

Instance space of the number partitioning problem

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2000 J. Phys. A: Math. Gen. 33 7265

(<http://iopscience.iop.org/0305-4470/33/41/301>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.123

The article was downloaded on 02/06/2010 at 08:33

Please note that [terms and conditions apply](#).

Instance space of the number partitioning problem

F F Ferreira and J F Fontanari

Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369,
13560-970 São Carlos SP, Brazil

Received 10 July 2000

Abstract. Within the replica framework we study analytically the instance space of the number partitioning problem. This classic integer programming problem consists of partitioning a sequence of N positive real numbers $\{a_1, a_2, \dots, a_N\}$ (the instance) into two sets such that the absolute value of the difference of the sums of a_j over the two sets is minimized. We show that, regardless of the distribution of the instance entries, there is an upper bound $\alpha_c N$ to the number of perfect random partitions (i.e. partitions for which that difference is zero). In particular, in the case where the two sets have the same cardinality (balanced partitions) we find $\alpha_c = \frac{1}{2}$. Moreover, in the case of unbalanced partitions, we show that perfect random partitions exist only if the difference between the cardinalities of the two sets scales like $mN^{1/2}$, where m is of the order of 1.

1. Introduction

Most statistical mechanics analyses of combinatorial optimization problems have concentrated on the characterization of average properties of the minima of a given cost function [1, 2]. Usually, the cost function depends on a large set of fixed parameters, termed the instance of the optimization problem (e.g. the distances between cities in the celebrated travelling salesman problem (TSP)) which, within the framework of statistical mechanics, are treated as quenched random variables distributed according to some known probability distribution. Furthermore, in order to consider the subspace of configurations with a given average cost, one defines a probability distribution on the space of configurations (e.g. the $N!/2N$ different tours or ordering of the cities in the TSP), namely, the Gibbs distribution with ‘temperature’ $T = 1/\beta$. The zero-temperature limit then singles out the configurations that minimize the cost function (ground states). Clearly, in this formulation the configurations are treated as fast, annealed variables.

Instead, in this work we explore the opposite viewpoint, namely, given a set of configurations we want to characterize the subspace of instances for which those configurations have a certain cost. The situation here is similar to the physics approach to neural networks. In a first stage, attention was given to the neural dynamics, while the coupling strengths between neurons were kept fixed according to some variant of the Hebb learning rule [3, 4]. (The neural dynamics itself can be viewed as a versatile heuristic in which the optimization problem is embedded in the neural couplings [5].) In the second stage which followed the seminal work of Gardner [6, 7], the focus was on the characterization of the couplings distribution that ensures the stability of a given set of neural states. Gardner’s formulation allowed a rich interchange of concepts and methods between the statistical physics and the computational learning theory communities [8].

The specific optimization problem we consider in this paper is the number partitioning problem (NPP) [9, 10] which has received considerable attention in the physics literature recently [11–13]. It is stated as follows. Given a sequence of N positive real numbers $\mathbf{a} = \{a_1, a_2, \dots, a_N\}$ (the instance), the NPP consists of partitioning them into two disjoint sets \mathcal{A}_1 and \mathcal{A}_2 such that the difference

$$\left| \sum_{a_j \in \mathcal{A}_1} a_j - \sum_{a_j \in \mathcal{A}_2} a_j \right| \quad (1)$$

is minimized. Alternatively, we can search for the Ising spin configurations $\mathbf{s} = \{s_1, \dots, s_N\}$ that minimize the cost function

$$E(\mathbf{s}) = \left| \sum_{j=1}^N a_j s_j \right| \quad (2)$$

where $s_j = 1$ if $a_j \in \mathcal{A}_1$ and $s_j = -1$ if $a_j \in \mathcal{A}_2$. Despite its simplicity, the NPP was shown to belong to the NP-complete class, which basically means that there is no known deterministic algorithm guaranteed to solve all instances of this problem within a polynomial time bound [1].

