunction of $s + 1$ 2-clauses $C_0, \cdots, C_s$ defined on s variables $\{x_1, x_2, \cdots, x_s\}$ such that $C_i = \overline{l_i} \vee l_{i+1}, 0 < i < s$, $C_0 = u \vee l_1$, and $C_s = \overline{l_s} \vee v$, where

1. l_i is either x_i or $\overline{x_i}$, and
2. u and v are from $\{x_i, \overline{x_i} \mid 1 \leq i \leq s\}$.

It can be shown that if a 2-SAT is unsatisfiable, then it must contain a bicycle ([4]).

Lemma 2. $F_3(n, rn)$ *is (1,0)-satisfiable w.h.p. when* $r < 1/3$.

Proof. For any fixed bicycle $B = C_0 \wedge \cdots \wedge C_s$, we consider the number of 3-CNF formulae F such tht $B \subset \pi(F)$. Let $\mathcal{C} = \{C_1, C_2, \cdots, C_{s-1}\}$. Since clauses in $\mathcal{C}$ are defined on distinct literals, no two clauses in $\mathcal{C}$ can be siblings with respect to the projection of any 3-clause. Similarly, no three clauses from B can be siblings with respect to a 3-clause. The only possible siblings are (C_0, C_i) and (C_s, C_i) for some $0 \leq i \leq s$.

Denote by $g(s, l)$ the number of 3-CNF formulas F such that $B \subset \pi(F)$, where $l = 0, 1$, or 2 is the number of clause pairs that belong to the projection of a same 3-clause in F. We have

$$g(s, l) = \binom{N - (s + 1 - l)}{m - (s + 1 - l)} \cdot (2(n - 2))^{s+1-2l}.$$

Let $p(s)$ denote the probability that a bicycle of length s over a given (ordered) set of s variables is part of $\pi(F)$. Then,

$$p(s) \leq \binom{N}{m}^{-1} (g(s, 0) + 2s \cdot g(s, 1) + g(s, 2))$$

$$\leq \binom{N}{m}^{-1} 2(s + 1) \binom{N - (s - 1)}{m - (s - 1)} \cdot (2(n - 2))^{s-3}$$

$$\leq \left(\frac{3r}{2(n - 1)}\right)^{s-1} \cdot \frac{s + 1}{2(n - 2)^2}$$

Let N_s denote the number of different bicycles of length of s and X be the number of bicycles in $\pi(F)$. As $N_s < n^s 2^s (2s)^2$, we have

$$\mathbb{E}[X] = \sum_{s=2}^{n} N_s p(s) \leq \frac{4n}{(n-2)^2} \sum_{s=2}^{n} s^2(s+1)(\frac{3rn}{n-1})^{s-1}$$

When $r < 1/3$, $\lim\limits_{n \to \infty} \mathbb{P}[X > 0] \leq \lim\limits_{n \to \infty} \mathbb{E}[X] = 0$.   □

A *snake* of length $t \geq 1$ is the conjunction of $2t$ 2-clauses $C_0, C_1, \cdots, C_{2t-1}$ and has following structure.

1. $C_i = (\bar{l}_i \vee l_{i+1}), 0 \leq i \leq 2t - 1$. $l_0 = l_{2t} = \bar{l}_t$
2. For any $0 < i, j < 2t - 1$, $l_i \neq l_j$ and $l_i \neq \bar{l}_j$.

If $\pi(F)$ contains a snake, then F is not (1,0)-satisfiable. We show that w.h.p. $\pi(F)$ contains a snake of length $\log_{3r} n$.

Lemma 3. $F_3(n, rn)$ is (1,0)-unsatisfiable w.h.p. when $r > 1/3$.

Proof. Let A be a snake of length t, X_A be the indicator variable that A occurs in F. Note that there only the two pairs, (C_0, C_{t-1}) and (C_t, C_{2t-1}), can potentially be siblings with respect to the projection of a 3-clause. Let $s = 2t - 1$ and let $p(s)$ be the probability that a snake of length t over a given set of variables is in $\pi(F)$. We have

$$p(s) = \binom{N}{m}^{-1} (g(s,0) + 2g(s,1) + g(s,2))$$

$$\approx \binom{N}{m}^{-1} 4g(s,2) \approx (\frac{3r}{2n})^{s-1} \frac{1}{n^2}$$

Let X denote the number of snakes of length t in $\pi(F)$. $\mathbb{E}[X] = \binom{n}{s} s! 2^s p(s) \approx (3r)^s/n$. When $r > 1/3$ and $t = \omega(\log_{3r} n)$, $\lim_{n \to \infty} \mathbb{E}[X] = \infty$.

