Biased random satisfiability problems: From easy to hard instances

A. Ramezanpour*
Institute for Advanced Studies in Basic Sciences, Zanjan 45195-1159, Iran

S. Moghimi-Araghi†
Department of Physics, Sharif University of Technology, P.O. Box 11365-9161, Tehran, Iran
(Received 17 November 2004; published 1 June 2005)

In this paper we study biased random K-satisfiability (K-SAT) problems in which each logical variable is negated with probability p. This generalization provides us a crossover from easy to hard problems and would help us in a better understanding of the typical complexity of random K-SAT problems. The exact solution of 1-SAT case is given. The critical point of K-SAT problems and results of replica method are derived in the replica symmetry framework. It is found that in this approximation $\alpha_c \propto p^{-(K-1)}$ for $p \to 0$. Solving numerically the survey propagation equations for $K=3$ we find that for $p < p^* \sim 0.17$ there is no replica symmetry breaking and still the SAT-UNSAT transition is discontinuous.

DOI: 10.1103/PhysRevE.71.066101 PACS number(s): 02.50.−r, 89.75.−k, 64.60.Cn, 75.10.Nr

I. INTRODUCTION

Optimization problems are subject of recent studies in the context of complex systems [1–4]. Random K-satisfiability (K-SAT) problems are well known examples of these problems which have their origin in computer science and complexity theory [5–11]. Finding the configuration of N logical variables which satisfies a formula of M clauses is a hard problem and indeed lies in the class of *NP*-complete problems for $K \geq 3$ [12].

From a physical point of view the interesting feature of random K-SAT problems is the presence of phase transitions in the thermodynamic limit where N and M approach infinity and $\alpha := M/N$ remains finite [2]. Here the transition is between SAT and UNSAT phases where a typical instance of the formula is satisfied or unsatisfied respectively with probability 1. It is around the phase transition that the time needed to find the solution of a typical instance grows exponentially with the size of the problem N.

In this paper we study a generalized version of the random K-SAT problems where each logical variable is negated with probability p rather than $1/2$ as in the original random K-SAT. The aim of this generalization is to go continuously from easy instances of the problem to hard ones. Clearly for $p=0$ we have an easy random K-SAT for all values of K. On the other hand as the studies indicate, the problem is hard for $p=1/2$ and $K \geq 3$. Thus one expects a crossover from easy to hard region by increasing p from zero. Is it possible to define a point beyond which one can say that the problem becomes hard? How the problem approaches the hard regime? What are the universal features of this crossover? These are some questions which can provide us a deeper understanding of the typical complexity of random K-SAT problems.

A similar problem to one that we are going to study is the Horn SAT problem [13] where all the clauses have only one negated variable. It is an easy problem and solved in a polynomial time. However, notice that as long as K is small fluctuations play an important role in our problem and this could give rise to significantly different behaviors for the problem. There is also another problem called $(2+p)$-*SAT* [3] which by tuning p goes from a 2-*SAT* to a 3-*SAT* problem. It is close to what we like to do in this paper but here we are able to study the easy-hard crossover for general K and it is a more general problem to this end.

In the following we first give the exact solution of 1-SAT problem by a statistical mechanics approach. We find the average number of unsatisfied clauses and the average number of solutions in the ground state of the system and explain the origin of their behaviors. Utilizing the cavity method and assuming the replica symmetry, we derive a relation for the critical point of K-SAT problems. It is found that in general $\alpha_c \propto p^{-(K-1)}$ as $p \to 0$. Next we obtain the free energy and the distribution of effective fields with the aid of replica method and in the replica symmetry approximation. Finally we resort to the numerical solution of survey propagation equations [15] for the case of $K=3$ and compare the extracted critical points with the predictions of replica symmetry assumption. It is found that for $p > p^* \sim 0.17$ the replica symmetry breaks at some point $\alpha_d < \alpha_c$ whereas for $p < p^*$ we are always in the easy-SAT phase if $\alpha < \alpha_c$.

The paper is organized as follows: In the next section we define the problem. Section III is devoted to the study of the 1-SAT problem. Assuming the presence of replica symmetry we give the results of cavity and replica methods in Secs. IV and V. Survey propagation equations for the case $K=3$ are numerically studied in Sec. VI. Section VII includes the conclusion remarks of the paper.

II. THE PROBLEM DEFINITION

We take N logical variables $\{x_i | i=1,\ldots,N\}$ where $x_i=1$ if the corresponding variable is true and otherwise $x_i=0$. Alternatively we can speak of N Ising variables $S_i := 2x_i - 1$. On the other side we have a formula which consists of M clauses

*Electronic address: ramezanpour@iasbs.ac.ir
†Electronic address: samanimi@sharif.edu

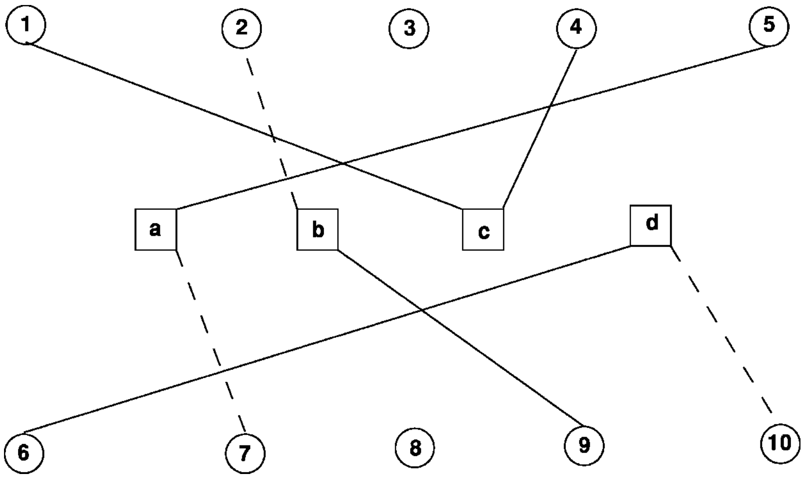

FIG. 1. Factor graph of the formula given in Eq. (1).

