On the Probabilistic Approach to the Random Satisfiability Problem

Giorgio Parisi

Dipartimento di Fisica, Sezione INFN, SMC and UdRm1 of INFM,
Università di Roma "La Sapienza",
Piazzale Aldo Moro 2, I-00185 Rome (Italy)
giorgio.parisi@roma1.infn.it,
http:chimera.roma1.infn.it

Abstract. In this note I will review some of the recent results that have been obtained in the probabilistic approach to the random satisfiability problem. At the present moment the results are only heuristic. In the case of the random 3-satisfiability problem a phase transition from the satisfiable to the unsatisfiable phase is found at $\alpha = 4.267$. There are other values of α that separates different regimes and they will be described in details. In this context the properties of the survey decimation algorithm will also be discussed.

1 Introduction

Recently many progresses [1,2] have been done on the analytic and numerical study of the random K-satisfiability problem [3,4,5,6], using the approach of survey-propagation that generalizes the more old approach based on the belief-propagation algorithm [1] [9,7,10,11] . Similar results have also been obtained for the coloring a of random graph [12].

In the random K-sat problem there are N variables $\sigma(i)$ that may be true of false (the index i will sometime called a node). An instance of the problem is given by a set of $M \equiv \alpha N$ clauses. In this note we will consider only the case $K = 3$. In this case each clause c is characterized by set of three nodes (i_1^c,i_2^c, i_3^c), that belong to the interval $1 - N$ and by three Boolean variables (b_1^c,b_2^c, b_3^c, i.e. the signatures in the clause). In the random case the i and b variables are random with flat probability distribution. Each clause c is true if the expression

$$E^c \equiv (\sigma(i_1^c) XOR\, b_1^c)\ OR\ (\sigma(i_2^c) XOR\, b_2^c)\ OR\ (\sigma(i_3^c) XOR\, b_3^c) \qquad (1)$$

is true [2].

[1] The belief propagation algorithm (sometimes called "Min-Sum") is the the zero temperature limit of the "Sum-Product" algorithm. In the statistical mechanics language [7] the belief propagation equations are the extension of the TAP equations for spin glasses [8,13] and the survey-propagation equations are the TAP equations generalized to the broken replica case.

[2] When all the b^c are false $E^c = \sigma(i_1^c)\ OR\ \sigma(i_2^c)\ OR\ \sigma(i_3^c)$ while when all the b^c are true $E^c = \overline{\sigma(i_1^c)}\ OR\ \overline{\sigma(i_2^c)}\ OR\ \overline{\sigma(i_3^c)}$.

E. Giunchiglia and A. Tacchella (Eds.): SAT 2003, LNCS 2919, pp. 203–213, 2004.
© Springer-Verlag Berlin Heidelberg 2004

The problem is satisfiable iff we can find a set of the variables σ such that all the clauses are true (i.e. a legal configuration); in other words we must find a truth value assignment. The entropy [11] of a satisfiable problem is the logarithm of the number of the different sets of the σ variables that make all the clauses true, i.e the number of legal configurations.

The goal of the analytic approach consists in finding for given α and for large values of N the probability that a random problem (i.e. a problem with random chosen clauses) is satisfiable. The $0-1$ law [4,6,16] is supposed to be valid: for $\alpha < \alpha_c$ all random systems (with probability one when N goes to infinity) are satisfiable and their entropy is proportional to N with a constant of proportionality that does not depend on the problem. On the other hand, for $\alpha > \alpha_c$ no random system (with probability one) is satisfiable. An heuristic argument[1,2] suggests that $\alpha_c = \alpha^* \approx 4.267$ where α^* can be computed using the survey-propagation equations defined later. There is already a proof [17] that the value of α^* computed with the techniques of survey-propagation is a rigorous upper bound to α_c (the proof has been obtained only for even K, the extension to odd K is technically difficult).

2 Results

Generally speaking we are interested to know not only the number of legal configurations, but also the properties of the set of all legal configurations. At this end it is convenient to say that two configurations are adjacent if their Hamming distance is less than ϵN, where ϵ is a small number.

