Can rare SAT formulae be easily recognized? On the efficiency of message-passing algorithms for *K*-SAT at large clause-to-variable ratios

Institute of Physics Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **40** (2007) 867–886 doi:10.1088/1751-8113/40/5/001

Can rare SAT formulae be easily recognized? On the efficiency of message-passing algorithms for *K*-SAT at large clause-to-variable ratios

Fabrizio Altarelli[1], **Rémi Monasson**[2] **and Francesco Zamponi**[2]

[1] Dipartimento di Fisica and INFM-SMC, Università di Roma 'La Sapienza', P.le A. Moro 2, 00185 Roma, Italy

[2] Laboratoire de Physique Théorique de l'École Normale Supérieure, 24 Rue Lhomond, 75231 Paris Cedex 05, France

Received 11 September 2006, in final form 21 November 2006
Published 17 January 2007
Online at stacks.iop.org/JPhysA/40/867

Abstract

For large clause-to-variable ratios, typical K-SAT instances drawn from the uniform distribution have no solution. We argue, based on statistical mechanics calculations using the replica and cavity methods, that rare satisfiable instances from the uniform distribution are very similar to typical instances drawn from the so-called planted distribution, where instances are chosen uniformly between the ones that admit a given solution. It then follows, from a recent article by Feige, Mossel and Vilenchik (2006 Complete convergence of message passing algorithms for some satisfiability problems *Proc. Random 2006* pp 339–50), that these rare instances can be easily recognized (in $O(\log N)$ time and with probability close to 1) by a simple message-passing algorithm.

PACS numbers: 89.20.Ff, 05.70.Fh, 75.10.Nr

1. Introduction

A lot of efforts have recently been devoted to the investigation of the computational complexity of hard computational problems under input model distributions. One popular case is the K-satisfiability (K-SAT) problem with uniform distribution where clauses are picked up uniformly at random from the set of K-clauses over N Boolean variables [1]. It is widely believed that there exists a phase transition when the number of clauses, M, and of variables, N, go to infinity at fixed ratio α. Instances with ratio α smaller than some critical value $\alpha_c(K)$ typically admit a solution, while instances with ratio $\alpha > \alpha_c(K)$ are almost surely not satisfiable. Rigorous studies combined with statistical physics methods have produced bounds and estimates to the value of $\alpha_c(K)$ and conjectured the existence of a rich structure of the space of solutions in the satisfiable phase [2–8].

If those results are certainly interesting from the random graph theory point of view, their relevance to computer science is a matter of debate. The major concern is that they are highly specific to one particular distribution of instances, with no obvious theoretical generality or usefulness for practical applications where instances are highly structured or extracted from unknown distributions. Recently, however, a strong motivation to the study of random K-SAT in computer science was pointed out by Feige [9]. Under the assumption that 3-SAT is hard on average under uniform distribution, Feige proved some worst-case hardness of approximation results for many different problems such as min bisection, dense K-subgraph, max bipartite clique, etc. The average-case hardness hypothesis can be informally stated as: *there is no fast algorithm capable of recognizing every satisfiable instance and most unsatisfiable instances for arbitrarily large (but bounded when $N \to \infty$) ratio α.*

In the present work, we point out that a similar but stronger hypothesis, with *every* replaced by *with probability p*, is wrong whatever $p < 1$. For large α (well above α_c), typical instances of the uniform distribution have no solution. We argue, based on statistical mechanics calculations using the replica and cavity methods, that rare satisfiable instances from the uniform distribution are very similar to typical instances drawn from the so-called planted distribution, where instances are chosen uniformly between the ones that admit a given solution. Our result then follows from a recent article by Feige, Mossel and Vilenchik who showed that, for K-SAT with the planted distribution, a simple message-passing algorithm is able to find the solution with probability $1 - \mathrm{e}^{-O(\alpha)}$ in polynomial time [10].

2. Definitions

We consider random K-SAT instances (formulae) with N Boolean variables and $M = \alpha N$ clauses. A clause has the form $C_i(X) = y_{i_1} \vee y_{i_2} \vee \cdots \vee y_{i_K}$, where $y \in \{x, \bar{x}\}$ represents the variable or its negation; the K-SAT problem consists in finding an assignment X such that $\wedge_j C_j(X) = \text{TRUE}$. Sometimes we will specialize to the case $k = 3$ for simplicity. We will consider the following distributions over the formulae F [10]:

- The uniform distribution $\mathcal{P}_{\text{unif}}[F]$ over all possible formulae with N variables and $M = \alpha N$ clauses made of K literals (variables or their negations) corresponding to *different* variables.
- The distribution $\mathcal{P}_{\text{sat}}[F]$ obtained from the distribution above by conditioning to satisfiability. In other words, $\mathcal{P}_{\text{sat}}[F]$ gives *uniform* probability to all satisfiable formulae and zero to the others.
- The *planted* distribution $\mathcal{P}_{\text{plant}}[F]$ which is constructed as follows: first one extracts with uniform probability one configuration X of the N variables, and then extracts with uniform probability one formula among the ones that admit X as a solution. Non-uniform variants of $\mathcal{P}_{\text{plant}}$ were studied in [12].

The number of formulae that have a given solution X is independent of X for symmetry reasons, $\mathcal{N}_f[X] = \mathcal{N}_f = \left[\binom{N}{K}(2^K - 1)\right]^M$. Define $\chi[F; X] = 1$ when X is a solution to F and $\chi[F; X] = 0$ otherwise, and $\mathcal{N}_s[F]$ the number of solutions to F. We have

$$\mathcal{P}_{\text{plant}}[F] = \frac{1}{2^N} \sum_X \frac{\chi[F; X]}{\mathcal{N}_f[X]} = \frac{\mathcal{N}_s[F]}{2^N \mathcal{N}_f}; \tag{1}$$

thus $\mathcal{P}_{\text{plant}}$ is *not* uniform over the satisfiable instances, but is proportional to the number of solutions to a given formula.

In [9], the following hardness hypothesis was introduced for formulae drawn from the uniform distribution $\mathcal{P}_{\text{unif}}$:

Hypothesis 1. *Even if α is arbitrarily large (but independent of N), there is no polynomial time algorithm that on most 3-SAT formulae outputs UNSAT, and always outputs SAT on a 3-SAT formula that is satisfiable.*

And this hypothesis is used to derive hardness of approximation results for various computational problems. A stronger form of hypothesis 1 is obtained by replacing *never* with *with probability p* (with respect to the uniform distribution over the formulae and possibly to some randomness built in the algorithm):

Hypothesis 1_p. *Even if α is arbitrarily large (but independent of N), there is no polynomial time algorithm that on most 3-SAT formulae outputs UNSAT, and outputs SAT with probability p on a 3-SAT formula that is satisfiable.*

We want to present some arguments supporting the idea that 1_p is false for any $p < 1$. Indeed, in [10] it has been shown that if the formulae are drawn with probability $\mathcal{P}_{\text{plant}}$, then a solution is found in polynomial time with probability $1 - \mathrm{e}^{-O(\alpha)}$ by a message-passing algorithm, called warning propagation (WP). WP is a simplified version of the zero-temperature belief propagation procedure, see [10, 11] for a presentation. It is important to note that WP is a constructive algorithm: when it declares a formula to be satisfiable, it provides a solution. This means that it never outputs SAT on a formula which is unsatisfiable. On the other hand, if the algorithm has not found a solution after a given number of iterations (which depends on N, see below), we declare the output to be UNSAT.

It is natural (and was already suggested in [10]) to try to extend this result to formulae drawn from the distribution $\mathcal{P}_{\text{sat}}$. The main ingredients that are needed in the proof of [10] are the following:

(i) at large α, formulae drawn from $\mathcal{P}_{\text{plant}}[F]$ typically have a single *cluster* of solutions with a large *core*: namely, there is a set $\mathcal{H}$ (the core) containing a fraction $1 - \mathrm{e}^{-O(\alpha)}$ of variables that have the same value in all the solutions of a given formula drawn from $\mathcal{P}_{\text{plant}}[F]$;

(ii) the *cavity fields* (or *variable-to-clause messages*) corresponding to the core variables, defined roughly as the number of clauses that are violated if one takes a solution to F and changes the value of a given core variable, are $O(\alpha)$;

(iii) the cavity fields for the core variables are $O(\alpha)$ even if they are computed with respect to a random configuration (see [10] for a precise definition); this is a consequence of the fact that if a variable x_i has value 1 in the solutions to F, then the probability of this variable appearing as x_i in a clause (according to $\mathcal{P}_{\text{plant}}$) is bigger than the probability of it appearing as $\bar{x}_i$ (and vice versa if the variable is 0 in the solutions).

We claim that formulae drawn from $\mathcal{P}_{\text{sat}}$ are very similar to the ones drawn from $\mathcal{P}_{\text{plant}}$, and in particular properties 1–3 hold for them. Moreover, we will show that the relative entropy (Kullback–Leibler divergence) of $\mathcal{P}_{\text{plant}}$ with respect to $\mathcal{P}_{\text{sat}}$ is $O(Ne^{-\alpha})$. In particular, property 1 implies that properties of the formulae drawn from $\mathcal{P}_{\text{sat}}$ (such as the distribution of the cavity fields) can be computed in a *replica symmetric* framework. We will indeed show that this is the case as the replica symmetric solution is stable for large α if one restricts to SAT formulae. In this way, we will compute the distribution of the cavity fields in a solution to show that: (i) only a fraction $\mathrm{e}^{-O(\alpha)}$ of the fields are zero (corresponding to non-core variables); (ii) the non-vanishing cavity fields are typically $O(\alpha)$; (iii) if the field corresponding to a variable x is (say) positive, then the number of clauses where the literal x appears is bigger than the number of clauses where $\bar{x}$ appears, the difference being $O(\alpha)$.

The validity of properties 1–3 together with the fact that the relative entropy of $\mathcal{P}_{\text{sat}}$ and $\mathcal{P}_{\text{plant}}$ is small strongly suggests that the analysis of [10] can be extended to $\mathcal{P}_{\text{sat}}$. Then WP will be efficient in finding solutions for satisfiable formulae in polynomial time, with probability close to 1 for large α, thus contradicting hypothesis 1_p (but *not* hypothesis 1).