which have been joined to each other by logical AND. Each clause in turn contains K logical variable selected randomly from the list of our N variables. These variables, which join to each other by logical OR, are negated with probability p. One obtains the original random K-SAT problem by choosing $p=1/2$. Here there is an example of a 2-SAT formula with 4 clauses and 10 logical variables

$$F \coloneqq (x_5 \vee \overline{x_7}) \wedge (\overline{x_2} \vee x_9) \wedge (x_1 \vee x_4) \wedge (x_6 \vee \overline{x_{10}}). \quad (1)$$

A very useful concept in these problems is the factor graph, a bipartite graph of variable nodes and function nodes. In Fig. 1 we have shown the factor graph of formula given above. In this figure the logical variables and the clauses have been represented by the circles (variable nodes) and the squares (function nodes) respectively. An edge in this graph only connects a variable node to a function node and its style gives the nature of the logical variable in the associated clause. Here dashed edges are used to indicate that the negated variable enters the clause. We can summarize the factor graph in matrix $C_{M\times N}$ with elements $C_{a,i} \in \{0,+1,-1\}$. In fact $C_{a,i}$ is +1 or −1 if clause a contains variable i or its negated respectively. Otherwise $C_{a,i}=0$.

III. THE SIMPLE CASE OF K=1

We start by giving the exact behavior of 1-SAT problem by keeping a statistical mechanics approach [9,10]. We define the energy of a formula as the number of violated clauses, that is

$$E[S,C] \coloneqq \sum_{a=1}^{M} \left[1 - \sum_{i=1}^{N} C_{a,i} S_i \right] \Big/ 2, \quad (2)$$

where S denotes the configuration of Ising variables. Note that by definition

$$\sum_{a=1}^{M} C_{a,i} = t_i - f_i, \quad (3)$$

where t_i and f_i give the number of full lines and dashed lines, respectively, emanating from variable node i. The set of $\{t_i, f_i\}$ only depends on the structure of the factor graph. Utilizing the above facts and summing over spin configurations the partition function reads

$$Z[C] \coloneqq \sum_{S} e^{-\beta E[S,C]} = \prod_{i} (e^{-\beta t_i} + e^{-\beta f_i}), \quad (4)$$

where $\beta=1/k_B T$. The free energy per variable, $-(1/\beta N)\ln(Z[C])$, still depends on the structure of the factor graph and we should take an average over this kind of disorder.

The probability to have the set $\{t_i, f_i\}$ is given by

$$P[\{t,f\}] = (M!/N^M) \prod_{i} \left(\frac{p^{f_i}(1-p)^{t_i}}{(t_i! f_i!)} \right). \quad (5)$$

Then the averaged free energy in the thermodynamic limit reads

$$f \coloneqq \overline{f[C]} = \alpha/2 - (1/\beta)\left(\ln(2) + e^{-\alpha} \sum_{n,m=-\infty}^{\infty} J_m(q\alpha) I_n(\alpha) \times \ln\{\cosh[\beta(m+n)/2]\} \right), \quad (6)$$

where we have defined $q \coloneqq 1-2p$. Moreover $J_m(\alpha)$ and $I_n(\alpha)$ are the Bessel functions of first kind. Now from Eq. (6) one can easily find the average energy per variable

$$e \coloneqq \overline{\langle E \rangle}/N = \alpha/2 - e^{-\alpha} \sum_{m,n=-\infty}^{\infty} J_m(q\alpha) I_n(\alpha)(m+n)/2 \tanh[\beta(m+n)/2]. \quad (7)$$

In the same way the entropy per variable is given by

$$s \coloneqq \overline{\langle S \rangle}/N = \ln(2) + e^{-\alpha} \sum_{m,n=-\infty}^{\infty} J_m(q\alpha) I_n(\alpha)(\ln\{\cosh[\beta(m+n)/2]\} - [\beta(m+n)/2]\tanh[\beta(m+n)/2]). \quad (8)$$

We are interested in the ground state properties of the problem and to this end we need to take the limit $\beta \to \infty$ in the above relations. After some simplifications we find for the ground state energy

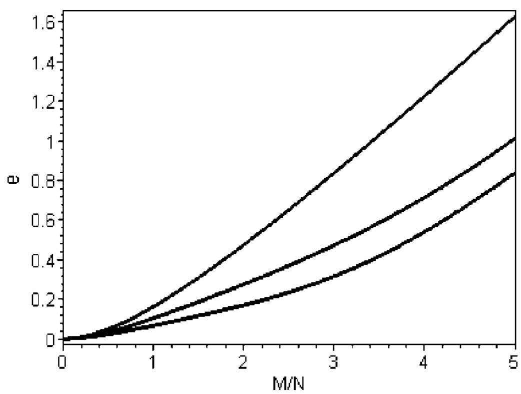

FIG. 2. Ground state energy of the 1-SAT problem from top to bottom for $p=1/2$, $p=1/4$, and $p=1/6$.

$$e_G = (1-p)\alpha - e^{-\alpha}\sum_{m+n>0}(m+n)J_m(q\alpha)I_n(\alpha). \tag{9}$$

In Fig. 2 we have shown the behavior of this quantity versus α and for some values of p. As expected the problem is always in the UNSAT phase where an infinite number of clauses are not satisfied in the thermodynamic limit. Indeed the probability that a new clause is not satisfied is given by $2p(1-p)\alpha$ and this quantity has always a nonzero value as long as p and α are nonzero.