We can argue that in the limit of large N:

1. In the interval $\alpha < \alpha_d \approx 3.86$ the set of all legal configurations is connected, i.e. there is a path of mutually adjacent configurations that joins two configurations of the set. In this region the belief-propagation equations (to be define later) have an unique solution.
2. In the interval $\alpha_d < \alpha < \alpha_c \approx 4.267$ the set of all the legal configurations breaks in an large number of different disconnected regions that are called with many different names in the physical literature [14,15] (states, valleys, clusters, lumps...). Roughly speaking the set of all the legal configurations can be naturally decomposed into clusters of proximate configurations, while configurations belonging to different clusters (or regions) are not close. This phenomenon is called in spontaneous replica symmetry breaking in the physical literature. The core of the approach of this note is the analysis of this phenomenon and of the methods used to tame its consequences [3]. The precise definition of these regions is rather complex [18]; roughly speaking we could say that two legal configurations belongs to the same region if they are

[3] Other models, where this phenomenon is not present, like random bipartite matching can be analyzed in a much simple way, although 15 years have been needed from the statements of the main result (i.e. the length of the shortest matching in the infinite N limit is $\zeta(2)$) to the rigorous proof of this fact.

in some sense adjacent, i.e. they belongs to a different region if their Hamming distance is greater than ϵN. In this way the precise definition of these regions depends on ϵ, however it can be argued that there is an interval in ϵ where the definition is non-trivial and is independent from the value of ϵ: for a rigorous definition of these regions see [20,21,19]. The number of these regions is given by $\exp(\Sigma^N(\alpha))$, where $\Sigma^N(\alpha)$ is the total complexity; for large N the total complexity is asymptotically given by $\Sigma^N(\alpha) = N\Sigma(\alpha)$ where $\Sigma(\alpha)$ is the complexity density. In this interval the belief-propagation equations have many solutions and each of these solution is associated to a different cluster. The statistical properties of the set of the solutions of the belief-propagation equations can be studied using the belief-propagation equations (to be defined later).

3. Only in the interval $\alpha_b \approx 3.92 < \alpha < \alpha_c$ there are literals σ that are frozen, i.e. they take the same value in all the legal configurations of a region [4]. We could say that the frozen variables form the backbone of a region. It is important to realize that a given clause may simultaneously belong to the backbone of one region and not belong to the backbone of an other region.

The arguments run as follow. Let us start with a given instance of the problem. We first write the belief propagation equations. For each clause that contains the node i (we will use the notation $c \in i$ although it may be not the most appropriate) $p_T(i,c)$ is defined to be the probability that the variable $\sigma(i)$ would be true in absence of the clause c when we average over the set of all the legal configuration ($p_F(i,c) = 1 - p_T(i,c)$ is the probability to be false). If the node i_1^c were contained in only one clause, we would have that

$$p_T(i_1^c) = u_T(p_T(i_2^c,c), p_T(i_3^c,c), b_1^c, b_2^c, b_3^c) \equiv u_T(i_1,c) \ ,$$
$$p_F(i_1^c) = 1 - u_T(i_1,c) \ , \tag{2}$$

where u_T is an appropriate function that is defined by the previous relation. An easy computation shows that when all the b are false, the variable $\sigma(i_1^c)$ must be true if both variable $\sigma(i_2^c)$ and $\sigma(i_3^c)$ are false, otherwise it can have any value. Therefore we have in this case that

$$u_T(i,c) = \frac{1}{2 - p_F(i_2^c,c)p_F(i_3^c,c)} \ . \tag{3}$$

In a similar way, if some of the b variable are true, we should exchange the indices T and F for the corresponding variables, i.e., if b_1^c is true, then $u_T(u_1)$ becomes $u_F(u_1)$. Finally we have that

[4] The distinction between α_d [5] a and α_b [1] is not usually done in the literature and sometimes it is wrongly assumed that $\alpha_b = \alpha_d$.

$$p_T(i,c) = \frac{\prod_{d\in i, d\neq c} u_T(i,d)}{Z_0(i,c)}$$
$$p_F(i,c) = \frac{\prod_{d\in i, d\neq c} u_F(i,d)}{Z_0(i,c)}$$
$$Z_0(i,c) = \prod_{d\in i, d\neq c} u_T(i,d) + \prod_{d\in i, d\neq c} u_F(i,d). \tag{4}$$

We note the previous formulae can be written in a more compact way if we introduce a two dimensional vector $\boldsymbol{p}$, with components p_T and p_F. We define the product of these vector

$$c_T = a_T\ b_T \qquad c_F = a_F\ b_F, \tag{5}$$

if $\boldsymbol{c} = \boldsymbol{a}\cdot\boldsymbol{b}$.