In order to apply the second moment method to X, we have to consider correlation between snakes. To satisfy a clause $(l_i \vee l_j)$, if $\bar{l}_i$ is false, then l_j must be true. This implication can be represented by two arcs $(\bar{l}_i, l_j)$, $(\bar{l}_j, l_i)$ in a digraph. The digraph for a snake of length t is a directed cycle $\bar{l}_t, l_1, l_2, \cdots, l_s, \bar{l}_t$. Two snakes are not independent if and only if there are some common arcs between the corresponding directed cycles. Let B be another snake of length t. Suppose B share i arcs with A and these arcs contain j vertices. Then, taking into consideration the fact that the dominating term is still the one where exactly two pairs in B are siblings in the projection of the formula, we have

$$\mathbb{P}[B|A] \leq \frac{\binom{N-2t-(2t-i)}{m-2t-(2t-i)} \cdot (2(n-2))^{2t} \cdot (2(n-2))^{2t-i}}{\binom{N-2t}{m-2t} \cdot (2(n-2))^{2t}}$$

$$\leq \left(\frac{m-2t}{N-2t} \cdot 2(n-2)\right)^{2t-i} \leq \left(\frac{3r}{2n}\right)^{2t-i}$$

It is clear that those common i arcs comprise $(j - i)$ directed paths. Fixing A, there are L_1 number of choices for the shared j vertices to occur in B, and there are L_2 number of choices for the remaining $2t - j$ vertices to occur in B.

$$L_1 = \left(2 \cdot \binom{2t}{2(j-i)}\right)^2 \cdot (j-i)! \le 4 \cdot (2t)^{4(j-i)}$$

$$L_2 \le \binom{n-j+1}{2t-j}(2t-j)! \cdot 2^{2t-j} \le (2(n-j+1))^{2t-j}$$

For a given A, let $\mathcal{A}(i,j)$ be the set of snakes sharing i arcs and j vertices with A, and write

$$p(i,j) = \sum_{B \in \mathcal{A}(i,j)} \mathbb{P}[B|A] = L_1 L_2 \mathbb{P}[B|A]$$

$$\le \left(\frac{3r}{2n}\right)^{2t-i} 4(2t)^{4(j-i)} (2(n-j+1))^{2t-j}.$$

If $i \le t$, then $i+1 \le j \le 2i$. If $t < i \le 2t$, then $i+1 \le j \le 2t$. Let $A \sim B$ denote the fact that A and B are dependent.

$$\sum_{A \sim B} \mathbb{P}[B|A] = \sum_{i=1}^{2t} \sum_{j=i+1}^{\min\{2i,2t\}} p(i,j) = \sum_{j=2}^{2t} \sum_{i=j/2}^{j-1} p(i,j)$$

$$\le \sum_{j=2}^{2t} (2(n-j+1))^{2t-j} 4 \sum_{i=j/2}^{j-1} \left(\frac{3r}{2n}\right)^{2t-i} (2t)^{4(j-i)}$$

$$\le \sum_{j=2}^{2t} (2(n-j+1))^{2t-j} 4 \cdot \frac{j}{2} \left(\frac{3r}{2n}\right)^{2t-j+1} (2t)^4$$

$$\le \sum_{j=2}^{2t} 2j \left(\frac{3r}{2n}\right)(2t)^4$$

$$\le \Theta(1) \cdot \frac{1}{n}t^6 = o(\frac{1}{n}(3r)^{2t}) = o(E[X]).$$

According to corollary 4.3.5 of [3], $\lim_{n \to \infty} \mathbb{P}[X > 0] = 1$.