The entropy of system at zero temperature is found from Eq. (8)

$$s_G = e^{-\alpha}I_0(\alpha\sqrt{1-q^2})\ln(2). \tag{10}$$

We have numerically computed this quantity and have shown its behavior with α in Fig. 3.

Clearly the number of solutions is the same for different values of p and small α. It is due to the fact that for small α each clause constrain a variable regardless of the nature of variable in that clause. However for large α the situation is different. In fact when a variable contributes in different forms in a number of clauses, with a larger probability it can take the values 0,1 without changing the number of violated clauses. This is the reason for the smaller entropy of smaller values of p in Fig. 3.

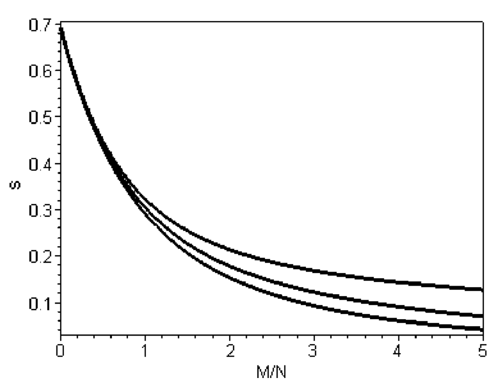

FIG. 3. Ground state entropy of the 1-SAT problem from top to bottom for $p=1/2$, $p=1/4$, and $p=1/6$.

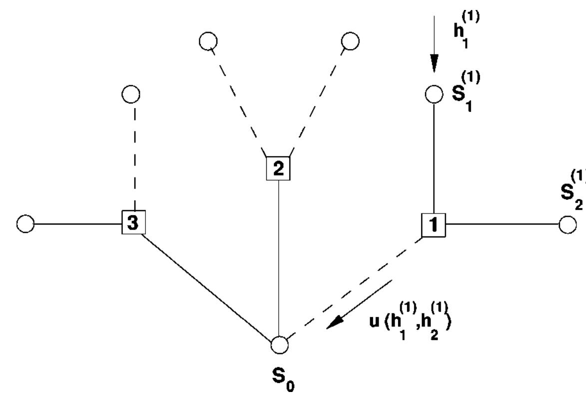

FIG. 4. S_0 experiences a cavity field determined from the fields experienced by the other neighbors of function nodes 1,2,3.

IV. CAVITY METHOD

In this section we apply the cavity method [11] to our problem. For simplicity we work only in the replica symmetric scheme and zero temperature.

Consider the variable node S_0 with $z=3$ neighbors in Fig. 4. It is known that the cavity field experienced by variable S_0 is

$$h_0 = \sum_{a=1}^{z} u(h_1^{(a)}, \ldots, h_{K-1}^{(a)}), \tag{11}$$

where

$$u(h_1, \ldots, h_{K-1}) = C_0\prod_{r=1}^{K-1}\theta(-C_r h_r). \tag{12}$$

Here $\theta(x)$ is the known step function which is 1 for $x>0$ and 0 otherwise. Note that u and h are stochastic quantities. In the replica symmetric framework the probability distribution of these quantities are given by [11]

$$Q(u) = c_0\delta(u) + c_-\delta(u+1) + c_+\delta(u-1),$$

$$P(h) = \sum_{l=-\infty}^{-1}P_-(h)\delta(l-h) + P_0\delta(h) + \sum_{l=1}^{\infty}P_+(h)\delta(l-h). \tag{13}$$

Then it is an easy exercise to show that for $h\geqslant 0$ the probability distribution of h is

$$P_+(h) = \sum_{z=h}^{\infty} f_{K\alpha}(z)\sum_{r=0}^{[(z-h)/2]}\binom{z}{2r+h}\binom{2r+h}{r}c_0^{z-2r-h}c_-^r c_+^{r+h}. \tag{14}$$

Here […] denotes the integer part of enclosed quantity and $f_{K\alpha}(z)$ is the degree distribution of variable nodes

$$f_{K\alpha}(z) = (z^{K\alpha}/z!)e^{-K\alpha}. \tag{15}$$

Similarly for $h\leqslant 0$, $P_-(-h)$ is obtained by interchanging the role of c_- and c_+ in the above equation. Simplifying the above relations one finds that