If the norm of a vector is defined by

$$|\boldsymbol{a}| = a_T + a_F\ , \tag{6}$$

the belief propagation equations are defined to be

$$\boldsymbol{p}(i,c) = \frac{\prod_{d\in i, d\neq c} \boldsymbol{u}(i,d)}{|\prod_{d\in i, d\neq c} \boldsymbol{u}(i,d)|}\ . \tag{7}$$

In total there are $3M$ variables $p_T(i,c)$ and $3M$ equations. These equations in the limit of large N should have an unique solution in the interval $\alpha < \alpha_d$ and the solution should give the correct values for the probabilities $p_T(i,c)$. In this interval the entropy (apart corrections that are subleading when N goes to infinity) is given by

$$S = -\sum_{i=1,N} \log(Z_1(i)) + 2\sum_{c=1,M} \log(Z_2(c))\ . \tag{8}$$

Here the first sum runs over the nodes and the second one runs over the clauses; $Z_2(c)$ is the probability that the clause c would be satisfied in a system where the validity of the clause c is not imposed. One finds that in the case where all the b variables are false

$$Z_2(c) = 1 - p_F(i_1^c)p_F(i_2^c)p_F(i_2^c)\ . \tag{9}$$

In a similar way $Z_1(i)$ is the probability that all we can find a legal configuration containing the site i starting from a configuration where the site i is not present and it is given by:

$$Z_1(i) = |\prod_{d\in i} \boldsymbol{u}(i,d)| \tag{10}$$

The belief propagation equations can also be written as :

$$\frac{\partial S}{\partial \boldsymbol{p}(i,c)} = 0\ . \tag{11}$$

The belief-propagation equations can be formally derived by a local analysis by assuming that, in the set of the legal configurations of the system where all the clauses $c \in i$ are removed, the variables $\sigma(k)$ that would enter in these clauses are not correlated. This cannot is not true for finite N, but this statement may be correct in the limit $N \to \infty$ with the appropriate qualifications.

Generally speaking for a given sample these equations may not have an exact solution, but they do have quasi-solutions [24] (i.e. approximate solutions, where the approximation becomes better and better when N goes to infinity [5]): these equations have been derived using a local analysis that is correct only in the limit $N \to \infty$.

In the interval $\alpha_d < \alpha < \alpha_b$ the variables p_F and p_T are different from 0 and 1; however in the region $\alpha_b < \alpha < \alpha_c$ there solutions (or quasi-solutions) of the belief equations have a fraction of the variables p_F and p_T that are equal to 0 or 1.

When the number of solutions of the belief propagation equations is large, the properties of the sets of solutions of the belief propagation equations can be obtained by computing the solution of the survey propagation equations defined as follows. In the general approach in each node we introduce the probability $(\mathcal{P}_{i,c}(p))$ to find a solution of the belief-propagation equations with $p_T(i,c) = p$. With some effort we can write down local equations for this probability. These are the full survey equations that allow the computation of the total entropy.

This approach is computationally heavy. As far as the computation of the complexity is concerned, we can use a simpler approach, where we keep only a small part of the information contained in the function $\mathcal{P}_{i,c}(p)$, i.e. the weight of the two delta function at $p = 0$ and $p = 1$. More precisely we introduce the quantity $s_T(i,c)$ that is defined as the probability of finding $p_T(i,c) = 1$, in the same way $s_F(i,c)$ is the probability of finding $p_T(i,c) = 0$ and $s_I(i,c)$ is the probability of finding $0 < p_T(i,c) < 1$. It is remarkable that it is possible to write closed equations also for these probabilities (these equations are usually called the survey propagation equations [1]).