2.4 Thresholds for the Solution Probability of $(1,0)$-k-SAT

Using Markov's inequality, the following upper bound on the threshold of the phase transition can be proved:

Theorem 4. *For all $k \ge 3$, $F_k(n, rn)$ is $(1,0)$-unsatisfiable w.h.p. when $r > \frac{2^k}{k+1} \ln 2$.*

In the rest of this section, we establish a lower bound on the threshold for $k > 3$ and show that the ratio of the lower bound over the upper bound goes to 1 as k goes to infinity. Our analysis uses the techniques introduced in [2] for proving lower bounds on the threshold for the phase transition of standard satisfying solutions of random SAT, but the calculation we have to deal with is even more complicated. The idea is to use a weighting scheme on satisfying assignments when applying the second moment method to prove lower bounds on the threshold.

For a clause c, denote by $\mathcal{S}(c)$ the set of $(1,0)$-super solutions of c, $\mathcal{S}^0(c)$ (resp. $\mathcal{S}^1(c)$) the set of assignments that satisfies exactly 0 (resp. 1) literal of c. Define $H(\sigma, c)$ be the number of satisfied literals minus the number of unsatisfied literals. For an event A, let $\mathbf{1}_A$ be the indicator variable that A occurs. The *weight* of σ w.r.t. c is defined as $w(\sigma, c) = \gamma^{H(\sigma,c)} \mathbf{1}_{\sigma \in \mathcal{S}(c)}$, $0 < \gamma < 1$ and is determined by k. These definitions extend naturally to a formula F: $w(\sigma, F) = \gamma^{H(\sigma,F)} \mathbf{1}_{\sigma \in \mathcal{S}(F)} = \prod_{c_i} w(\sigma, c_i)$. Let $X = \sum_{\sigma} w(\sigma, F)$. F is $(1,0)$-satisfiable if and only if $X > 0$.

Note that by viewing an instance of $(1,0)$-k-SAT as a generalized Boolean satisfiability problem (Boolean CSP) and applying the conditions established in [5], random $(1,0)$-k-SAT has a sharp threshold. Therefore, to show $X > 0$ w.h.p., it is sufficient to prove that $\mathbb{P}[X > 0]$ is greater than some constant.

For a fixed σ and a random k-clause c, since σ $(1\text{-}0)$-satisfies c if at least two literals in c evaluate to true under σ, we have

$$\mathbb{E}[w(\sigma, c)] = \mathbb{E}\left[\gamma^{H(\sigma,c)}(1 - \mathbf{1}_{\sigma \in \mathcal{S}^0(c)} - \mathbf{1}_{\sigma \in \mathcal{S}^1(c)})\right]$$

$$= (\frac{\gamma + \gamma^{-1}}{2})^k - 2^{-k}\gamma^{-k} - k2^{-k}\gamma^{-k+2} = \phi(\gamma)$$

Thus, $\mathbb{E}[X] = \sum_{\sigma} \prod_{c_i} \mathbb{E}[w(\sigma, c)] = (2\phi(\gamma)^r)^n$.

We now consider $\mathbb{E}[X^2]$. Fix a pair of assignments σ, τ such that they overlap each other on $z = \alpha n$ variables. Consider a random k-clause c and write

$$f(\alpha) = \mathbb{E}[w(\sigma, c)w(\tau, c)] = \mathbb{E}\left[\gamma^{H(\sigma,c)+H(\tau,c)}\mathbf{1}_{\sigma, \tau \in \mathcal{S}(c)}\right].$$

We have the following equations for relevant events

$$\mathbf{1}_{\sigma, \tau \in \mathcal{S}(c)} = 1 - \mathbf{1}_{\sigma \notin \mathcal{S}(c)} - \mathbf{1}_{\tau \notin \mathcal{S}(c)} + \mathbf{1}_{\sigma, \tau \notin \mathcal{S}(c)},$$

$$\mathbf{1}_{\sigma \notin \mathcal{S}(c)} = \mathbf{1}_{\sigma \in \mathcal{S}^0(c)} + \mathbf{1}_{\sigma \in \mathcal{S}^1(c)},$$

$$\mathbf{1}_{\sigma, \tau \notin \mathcal{S}(c)} = \mathbf{1}_{\sigma \in \mathcal{S}^0(c), \tau \in \mathcal{S}^0(c)} + \mathbf{1}_{\sigma \in \mathcal{S}^0(c), \tau \in \mathcal{S}^1(c)}$$
$$+ \mathbf{1}_{\sigma \in \mathcal{S}^1(c), \tau \in \mathcal{S}^0(c)} + \mathbf{1}_{\sigma \in \mathcal{S}^1(c), \tau \in \mathcal{S}^1(c)},$$