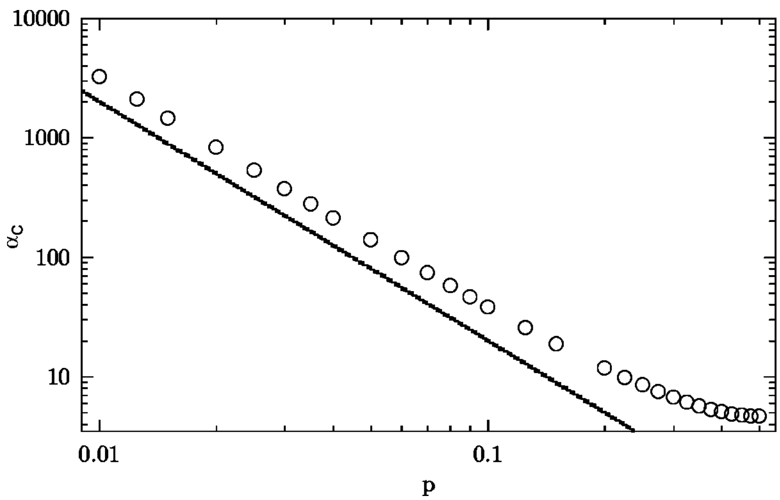

FIG. 5. α_c versus p for $K=3$ and in the replica symmetric approximation. The line shows a power law of exponent 2.

$$P_0 = e^{-K\alpha(1-c_0)} I_0(2K\alpha\sqrt{c_- c_+}), \tag{16}$$

and after some straightforward algebra for $P_+ = \Sigma_{h=1}^{\infty} P_+(h)$ one obtains

$$P_+ + P_0 = e^{-K\alpha c_+} \int_{K\alpha c_-}^{\infty} dt\, e^{-t} I_0(2\sqrt{K\alpha c_+ t}). \tag{17}$$

Equations (16) and (17) are two independent relations which along with the normalization condition $P_- + P_0 + P_+ = 1$ determine $P(h)$ and $Q(u)$. However we still need to derive the relations between $\{c_0, c_-, c_+\}$ and $\{P_0, P_-, P_+\}$. To this end note that from Eq. (12) one has

$$c_0 = 1 - [pP_+ + (1-p)P_-]^{K-1} \coloneqq 1 - c_p,$$
$$c_- = pc_p, \quad c_+ = (1-p)c_p. \tag{18}$$

Summing the above relations one can use Eqs. (16) and (17) to find the following equation for c_p:

$$c_p = [g(c_p)]^{K-1}, \tag{19}$$

where

$$g(c_p) \coloneqq 1 - p - (1-p)P_0 - (1-2p)P_+. \tag{20}$$

For $p=1/2$ we recover the known relation for the effective field distribution

$$P(h) = e^{-K\alpha c_p} I_h(K\alpha c_p), \tag{21}$$

where now c_p satisfies

$$c_p = \left[\frac{1 - e^{-K\alpha c_p} I_0(K\alpha c_p)}{2}\right]^{K-1}. \tag{22}$$

Returning to our general problem we find from Eq. (20) that $g(0)=0$ and

$$\left.\frac{dg(c_p)}{dc_p}\right|_{c_p=0} = 2K\alpha p(1-p). \tag{23}$$

Thus as expected Eq. (19) suggests a continuous transition for $K=2$ and discontinuous transitions for $K \geqslant 3$. Indeed for $K=2$ the critical value of α is given by

$$\alpha_c = \frac{1}{4p(1-p)}. \tag{24}$$

Due to the absence of replica symmetry breaking the above results are exact for $K=2$. As expected for $p=1/2$ we get the known value of $\alpha_c=1$ and $\alpha_c \to \infty$ for $p \to 0,1$.

What can be said about α_c for general K? Let us focus on the behavior of α_c as $p \to 0$. First note that just above the critical point P_+ takes a finite value and from the definition of c_p, Eq. (18), one obtains $c_p \propto p^{K-1}$. On the other hand in $g(c_p)$, c_p always appears along with α as αc_p. Expanding $g(c_p)$ one finds that only for a finite αc_p the two sides of Eq. (19) would have the same scaling with p. This in turn suggests that the critical value of α should scale as $p^{-(K-1)}$.

In Fig. 5 we have solved Eq. (19) numerically for $K=3$. Indeed for $K \geqslant 3$ the replica symmetric predictions provide an upper bound for α_c [16]. For $K=3$ we found that as expected α_c approaches to infinity like p^{-2} as $p \to 0$.

V. REPLICA APPROACH

In the following we will keep the same lines as Ref. [10] to calculate the free energy of biased random K-SAT problems in the replica formalism. As before we can write energy or the number of violated clauses as

$$E[S,C] = \sum_{l=1}^{M} \delta\left(\sum_{i=1}^{N} C_{l,i} S_i + K\right). \tag{25}$$

Our goal is to find $\overline{\ln(Z[C])}$ and to this end we need to obtain

$$Z_n \coloneqq \overline{Z[C]^n} = \sum_{S^1,\ldots,S^n} \overline{e^{-\beta \Sigma_{a=1}^{n} E[S^a,C]}}. \tag{26}$$

The overline denotes averaging with respect to the random structure of the factor graph. Due to the independent nature of clauses one can write

$$Z_n = \sum_{S^1,\ldots,S^n} \zeta^M, \tag{27}$$

where

$$\zeta = \overline{e^{-\beta\Sigma_a \delta(\Sigma_i C_{l,i} S_i^a + K)}}. \tag{28}$$