We can use a more compact notation by introducing a three dimensional vector $\boldsymbol{s}$ given by

$$\boldsymbol{s} = \{s_T, s_I, s_F\} \ . \tag{12}$$

Everything works as before with the only difference that we have a three component vector instead of a two component vector. Generalizing the previous arguments one can introduce the quantity $\boldsymbol{u}(i,c)$ that is the value that the survey at i would take if only the clause c would be present in i (in other words $\boldsymbol{u}(i,c)$ is the message that arrives to the site i coming from the clause c). In the case where all the b are false, a simple computation gives

[5] More precisely if we have N equations $E_i[\sigma] = 0$ for $i = 1, N$ a solution σ of this system of equation satisfies the condition $N^1 sum_{i=1,N})(E_i[\sigma])^2 = 0$; a quasi-solution satisfies the weaker condition $N^1 sum_{i=1,N})(E_i[\sigma])^2 < h(N)$, where $h(N)$ is a function that goes to zero when N goes to infinity. The definition of a quasi-solution depends on the properties of the function $h(N)$ and this point must be further investigated: it may turn out at the end that quasi-solutions are not needed.

$$\boldsymbol{u}(i,c) = \{s_F(i_2^c,c)s_F(i_3^c,c),\ 1 - s_F(i_2^c,c)s_F(i_3^c,c),\ 0\}\ . \tag{13}$$

The formula can generalized as before [6] to the case of different values of b. One finally finds the survey propagation equations:

$$\boldsymbol{s}(i,c) = \frac{\prod_{d\in i, d\neq c}\boldsymbol{u}(i,d)}{|\prod_{d\in i, d\neq c}\boldsymbol{u}(i,d)|}\ , \tag{14}$$

where we have defined product in such a way that

$$\boldsymbol{ab} = \{a_T b_T + a_I b_T + a_T b_I, a_I b_I,\ a_F\ b_F + a_I\ b_F + a_F\ b_I\}. \tag{15}$$

It is convenient to introduce the reduced complexity ($\Sigma_R(\alpha)$), that counts the number of solutions of the belief equations where two different solutions are considered equals if they coincide in the points where the beliefs are 0 or 1 [7]. In other words two solutions of the beliefs equations with an identical backbone enters only once in the counting that leads to the reduced complexity.

If there is an unique solution to the survey propagation equations, it is possible to argue that the reduced total complexity should be given by

$$\Sigma_R = -\sum_{i=1,N}\ln(Z_1(i)) + 2\sum_{c=1,M}\ln(Z_2(c)) \tag{16}$$

where now the definition of the Z's is changed and it is done using the surveys, not the beliefs:

$$Z_1(i) = \ln(|\prod_{d\in i}\boldsymbol{u}(i,d)|), \quad Z_2(c) = \ln(|\boldsymbol{s}(i,c)\boldsymbol{u}(i,c)|) \tag{17}$$

The reduced complexity $\Sigma_R(\alpha)$ it is particularly interesting because it hat been conjecture that it should vanishes at the critical point α_c. This allow the computation of the point α_c.

It is interesting that also in this case the survey propagation equations can be written in a simple form:

$$\frac{\partial\Sigma_R}{\partial\boldsymbol{s}(i,c)} = 0\ . \tag{18}$$

One finally finds that the survey-propagation equations do not have an unique solution when $\alpha > \alpha_U \approx 4.36$. The fact is not of direct importance because $\alpha_U > \alpha_c$. Indeed in the region $\alpha > \alpha_c$ the complexity is negative so that the there are no solutions of the belief-propagation equations associated to the solution of the survey-propagation equation.

[6] It always happens that the vector $\boldsymbol{u}$ has only one zero component ($u_T u_F = 0$). This fact may be used to further simplify the analysis.

[7] It is not clear at the present moment if there are different solutions of the belief equations that coincide in all the points where the probability is 0 or 1: in other words we would like to know if $\Sigma_R(\alpha) = \Sigma(\alpha)$, where $\Sigma(\alpha)$ counts the total number of solutions of the belief equations.

It is evident that for $\alpha > \alpha_c$ there are no more legal configurations and $\Sigma_R(\alpha)$ is not well defined. A negative value $\Sigma_R(\alpha)$ can be interpreted by saying that the probability of finding a legal configuration goes to zero exponentially with N. We stress that the entropy density remains finite at α_c, the conjecture that $\Sigma_R(\alpha)$ vanishes at α_c implies that the reduced complexity is captures the number of essentially different regions of legal configuration. A priori a finite value $\Sigma_R(\alpha_c)$ cannot be excluded.