In the proposed framework, we aim to characterize the subspace of instances $\{\mathbf{a}\}$ for which the fixed set of partitions $\{\mathbf{s}^l\}_{l=1, \dots, P}$, are perfect, i.e. $E(\mathbf{s}^l) = 0 \forall l$. To achieve this we define the energy in the instance space as

$$\mathcal{H}(\mathbf{a}) = \sum_{l=1}^P \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N a_j s_j^l \right)^2 \quad (3)$$

so that the P partitions are perfect only if $\mathcal{H} = 0$. Henceforth we will assume that P increases linearly with N , i.e. $P = \alpha N$. In addition, we will assume that the components s_j^l are statistically independent random variables drawn from the probability distribution

$$\mathcal{P}(s_j^l) = \frac{1}{2} \left(1 + \frac{m}{\sqrt{N}} \right) \delta(s_j^l - 1) + \frac{1}{2} \left(1 - \frac{m}{\sqrt{N}} \right) \delta(s_j^l + 1) \quad (4)$$

where the weights of the Dirac delta functions are chosen so that $\langle s_j^l \rangle = m/\sqrt{N}$. The motivation for this choice is twofold. First, the exhaustive search in the Ising configuration space for $N \leq 26$ as well as the analytical solution of the linear relaxation of the NPP indicate that the average difference between the cardinalities of sets \mathcal{A}_1 and \mathcal{A}_2 ,

$$\hat{m} = \sum_{j=1}^N s_j \quad (5)$$

scales like $N^{1/2}$ for large N [13]. Second, the scaling of the bias in the distribution (4) yields a non-trivial thermodynamic limit, $N \rightarrow \infty$, for the average free-energy density associated with the Hamiltonian (3). Of course, for any realistic instance \mathbf{a} the different components s_i of a perfect partition are probably not statistically independent, i.e. the perfect partition \mathbf{s} is not random. Restricting the analysis to random partitions is a concession we must make in order to tackle the problem analytically. A similar arbitrary assumption is made in the ‘direct’ approach where the components a_i are chosen as random independent variables [9, 11, 12]. However, in the concluding section we will comment on the relevance of our results to non-random partitions as well.

In this paper we will apply standard statistical mechanics techniques to study analytically the ground states of the Hamiltonian (3). We concentrate our analysis on the zero-energy

instances (i.e. instances for which perfect random partitions exist) only, since the properties of the non-zero energy instances depend strongly on the rather arbitrary choice of the energy (3). Moreover, perfect random partitions are important from a practical viewpoint as they may have code-breaking implications [14] and so it may be of interest to estimate the maximum number of those partitions that can be encoded in an arbitrary instance, as well as to characterize those instances that maximize the number of coded perfect random partitions.

The rest of this paper is organized in the following way. In section 2 we use the replica method to evaluate the average free-energy density in the thermodynamic limit and to derive the replica-symmetric order parameters that describe the statistical properties of the instance space. In particular, we show that there is a critical value, $\alpha_c(m)N$, which limits the number of perfect random partitions. Also in that section, we study the stability of the replica-symmetric solution with respect to replica symmetry breaking and show that the zero-energy instances can reliably be described by the replica-symmetric order parameters. In section 3 we calculate the probability density for a given entry, say a_k , to have value a . This is achieved by integrating the joint probability distribution (the Gibbs distribution) over all entries except a_k . Finally, in section 4 we present some concluding remarks.

2. Replica approach

Following the standard prescription of performing quenched averages on extensive quantities only [2], we define the average free-energy density f as

$$-\beta f = \lim_{N \rightarrow \infty} \frac{1}{N} \langle \ln Z \rangle \quad (6)$$

where

$$Z = \int_0^\infty \prod_j da_j \delta \left(R - \frac{1}{N} \sum_j a_j \right) e^{-\beta \mathcal{H}(\mathbf{a})} \quad (7)$$

is the partition function and $\beta = 1/T$ is the inverse temperature. Taking the limit $T \rightarrow 0$ in equation (7) ensures that only the instances that minimize $\mathcal{H}(\mathbf{a})$ will contribute to Z . Here $\langle \dots \rangle$ denotes the average over the partitions s^l ($l = 1, \dots, P$). The constraint on the mean of the instance vector is needed in order to exclude the trivial solution $\mathbf{a} = 0$. Fortunately, the arbitrary parameter R does not play a relevant role in the theory, giving only the scale of the order parameters of the model.