Then for a single clause we can use the fact that

$$\delta\left(\sum_i C_{l,i} S_i + K\right) = \prod_{j=1}^{K} \delta(S_{i_j}^a + C_j), \tag{29}$$

to write the following expression for ζ

$$\zeta = \frac{1}{\binom{N}{K}} \sum_{C_1,\ldots,C_K=\pm 1} p^{\nu[C]}(1-p)^{K-\nu[C]} \sum_{i_1,\ldots,i_K=1}^{N} e^{-\beta\Sigma_a \Pi_{j=1}^K \delta(S_{i_j}^a + C_j)} \times (1 + O(1/N) + \cdots), \tag{30}$$

where $\nu[C]$ is the number of minus elements in the set $\{C_j | j=1,\ldots,K\}$. Let us also define

$$x(\vec{\sigma}) \coloneqq (1/N)\sum_{i=1}^{N} \delta(\vec{S}_i - \vec{\sigma}), \tag{31}$$

where $\vec{\sigma}$ and $\vec{S}_i$ are vectors of n Ising elements in the replica space. Then apart from some irrelevant terms and constants we obtain

$$\zeta = \sum_{C_1,\ldots,C_K=\pm 1} p^{\nu[C]}(1-p)^{K-\nu[C]} \times \sum_{\vec{\sigma}_1,\ldots,\vec{\sigma}_K} x(-C_1\vec{\sigma}_1)\cdots x(-\vec{C}_K \sigma_K) e^{-\beta\Sigma_a \Pi_{j=1}^K \delta(\sigma_j^a - 1)} \tag{32}$$

and thus

$$Z_n \sim \sum_{x(\vec{\sigma})} e^{-N\tilde{f}(x)}, \quad \tilde{f}(x) \coloneqq -\alpha \ln[W(x)] + \sum_{\vec{\sigma}} x(\vec{\sigma}) \ln[x(\vec{\sigma})], \tag{33}$$

where $\alpha \coloneqq M/N$ and

$$W(x) \coloneqq \sum_{\nu|K} p^{\nu}(1-p)^{K-\nu} \sum_{\vec{\sigma}_1,\ldots,\vec{\sigma}_K} \prod_{j=1}^{\nu} x(\vec{\sigma}_j) \times \prod_{j=\nu+1}^{K} x(-\vec{\sigma}_j) e^{-\beta\Sigma_a \Pi_{j=1}^K \delta(\sigma_j^a - 1)}. \tag{34}$$

In this equation $\Sigma_{\nu|K}$ means a sum over all the selections of ν variables from K ones and in the same time it orders these selected variables in the beginning of a K-member list.

Now we should find a form for $x(\vec{\sigma})$ which minimizes $\tilde{f}(x)$. As in previous studies [10] we use the following ansatz in the replica symmetric scheme

$$x(\vec{\sigma}) = \int_{-1}^{1} dm P(m) \prod_{a=1}^{n} \left(\frac{1 + m\sigma^a}{2}\right). \tag{35}$$

Note that for $p \neq 1/2$ we do not have the symmetry relation $x(-\vec{\sigma}) = x(\vec{\sigma})$ and so $P(m)$ is not an even function.

Applying the above ansatz we find

$$W(x) = \sum_{\nu=0}^{K} B(\nu,K;p) \int \prod_{j=1}^{K} dm_j P(m_j) \prod_a A_{\nu,K-\nu}(m), \tag{36}$$

where $B(\nu,K;p)$ is the binomial distribution

$$B(\nu,K;p) = \binom{K}{\nu} p^{\nu}(1-p)^{K-\nu}, \tag{37}$$

and

$$A_{\nu,K-\nu}(m) \coloneqq \sum_{\sigma_1^a,\ldots,\sigma_K^a} \prod_{j=1}^{\nu} \left(\frac{1 + m_j\sigma_j^a}{2}\right) \prod_{j=\nu+1}^{K} \left(\frac{1 - m_j\sigma_j^a}{2}\right) e^{-\beta\Pi_{j=1}^K \delta(\sigma_j^a - 1)}. \tag{38}$$

Doing the sum over σ one obtains

$$A_{\nu,K-\nu}(m) = 1 + (e^{-\beta} - 1) \prod_{j=1}^{\nu} \left(\frac{1 + m_j}{2}\right) \prod_{j=\nu+1}^{K} \left(\frac{1 - m_j}{2}\right). \tag{39}$$

If we optimize $\tilde{f}(x)$ with respect to $x(\vec{\sigma})$ we find that

$$x(\vec{\sigma}) = \Lambda e^{\alpha W'(x)/W(x)}, \tag{40}$$

where

$$W'(x) \coloneqq \frac{\delta}{\delta x(\vec{\sigma})} W(x), \tag{41}$$

and Λ is determined from the normalization condition. After some calculations one finds the following relation for $W'(x)$

$$W'(x) = \sum_{\nu=0}^{K} B(\nu,K;p) \int \prod_{j=1}^{K} dm_j P(m_j) \left[\nu \prod_{a_+} A_{\nu-1,K-\nu}(m) + (K-\nu) \prod_{a_-} A_{\nu,K-\nu-1}(m) \right], \tag{42}$$

where $\Pi_{a_\pm}$ denotes a product over the indices a for them $\sigma^a = \pm 1$. Moreover in this equation

$$A_{\nu-1,K-\nu}(m) = 1 + (e^{-\beta} - 1) \prod_{j=1}^{\nu-1} \left(\frac{1 + m_j}{2}\right) \prod_{j=\nu}^{K-1} \left(\frac{1 - m_j}{2}\right),$$

$$A_{\nu,K-\nu-1}(m) = 1 + (e^{-\beta} - 1) \prod_{j=1}^{\nu} \left(\frac{1 + m_j}{2}\right) \prod_{j=\nu+1}^{K-1} \left(\frac{1 - m_j}{2}\right). \tag{43}$$