and for mathematical expectations

$$\mathbb{E}\left[\gamma^{H(\sigma,c)+H(\tau,c)}\mathbf{1}\right] = (\alpha(\frac{\gamma^2+\gamma^{-2}}{2})+1-\alpha)^k,$$

$$\mathbb{E}\left[\gamma^{H(\sigma,c)+H(\tau,c)}\mathbf{1}_{\sigma\notin\mathcal{S}(c)}\right] = 2^{-k}((\alpha\gamma^{-2}+1-\alpha)^k$$
$$+ k(\alpha\gamma^{-2}+1-\alpha)^{k-1}(\alpha\gamma^2+1-\alpha)),$$

$$\mathbb{E}\left[\gamma^{H(\sigma,c)+H(\tau,c)}\mathbf{1}_{\sigma,\tau\notin\mathcal{S}(c)}\right] = 2^{-k}(\alpha^k\gamma^{-2k}+2k\gamma^{-2k+2}\alpha^{k-1}(1-\alpha)$$
$$+ \gamma^{-2k+4}(k\alpha^k+k(k-1)\alpha^{k-2}(1-\alpha)^2)).$$

Therefore, the expectation of X^2 can be written as

$$\mathbb{E}\left[X^2\right] = \sum_{\sigma,\tau}\mathbb{E}[w(\sigma,F)w(\tau,F)]$$

$$= \sum_{\sigma,\tau}\prod_{c_i}\mathbb{E}[w(\sigma,c_i)w(\tau,c_i)] = 2^n\sum_{z=0}^{n}\binom{n}{z}f(z/n)^{rn}$$

The following lemma from [1] enables us to consider the dominant part of $\mathbb{E}\left[X^2\right]$.

Lemma 4. *Let h be a real analytic positive function on $[0,1]$ and define $g(\alpha) = h(\alpha)/(\alpha^\alpha(1-\alpha)^{1-\alpha})$, where $0^0 \equiv 1$. If g has exactly one maximum at $g(\beta)$, $\beta \in (0,1)$, and $g''(\beta) < 0$, then there exists constant $C > 0$ such that for all sufficient large n, $\sum_{z=0}^{n}\binom{n}{z}h(z/n)^n \leq C \times g(\beta)^n$.*

Define $g_r(\alpha) = f(\alpha)^r/(\alpha^\alpha(1-\alpha)^{1-\alpha})$ and say $g_r(\alpha)$ satisfies the *dominant condition* if $g_r''(1/2) < 0$ and $g_r(1/2)$ is the unique global maximum. According to lemma 4 and $\phi(\gamma)^2 = f(1/2)$, if $g_r(\alpha)$ satisfies the dominant condition, then

$$\mathbb{P}[X > 0] > \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} = \frac{4^n f(1/2)^{rn}}{\mathbb{E}[X^2]}$$
$$> \frac{(2g_r(1/2))^n}{C \cdot (2g_r(1/2))^n} = \frac{1}{C},$$

where C is a constant when k is fixed.

If we can find suitable γ and r so that $g_r(\alpha)$ satisfies the dominant condition, then $X > 0$ w.h.p.. It is clear that the dominant condition implies $f'(1/2) = 0$. According to [2], a necessary condition for $f'(1/2) = 0$ is that the sum of vectors scaled by their corresponding weight is 0, i.e., $\sum_{\mathbf{v}\in\{0,1\}^k} w(\mathbf{v})\mathbf{v} = \mathbf{0}$. For $(1,0)$-k-SAT, this is $\sum_{i=1}^{k}\binom{k}{i}\gamma^{2i}(2i-k) = 0$. When $k = 4$, this equation requires $\gamma = 0$. Thus, the weighting scheme is not meaningful when $k = 4$. Therefore, we consider the case of $k > 4$ first and then the case of $k = 4$ in a different way.