3 Methods

We now show how to obtain the above results on the solutions of the belief propagation equations (and of the survey propagation equations) for a large random system in the limit of large N. These equations are interesting especially in the infinite N limit where the factor graph does not contain short loop. For finite N in the random case, (or in the infinite N limit for a non-random case) the belief equations may have solutions, but the properties of these solutions do not represent exactly the properties of the systems. If the number of short loops is small, perturbative techniques may be used to compute the corrections to the belief equations. If short loops are very common (e.g. if the literals are on the sites of an f.c.c. lattice and the clauses are on faces of the same lattice), it is rather likely that the beliefs equations are useless and they could only used as starting point of much more sophisticated approaches.

We attack this problem by studying the solution of the belief propagation equations (and of the survey propagation equations) on an random infinite tree. Sometimes the solution is unique, i.e. it does not depends on the boundary conditions at infinity, and sometimes it is not not unique. The statistical properties of the solution of these equations can be studied with probabilistic method. One arrives to integral equations that can be solved numerically using the method of population dynamics [1,2,14,15]. The numerical solutions of these integral equations can be used to compute α_d, α_b, α_c, and α_U.

The generalization of the Aldous construction [22,23] of an infinite rooted tree associated to a graph can play the role of a bridge between a finite instance of the problems and the infinite random tree where analytic computations [1,2,14,15] are done. For example it could be used to prove the existence and the uniqueness of the beliefs and survey propagation equation in the appropriate intervals.

In this way one can argue that the properties on an infinite random tree are relevant for the behaviour of a given random system in the limit of large N.

We can check that the results we have obtained in these way for the solution of the belief and survey propagation equations are correct by computing in an explicit way the solution (when it is unique) of the equations for a given sample for large N (e.g $N = 10^4 - 10^6$). For example we may compare the distribution of the beliefs or of the surveys in a large system with the one obtained by solving the integral equations for the probabilities: the agreement is usually very good.

In the same spirit the validity of the result for α_d may be checked by studying the convergence of the iterative procedure for finding the solution of the

belief-propagation equations on a given large problem One finds that just at α_d the iterative procedure for finding a solution does not converge anymore and this is likely a sign of the existence of many solutions to the belief-propagation equations. In a similar way we can check the correctness of α_U.

4 Survey Decimation Algorithm

The survey decimation algorithm has been proposed [1,2,25,26,27] for finding the solution of the random K-satisfiability problem [4,5,6].

We start by solving the survey propagation equation. If a survey $(\boldsymbol{s}(i))$ is very near to $(1,0,0)$ (or to $(0,0,1)$) in most of the legal solutions of the beliefs equations (and consequently in the legal configurations) the corresponding local variables will be true (or false).

The main step in the decimation procedure consists is starting from a problem with N variables and to consider a problem with $N-1$ variables where $\boldsymbol{s}(i)$ is fixed to be true (or false). We denote

$$\Delta(i) = \Sigma^N - \Sigma^{N-1} \ . \tag{19}$$

If $\Delta(i)$ is small, the second problem it is easier to solve: it has nearly the same number of solutions of the belief equations and one variable less. (We assume that the complexity can be computed by solving the survey propagation equations).

The decimation algorithm proceeds as follows. We reduces by one the number of variables choosing the node i in the appropriate way, e.g. by choosing the node with minimal $\Delta(i)$. We recompute the solutions of the survey equations and we reduce again the number of variables. At the end of the day two things may happen:

1. We arrive to a negative complexity (in this case the reduced problem should have no solutions and we are lost),
2. The denominator in equation (14) becomes zero, signaling the presence of a contradiction (and also in this case we are lost),
3. The non-trivial solution of the survey equation disappears. If this happens the reduced problem is now easy to be solved.

The quantity $\Delta(i)$ may be estimated analytically so that it is possible to choose the variable with minimal $\Delta(i)$. A careful analysis of the results for large, but finite N [27] shows that the algorithm works in the limit of infinite N up to $\alpha_A \approx 4.252$, that is definite less, but very near to α_c.

Unfortunately at the present moment this result for α_A can be obtained only analyzing how the argument works on a finite sample and we are unable to write down integral equations for the probability distributions of the solution of the survey propagation equations. This drawback leads to the impossibility of computing analytically α_A: it is a rather difficult task to understand in details why for α_A it so near to α_c. It is interesting to note that for $\alpha < \alpha_A$ the survey decimation algorithm takes a time that is polynomial in N and using a smart

implementation the time is nearly linear in N. It is important to stress that survey algorithm is an incomplete search procedure which may not be able to find any solution to a satisfiable instance. This actually happens with a non-negligible probability e.g. for sizes of the order of a few thousands of variables also when $\alpha < \alpha_A$, however in the limit $N \to \infty$, it should work as soon $\alpha < \alpha_A$.