3. Statistical physics analysis of $\mathcal{P}_{\text{sat}}$

3.1. From $\mathcal{P}_{\text{unif}}$ to $\mathcal{P}_{\text{sat}}$: the replica calculation

We want to compute properties of the satisfiable formulae drawn from the uniform distribution $\mathcal{P}_{\text{sat}}[F]$ using the replica method. Following [5], we introduce a cost function

$$E[X] = \sum_{i=1}^{M} \delta[C_i(X); \text{FALSE}], \tag{2}$$

where the function $\delta[C_i(X); \text{FALSE}]$ is 1 if clause C_i is false in the assignment X and 0 otherwise (i.e. E counts the number of violated clauses). The replicated and disorder-averaged (i.e. averaged over the distribution $\mathcal{P}_{\text{unif}}[F]$ of the formulae) partition function is

$$\overline{Z(\beta)^n} = \overline{\left[\sum_X \mathrm{e}^{-\beta E[X]}\right]^n} = \overline{[g_0\,\mathrm{e}^{-\beta E_0} + g_1\,\mathrm{e}^{-\beta E_1} + \cdots]^n} \tag{3}$$

where E_0 is the energy of the ground state (i.e. the minimal number of unsatisfied clauses in F) and g_0 its degeneracy. In the limit $\beta \to \infty$, $n \to 0$ with fixed product $\nu = n\beta$, defining

$$P(E_0 = Ne_0) = \mathrm{e}^{N\omega(e_0)+o(N)} \tag{4}$$

the distribution of the ground-state energy with respect to $\mathcal{P}_{\text{unif}}[F]$, we have

$$\overline{Z(\beta)^n} \sim \overline{g_0^n\,\mathrm{e}^{-n\beta E_0}} = \int \mathrm{d}e_0\,\mathrm{e}^{N[\omega(e_0)-\nu e_0]} + O(\mathrm{e}^{-\beta}) = \mathrm{e}^{N\mathcal{F}(\nu)} \tag{5}$$

since g_0 is independent of n and therefore disappears for $n \to 0$. The function $\mathcal{F}(\nu)$ is defined by the saddle-point condition $\mathcal{F}(\nu) = \max_{e_0}[\omega(e_0) - \nu e_0]$. We will verify later that $\mathcal{F}(\nu)$ is convex (for sufficiently large α, ν), so that $\mathcal{F}(\nu)$ and $\omega(e_0)$ are the Legendre transforms of each other. The dominant contribution to the integral in (5) comes from formulae with ground-state energy density e_0 given by the equation $e_0(\nu) = -\partial_\nu \mathcal{F}(\nu)$. As we will see from the calculation of $\mathcal{F}(\nu)$, for large ν we have $e_0(\nu) \sim \mathrm{e}^{-\nu}$, so that to have $e_0(\nu) = 0$ we have to take the limit $\nu \to \infty$. By imposing this limit we implement the constraint $e_0 = 0$ and obtain information on $\mathcal{P}_{\text{sat}}[F]$. Note that we cannot implement the exact constraint of satisfiability, $E_0 = 0$, but only $\lim_{N\to\infty} E_0/N = 0$, as usual in most statistical mechanics computations. All our results are then affected by corrections vanishing only for $N \to \infty$.

In a replica symmetric framework, the free energy $\mathcal{F}(\nu)$ is obtained by maximizing a functional $\mathcal{F}[R(z), \nu]$ over a functional order parameter $R(z)$ (appendix A). This order parameter is the probability distribution of a random field z_i acting on each variable. The latter is the difference between the minimal number of violated clauses when the variable x_i is set to TRUE and the same quantity for x_i = FALSE. The distribution $R(z)$ is determined by the saddle-point equations $\delta\mathcal{F}[R(z), \nu]/\delta R(z) = 0$, which admit a solution in which the fields z are integer valued, as expected,

$$R(z) = \sum_{n=-\infty}^{+\infty} r_n \delta(z - n). \tag{6}$$

The coefficients r_n are obtained by substituting this expression into the saddle-point equations and solving them (appendix B). In the limit $\nu \to \infty$, the saddle-point equations give a self-consistency equation for $\rho_0 = \lim_{\nu\to\infty} r_0$:

$$\rho_0 = \frac{1}{2\,\mathrm{e}^{\mathcal{G}(\rho_0)} - 1}, \qquad \mathcal{G}(\rho_0) = \frac{\alpha K}{2} \frac{\left(\frac{1-\rho_0}{2}\right)^{K-1}}{1 - \left(\frac{1-\rho_0}{2}\right)^{K}}, \tag{7}$$

while all the other coefficients are given by a Poissonian distribution

$$\rho_n = \lim_{\nu\to\infty} r_n = \frac{\mathcal{G}(\rho_0)^{|n|}}{|n|!} \frac{1}{2\,\mathrm{e}^{\mathcal{G}(\rho_0)} - 1}. \tag{8}$$

Using the above expressions, it is possible to show that the ground-state energy vanishes exponentially for $\nu \to \infty$, $\overline{e_0}(\nu) = -\partial_\nu \mathcal{F}(\nu) = C\mathrm{e}^{-\nu}$ (appendix C).

We shall be interested in the value of $\omega(e_0 = 0)$ which is related to the probability of a formula extracted from $\mathcal{P}_{\text{unif}}[F]$ being satisfiable; setting from (C.12) $e_0 \sim \mathrm{e}^{-\nu}$ we obtain that $\omega(e_0(\nu)) = \mathcal{F}(\nu) + \nu e_0(\nu) \Rightarrow \omega(0) = \mathcal{F}(\infty) + O(\nu\mathrm{e}^{-\nu})$, and using the result for $\mathcal{F}(\nu)$ given in appendix D we get

$$\omega(0) = \mathcal{F}(\infty) = \log[2\mathrm{e}^{\mathcal{G}} - 1] - \frac{2\mathcal{G}\,\mathrm{e}^{\mathcal{G}}}{2\mathrm{e}^{\mathcal{G}} - 1} + \alpha \log\left[1 - \left(\frac{1-\rho_0}{2}\right)^{K}\right], \tag{9}$$

where $\mathcal{G} \equiv \mathcal{G}(\rho_0)$.

To conclude this section, let us discuss the stability of the replica symmetric solution for $\nu \to \infty$, i.e. for satisfiable formulae. The eigenvalues of the stability matrix around the saddle point are calculated in appendix E. We show that all the eigenvalues are negative for large enough α and $\nu \to \infty$. This implies that the replica symmetric solution is *locally* stable, but does not exclude the existence of a first-order transition to a different solution. In appendix F, we show that if one considers a function $R(z)$ different from (6) that allows also non-integer values of the fields, the weights of the non-integer fields vanish for $\alpha > \alpha_s$ (for some constant α_s which depends on K) and one gets back equation (6). This result rules out the existence of a first-order phase transition in the replica symmetric subspace. To exclude the possibility of a first-order transition with replica symmetry breaking one should perform the full 1RSB computation, that we leave for future work.

3.2. Interpretation of the self-consistency equations for the field distribution

Self-consistency equations (7) can be found back within the cavity method. Consider a formula over $N-1$ variables, and add a new variable x, by connecting it to the others $N-1$ variables through ℓ_+ clauses where x enters and ℓ_- clauses where $\bar{x}$ enters. We assume that ℓ_+, ℓ_- are independent Poissonian variables, with probabilities

$$p_L(\ell_\pm) = \frac{(\alpha' K/2)^{\ell_\pm}}{\ell_\pm!}\,\mathrm{e}^{-\alpha' K/2} \tag{10}$$

where α' is some constant to be determined later. The signs of the other $K-1$ variables in each clause are chosen uniformly at random. Then the probability that a new clause constrains the value of x (the clause sends a message to x in WP language [10, 11]) is equal to

$$q = \left(\frac{1-\rho_0}{2}\right)^{K-1}, \tag{11}$$

as $(1-\rho_0)/2$ is the probability that the field on an 'old' variable due to the existing formula is in contradiction with its sign in the new clause: e.g., if $C = x \vee x_1 \vee \cdots \vee x_{K-1}$ this is the probability that all the fields on $x_1, \ldots, x_{K-1}$ are negative, so that C sends a message ('be 1') to x. Let us call m_+, m_- the numbers of clauses containing, respectively, $x, \bar{x}$ and sending messages to the new variable. These are stochastic independent variables with probabilities

$$p_M(m_\pm) = \sum_{\ell=m_\pm}^{\infty} p_L(\ell) \binom{\ell}{m_\pm} q^{m_\pm}(1-q)^{\ell-m_\pm} = \frac{(\alpha' K q/2)^{m_\pm}}{m_\pm!}\,\mathrm{e}^{-\alpha' K q/2}. \tag{12}$$

We will also need later on the weighted distribution of $m_\pm$, where the weight is the number of clauses of type ±,

$$p_M^{(w)}(m_\pm) = \sum_{\ell=m_\pm}^{\infty} \ell p_L(\ell) \binom{\ell}{m_\pm} q^{m_\pm}(1-q)^{\ell-m_\pm} = p_M(m_\pm) \times \left(m_\pm + \frac{\alpha' K}{2}(1-q)\right). \tag{13}$$

Given m_+, m_- the best value for the variable x is TRUE if $m_+ > m_-$ and FALSE otherwise. The minimal number of violated clauses in the formula therefore increases by $E = \min(m_-, m_+)$. The field acting on the variable x is the difference between the number of violated clauses when x is TRUE and when it is FALSE, $z = m_+ - m_-$. In particular, the formula keeps being satisfiable upon inclusion of the new clauses if m_- or m_+ is equal to zero. The joint probability of the increase in energy and of the field reads

$$P(E, z) = \sum_{m_+=0}^{\infty} p_M(m_+) \sum_{m_-=0}^{\infty} p_M(m_-) \delta_{E,\min(m_-,m_+)} \delta_{z,(m_+-m_-)}. \tag{14}$$

Introducing the chemical potential ν of section 3.1 we weight each formula with a factor $\exp(-\nu E)$. The probability that the new variable is subjected to a field equal to $z = n$ is thus

$$r_n(\nu) = \frac{\sum_{E \geqslant 0} P(E, n)\,\mathrm{e}^{-\nu E}}{\sum_{E \geqslant 0} \sum_m P(E, m)\,\mathrm{e}^{-\nu E}} \tag{15}$$

where the denominator takes into account the proper reweighting of all formulae at fixed chemical potential ν.

We now restrict to the case of satisfiable formulae $\nu \to \infty$. From (14), (15) we obtain the probability of having zero field on x given that the formula is satisfiable,

$$\rho_0 = \lim_{\nu \to \infty} r_0(\nu) = \frac{P(0, 0)}{\sum_m P(0, m)} = \frac{1}{2\,\mathrm{e}^{\frac{\alpha' K q}{2}} - 1}. \tag{16}$$

We have now to take into account the fact that we want each variable to appear in αK clauses on average. Even if $\ell_\pm$ are distributed according to a Poissonian with average $\alpha' K/2$, the fact that when we add the clauses we must discard the possibilities in which the variable x receives contradictory messages makes the average number of clauses we add smaller than $\alpha' K$. It is easy to check, using (11) and (16), that

$$\begin{aligned}\langle \ell_+ + \ell_- \rangle &= \frac{\sum_{m_+,m_-} \left(p_M^{(w)}(m_+) p_M(m_-) + p_M(m_+) p_M^{(w)}(m_-)\right) \delta_{0,\min(m_-,m_+)}}{\sum_m P(0, m)} \\ &= \alpha' K \left[1 - \left(\frac{1-\rho_0}{2}\right)^K\right]. \end{aligned} \tag{17}$$

Then if we want $\langle \ell_+ + \ell_- \rangle = \alpha K$ we have to renormalize α' as

$$\alpha' = \frac{\alpha}{1 - \left(\frac{1-\rho_0}{2}\right)^K}, \tag{18}$$

and substituting into (16) we get back equation (7) as $\frac{\alpha' K q}{2} = \mathcal{G}(\rho_0)$.