It is too complicated to directly prove that $g_r(\alpha)$ satisfies the dominant condition, at least for small k. Therefore, we plot figures to show how $g_r(\alpha)$ changes when k is fixed. Figure 1 shows the case when $k = 5$. For each k, when r is smaller than some r_k^*, $g_r(\alpha)$ satisfies the dominant condition and $F_r(n,rn)$ is $(1,0)$-satisfiable w.h.p. Thus r_k^* is a lower bound for $F_k(n,rn)$. We do this analysis for k up to 11 and show the values in Table 1. We can see that the ratio of

the lower bound over the upper bound of thresholds of $F_k(n, rn)$ goes to 1 as k becomes large. We still have to solve the case $k = 4$ separately. The weighting scheme, $w(\sigma, c) = \gamma^{H(\sigma,c)} \mathbf{1}_{\sigma \in \mathcal{S}(c)}$, does not work for any $\gamma > 0$. This is because $H(\sigma, F)$ is either 0 or positive. Thus, a compromise is to consider only those assignments which satisfy $H(\sigma, F) = 0$. Specifically, for each clause of F, exactly two literals are satisfied and exactly two literals are unsatisfied. And every satisfying assignment has the same weight, 1. By doing this, the likelihood for an assignment not to be in X is doubled. Therefore, the upper bound for such solutions becomes $\frac{2^{k-1}}{1+k} \ln 2$, half of the upper bound for $(1,0)$-4-SAT. The remaining analysis is similar to the analysis of $k > 4$. The r_4^* we found is 0.602.
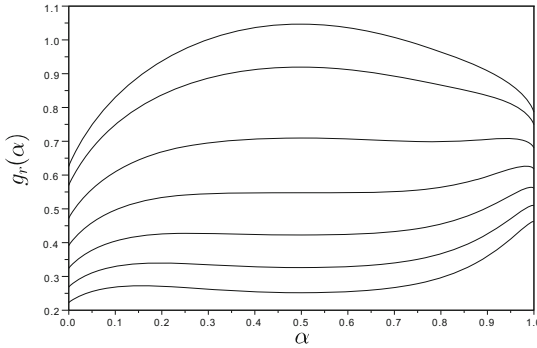

Fig. 1. $k = 5$, $r = 1, 1.2, 1.6, 2, 2.4, 2.8, 3.2$ (top down)

Table 1. Upper bound and lower bound for different k

k	4	5	6	7	8	9	10	11
Upper Bound	2.2	3.6	6.3	11.1	19.7	35.5	64.5	118.3
Lower Bound	0.6	1.6	3.7	7.8	15.8	30.9	59.3	113.4

3   Super Solutions for Random Binary CSPs

We consider random binary CSPs defined on a domain D of size $|D| = d$. A binary CSP $\mathcal{C}$ consists of a set of variables $X = \{x_1, \cdots, x_n\}$ and a set of binary constraints $(C_1, \cdots, C_m)$. Each constraint C_i is specified by its *constraint scope*, an unordered pair of two variables in X, and a *constraint relation* R_{C_i} that defines a set of incompatible value tuples in the binary relation $D \times D$ for the scope variables. An incompatible value tuple is also called a *restriction*. The constraint graph of a binary CSP is a graph whose vertices correspond to the set of variables and edges correspond to the set of constraint scopes. We use the following random CSP model $\mathcal{B}_{n,m}^{d,q}$ where the domain size is allowed to increase with the number of variables.

1. Its constraint graph is a random graph $G(n, m)$ where the m edges are selected uniformly at randomly from all the possible $\binom{n}{2}$ edges.
2. For each edge, its constraint relation is determined by choosing each value tuple in $D \times D$ as a restriction independently with probability q.

Proposed and studied in a series of papers by Xu, et al., this class of models for random CSPs is known as the Model RB [11–13]. In particular, the exact threshold of the phase transition of standard satisfiability has been established in [12] and the (resolution) complexity of random instances at the phase transition has been analyzed in [13].

Denote by $H(\sigma_1, \sigma_2)$ the set of variables being assigned different values by σ_1 and σ_2, i.e., $H(\sigma_1, \sigma_2) = \{x_i | \sigma_1(x_i) \neq \sigma_2(x_i), 1 \leq i \leq n\}$. Let σ be a fixed assignment and I be a random $\mathcal{B}^{d,q}_{n,m}$ instance. Define the following three events:

1. $S(\sigma)$: σ is a solution for I.
2. $S_i(\sigma)$: there exists another solution σ' for I such that $H(\sigma, \sigma') = \{x_i\}$.
3. $T(\sigma)$: σ is a $(1, 0)$-super solution for I.