In the interval $\alpha_A < \alpha < \alpha_c$ the decimation procedure leads to a regime of negative complexity so that the algorithm does not work. Unfortunately there is no analytic computation of α_A. It is likely that the fact that $\alpha_U = 4.36$ is not far from α_c is related to the fact the $\alpha_c - \alpha_A$ iss small.

It would be very interesting to understand better if such a relation is true and to put in a quantitative form. In this regard it would be be important to study the K dependence of $\alpha_c - \alpha_A$ and $\alpha_U - \alpha_c$. This analysis may give some hints why $\alpha_c - \alpha_A$ is so small in 3-SAT.

5 Conclusions

Similar problems have been studied by physicists in the case of infinite range spin glasses [7]: here the problem consists in finding the minimum E_J of the quantity:

$$H_J[\tau] \equiv \sum_{i,k=1,N} J_{i,k}\tau_i\tau_k \tag{20}$$

where the minimum is done respect to the variables $\tau_i = \pm 1$ and the J are independent Gaussian variables with zero average and variance $N^{-1/2}$. Physical intuition tells us that in the limit N goes to infinity the intensive quantity

$$e_J = \frac{E_J}{N} \tag{21}$$

should be (with probability 1) independent from N and it will be denoted by e_∞. In 1979 it was argued using the so called replica method (that will be not discussed in this note) that e_∞ was equal to the maximum of a certain functional $F[q]$, where $q(x)$ is a function defined in the interval $[0-1]$. Later on, 1985 the same results were rederived using heuristical probabilistic consideration, similar to those presented here (but much more complex). In this note we have introduced a hierarchical construction, where three levels (configurations, beliefs, surveys) are presents: in the case of spin glasses an infinite number of levels is needed (in the spin glass case the survey equations do not have an unique solution and we have to consider higher and higher levels of abstraction). Only very recently Talagrand [28], heavily using Guerra's ideas and results, was able to prove that the value for e_∞, computed 24 year before, was correct.

The possibility of using these techniques for deriving eventually exact results on the K-SAT problem is a very recent one: only a few year ago [14,15] the previous techniques has been extended to more complex spin glass models where the matrix J is sparse (it has an average number of elements per raw that does not increase with N).

The field is very young and rigorous proofs of many steps of the construction (also those that would likely be relatively simple) are lacking. We only know, as a consequence of general theorems, that this methods give an upper bound to the value of α_c, and this upper bound should be computed by maximizing an appropriate functional of the probability distribution of the surveys. This upper bound is rigorous one [17]: it essentially use positivity arguments (the average of a non-negative function is non-negative) in a very smart way and it does not depend on the existence or uniqueness of the solutions of the equations for the probability distribution of the survey. On the contrary the way we have followed to compute this upper bound (i.e. α^*) require some extra work before becoming fully rigorous. I stress that this upper bounds and the Talagrand's exact result do not need in any way considerations on the solutions of the survey propagation equations (or of their generalization) on a finite sample. The survey propagation equations are crucial for giving an intuitive image of the situation (i.e. at a metaphoric level) and for constructing the survey decimation algorithm. The heuristic derivation could have been done using the replica method, where survey propagation equations are never mentioned, but the argument is much more difficult to follow and to transform in a rigorous one.

The other results come from empirical (sometimes very strong) numerical evidence and from heuristic arguments. For example at my knowledge there is no proof that the integral equations for the probability of the surveys (or of the beliefs) have an unique solution and that the population dynamics algorithm converges (to that unique solution). Proofs in this direction would be very useful and are a necessary step to arrive to a rigorous quantitative upper bounds and eventually exact results. On the other hand the proof of existence of an unique solutions (or quasi-solutions) of the surveys (or beliefs) propagation equations in the large N limit is lacking for any value of α, although the analysis with the population dynamics (whose precise mathematical properties have to be clarified) tell us which should the maximum values (α_U and α_b respectively, below which these uniqueness properties hold. Many steps have to be done, but a rigorous determination of α_c seems to be a feasible task in a not too far future.