The generating function $G(x)$ of the distribution of variable occurrences $\ell = \ell_+ + \ell_-$ can be easily computed by adding a weight x^ℓ in (12) and summing over $m_\pm$ with the constraint $\min(m_+, m_-) = 0$; after a correct normalization one obtains

$$G(x) = \mathrm{e}^{\alpha' K(x-1)(1-q)} \frac{2\,\mathrm{e}^{\alpha' K x q/2} - 1}{2\,\mathrm{e}^{\alpha' K q/2} - 1}, \tag{19}$$

that generates the difference of two Poissonians distribution with different parameters. This distribution differs from the normal Poissonian distribution of occurrences. However, the difference is exponentially small in α for all values of x, indeed for large α we get (recalling that $\rho_0 \to 0$)

$$G(x) = \exp(\alpha' K(x-1)(1-q/2)) + \mathrm{e}^{-O(\alpha)} = \mathrm{e}^{\alpha K(x-1)} + \mathrm{e}^{-O(\alpha)}, \tag{20}$$

which is the generating function of a Poissonian with parameter αK, consistently with (17). The fact that the distribution is Poissonian implies that the cavity fields have the same distribution of the true fields; the latter has been obtained from the replica method in section 3.1.

4. Comparison of $\mathcal{P}_{\mathrm{sat}}$ and $\mathcal{P}_{\mathrm{plant}}$ at large ratio α

For large α, the solution to (7) is well approximated by

$$\rho_0 = \frac{1}{2\,\mathrm{e}^{\gamma} - 1}, \qquad \gamma \equiv \mathcal{G}(0) = \frac{\alpha K}{2^K - 1}. \tag{21}$$

The distribution of the fields becomes

$$\rho_n = \frac{\gamma^{|n|}}{|n|!} \frac{1}{2\,\mathrm{e}^{\gamma} - 1}. \tag{22}$$

This result implies that the solutions of formulae extracted from $\mathcal{P}_{\mathrm{sat}}[F]$ are very similar to each other. They differ by a fraction $\mathrm{e}^{-O(\alpha)}$ of the variables only (this is in fact the weight r_0 of the field $z = 0$). The remaining fields $z \neq 0$ have typically values $O(\gamma) = O(\alpha)$, so that the variables in the core (with the same assignments in all the solutions) have strong cavity fields pointing to their correct assignments. Moreover, there is a correlation between cavity fields (or equivalently, between values of the variable in the solutions) and occurrence of the variables in the formula, as discussed in the next section. We will in addition show that the results (22) coincide with those obtained from the planted distribution, thus indicating that the two distributions coincide with errors $\mathrm{e}^{-O(\alpha)}$; indeed, we will compute the relative (extensive) entropy of the distribution and show that it is of the order of $N\mathrm{e}^{-O(\alpha)}$.

4.1. Distribution of the fields

To show that the Poissonian distribution (22) is the same distribution that is obtained from the planted distribution recall that ρ_n is the probability of violating $|n|$ clauses when a variable x_i is flipped from the correct value it has in the ground state to the opposite value. The sign of n is positive if x_i = TRUE in the solution and negative otherwise.

The planted distribution is constructed by extracting at random a configuration X, called root, and giving the same probability to all the choices of the clauses for which X is a solution. Then, if we choose at random a set of K indices $(i_1, \ldots, i_K)$, and consider all the possible 2^K equations we can construct with these indices, we see that only one of the choices is not allowed (e.g. if the configuration X is such that $(x_{i_1}, \ldots, x_{i_K}) = (1, \ldots, 1)$, only the choice $\bar{x}_{i_1} \vee \cdots \vee \bar{x}_{i_K}$ is not allowed).

The probability of violating $|n|$ clauses can be computed as follows. For simplicity, we choose the root $X = (1, \ldots, 1)$. Then we want to know how many clauses are violated when we flip one variable, e.g. $x_1 \to 0$. The clauses that are violated have the form

$$x_1 \vee \bar{x}_{i_2} \vee \cdots \vee \bar{x}_{i_K}. \tag{23}$$

The probability $p(|n|)$ that $|n|$ such clauses appear in a formula is a binomial distribution with parameter p given by the product of the probability that the variable x_1 appears in the clause, which is K/N, times the probability that the signs of the variables in the clause make it unsatisfied by the flipped assignment, which is $1/(2^K - 1)$:

$$p = \frac{K}{N}\frac{1}{2^K - 1}. \tag{24}$$

The number of equations is $M = \alpha N$ and they are independent so the probability of violating $|n|$ equations is

$$p(|n|) = \binom{M}{|n|} p^{|n|}(1-p)^{M-|n|} \sim \frac{\gamma^{|n|}}{|n|!}\,\mathrm{e}^{-\gamma}. \tag{25}$$

In a generic root X almost half of the variables are TRUE, giving rise to a positive field, while the other half are FALSE and correspond to negative fields; then we have

$$\rho_{\mathrm{plant}}(n) = \mathrm{e}^{-\gamma}\delta_{n0} + \frac{\gamma^{|n|}}{2|n|!}\,\mathrm{e}^{-\gamma}(1 - \delta_{n0}). \tag{26}$$

This distribution differ from (22) by $\mathrm{e}^{-O(\alpha)}$.

4.2. Correlation between fields and occurrences of the negations

Most variables are typically subject to strong fields, of the order of α in absolute value, in ground-state assignments. We now show that the sign of the field z associated with a variable, say, x, is strongly correlated to the numbers of occurrences of literals x and $\bar{x}$ in the formula.

Consider the cavity derivation of the self-consistent equations for the fields exposed in section 3.2. Suppose $z > 0$, and require the formula to be satisfiable ($\nu \to \infty$). Then the number of messages coming from ±-type clauses are $m_- = 0$, $m_+ \geqslant 1$. We define the average values of the number of ±-type clauses, $\langle \ell_\pm \rangle_{z>0}$, as follows:

$$\langle \ell_\pm \rangle_{z>0} = \frac{\sum_{m_+ \geqslant 1} p_M^{(w)}(m_+)\, p_M(0)}{\sum_{m_+ \geqslant 1} p_M(m_+)\, p_M(0)}. \tag{27}$$

We get

$$\begin{aligned}
\langle \ell_+ \rangle_{z>0} &= \frac{\alpha' K}{2}\left[\frac{1 - (1-q)\,\mathrm{e}^{-\mathcal{G}}}{1 - \mathrm{e}^{-\mathcal{G}}}\right] = \frac{\alpha K}{2}\frac{1}{1 - 2^{-K}} + \mathrm{e}^{-O(\alpha)},\\
\langle \ell_- \rangle_{z>0} &= \frac{\alpha' K}{2}(1-q) = \frac{\alpha K}{2}\frac{1 - 2^{-(K-1)}}{1 - 2^{-K}} + \mathrm{e}^{-O(\alpha)},
\end{aligned} \tag{28}$$

and finally

$$\frac{\langle \ell_+ \rangle_{z>0} - \langle \ell_- \rangle_{z>0}}{\langle \ell_+ \rangle_{z>0} + \langle \ell_- \rangle_{z>0}} = \frac{1}{2^K - 1} + \mathrm{e}^{-O(\alpha)}. \tag{29}$$

The average value of the bias between the numbers of positive and negative occurrences found for $\mathcal{P}_{\mathrm{sat}}$ coincides with its counterpart for $\mathcal{P}_{\mathrm{plant}}$ at large ratios α. Consider again the planted distribution with respect to $X = (1, \ldots, 1)$. It is easy to show that variables occur more frequently non-negated [12]. Indeed, if x enters a given clause C, there are $2^K - 1$

possible assignments of the negations, among which 2^{K-1} contain x and $2^{K-1}-1$ contain $\bar{x}$ (because the assignment in which all variables are negated is forbidden). Then, it is clear that

$$\frac{\langle \ell_+ \rangle_{\text{plant}} - \langle \ell_- \rangle_{\text{plant}}}{\langle \ell_+ \rangle_{\text{plant}} + \langle \ell_- \rangle_{\text{plant}}} = \frac{1}{2^K - 1}, \tag{30}$$

as in (29).

4.3. Relative entropy

The relative entropy of $\mathcal{P}_{\text{plant}}$ to $\mathcal{P}_{\text{sat}}$ is given, using (1), by

$$\sigma = -\sum_F \mathcal{P}_{\text{sat}}[F] \log \frac{\mathcal{P}_{\text{plant}}[F]}{\mathcal{P}_{\text{sat}}[F]} = -\log \mathcal{N}_{\text{sat}} + N \log 2 + \log \mathcal{N}_f - \sum_F \mathcal{P}_{\text{sat}}(F) \log \mathcal{N}_s[F] \tag{31}$$

where $\mathcal{N}_{\text{sat}}$ denotes the number of satisfiable formulae. We have

$$\mathcal{N}_{\text{sat}} = \mathrm{e}^{N\omega(0)} \times \mathcal{N} \qquad \text{where} \quad \mathcal{N} = \left[\binom{N}{K} 2^K \right]^M \tag{32}$$

is the total number of formulae. Using $\rho_0 \sim \mathrm{e}^{-\gamma}/2$, $\mathcal{G}(\rho_0) \sim \gamma + O(\mathrm{e}^{-\gamma})$, $\rho_0\, \mathrm{e}^{\mathcal{G}(\rho_0)} \sim 1/2 + O(\mathrm{e}^{-\gamma})$, we get

$$\omega(0) \sim \log 2 + \alpha \log \left(1 - \frac{1}{2^K}\right) + \frac{1}{2}\gamma\, \mathrm{e}^{-\gamma} + O(\mathrm{e}^{-\gamma}). \tag{33}$$

The value of the number of formulae sharing a common root, $\mathcal{N}_f$, is given in section 2. The last term in (31) represents the average entropy of satisfiable formulae. It is bounded from above by $N\rho_0 \log 2 \sim N\, \mathrm{e}^{-\gamma}$, because ρ_0 is an upper bound to the fraction of variables that can change values from solution to solution (inside the unique cluster).

Gathering all contributions we get the following expression for the relative entropy, valid for large ratios α,

$$\sigma = \tfrac{1}{2} N\gamma\, \mathrm{e}^{-\gamma} + O(N\, \mathrm{e}^{-\gamma}). \tag{34}$$

Hence, σ is extensive in N and decreases exponentially with α.