It is clear that $\mathbb{P}[T(\sigma)] = \mathbb{P}[S(\sigma)]\mathbb{P}[\cap_{1 \leq i \leq n} S_i(\sigma)|S(\sigma)]$. Estimating the probability of a $(1, 0)$-super solution for a random CSP instance is, however, more complicated than estimating the probability of a satisfying assignment, largely due to the fact that the events $S_i(\sigma), 1 \leq i \leq n$, are not independent. This is the major hurdle we need to overcome. Note that in a random CSP instance, the selection of constraints and the selection of restrictions for each constraint are independent. Let $C \subset (X \times X)^m$ be the collection of all possible sets of m unordered pairs of variables. For a given set $e \in C$ of m unordered pairs, denote by $E(e)$ the event that e is selected as the set of constraints of the random instance I. Let m_i be the number of constraints x_i is involved with. Considering an assignment σ', $H(\sigma, \sigma') = \{x_i\}$, it is clear that $\mathbb{P}\left[\overline{S(\sigma')}|S(\sigma) \cap E(e)\right] = 1 - (1 - q)^{m_i}$. Let $D' = D \setminus \{\sigma(x_i)\}$, $\sigma'(x_i) = y$, $p = 1 - q$, then

$$\mathbb{P}[S_i(\sigma)|S(\sigma) \cap E(e)] = \mathbb{P}\left[\cup_{y \in D'} S(\sigma')|S(\sigma) \cap E(e))\right]$$
$$= \mathbb{P}\left[\overline{\cap_{y \in D'} \overline{S(\sigma')}}|S(\sigma) \cap E(e))\right]$$
$$= 1 - \mathbb{P}\left[\cap_{y \in D'} \overline{S(\sigma')}|S(\sigma) \cap E(e))\right]$$
$$= 1 - (1 - (1 - q)^{m_i})^{d-1}.$$

This shows that, conditioned on $S(\sigma)$ and fixed constraint sets e, $S_i(\sigma)$ and $S_j(\sigma)$ are independent for any $i \neq j$.

$$\mathbb{P}[T(\sigma)] = \mathbb{P}[\cup_{e \in C} (E(e) \cap S(\sigma) \cap (\cap_{1 \leq i \leq n} S_i(\sigma)))]$$
$$= \sum_{e \in C} \mathbb{P}[E(e)]\mathbb{P}[S(\sigma)|E(e)]\mathbb{P}\left[\cap_{i \in [n]} S_i(\sigma)|S(\sigma) \cap E(e)\right]$$
$$= \binom{\binom{n}{2}}{m}^{-1} p^m \sum_{e \in C} \prod_{i=1}^{n} \left(1 - (1 - p^{m_i})^{d-1}\right). \tag{1}$$

Let Y_σ be an indicator variable of T_σ and $Y = \sum_\sigma Y_\sigma$ be the number of $(1, 0)$-super solutions. We have

$$\mathbb{E}[Y] = d^n \cdot \mathbb{E}[Y_\sigma] = d^n \cdot \mathbb{P}[T_\sigma]$$

We have the following lower and upper bounds on the threshold of solution probability.

Theorem 5. *Consider the random CSP $\mathcal{B}_{n,m}^{d,q}$ with $d = \sqrt{n}$, $p = 1 - q$, $m = c \cdot n \ln n$ where c is a positive constant.*

- *If $c > -\frac{1}{3 \ln p}$, $\lim\limits_{n \to \infty} \mathbb{E}[Y] = 0$ and thus $\mathcal{B}_{n,m}^{d,q}$ is $(1,0)$-unsatisfiable w.h.p.*
- *If $c < -\frac{1}{10 \ln p}$ and $q = 1 - p < 0.43$, $\lim\limits_{n \to \infty} \mathbb{E}[Y] \to \infty$.*

Proof. Due to space limit, we give a brief proof of the first part of the conclusion and omit the proof of the second part of the conclusion.