We are interested to limit $n \to 0$ where $x(\vec{\sigma})$ can be written as

$$x(\vec{\sigma}) = \int dm P(m) e^{u \ln((1+m)/(1-m))}. \tag{44}$$

Now doing the standard algebra [10] we find the following self consistency relation for $P(m)$

$$P(m) = \frac{2}{1-m^2}\int_{-\infty}^{\infty} du e^{-iu \ln((1+m)/(1-m))} e^{-\alpha K + \alpha W'(iu)}, \tag{45}$$

where

$$W'(u) = \sum_{\nu=0}^{K-1} B(\nu,K;p) \int \prod_{j=1}^{K-1} dm_j P(m_j)[\nu e^{u \ln(A_{\nu-1,K-\nu}(m))} + (K-\nu) e^{-u \ln(A_{\nu,K-\nu-1}(m))}]. \tag{46}$$

Similarly for the free energy we find

$$\beta f = -\ln(2) - \alpha(1-K)\sum_{\nu=0}^{K} B(\nu,K;p) \times \int \prod_{j=1}^{K} dm_j P(m_j) \ln(A_{\nu,K-\nu}) - (\alpha/2)\sum_{\nu=0}^{K} B(\nu,K;p) \int \prod_{j=1}^{K-1} dm_j P(m_j) \times [\nu \ln(A_{\nu-1,K-\nu}) + (K-\nu)\ln(A_{\nu,K-\nu-1})] + \tfrac{1}{2}\int dm P(m) \ln(1-m^2). \tag{47}$$

Equations (47) and (45) return the known relations for $p=1/2$ [10] when $P(m)$ is an even function. Finally let us consider the limit $\beta \to \infty$ of Eqs. (45) and (47). To this end we should work with effective fields z given by $m=\tanh(\beta z/2)$ [9]. Then for $\beta=\infty$ we get

$$R(z) = e^{-\alpha K}\int_{-\infty}^{\infty} du e^{-iuz} e^{\alpha W'(iu)}. \tag{48}$$

Now the relation that gives $W'(u)$ reads

$$W'(u) = K - K\left[p\int_0^{\infty} dz R(z) + (1-p)\int_{-\infty}^{0} dz R(z)\right]^{K-1} + \sum_{\nu=0}^{K-1} B(\nu,K;p)\left[\nu \int D_{\nu-1,K-\nu} e^{-u \min(1,z_1,\ldots,z_{\nu-1},-z_\nu,\ldots,-z_{K-1})} + (K-\nu)\int D_{\nu,K-\nu-1} e^{u \min(1,z_1,\ldots,z_\nu,-z_{\nu+1},\ldots,-z_{K-1})}\right], \tag{49}$$

where

$$D_{\nu-1,K-\nu} \coloneqq \int_0^{\infty} \prod_{j=1}^{\nu-1} dz_j R(z_j) \int_{-\infty}^{0} \prod_{j=\nu}^{K-1} dz_j R(z_j),$$

$$D_{\nu,K-\nu-1} \coloneqq \int_0^{\infty} \prod_{j=1}^{\nu} dz_j R(z_j) \int_{-\infty}^{0} \prod_{j=\nu+1}^{K-1} dz_j R(z_j). \tag{50}$$

For the free energy in this limit we have

$$f = \alpha(1-K)\sum_{\nu=0}^{K} B(\nu,K;p) \int D_{\nu,K-\nu} \min(1,z_1,\ldots,z_\nu,-z_{\nu+1},\ldots,-z_K) + (\alpha/2)\sum_{\nu=0}^{K} B(\nu,K;p) \times\left[\nu \int D_{\nu-1,K-\nu} \min(1,z_1,\ldots,z_{\nu-1},-z_\nu,\ldots,-z_{K-1}) + (K-\nu)\int D_{\nu,K-\nu-1} \min(1,z_1,\ldots,z_\nu,-z_{\nu+1},\ldots,-z_{K-1})\right] + \tfrac{1}{2}\left[\int_{-\infty}^{0} dz R(z) z - \int_0^{\infty} dz R(z) z\right]. \tag{51}$$

Considering the simple case of $K=1$ the effective field distribution reads

$$R(z) = e^{-\alpha}\sum_{m=-\infty}^{\infty}\left(I_0(q\alpha) I_m(\alpha) + \sum_{n=1}^{\infty}(-1)^n I_{2n}(q\alpha) \times [I_{2n-m}(\alpha) + I_{2n+m}(\alpha)]\right)\delta(z-m), \tag{52}$$

which for $p=1/2$ returns

$$R(z) = e^{-\alpha}\sum_{n=-\infty}^{\infty} I_n(\alpha)\delta(z-n). \tag{53}$$

Compare the above relation with Eq. (21) which gives the effective field distribution in the cavity method and in the replica symmetric approximation. In fact the two distributions are the same as they should be as long as we use an ansatz in which the effective fields take integer values.