5. Finite energy results

The previous results extend to formulae having a small minimal fraction of unsatisfied clauses. This point is interesting since the relationship between approximation hardness and average-case complexity can be deduced from a weaker form of hypothesis 1 [9].

Hypothesis 2. *For every fixed $\epsilon > 0$, for α arbitrarily large (but independent of N), there is no polynomial time algorithm that on most 3-SAT formulae outputs typical, and never outputs typical on 3-SAT formulae with $(1-\epsilon)M$ satisfiable clauses.*

If we choose ν to be a large, finite number we find from the above replica calculation that the ground-state energy (C.12) dominating the integral (5) becomes

$$e_0(\nu) \sim \frac{\gamma}{K}[1 + O(\gamma^2\, \mathrm{e}^{-\gamma})]\, \mathrm{e}^{-\nu} \tag{35}$$

for large α. As in the $\nu \to \infty$ case, most cavity fields are non-zero and typically of the order of α. In addition, using the calculation of section 4.2, we can extend the calculation of the average difference between the number of ± occurrences of a variable with positive field, see section 4.2, to the case of large but finite ν with the result,

$$\frac{\langle \ell_+\rangle_{z>0} - \langle \ell_-\rangle_{z>0}}{\langle \ell_+\rangle_{z>0} + \langle \ell_-\rangle_{z>0}} = \frac{1}{2^K - 1} - \frac{\alpha K}{2(2^K - 1)^2}\,\mathrm{e}^{-\nu} + O(\alpha^{-1}), \tag{36}$$

to first order in $\mathrm{e}^{-\nu}$. Eliminating ν between (35) and (36), we obtain

$$\frac{\langle \ell_+\rangle_{z>0} - \langle \ell_-\rangle_{z>0}}{\langle \ell_+\rangle_{z>0} + \langle \ell_-\rangle_{z>0}} = \frac{1}{2^K - 1}\left[1 - eK2^K\left(\frac{1}{2} - \frac{1}{2^{K+1} - 2} - \frac{2^K}{\alpha K}\right)\right], \tag{37}$$

to first order in the ground-state energy density, e. This also suggests that formulae that are not exactly satisfiable but have few violated clauses ($e \ll 2^{-K+1}/K$) can be detected by the WP algorithm. A consequence is that a weaker version of hypothesis 2 in which 'never' is replaced with 'with probability p' should also be false for any $p > 0$.

We have checked the validity of this prediction on the following distribution of formulae, referred to as $\mathcal{P}^{(E)}_{\text{plant}}$ hereafter. Pick up uniformly at random a configuration X of the variables and choose M times independently a set of K indices uniformly over the $\binom{N}{K}$ possible ones to build M clauses. For the first E clauses, the negations of the variables are chosen such that the clause is violated by X (there is only one such assignment), while for the remaining $M - E$ clauses the negations are chosen such that the clause is satisfied in X (there are $2^K - 1$ such assignments and we choose one of them at random, as in the planted distribution). A simple calculation similar to the one of section 4.3 shows that the relative entropy, $\sigma\left(e = \frac{E}{N}\right)$, of $\mathcal{P}^{(E)}_{\text{plant}}$ and $\mathcal{P}_{\text{unif}}$ constrained to formulae with ground-state energy E is $\sigma(e) = N[\mathrm{e}^{-O(\alpha)}(1 + O(e)) + O(e^2)]$. Thus, both distributions are similar in the large α limit, at least for small enough e.

From a numerical point of view, we extracted 3-SAT formulae from $\mathcal{P}^{(E)}_{\text{plant}}$ with $N = 200$ variables, $M = 2000$ (i.e., $\alpha = 10$) and studied the convergence of the WP algorithm as a function of E. When the algorithm converges, it returns a partial assignment of the variables [10], the unassigned variables having a zero cavity field. Without entering into a detailed numerical investigation, we roughly observed that for $E < 10$ the algorithm behaves essentially as for $E = 0$: it converges after few iterations, and in the returned partial assignment most of the variables ($\sim$197 $\sim N(1 - \mathrm{e}^{-\gamma})$) have the same value they have in the reference configuration X and the rest of the variables are unassigned. After optimization over the unassigned variables, the energy of the resulting configuration differs from E by $\sim N\mathrm{e}^{-\gamma} \sim 3$ at most. Note that $E = 10$ corresponds to $e = 10/200 = 0.05$ and is compatible with the value $e_c \sim 2^{-K+1}/K \sim 0.08$ we found above (37) (there are corrections proportional to $N^{-1/2}$). Above $E \sim 15$, the probability of convergence decreases and the number of unassigned variables increases, but when the algorithm converges and one optimizes over the unassigned variables, the resulting configuration has an energy close to E by $\sim$3. Above $E \sim 50$, the algorithm almost never converges. Finally, it is interesting to observe that when the algorithm converges it does so after $\sim \log N \sim 6$ iterations, as predicted in [10] for $E = 0$, independent of the value of E. If convergence is not attained after $\sim$10 iterations, it is very likely that the algorithm will not converge in the following iterations. This allows one to put a cut-off on the number of required iterations *a priori*.

6. Conclusions

The present work supports the claim that satisfiable formulae from the uniform distribution can be recognized in polynomial time with probability close to unity provided the ratio of clause-per-variable is made large enough. In other words, WP should be efficient to solve the random

K-SAT problem at large α. This claim comes from the closeness of the two distributions $\mathcal{P}_{\text{sat}}$ and $\mathcal{P}_{\text{plant}}$ for large but finite ratios α. More precisely, both distributions produce formulae that (i) have a single cluster of solutions, in which (ii) a large fraction $1 - \mathrm{e}^{-O(\alpha)}$ of variables are strongly constrained (they have the same value in all solutions and a cavity field $O(\alpha)$) and a small fraction $\mathrm{e}^{-O(\alpha)}$ is free to change its value (zero cavity field). Moreover, (iii) as shown in section 4.2, a positively constrained variable x (i.e. TRUE in all the solutions) is very likely to appear more times as x than as $\bar{x}$ in the formula. The efficiency of WP on $\mathcal{P}_{\text{plant}}$ relies on these properties, and therefore extends to $\mathcal{P}_{\text{sat}}$, then to $\mathcal{P}_{\text{unif}}$ once a cut-off (of the order of $\log N$) is imposed on the number of iterations. Furthermore, these results extend to the case of a small but finite energy. Formulae with a minimal fraction of unsatisfied clauses larger than zero but much smaller than 2^{-K} (the typical value at large α) can be recognized with large probability by WP in polynomial time.

Yet the above findings are somewhat unsatisfactory for the following reason. It is easy to repeat the statistical mechanics calculations presented here for other Boolean functions expressing the truth values of clauses from the variables, e.g., the XORSAT model [13]. The outcome is that at large ratios properties (i) and (ii) hold quite generally but property (iii) does not. Hence, while from a probabilistic point of view the solution spaces of satisfiable SAT and XORSAT formulae far above the threshold are similar, they are not from an algorithmic point of view. More precisely, WP cannot find out whether a XORSAT formula is typical (and has a minimal fraction of unsatisfiable clauses close to $\frac{1}{2}$) or exceptional (minimal fraction $e \ll \frac{1}{2}$). It would be thus interesting to devise an algorithm capable of performing this task. What implications this would have on hypothesis 2 remains to be clarified too.

Acknowledgments

We wish to thank Giorgio Parisi for many discussions and for a careful reading of the manuscript, and Uriel Feige for useful comments. This work has been supported by the EU Research Training Network STIPCO (HPRN-CT-2002-00319).

Appendix A. Replicated free energy

Here, we sketch the derivation of the replicated free energy following [5]. The partition function can be written as $Z(\beta) = \sum_X \prod_{i=1}^M e_i(X)$, where $e_i(X) = 1$ if the clause C_i = TRUE in configuration X and $\mathrm{e}^{-\beta}$ otherwise. Then,

$$\begin{aligned}\overline{Z(\beta)^n} &= \overline{\sum_{X_1\cdots X_n}\prod_{i=1}^{M} e_i(X_1)\cdots e_i(X_n)} = \sum_{X_1\cdots X_n}\prod_{i=1}^{M}\overline{e_i(X_1)\cdots e_i(X_n)}\\ &= \sum_{X_1\cdots X_n}[\overline{e(X_1)\cdots e(X_n)}]^M,\end{aligned} \tag{A.1}$$

as the clauses are all chosen independently and with the same probability distribution. It is convenient to represent the variables x_i as spins, i.e. $x_i = 0 \leftrightarrow \sigma_i = -1$ and $x_i = 1 \leftrightarrow \sigma_i = 1$; then σ_i^a denotes the value of the spin at site i for replica a, $\vec{\sigma}_i$ is the n-component vector of the replicas of site i, $\underline{\sigma}^a$ is the N-component vector of the configuration of replica a and $\underline{\vec{\sigma}}$ is the full replicated configuration. Then, we can compute

$$\overline{e(\underline{\sigma}^1)\cdots e(\underline{\sigma}^n)} = \frac{1}{\binom{N}{K}}\sum_{i_1<\cdots<i_K}^{1,N}\frac{1}{2^K}\sum_{q_1\cdots q_K}^{-1,1}\prod_{a=1}^{n}\left\{1 + (\mathrm{e}^{-\beta}-1)\prod_{\ell=1}^{K}\delta\big[\sigma_{i_\ell}^{a}, q_\ell\big]\right\}, \tag{A.2}$$

where the variables q_ℓ correspond to the random choice of the negation in the clause C ($q_\ell = 1$ means that the variable x_{i_ℓ} is negated in C). To leading order in N we can neglect the constraint that all i's have to be different and replace $\binom{N}{K}^{-1} \sum_{i_1<\cdots<i_K}^{1,N}$ with $N^{-K} \sum_{i_1,\ldots,i_K}^{1,N}$.