At this point we note that, in the neural networks context, this problem is identical to that of determining a vector of N neural couplings \mathbf{a} that is orthogonal to all P binary patterns s^l , i.e. $\mathbf{a} \cdot s^l = 0 \forall l$, which contrasts with the usual stability requirement in that context, namely, $\mathbf{a} \cdot s^l > 0 \forall l$ [6–8]. Interestingly, the problem of minimizing the Hamiltonian (3) subject to a fixed mean R but with a_i limited to the simplex $[0, 1]$ was investigated so as to improve the storage performance of analogue attractor neural networks by reducing the interference between stored patterns [15]. In addition, the disordered model considered here may be viewed as a variant of the model of replicators with random interactions studied by Diederich and Oppen [16] (see also [17]) in which the fitness functional is given by

$$\mathcal{H}(\mathbf{a}) = \sum_{ij} J_{ij} a_i a_j \quad (8)$$

with the couplings given by the Hebb rule $J_{ij} = \frac{1}{N} \sum_l^P s_i^l s_j^l$ [3]. In this sense, the statistical mechanics analysis of the Hamiltonian (3) does not present any new technical difficulty, and so we will only sketch it in the following.

As usual, the quenched average in equation (6) is evaluated through the replica method: using the identity

$$\langle \ln Z \rangle = \lim_{n \rightarrow 0} \frac{1}{n} \ln \langle Z^n \rangle \quad (9)$$

we first evaluate $\langle Z^n \rangle$ for *integer* n and then analytically continue to $n = 0$. Using standard techniques [7] we obtain, in the thermodynamic limit

$$-\beta f = \lim_{n \rightarrow 0} \frac{1}{n} \text{extr} \left\{ \frac{R^2}{2} \sum_{\rho} Q_{\rho} \hat{Q}_{\rho} - R^2 \sum_{\rho < \delta} q_{\rho\delta} \hat{q}_{\rho\delta} + R \sum_{\rho} \hat{R}_{\rho} \right. \\ \left. + \alpha \ln G_1(q_{\rho\delta}, Q_{\rho}) + \ln G_2(\hat{q}_{\rho\delta}, \hat{R}_{\rho}, \hat{Q}_{\rho}) \right\} \quad (10)$$

where

$$G_1 = \int_{-\infty}^{\infty} \prod_{\rho} \frac{d\tilde{x}_{\rho}}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \sum_{\rho} (1 + 2\beta R^2 Q_{\rho}) \tilde{x}_{\rho}^2 - 2\beta R^2 \sum_{\rho < \delta} \tilde{x}_{\rho} \tilde{x}_{\delta} q_{\rho\delta} \right. \\ \left. + im \sqrt{2\beta R^2} \sum_{\rho} \tilde{x}_{\rho} \right] \quad (11)$$

and

$$G_2 = \int_0^{\infty} \prod_{\rho} da_{\rho} \exp \left(-\frac{1}{2} \sum_{\rho} \hat{Q}_{\rho} a_{\rho}^2 + \sum_{\rho < \delta} \hat{q}_{\rho\delta} a_{\rho} a_{\delta} - \sum_{\rho} \hat{R}_{\rho} a_{\rho} \right). \quad (12)$$

The extremum in equation (10) is taken over all saddle-point parameters $(\hat{q}_{\rho\delta}, \hat{R}_{\rho}, \hat{Q}_{\rho}, q_{\rho\delta}, Q_{\rho})$. The physical order parameters

$$q_{\rho\delta} = \left\langle \frac{1}{NR^2} \sum_{i=1}^N \langle a_i^{\rho} \rangle_T \langle a_i^{\delta} \rangle_T \right\rangle \quad \rho < \delta \quad (13)$$

and

$$Q_{\rho} = \left\langle \frac{1}{NR^2} \sum_{i=1}^N \langle (a_i^{\rho})^2 \rangle_T \right\rangle \quad (14)$$

measure the overlap between a pair of different equilibrium instances \mathbf{a}^{ρ} and \mathbf{a}^{δ} , and the overlap of an equilibrium instance \mathbf{a}^{ρ} with itself, respectively. Here, $\langle \cdot \cdot \rangle_T$ denotes a thermal average.

2.1. Replica-symmetric solution

To proceed further we make the replica symmetric ansatz, i.e. we assume that the values of the order parameters are independent of their replica indices

$$\begin{aligned} q_{\rho\delta} = q & \quad \text{and} & \quad \hat{q}_{\rho\delta} = \hat{q} & \quad \forall \rho < \delta \\ Q_{\rho} = Q & \quad \text{and} & \quad \hat{Q}_{\rho} = \hat{Q} & \quad \forall \rho \\ \hat{R}_{\rho} = \hat{R} & \quad \forall \rho. \end{aligned} \quad (15)$$