VI. SURVEY PROPAGATION EQUATIONS

In this section we study the behavior of 3-SAT problem by means of numerical solution of survey propagation equations [14,15]. Let us first write the general form of these equations. We define $\eta_{a\to i}$ as the probability that in a state selected randomly from the existing states of the problem, the clause a sends a warning to variable i to take the value that satisfies it. This warning is sent if the other members of a do not satisfy this clause. We denote by $V(a)$ the set of neighbors of a. Then assuming a tree structure for the factor graph we have

$$\eta_{a\to i} = \prod_{j\in V(a)|i} P_a^u(j), \tag{54}$$

where the product is over all the neighbors of a excluding i and $P_a^u(j)$ is the probability that variable j does not satisfy clause a. Let us denote by $V_a^s(j)$ the set of clauses that variable j appears in them as it appears in clause a, Fig. 6. The remaining set of clauses are denoted by $V_a^u(j)$. With these definitions $P_a^u(j)$ is given by [15]

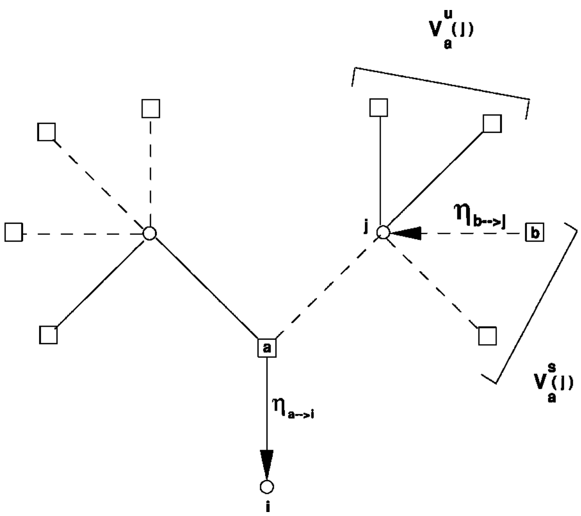

FIG. 6. The survey warning $\eta_{a\to i}$ is determined by the set of surveys $\eta_{b\to j}$.

$$P_a^u(j) = \frac{\Pi^u_{j\to a}}{\Pi^s_{j\to a} + \Pi^0_{j\to a} + \Pi^u_{j\to a}}, \tag{55}$$

where

$$\Pi^u_{j\to a} = \left[1 - \prod_{b\in V_a^u(j)} (1-\eta_{b\to j})\right] \prod_{b\in V_a^s(j)} (1-\eta_{b\to j}),$$

$$\Pi^s_{j\to a} = \left[1 - \prod_{b\in V_a^s(j)} (1-\eta_{b\to j})\right] \prod_{b\in V_a^u(j)} (1-\eta_{b\to j}),$$

$$\Pi^0_{j\to a} = \prod_{b\in V(j)|a} (1-\eta_{b\to j}). \tag{56}$$

Now starting from an arbitrary configuration for the warnings sent along the edges of the factor graph one obtains the new values of η's from Eqs. (54)–(56) and repeat this procedure until reaches to a stationary state. It is believed that in the whole region of SAT phase the above equations result in the correct solutions of random K-SAT problems [14]. Here we apply the same procedure to 3-SAT problem to compute Σ, the complexity of our problems. The complexity of a formula is the logarithm of the number of states and reads [15]

$$\Sigma = \sum_{a=1}^{M} \Sigma_a - \sum_{i=1}^{N} (z_i - 1)\Sigma_i, \tag{57}$$

where

$$\Sigma_a = \log\left[\prod_{j\in V(a)} (\Pi^s_{j\to a} + \Pi^0_{j\to a} + \Pi^u_{j\to a}) - \prod_{j\in V(a)} \Pi^u_{j\to a}\right],$$

$$\Sigma_i = \log[\Pi_i^- + \Pi_i^0 + \Pi_i^+], \tag{58}$$

and

$$\Pi_i^- = \left[1 - \prod_{a\in V_-(i)} (1-\eta_{a\to i})\right] \prod_{a\in V_+(i)} (1-\eta_{a\to i}),$$

$$\Pi_i^+ = \left[1 - \prod_{a\in V_+(i)} (1-\eta_{a\to i})\right] \prod_{a\in V_-(i)} (1-\eta_{a\to i}),$$

$$\Pi_i^0 = \prod_{a\in V(i)} (1-\eta_{a\to i}). \tag{59}$$

In these equations $V(i)$ denotes the set of z_i neighbors of variable node i, $V_+(i)$ is the set of function nodes in $V(i)$ that have been connected to i by a full line and $V_-(i)$ gives the complementary subset.

It is known that Σ is zero in the replica symmetric and UNSAT phases and nonzero in the hard-SAT phase [15]. Increasing α one first encounters the replica symmetry breaking point at α_d where Σ takes discontinuously its maximum value Σ_m. After this stage Σ decreases and finally vanishes at the critical point α_c. One can use these properties of Σ to compute α_d and α_c.