Introducing the order parameter

$$\rho(\vec{\tau}|\underline{\vec{\sigma}}) = \frac{1}{N} \sum_{i=1}^{N} \prod_{a=1}^{n} \delta\big[\tau^a, \sigma_i^a\big], \tag{A.3}$$

that counts the number of sites i such that $\vec{\sigma}_i = \vec{\tau}$, we can write

$$\overline{e(\underline{\sigma}^1)\cdots e(\underline{\sigma}^n)} = \sum_{\vec{\tau}_1\cdots\vec{\tau}_K} \rho(\vec{\tau}_1|\underline{\vec{\sigma}})\cdots\rho(\vec{\tau}_K|\underline{\vec{\sigma}})\mathcal{E}(\vec{\tau}_1,\ldots,\vec{\tau}_K), \tag{A.4}$$

where

$$\mathcal{E}(\vec{\tau}_1,\ldots,\vec{\tau}_K) = \frac{1}{2^K} \sum_{q_1\cdots q_K}^{-1,1} \prod_{a=1}^{n} \left[1 + (\mathrm{e}^{-\beta} - 1) \prod_{\ell=1}^{K} \delta\big[\tau_\ell^a, q_\ell\big]\right]. \tag{A.5}$$

Finally, we write

$$\begin{aligned}\overline{Z(\beta)^n} &= \sum_{\underline{\vec{\sigma}}} \exp\left(M \log \sum_{\vec{\tau}_1\cdots\vec{\tau}_K} \rho(\vec{\tau}_1|\underline{\vec{\sigma}})\cdots\rho(\vec{\tau}_K|\underline{\vec{\sigma}})\mathcal{E}(\vec{\tau}_1,\ldots,\vec{\tau}_K)\right)\\ &= \int_0^1 \delta c(\vec{\tau}) \exp\left(N\alpha \log \sum_{\vec{\tau}_1\cdots\vec{\tau}_K} c(\vec{\tau}_1)\cdots c(\vec{\tau}_K)\mathcal{E}(\vec{\tau}_1,\ldots,\vec{\tau}_K)\right)\\ &\quad\times \sum_{\underline{\vec{\sigma}}} \prod_{\vec{\tau}} \delta[c(\vec{\tau}) - \rho(\vec{\tau}|\underline{\vec{\sigma}})],\end{aligned} \tag{A.6}$$

and observing that

$$\sum_{\underline{\vec{\sigma}}} \prod_{\vec{\tau}} \delta[c(\vec{\tau}) - \rho(\vec{\tau}|\underline{\vec{\sigma}})] = \frac{N!}{\prod_{\vec{\tau}}[Nc(\vec{\tau})]!} \sim \exp\left(-N \sum_{\vec{\tau}} c(\vec{\tau}) \log c(\vec{\tau})\right), \tag{A.7}$$

we finally obtain

$$\begin{aligned}\overline{Z^n[\beta]} &= \int_0^1 dc(\vec{\tau})\, \mathrm{e}^{N\mathcal{F}[c(\vec{\tau}),n,\beta]},\\ \mathcal{F}[c(\vec{\tau}), n, \beta] &= -\sum_{\vec{\tau}} c(\vec{\tau}) \log c(\vec{\tau}) + \alpha \log\left[\sum_{\vec{\tau}_1\cdots\vec{\tau}_K} c(\vec{\tau}_1)\cdots c(\vec{\tau}_K)\mathcal{E}(\vec{\tau}_1,\ldots,\vec{\tau}_K)\right].\end{aligned} \tag{A.8}$$

The partition function (A.8) can then be evaluated by a saddle point, and the saddle-point value of $c(\vec{\tau})$ is the average of the order parameter $\rho(\vec{\tau}|\underline{\vec{\sigma}})$. For symmetry reasons we expect that $c(\vec{\tau}) = c(-\vec{\tau})$ at the saddle point so the average over the signs $(q_1, \ldots, q_K)$ in (A.8) can be dropped setting $q_\ell \equiv 1$.

The replica symmetric *ansatz* amounts to choose

$$c(\vec{\tau}) = C\left[\sum_a \tau^a\right] = \int_{-\infty}^{\infty} \mathrm{d}z\, R(z) \frac{\mathrm{e}^{\frac{\beta z}{2}\sum_a \tau^a}}{[2\cosh(\beta z/2)]^n}, \tag{A.9}$$

where the last expression is a reparametrization of $c(\vec{\tau})$ in terms of a new function $R(z)$ thus defined, and which must satisfy $R(z) = R(-z)$. The normalization $\sum_{\vec{\tau}} c(\vec{\tau}) = 1$ implies

$\int \mathrm{d}z\, R(z) = 1$. Substituting into (A.8) we get, in the limit $\beta \to \infty$, $n \to 0$, $\nu = n\beta$,

$$\mathcal{F}[R(z), \nu] = -\int \frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi}\, \mathrm{e}^{\mathrm{i}x\hat{x}+\frac{1}{2}\nu|\hat{x}|}\varphi(x)\log\varphi(x) + \alpha \log \int_{-\infty}^{+\infty} \mathrm{d}z_1 \ldots \mathrm{d}z_K\, R(z_1)\ldots R(z_K)\, \mathrm{e}^{\nu\Phi(\mathbf{z})}, \tag{A.10}$$

where

$$\varphi(x) = \int \mathrm{d}z\, \mathrm{e}^{-\mathrm{i}xz-\frac{1}{2}\nu|z|}R(z), \tag{A.11}$$

$$\Phi(\mathbf{z}) = \max_{\sigma\in\{-1,1\}^K} \frac{1}{2}\sum_j(\sigma_j z_j - |z_j|) - \mathbb{1}_{\sigma,\mathbf{1}} = \begin{cases} -\min(1, z_1, \ldots, z_K) & \text{if} \quad (z_j > 0\,\forall j) \\ 0 & \text{otherwise.} \end{cases} \tag{A.12}$$

Appendix B. Saddle-point equation

Differentiating (A.10) with respect to $R(z)$, with the constraint $\int \mathrm{d}z R(z) = 1$, we get ($\mathbf{z} = (z, z_2, \ldots, z_K)$)

$$\begin{aligned} 0 &= \frac{\delta}{\delta R(z)}\left\{\mathcal{F}[R(\cdot), \nu] + \lambda\left[\int R(z')\,\mathrm{d}z' - 1\right]\right\} \\ &= -\int \frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi} \exp\left(\mathrm{i}x\hat{x} + \frac{1}{2}\nu|\hat{x}| - \mathrm{i}xz - \frac{1}{2}\nu|z|\right)[1 + \log\varphi(x)] \\ &\quad + \frac{\alpha K}{\mathcal{D}[R(\cdot)]}\int_{-\infty}^{+\infty} \mathrm{d}z_2\cdots\mathrm{d}z_K\, R(z_2)\cdots R(z_K)\, \mathrm{e}^{\nu\Phi(\mathbf{z})} + \lambda \end{aligned} \tag{B.1}$$

where

$$\mathcal{D}[R(\cdot)] = \int_{-\infty}^{+\infty} \mathrm{d}z_1\cdots\mathrm{d}z_k\, R(z_1)\cdots R(z_k)\, \mathrm{e}^{\nu\Phi(z_1,\ldots,z_K)}. \tag{B.2}$$

The function $R(z)$ is even, $R(z) = R(-z)$: in principle we should add a Lagrange multiplier to enforce this constraint, however this is equivalent to consider the equation above for $z \geqslant 0$ only.

In the last term, using the normalization of $R(z)$ and the definition of Φ we can write

$$\begin{aligned} &\int_{-\infty}^{\infty} \mathrm{d}z_2\cdots\mathrm{d}z_K\, R(z_2)\cdots R(z_K)\, \mathrm{e}^{\nu\Phi(\mathbf{z})} \\ &\quad = 1 - \frac{1}{2^{K-1}} + \int_0^{\infty} \mathrm{d}z_2\cdots\mathrm{d}z_K\, R(z_2)\cdots R(z_K)\, \mathrm{e}^{\nu\Phi(\mathbf{z})} \\ &\quad = 1 - \frac{1}{2^{K-1}} + \int \frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi}\exp(-\nu\min(z,\hat{x}) - \mathrm{i}x\hat{x}) \\ &\qquad \times \int_0^{\infty} \mathrm{d}z_2\cdots\mathrm{d}z_K\, R(z_2)\cdots R(z_K)\exp(\mathrm{i}x\min(1, z_2, \ldots, z_K)). \end{aligned} \tag{B.3}$$

Defining

$$Q(x) = \int_0^{\infty} \mathrm{d}z_2\cdots\mathrm{d}z_K\, R(z_2)\cdots R(z_K)\exp(\mathrm{i}x\min(1, z_2, \ldots, z_K)) \tag{B.4}$$

and using the relation $\min(z, \hat{x}) = -\frac{1}{2}[|z - \hat{x}| - z - \hat{x}]$, the last integral in (B.3) can be written as

$$\int \frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi}\exp(-\nu\min(z,\hat{x})-\mathrm{i}x\hat{x})Q(x)$$
$$=\int\frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi}\exp\left(-\mathrm{i}xz+\mathrm{i}x\hat{x}-\frac{\nu}{2}z+\frac{\nu}{2}|\hat{x}|\right)Q\left(x+\frac{\mathrm{i}\nu}{2}\right)$$
$$=\int\mathrm{d}x\,K(z,x)Q\left(x+\frac{\mathrm{i}\nu}{2}\right), \tag{B.5}$$

having defined the kernel $K(z,x)=\int\frac{\mathrm{d}\hat{x}}{2\pi}\exp\left(-\mathrm{i}xz+\mathrm{i}x\hat{x}-\frac{\nu}{2}z+\frac{\nu}{2}|\hat{x}|\right)$, that appears also in equation (B.1) for $z\geqslant 0$; note that $\int\mathrm{d}x\,K(z,x)=1$. The saddle-point equation (B.1) then becomes, for $z\geqslant 0$,

$$0=\int\mathrm{d}x\,K(z,x)\left\{\lambda-1-\log\varphi(x)+\frac{\alpha K}{\mathcal{D}[R(\cdot)]}\left[1-\frac{1}{2^{K-1}}+Q\left(x+\frac{\mathrm{i}\nu}{2}\right)\right]\right\}. \tag{B.6}$$

A solution of this equation is obtained when the term in curly brackets vanishes. Inverting equation (A.11), $R(z)=\int\frac{\mathrm{d}x}{2\pi}\,\mathrm{e}^{\mathrm{i}xz+\frac{\nu}{2}|z|}\varphi(x)$, and expressing $\varphi(x)$ using (B.6) we get

$$R(z)=\int_{-\infty}^{+\infty}\frac{\mathrm{d}x}{2\pi}\exp\left\{\mathrm{i}xz+\frac{\nu}{2}|z|+\lambda-1+\frac{\alpha K}{\mathcal{D}[R(\cdot)]}\left[1-\frac{1}{2^{K-1}}+Q\left(x+\frac{\mathrm{i}\nu}{2}\right)\right]\right\}. \tag{B.7}$$

Substituting

$$R(z)=\sum_{n=-\infty}^{+\infty}r_n\delta(z-n) \tag{B.8}$$

into (B.7) we obtain the coefficients

$$r_n=\frac{\mathrm{e}^{\frac{1}{2}\nu|n|}I_n(\alpha KB)}{\sum_{n'=-\infty}^{+\infty}\mathrm{e}^{\frac{1}{2}\nu|n'|}I_{n'}(\alpha KB)}, \tag{B.9}$$

$$B=\frac{\left(\frac{1-r_0}{2}\right)^{K-1}\mathrm{e}^{-\frac{1}{2}\nu}}{1+\left(\frac{1-r_0}{2}\right)^{K}(\mathrm{e}^{-\nu}-1)}, \tag{B.10}$$

where the denominator of B is $\mathcal{D}[R(\cdot)]$ from equation (B.2) and $I_n(x)$ is the modified Bessel function of order n.