References

1. M. Mézard, G. Parisi and R. Zecchina, Science **297**, 812 (2002).
2. M. Mézard and R. Zecchina *The random K-satisfiability problem: from an analytic solution to an efficient algorithm* cond-mat 0207194.
3. S.A. Cook, D.G. Mitchell, *Finding Hard Instances of the Satisfiability Problem: A Survey*, In: *Satisfiability Problem: Theory and Applications*. Du, Gu and Pardalos (Eds). DIMACS Series in *Discrete Mathematics and Theoretical Computer Science*, Volume 35, (1997)
4. S. Kirkpatrick, B. Selman, *Critical Behaviour in the satisfiability of random Boolean expressions*, Science 264, 1297 (1994)
5. Biroli, G., Monasson, R. and Weigt, M. *A Variational description of the ground state structure in random satisfiability problems, Euro. Phys. J.* **B 14** 551 (2000),

6. Dubois O. Monasson R., Selman B. and Zecchina R. (Eds.), *Phase Transitions in Combinatorial Problems*, Theoret. Comp. Sci. 265, (2001), G. Biroli, S. Cocco, R. Monasson, Physica A 306, 381 (2002).
7. Mézard, M., Parisi, G. and Virasoro, M.A. *Spin Glass Theory and Beyond*, World Scientific, Singapore, (1987).
8. D.J. Thouless, P.A. Anderson and R. G. Palmer, *Solution of a 'solvable' model,* Phil. Mag. 35, 593 (1977)
9. J.S. Yedidia, W.T. Freeman and Y. Weiss, *Generalized Belief Propagation*, in *Advances in Neural Information Processing Systems 13* eds. T.K. Leen, T.G. Dietterich, and V. Tresp, MIT Press 2001, pp. 689-695.
10. F.R. Kschischang, B.J. Frey, H.-A. Loeliger, *Factor Graphs and the Sum-Product Algorithm*, *IEEE Trans. Infor. Theory* **47**, 498 (2002).
11. Monasson, R. and Zecchina, R. *Entropy of the K-satisfiability problem*, *Phys. Rev. Lett.* **76** 3881–3885 (1996).
12. R. Mulet, A. Pagnani, M. Weigt, R. Zecchina *Phys. Rev. Lett.* **89**, 268701 (2002).
13. C. De Dominicis and Y. Y. Goldschmidt: *Replica symmetry breaking in finite connectivity systems: a large connectivity expansion at finite and zero temperature*, J. Phys. A (Math. Gen.) **22**, L775 (1989).
14. M. Mézard and G. Parisi: Eur.Phys. J. B **20** (2001) 217;
15. M. Mézard and G. Parisi: *'The cavity method at zero temperature'*, cond-mat/0207121 (2002) to appear in J. Stat. Phys.
16. O. Dubois, Y. Boufkhad, J. Mandler, *Typical random 3-SAT formulae and the satisfiability threshold*, in *Proc. 11th ACM-SIAM Symp. on Discrete Algorithms*, 124 (San Francisco, CA, 2000).
17. S. Franz and M. Leone, *Replica bounds for optimization problems and diluted spin systems*, cond-mat/0208280.
18. G. Parisi, *Glasses, replicas and all that* cond-mat/0301157 (2003).
19. S. Cocco, O. Dubois, J. Mandler, R. Monasson. Phys. Rev. Lett. **90**, 047205 (2003).
20. M. Talagrand, *Rigorous low temperature results for the p-spin mean field spin glass model*, *Prob. Theory and Related Fields* **117**, 303–360 (2000).
21. Dubois and Mandler, FOCS 2002, 769.
22. D. Aldous, *The zeta(2) Limit in the Random Assignment Problem,* Random Structures and Algorithms **18** (2001) 381-418.
23. G. Parisi: cs.CC/0212047 *On local equilibrium equations for clustering states* (2002).
24. G. Parisi: cs.CC/0212009 *On the survey-propagation equations for the random K-satisfiability problem* (2002).
25. A. Braustein, M. Mezard, M. Weigt, R. Zecchina: cond-mat/0212451 *Constraint Satisfaction by Survey Propagation* (2002).
26. A. Braunstein, M. Mezard, R. Zecchina; cs.CC/0212002 *Survey propagation: an algorithm for satisfiability* (2002).
27. G. Parisi: cs.CC/0301015 *Some remarks on the survey decimation algorithm for K-satisfiability* (2003).
28. M. Talagrand, private communication (2003).