To solve the survey propagation equations we used the software given in [17]. In Fig. 7 we have shown the results of this computation for α_c and α_d and compared α_c with the predictions of replica symmetric case. As the figure shows the behavior of α_c with p is qualitatively similar to the one obtained with the replica symmetry assumption. The represented data have been restricted to relatively large values of

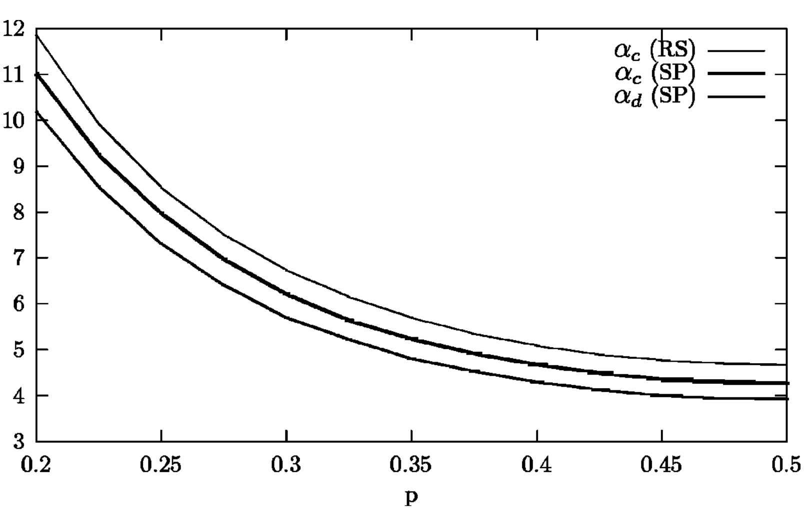

FIG. 7. From top to bottom: the replica symmetry predictions for α_c (RS), survey propagation predictions of α_c (SP) and α_d (SP) for $K=3$ and $N=10\,000$. The numerical results have been obtained for one realization with the convergence limit equal to 0.001.

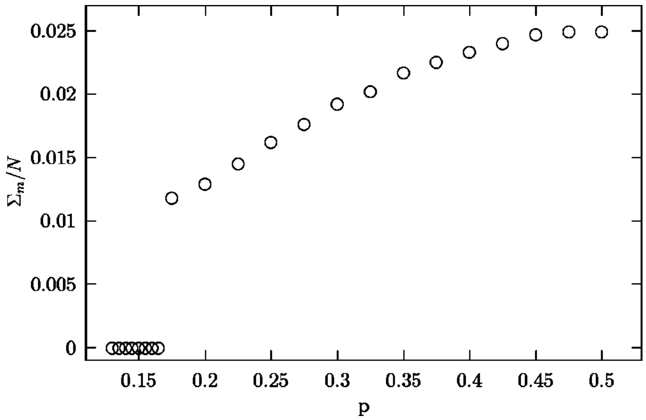

FIG. 8. Maximum complexity of the 3-SAT in terms of p. The parameters are the same as Fig. 7.

p. It is due to the fact that for smaller values of p the complexity vanishes and we are not able to identify α_c by looking at Σ.

In Fig. 8 we also showed the behavior of Σ_m versus p. It is seen that around $p^*=0.17$ the maximum complexity vanishes discontinuously. Then it can be concluded that for $p < p^*$ we have a simple problem as in the regime of easy-SAT phase.

VII. CONCLUSION

In summery we studied biased random K-SAT problems in which a variable is negated with probability p. This definition enables us to go continuously from easy random K-SAT problems to the hard ones. Certainly this can help us in a better understanding of the typical complexity of random K-SAT problems. In this paper we gave the exact solution of 1-SAT case and the full picture of general K-SAT problems in the replica symmetry approximation. From these results, which are exact for $K=2$, one can obtain an upper bound for the critical value of $\alpha_c(p,K)$. We found that $\alpha_c(p,K)$ has a power law behavior $p^{-\tau_K}$ for $p\to 0$ where $\tau_K=K-1$. We studied 3-SAT problems with the help of numerical solution of the survey propagation equations and found no replica symmetry breaking transition for $p<p^*\sim 0.17$. However in contrast to the tricritical point of $2+p$-SAT problem we found that in both sides of p^* the SAT-UNSAT transition is discontinuous. This phenomenon does not support the current belief that hardness of a problem may stem from the discontinuous nature of its transition. Certainly it still demands more studies to have a clear picture of the origins of typical complexity in these problems.

ACKNOWLEDGMENTS

We thank M. Mezard for helpful suggestions in writing this paper. A. R. is grateful to V. Karimipour for the introduction of this beautiful field to him.

[1] Y. Fu and P. W. Anderson, J. Phys. A **19**, 1605 (1986).
[2] S. Kirkpatrick and B. Selman, Science **264**, 1297 (1994).
[3] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, Nature (London) **400**, 133 (1999).
[4] M. Mezard, Science **301**, 1685 (2003).
[5] C. H. Papadimitriou, *Computational Complexity* (Addison-Wesley, Reading, 1994).
[6] M. Mezard, cond-mat/0212448.
[7] G. Parisi, cs.CC/0312011.
[8] O. C. Martin, R. Monasson, and R. Zecchina, Theor. Comput. Sci. **265**, 3 (2001).
[9] R. Monasson and R. Zecchina, Phys. Rev. Lett. **76**, 3881 (1996).
[10] R. Monasson and R. Zecchina, Phys. Rev. E **56**, 1357 (1997).
[11] M. Mezard and R. Zecchina, Phys. Rev. E **66**, 056126 (2002).
[12] S. Cook, in *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, Assoc. Comput. Mach., New York, 1971, p. 151.
[13] T. Walsh, APES Report No. APES-37-2002.
[14] A. Braunstein, M. Mezard, M. Weigt, and R. Zecchina, cond-mat/0212451.
[15] A. Braunstein, M. Mezard, and R. Zecchina, cs.CC/0212002.
[16] S. Franz and M. Leone, J. Stat. Phys. **111**, 535 (2003).
[17] http://www.ictp. trieste.it/zecchina/SP