Appendix C. Ground-state energy

We want to compute the ground-state energy

$$\overline{e_0}(\nu)=-\frac{\partial}{\partial\nu}\mathcal{F}[R(\cdot),\nu]=\frac{1}{2}\int\frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi}\,\mathrm{e}^{\mathrm{i}x\hat{x}+\frac{1}{2}\nu|\hat{x}|}\Bigg\{|\hat{x}|\varphi(x)\log\varphi(x)-[1+\log\varphi(x)]$$
$$\times\int_{-\infty}^{+\infty}\mathrm{d}z\,\mathrm{e}^{-\mathrm{i}xz-\frac{1}{2}\nu|z|}|z|R(z)\Bigg\}-\alpha\int_{-\infty}^{+\infty}\mathrm{d}z_1\cdots\mathrm{d}z_K\frac{R(z_1)\cdots R(z_K)}{\mathcal{D}[R(\cdot)]}\Phi(\mathbf{z})\,\mathrm{e}^{\nu\Phi(\mathbf{z})}.$$

We can use the saddle-point equations (B.1) and (B.7) to eliminate the integrals $\mathrm{d}x\,\mathrm{d}\hat{x}$ and obtain

$$\overline{e_0}(\nu)=-\int_0^\infty\mathrm{d}z\,zR(z)+\frac{\alpha K}{4}\int_0^\infty\mathrm{d}z_2\cdots\mathrm{d}z_K\frac{R(z_2)\ldots R(z_K)}{\mathcal{D}[R(\cdot)]}\min(1,z_2,\ldots,z_K)$$
$$+\alpha\int_0^\infty\mathrm{d}z_1\cdots\mathrm{d}z_K\frac{R(z_1)\cdots R(z_K)}{\mathcal{D}[R(\cdot)]}\left[\frac{K}{2}\min(1,z_2,\ldots,z_K)\right.$$
$$\left.+(1-K)\min(1,z_1,\ldots,z_K)\right]\exp(-\nu\min(1,z_1,\ldots,z_K)).$$

Using equation (6) for $R(z)$ we have

$$\overline{e_0}(\nu) = -\sum_{n=1}^{\infty} n r_n + \frac{\alpha K}{2} \frac{\left(\frac{1-r0}{2}\right)^{K-1}\left(\frac{1+r_0}{2}\right)}{1+\left(\frac{1-r0}{2}\right)^{K}(\mathrm{e}^{-\nu}-1)} + \alpha\left(1-\frac{K}{2}\right)\frac{\left(\frac{1-r0}{2}\right)^{K}\mathrm{e}^{-\nu}}{1+\left(\frac{1-r0}{2}\right)^{K}(\mathrm{e}^{-\nu}-1)} \tag{C.1}$$

and from the expressions (B.9) and (B.10) for r_n and B we obtain

$$\overline{e_0}(\nu) = -\frac{\partial}{\partial\nu}\log\mathcal{I}(\alpha K B,\nu) + \frac{\alpha K B}{2}\frac{1+r_0}{2}\,\mathrm{e}^{\frac{1}{2}\nu} + \alpha\left(1-\frac{K}{2}\right)B\frac{1-r_0}{2}\,\mathrm{e}^{-\frac{1}{2}\nu} \tag{C.2}$$

$$\mathcal{I}(z,\nu) \equiv \sum_{n=-\infty}^{+\infty} \mathrm{e}^{\frac{1}{2}\nu|n|} I_n(z). \tag{C.3}$$

For $\nu > 0$ this sum is always converging, as can be seen from equation (D.7).

C.1. The limit $\nu \to \infty$

We are interested in the limit $\nu \to \infty$ as in this limit $\overline{e_0}(\nu) \to 0$ as we will show. Let us define $\epsilon \equiv \mathrm{e}^{-\nu}$ and from (B.10) write

$$\begin{aligned} G &\equiv \frac{\alpha K B \mathrm{e}^{\frac{1}{2}\nu}}{2} = \frac{\alpha K}{2}\frac{\left(\frac{1-r_0}{2}\right)^{K-1}}{1-\left(\frac{1-r_0}{2}\right)^{K}(1-\epsilon)} = \frac{\alpha K}{2}\frac{\left(\frac{1-r_0}{2}\right)^{K-1}}{1-\left(\frac{1-r_0}{2}\right)^{K}}\left[1-\epsilon\frac{\left(\frac{1-r_0}{2}\right)^{K}}{1-\left(\frac{1-r_0}{2}\right)^{K}}\right] \\ &\equiv \mathcal{G}\left[1-\epsilon\frac{2\mathcal{G}}{\alpha K}\frac{1-r_0}{2}\right], \\ \mathcal{G} &= \frac{\alpha K}{2}\frac{\left(\frac{1-r_0}{2}\right)^{K-1}}{1-\left(\frac{1-r_0}{2}\right)^{K}}. \end{aligned} \tag{C.4}$$

Using the small-z expansion of the Bessel functions $I_n(z)$ $(n \geqslant 0)$

$$I_n(z) \sim \frac{z^n}{2^n n!}\left[1+\frac{z^2}{4(n+1)}+O(z^4)\right] \tag{C.5}$$

and $I_{-n}(z) = I_n(z)$ we have, using the identities

$$\sum_{n=-\infty}^{\infty}\frac{G^{|n|}}{|n|!} = 2\,\mathrm{e}^G - 1, \qquad G^2\sum_{n=-\infty}^{\infty}\frac{G^{|n|}}{(|n|+1)!} = 2G\,\mathrm{e}^G - G^2 - 2G, \tag{C.6}$$

that

$$\begin{aligned} \mathcal{I}(\alpha K B,\nu) \sim \sum_{n=-\infty}^{\infty}\frac{G^{|n|}}{|n|!}\left[1+\frac{\epsilon G^2}{|n|+1}+O(\epsilon^2)\right] = 2\,\mathrm{e}^G - 1 \\ +\epsilon(2G\,\mathrm{e}^G - G^2 - 2G) + O(\epsilon^2). \end{aligned} \tag{C.7}$$

The equation for r_0 is from (B.9), (C.5):

$$\begin{aligned} r_0 &= \frac{I_0\left(2G\,\mathrm{e}^{\frac{1}{2}\nu}\right)}{2\,\mathrm{e}^G - 1 + \epsilon(2G\,\mathrm{e}^G - G^2 - 2G) + O(\epsilon^2)} = \frac{1+\epsilon G^2 + O(\epsilon^2)}{2\,\mathrm{e}^G - 1 + \epsilon(2G\,\mathrm{e}^G - G^2 - 2G) + O(\epsilon^2)} \\ &= \frac{1}{2\,\mathrm{e}^{\mathcal{G}}-1}\left\{1+\epsilon\mathcal{G}^2+\frac{\epsilon}{2\,\mathrm{e}^{\mathcal{G}}-1}\left[2\,\mathrm{e}^{\mathcal{G}}\frac{2\mathcal{G}}{\alpha K}\frac{1-r_0}{2} - 2\mathcal{G}\,\mathrm{e}^{\mathcal{G}} + \mathcal{G}^2 + 2\mathcal{G}\right] + O(\epsilon^2)\right\} \\ &= F_0(r_0) + \epsilon F_1(r_0) + O(\epsilon^2). \end{aligned} \tag{C.8}$$

The solution to the previous equation is

$$r_0 = \rho_0 + \epsilon\rho_1, \qquad \rho_0 = \frac{1}{2\,\mathrm{e}^{\mathcal{G}(\rho_0)} - 1}, \qquad \rho_1 = \frac{F_1(\rho_0)}{1-F_0'(\rho_0)}. \tag{C.9}$$

To write the energy we need to compute

$$\sum_{n=1}^{\infty} n\, \mathrm{e}^{\frac{1}{2}n\nu} I_n(\alpha KB) = \sum_{n=1}^{\infty} \frac{G^n}{(n-1)!}\left[1 + \frac{\epsilon G^2}{n+1} + O(\epsilon^2)\right]$$
$$= G\,\mathrm{e}^G + \epsilon G(1 - \mathrm{e}^G + G\,\mathrm{e}^G) + O(\epsilon^2). \qquad \text{(C.10)}$$

The energy (C.2) is then given by, neglecting $O(\epsilon^2)$,

$$\overline{e_0}(\nu) = -\frac{G\,\mathrm{e}^G + \epsilon G(1 - \mathrm{e}^G + G\,\mathrm{e}^G)}{2\,\mathrm{e}^G - 1 + \epsilon(2G\,\mathrm{e}^G - G^2 - 2G)} + G\frac{1+r_0}{2} + \epsilon\left(\frac{2}{K} - 1\right) G\frac{1-r_0}{2}. \qquad \text{(C.11)}$$

Given that $\overline{e_0}(\nu) = -\partial_\nu F$ and that r_0 is the solution of $\partial_{r_0} F = 0$, the term $\epsilon\rho_1$ in r_0 should not contribute to $\overline{e_0}(\nu)$ at first order in ϵ. Then, we can write

$$\begin{aligned}
\overline{e_0}(\nu) &= -\mathcal{G}\,\mathrm{e}^{\mathcal{G}}\rho_0\left\{1 - \epsilon\frac{2\mathcal{G}^2}{\alpha K}\frac{1-\rho_0}{2}\left[\frac{1}{\mathcal{G}} - \rho_0\right] + \epsilon[\mathrm{e}^{-\mathcal{G}} - 1 + \mathcal{G}] - \epsilon\rho_0[2\mathcal{G}\,\mathrm{e}^{\mathcal{G}} - \mathcal{G}^2 - 2\mathcal{G}]\right\}\\
&\quad + \mathcal{G}\rho_0\,\mathrm{e}^{\mathcal{G}}\left[1 - \epsilon\frac{2\mathcal{G}}{\alpha K}\frac{1-\rho_0}{2}\right] + \epsilon\left(\frac{2}{K} - 1\right)\mathcal{G}\frac{1-\rho_0}{2}\\
&= \epsilon\mathcal{G}\mathrm{e}^{\mathcal{G}}\rho_0\left\{-\frac{\mathcal{G}^2\rho_0(1-\rho_0)}{\alpha K} - [\mathrm{e}^{-\mathcal{G}} - 1 + \mathcal{G}] + \rho_0[2\mathcal{G}\mathrm{e}^{\mathcal{G}} - \mathcal{G}^2 - 2\mathcal{G}]\right.\\
&\quad \left. + \frac{1}{\rho_0\,\mathrm{e}^{\mathcal{G}}}\left(\frac{2}{K} - 1\right)\frac{1-\rho_0}{2}\right\}
\end{aligned} \qquad \text{(C.12)}$$

and $\overline{e_0}(\nu) \sim \mathrm{e}^{-\nu}$ for large ν. The latter expression is complicated, but it simplifies considerably in the limit $\alpha \to \infty$.