The right hand side of Eq. (1), subject to $\sum_{i=1}^{n} m_i = 2m$, achieves the global maximum when $m_i = \frac{2m}{n}$, $1 \le i \le n$. This can be proved by the method of Lagrange multipliers. Let $c = c' \cdot -\frac{1}{\ln p}$, then

$$
\begin{aligned}
\mathbb{E}[Y] &\le (d \cdot p^{c \ln n} \cdot (1 - (1 - p^{2c \ln n})^{d-1}))^n \\
&= (d \cdot n^{c \ln p} \cdot (1 - (1 - n^{2c \ln p})^{d-1}))^n \\
&\approx (n^{1/2 - c'} \cdot (1 - (1 - n^{-2c'})^{n^{1/2}}))^n.
\end{aligned}
$$

For any a, b satisfying $0 \le a \le 1$ and $ab < 1$, $(1 - a)^b \ge 1 - ab$. If $c' > 1/3$, then

$$
\mathbb{E}[Y] \le (n^{1/2 - c'} \cdot n^{-2c'} n^{1/2})^n = (n^{1 - 3c'})^n \to 0.
$$

4   Conclusions

To the best of our knowledge, we have conducted (for the first time) a probabilistic analysis of super solutions of random instances of SAT and CSPs. While we have focused on the special (but already challenging) case of $(1,0)$-super solutions, some of our analysis extends to the case of $(a,0)$-super solutions for $a > 1$. For random instances of CSPs, new analytical methods and ideas are needed to obtain a more detailed characterization of the behavior of the super solutions, and we leave this as a future work. It is also highly interesting to conduct a systematic empirical analysis to fully understand the hardness of solving random instances of $(1,0)$-k-SAT as well as the hardness of solving the projected standard SAT instances, which may serve as suite of SAT benchmark with a unique structural properties. Finally, we wonder if our analysis can be extended to random instances of other problems such as graphical games where solution concepts similar to super solutions have been used.

References

1. Achlioptas, D., Moore, C.: The asymptotic order of the random k-sat threshold. In: Proceedings of the FOCS, pp. 779–788 (2002)
2. Achlioptas, D., Peres, Y.: The threshold for random k-sat is 2k (ln 2 - o(k)). In: STOC 2003, pp. 223–231. ACM, New York (2003)

3. Alon, N., Spencer, J.H.: The Probabilistic Method. Wiley, New York (2008)
4. Chvatal, V., Reed, B.: Mick gets some (the odds are on his side) (satisfiability). In: SFCS 1992, pp. 620–627. IEEE Computer Society, Washington (1992)
5. Creignou, N., Daud, H.: Combinatorial sharpness criterion and phase transition classification for random CSPS. Inf. Comput. **190**(2), 220–238 (2004)
6. Culberson, J., Gent, I.: Frozen development in graph coloring. Theoret. Comput. Sci. **265**(1–2), 227–264 (2001)
7. Gao, Y., Culberson, J.: Consistency and random constraint satisfaction models. J. Artif. Intell. Res. **28**, 517–557 (2007)
8. Ginsberg, M.L., Parkes, A.J., Roy, A.: Supermodels and robustness. In: AAAI 1998/IAAI 1998, Menlo Park, CA, USA, pp. 334–339 (1998)
9. Gomes, C., Walsh, T.: Randomness and structure. In: Rossi, F., van Beek, P., Walsh, T. (eds.) Handbook of Constraint Programming, pp. 639–664. Elsevier, Amsterdam (2006)
10. Hebrard, E., Hnich, B., Walsh, T.: Super solutions in constraint programming. In: Régin, J.-C., Rueher, M. (eds.) CPAIOR 2004. LNCS, vol. 3011, pp. 157–172. Springer, Heidelberg (2004)
11. Xu, K., Boussemart, F., Hemery, F., Lecoutre, C.: Random constraint satisfaction: easy generation of hard (satisfiable) instances. Artif. Intell. **171**, 514–534 (2007)
12. Xu, K., Li, W.: Exact phase transitions in random constraint satisfaction problems. J. Artif. Intell. Res. **12**, 93–103 (2000)
13. Xu, K., Li, W.: Many hard examples in exact phase transitions. Theoret. Comput. Sci. **355**, 291–302 (2006)