Appendix D. Free energy of the RS solution

Finally, we can compute the free energy corresponding to the solution to (B.9). We begin by calculating

$$\varphi(x) = \int \mathrm{d}z\, \mathrm{e}^{-\mathrm{i}xz - \frac{1}{2}\nu|z|} \sum_{n=-\infty}^{\infty} \frac{\mathrm{e}^{\frac{1}{2}\nu|n|} I_n(\alpha KB)}{\mathcal{I}(\alpha KB, \nu)}\delta(z - n) = \frac{\mathrm{e}^{\alpha KB\cos x}}{\mathcal{I}(\alpha KB, \nu)}. \qquad \text{(D.1)}$$

Then using $\varphi(x)\log\varphi(x) = \left[\frac{\partial}{\partial p}\varphi(x)^p\right]_{p=1}$, we rewrite the first term in (A.10) as

$$\begin{aligned}
\int \frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi}\,\mathrm{e}^{\mathrm{i}x\hat{x} + \frac{1}{2}\nu|\hat{x}|}\varphi(x)\log\varphi(x) &= \left.\frac{\partial}{\partial p}\right|_{p=1}\int \frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi}\,\mathrm{e}^{\mathrm{i}x\hat{x} + \frac{1}{2}\nu|\hat{x}|}\varphi(x)^p\\
&= \left.\frac{\partial}{\partial p}\right|_{p=1}\int \frac{\mathrm{d}x\,\mathrm{d}\hat{x}}{2\pi}\,\mathrm{e}^{\mathrm{i}x\hat{x} + \frac{1}{2}\nu|\hat{x}|} \times \frac{\mathrm{e}^{p\alpha KB\cos x}}{\mathcal{I}(\alpha KB)^p}\\
&= \left.\frac{\partial}{\partial p}\right|_{p=1}\int \mathrm{d}\hat{x}\,\frac{\mathrm{e}^{\frac{1}{2}\nu|\hat{x}|}}{\mathcal{I}(\alpha KB)^p}\int \frac{\mathrm{d}x}{2\pi}\,\mathrm{e}^{p\alpha KB\cos x}\cos(\hat{x}x)
\end{aligned} \qquad \text{(D.2)}$$

where all integrals are between $-\infty$ and $+\infty$. The dx integral is of the form

$$f(\hat{x}) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{2\pi}\cos(\hat{x}x)\psi(\cos x) = \sum_{n=-\infty}^{\infty} f_n\delta(n - \hat{x}) \qquad \text{(D.3)}$$

where

$$f_n = \int_0^{2\pi}\frac{\mathrm{d}t}{2\pi}\,\mathrm{e}^{\mathrm{i}tn}\psi(\cos t) = \frac{1}{\pi}\int_0^{\pi}\mathrm{d}t\,\mathrm{e}^{p\alpha KB\cos t}\cos(nt) = I_n(p\alpha KB). \qquad \text{(D.4)}$$

In this way, we obtain for the double integral $\mathrm{d}x\,\mathrm{d}\hat{x}$

$$\left.\frac{\partial}{\partial p}\right|_{p=1} \frac{\sum_{n=-\infty}^{\infty} \mathrm{e}^{\frac{1}{2}\nu|n|} I_n(p\alpha KB)}{\mathcal{I}(\alpha KB,\nu)^p} = \frac{\alpha KB \mathcal{I}^{(1,0)}(\alpha KB,\nu)}{\mathcal{I}(\alpha KB,\nu)} - \log \mathcal{I}(\alpha KB,\nu). \tag{D.5}$$

The energy term in the free energy is just $\alpha \log \mathcal{D}[R(\cdot)]$, and we obtain

$$\mathcal{F}(\nu) = -\alpha KB \frac{\mathcal{I}^{(1,0)}(\alpha KB,\nu)}{\mathcal{I}(\alpha KB,\nu)} + \log \mathcal{I}(\alpha KB,\nu) + \alpha \log\left[1 + \left(\frac{1-r_0}{2}\right)^K (\mathrm{e}^{-\nu} - 1)\right]. \tag{D.6}$$

The sums can be expressed in terms of fast-converging series (for $\nu > 0$):

$$\mathcal{I}(z,\nu) = 2\,\mathrm{e}^{z\cosh(\frac{1}{2}\nu)} - I_0(z) - 2\sum_{n=1}^{\infty} \mathrm{e}^{-\frac{1}{2}\nu n} I_n(z), \tag{D.7}$$

$$\mathcal{I}^{(1,0)}(z,\nu) = 2\cosh\left(\frac{1}{2}\nu\right) \mathrm{e}^{z\cosh(\frac{1}{2}\nu)} - \mathrm{e}^{-\frac{1}{2}\nu} I_0(z) - 2\cosh\left(\frac{1}{2}\nu\right) \sum_{n=1}^{\infty} \mathrm{e}^{-\frac{1}{2}\nu n} I_n(z). \tag{D.8}$$

Appendix E. Eigenvalues of the stability matrix in the RS solution

Differentiation of the free energy (A.8) gives

$$M_{\vec{\sigma}\vec{\tau}} \equiv \frac{\partial^2 \mathcal{F}}{\partial c(\vec{\sigma})\partial c(\vec{\tau})} = -\frac{1}{c(\vec{\sigma})}\delta_{\vec{\sigma}\vec{\tau}} + \frac{\alpha K(K-1)\sum_{\vec{\sigma}_3\cdots\vec{\sigma}_K} c(\vec{\sigma}_3)\cdots c(\vec{\sigma}_K)\mathcal{E}(\vec{\sigma},\vec{\tau},\vec{\sigma}_3,\ldots,\vec{\sigma}_K)}{\sum_{\vec{\sigma}_1\cdots\vec{\sigma}_K} c(\vec{\sigma}_1)\cdots c(\vec{\sigma}_K)\mathcal{E}(\vec{\sigma}_1,\ldots,\vec{\sigma}_K)}$$
$$- \frac{\alpha K^2 \sum_{\vec{\sigma}_2\cdots\vec{\sigma}_K} c(\vec{\sigma}_2)\cdots c(\vec{\sigma}_K)\mathcal{E}(\vec{\sigma},\vec{\sigma}_2,\ldots,\vec{\sigma}_K) \sum_{\vec{\sigma}'_2\cdots\vec{\sigma}'_K} c(\vec{\sigma}'_2)\cdots c(\vec{\sigma}'_K)\mathcal{E}(\vec{\tau},\vec{\sigma}'_2,\ldots,\vec{\sigma}'_K)}{\left[\sum_{\vec{\sigma}_1\cdots\vec{\sigma}_K} c(\vec{\sigma}_1)\cdots c(\vec{\sigma}_K)\mathcal{E}(\vec{\sigma}_1,\ldots,\vec{\sigma}_K)\right]^2}, \tag{E.1}$$

with $\mathcal{E}$ as in (A.5). The function $c(\vec{\sigma})$ can be computed at the saddle point using (A.9), (6) and (8); defining $s = \frac{1}{n}\sum_a \sigma_a$ we have

$$c(\vec{\sigma}) = \frac{1}{2\,\mathrm{e}^{\mathcal{G}} - 1}\left[\mathrm{e}^{\mathcal{G}\mathrm{e}^{-\frac{\nu(1-s)}{2}}} + \mathrm{e}^{\mathcal{G}\mathrm{e}^{\frac{-\nu(1+s)}{2}}} - 1\right] \sim \frac{1}{2\,\mathrm{e}^{\mathcal{G}} - 1}\{\mathrm{e}^{\mathcal{G}}\delta[|s|,1] + (1 - \delta[|s|,1])\}, \tag{E.2}$$

where the last equality holds for $\nu \to \infty$. For large α, $\mathcal{G}$ becomes large and the expression for $c(\vec{\sigma})$ further simplifies:

$$c(\vec{\sigma}) = \tfrac{1}{2}\delta(|s|,1) + O(\mathrm{e}^{-\alpha}). \tag{E.3}$$

This allows a straightforward calculation of the sums appearing in (E.1). In order to do this, we observe that $\mathcal{E}(\vec{\sigma}_1,\ldots,\vec{\sigma}_K)$, defined in equation (A.5), is equal (in the limit $\beta \to \infty$) to $1/2^K$ times the number of vectors in $\{-1,1\}^K$ that are not equal to any of the columns of the matrix whose rows are the vectors $\vec{\sigma}_1,\ldots,\vec{\sigma}_K$. Then, to $o(1)$ in α,

$$\mathcal{D}_K \equiv \sum_{\vec{\sigma}_1,\ldots,\vec{\sigma}_K} c(\vec{\sigma}_1)\cdots c(\vec{\sigma}_K)\mathcal{E}(\vec{\sigma}_1,\ldots,\vec{\sigma}_K) \tag{E.4}$$

$$= 2^K \times \frac{1}{2^K} \times \frac{1}{2^K}(2^K - 1) = 1 - \frac{1}{2^K} \tag{E.5}$$

since the only terms that contribute to the sum are those with $\vec{\sigma}_i = (1,1,\ldots,1)$ or $(-1,-1,\ldots,-1)$, and the corresponding matrices have all the columns equal (so that there are $2^K - 1$ vectors that are not equal to any column). In the same way, we obtain

$$\mathcal{D}_{K-1}(\vec{\sigma}) \equiv \sum_{\vec{\sigma}_2,\ldots,\vec{\sigma}_K} c(\vec{\sigma}_2)\cdots c(\vec{\sigma}_K)\mathcal{E}(\vec{\sigma},\vec{\sigma}_2,\ldots,\vec{\sigma}_K) \tag{E.6}$$

$$= 2^{K-1} \times \frac{1}{2^{K-1}} \times \frac{1}{2^K}[2^K - (2 - \delta(|s|,1))] = 1 - \frac{1}{2^{K-1}} + \frac{\delta(|s|,1)}{2^K} \tag{E.7}$$

since if $|s| = 1$ all the columns are equal, while if $|s| < 1$ there will be two different column types, and

$$\mathcal{D}_{K-2}(\vec{\sigma},\vec{\tau}) \equiv \sum_{\vec{\sigma}_3,\ldots,\vec{\sigma}_K} c(\vec{\sigma}_3)\cdots c(\vec{\sigma}_K)\mathcal{E}(\vec{\sigma},\vec{\tau},\vec{\sigma}_3,\ldots,\vec{\sigma}_K) \tag{E.8}$$

$$= 2^{K-2} \times \frac{1}{2^{K-2}} \times \frac{1}{2^K}[2^K - D(\vec{\sigma},\vec{\tau})] = 1 - \frac{D(\vec{\sigma},\vec{\tau})}{2^K} \tag{E.9}$$

where the function $D(\vec{\sigma},\vec{\tau})$ counts the number of different pairs, among the possible four $(-1,-1)$, $(-1,1)$, $(1,-1)$ and $(1,1)$, that actually occur in $\{(\sigma^a,\tau^a), a = 1,\ldots,n\}$. It is straightforward to verify the recursion relations

$$\mathcal{D}_K = \sum_{\vec{\sigma}} c(\vec{\sigma})\mathcal{D}_{K-1}(\vec{\sigma}), \tag{E.10}$$

$$\mathcal{D}_{K-1}(\vec{\sigma}) = \sum_{\vec{\tau}} c(\vec{\tau})\mathcal{D}_{K-2}(\vec{\sigma},\vec{\tau}). \tag{E.11}$$

Then we get, neglecting $\mathrm{e}^{-O(\alpha)}$,

$$M_{\vec{\sigma}\vec{\tau}} = -\frac{2\mathrm{e}^{\mathcal{G}}\delta_{\vec{\sigma}\vec{\tau}}}{\mathrm{e}^{\mathcal{G}}\delta[|s|,1] + (1-\delta[|s|,1])} + \frac{\alpha K(K-1)[2^K - D(\vec{\sigma},\vec{\tau})]}{2^K - 1} \\ - \frac{\alpha K^2(\delta[|s|,1] + 2^K - 2)(\delta[|t|,1] + 2^K - 2)}{(2^K-1)^2}. \tag{E.12}$$

This matrix is invariant under permutations of the replicas so it preserves the symmetry of the vectors under permutations. This means that it can be block-diagonalized in subspaces of given replica symmetry.

First we have to take into account the constraint $\sum_{\vec{\sigma}} c(\vec{\sigma}) = 1$. This can be done by considering $\mathcal{F}[c(\vec{\sigma})] = \mathcal{F}\big[1 - \sum_{\vec{\sigma}} c'(\vec{\sigma}), c'(\vec{\sigma})\big]$ where $c'(\vec{\sigma}) = c(\vec{\sigma})$ for $\vec{\sigma} \neq (+1,\ldots,+1) \equiv \vec{1}$ and $c'(\vec{\sigma})$ has no $\vec{1}$ component. Then it is easy to show that Hessian matrix of $\mathcal{F}$ with respect to c' is $(\vec{\sigma},\vec{\tau} \neq \vec{1})$

$$M'_{\vec{\sigma}\vec{\tau}} = M_{\vec{\sigma}\vec{\tau}} - M_{\vec{\sigma}\vec{1}} - M_{\vec{1}\vec{\tau}} + M_{\vec{1}\vec{1}} = -\frac{2\mathrm{e}^{\mathcal{G}}\delta_{\vec{\sigma}\vec{\tau}}}{\mathrm{e}^{\mathcal{G}}\delta[|s|,1] + (1-\delta[|s|,1])} - 2 \\ - \frac{\alpha K(K-1)}{2^K - 1}[1 + D(\vec{\sigma},\vec{\tau}) - D(\vec{1},\vec{\tau}) - D(\vec{\sigma},\vec{1})] \\ - \frac{\alpha K^2(\delta[|s|,1] - 2)(\delta[|t|,1] - 2)}{(2^K-1)^2}. \tag{E.13}$$

Let us start with the non-symmetric subspaces. In these subspaces, $|s| \neq 1$ and $|t| \neq 1$, then

$$M'_{\vec{\sigma}\vec{\tau}} = -2\,\mathrm{e}^{\mathcal{G}}\delta_{\vec{\sigma}\vec{\tau}} - \frac{\alpha K(K-1)}{2^K - 1}[1 + D(\vec{\sigma},\vec{\tau}) - D(\vec{1},\vec{\tau}) - D(\vec{\sigma},\vec{1})] - 2 - \frac{4\alpha K^2}{(2^K-1)^2}. \tag{E.14}$$

The diagonal term is $O(\mathrm{e}^{\alpha})$ while the off-diagonal part is $O(\alpha)$. This means that even in the properly symmetrized basis the matrix elements will have a diagonal part of $O(\mathrm{e}^{\alpha})$ while the off-diagonal elements will be $O(\alpha)$. Then it is easy to show (e.g., in perturbation theory) that the off-diagonal terms can change the eigenvalues at most by a quantity $O(\alpha 2^n)$, so it cannot change the sign of the eigenvalues. In this space, the matrix M has then only negative eigenvalues and $\mathcal{F}$ has a maximum.

In the symmetric subspace, we can use the same argument for all the eigenvalues but the diagonal element corresponding to $\vec{\sigma}, \vec{\tau} = -\vec{1}$ which is not $O(e^{\alpha})$. However, we can write

$$M'_{\vec{\sigma}\vec{\tau}} = -2\,\mathrm{e}^{\mathcal{G}}\delta_{\vec{\sigma}\vec{\tau}}(1 - \delta[|s|, 1]) + V_{\vec{\sigma}\vec{\tau}}, \tag{E.15}$$

and treat V as a perturbation. In the dangerous subspace $\vec{\sigma} = \vec{\tau} = -\vec{1}$ where the diagonal part has zero eigenvalue, the matrix element of the perturbation is

$$V_{-\vec{1},-\vec{1}} = M'_{-\vec{1},-\vec{1}} = M_{-\vec{1},-\vec{1}} - M_{-\vec{1},\vec{1}} - M_{\vec{1},-\vec{1}} + M_{\vec{1},\vec{1}} = -4 < 0, \tag{E.16}$$

so the eigenvalues of M' are all negative for α large enough.

Appendix F. Solutions with non-integer fields in the $\nu \to \infty$ limit

We look for a solution with rational-valued fields,

$$R(z) = \sum_{n=-\infty}^{+\infty} r_n \delta\left(z - \frac{n}{p}\right) \tag{F.1}$$

where p is an integer $\geqslant 1$. As the fields are expected to be integer valued, the existence of such a solution would be an indication for an instability of the replica symmetric solution. We plug this *ansatz* in the self-consistent equation for R (B.7) and find a self-consistent equation for the p variables $r_0, r_1, \ldots, r_{p-1}$:

$$\sum_{n=-\infty}^{+\infty} r_n \cos\left(x\frac{n}{p}\right) \mathrm{e}^{-\nu|n|/(2p)} = \exp\left(\mu + \alpha K \sum_{q=1}^{p} A_q \cos\left(x\frac{q}{p}\right) \mathrm{e}^{-\nu q/(2p)}\right) \tag{F.2}$$

which must be true for any x. In the above equation we have defined

$$\begin{aligned} A_1 &= \frac{w^{K-1} - (w - r_1)^{K-1}}{1 - w^K}, \\ A_q &= \frac{(w - r_{q-1})^{K-1} - (w - r_q)^{K-1}}{1 - w^K} (2 \leqslant q \leqslant p-1), \ldots, A_p = \frac{(w - r_{p-1})^{K-1}}{1 - w^K}, \end{aligned} \tag{F.3}$$

where $w = (1 - r_0)/2$. To calculate the constant μ we set $x = \mathrm{i}\frac{\nu}{2}$ and send $\nu \to \infty$ to obtain

$$\sum_{n=-\infty}^{+\infty} r_n \left(\frac{1 + \delta_{n,0}}{2}\right) = \exp\left(\mu + \frac{1}{2}\alpha K \sum_{q=1}^{p} A_q\right). \tag{F.4}$$

In addition setting $x = 0$ and sending $\nu \to \infty$ we have $r_0 = e^{\mu}$. Combining the two equations above, we obtain the self-consistent equation

$$r_0 = \frac{1}{2\exp\left(\frac{\alpha}{2} K \frac{w^{K-1}}{1 - w^K}\right) - 1} \tag{F.5}$$

which is the same equation as in the case of integer fields only ($q = 1$). The equation for the weight of the smallest non-zero field reads

$$r_1 = r_0 \frac{A_1}{2}, \tag{F.6}$$

and for $r_1 \neq 0$ can be written equivalently, with $y = r_1/w$, as

$$\frac{y}{1-(1-y)^{K-1}} = \alpha K \left(\frac{1}{2} - w\right) \frac{w^{K-2}}{1-w^K}. \tag{F.7}$$

As r_0 ranges from 0 to 1, w ranges from 0 to $\frac{1}{2}$. Moreover, note that $r_1 \leqslant w$ because w is the probability of having a positive field; then y ranges from 0 to 1. When α is large the rhs is $\sim\gamma\, \mathrm{e}^{-\gamma}$ which is very small while the lhs is larger than $1/(K-1)$ (minimal value in $y = 0$). Therefore, the latter equation has no solution and the only solution to (F.6) is $r_1 = 0$.

References

[1] Mitchell D, Selman B and Levesque H 1992 Hard and easy distributions of SAT problems *Proc. 10th Natl Conf. on Artificial Intelligence (AAAI-92)* (Cambridge, MA: The AAAI Press/MIT Press) pp 440–6

[2] Friedgut E 1999 Sharp thresholds of graph properties, and the k-sat problem *J. Am. Math. Soc.* **12** 1017

[3] For a review of rigorous works on upper and lower bounds to the threshold ratio, see the special issue of 2001 *Theor. Comput. Sci.* **265** 109–30

[4] Achlioptas D, Naor A and Peres Y 2005 Rigorous location of phase transitions in hard optimization problems *Nature* **435** 759–64

[5] Monasson R and Zecchina R 1997 Statistical mechanics of the random K-satisfiability model *Phys. Rev.* E **56** 1357

[6] Biroli G, Monasson R and Weigt M 2000 A variational description of the ground state structure in random satisfiability problems *Eur. Phys. J.* B **14** 551

[7] Mézard M, Parisi G and Zecchina R 2002 Analytic and algorithmic solutions of random satisfiability problems *Science* **297** 812

[8] Mertens S, Mézard M and Zecchina R 2005 Threshold values of random K-SAT from the cavity method *Random Struct. Algorithms* **28** 340–73

[9] Feige U 2002 Relations between average case complexity and approximation complexity *Proc. 4th STOC* pp 534–543 (Available online at http://www.wisdom.weizmann.ac.il/feige/Approx/r3sat.ps)

[10] Feige U, Mossel E and Vilenchik D 2006 Complete convergence of message passing algorithms for some satisfiability problems *Proc. Random 2006* pp 339–50 (Available online at http://research.microsoft.com/research/theory/feige/homepagefiles/WP_9_14.ps)

[11] MacKay D J C 2003 *Information Theory, Inference, and Learning Algorithms* (Cambridge: Cambridge University Press)

[12] Bartel W *et al* 2002 Hiding solutions in random satisfiability problems: a statistical mechanics approach *Phys. Rev. Lett.* **88** 188701

[13] Creignou N and Daudé H 1999 Satisfiability threshold for random XOR-CNF formulas *Discr. Appl. Math.* **96–97** 41