Evaluation of equations (11) and (12) with this ansatz is straightforward. In order to write the replica symmetric average free-energy density it is convenient to introduce the new variables

$$\eta = R^2(\hat{Q} + \hat{q}) \quad \tau = \frac{\hat{R}}{\sqrt{2(\hat{Q} + \hat{q})}} \quad \theta = \frac{\hat{q}}{2(\hat{Q} + \hat{q})} \quad (16)$$

and rescale the temperature $\beta' = \beta R^2$ so that

$$\begin{aligned} -\frac{\beta'}{R^2} f_{rs} &= \frac{1}{2} \ln \left(\frac{\pi R^2}{2} \right) + \frac{1}{2} \eta [Q - 2\theta(Q - q)] + \tau \sqrt{2\eta} \\ &- \frac{\alpha}{2} \ln [1 + 2\beta'(Q - q)] - \beta' \alpha \frac{m^2 + q}{1 + 2\beta'(Q - q)} \\ &- \frac{1}{2} \ln \eta + \int_{-\infty}^{\infty} Dz \ln (e^{\Xi_z^2} \operatorname{erfc} \Xi_z) \end{aligned} \quad (17)$$

where

$$\Xi_z = \tau + z\theta^{1/2} \quad (18)$$

and

$$Dz = \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \quad (19)$$

is the Gaussian measure. Thus it is clear from equation (17) that the parameter R yields the scales of the temperature and free energy, not affecting in any significant way the physical, replica-symmetric order parameters

$$q = \left\langle \frac{1}{NR^2} \sum_{i=1}^N \langle a_i \rangle_T^2 \right\rangle \quad (20)$$

and

$$Q = \left\langle \frac{1}{NR^2} \sum_{i=1}^N \langle a_i^2 \rangle_T \right\rangle. \quad (21)$$

The replica-symmetric average energy density $\epsilon_{rs} = \partial (\beta f_{rs}) / \partial \beta$ is given by

$$\epsilon_{rs} / \alpha R^2 = \frac{q + m^2}{[1 + 2\beta'(Q - q)]^2} + \frac{Q - q}{1 + 2\beta'(Q - q)} \quad (22)$$

which vanishes in the limit $\beta' \rightarrow \infty$ provided that $q < Q$. As justified in section 1 we will focus on this limit only. After some algebra, the saddle-point equations in this limit are written as

$$\theta = \frac{q + m^2}{2(Q - q)} \quad (23)$$

$$\eta = \frac{\alpha}{Q - q} \quad (24)$$

$$\tau = \sqrt{\frac{Q - q}{2\alpha}} \left(1 - \alpha + \frac{\alpha m^2}{Q - q} \right) \quad (25)$$

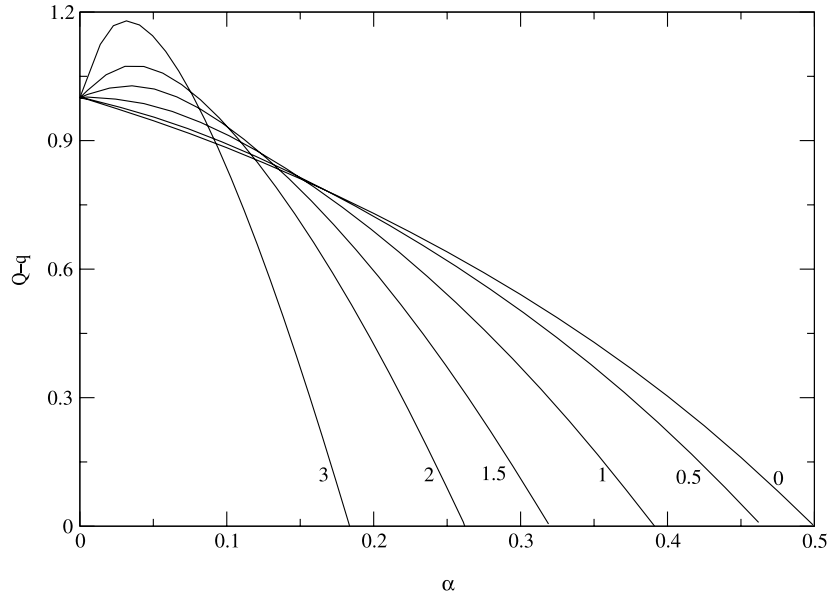


Figure 1. Average variance $Q - q$ of the zero-energy instance entries as a function of α for $m = 0, 0.5, 1, 1.5, 2$ and 3 . The value of α at which the variance vanishes (α_c) gives an upper bound to the number of perfect random partitions.

$$\sqrt{2\eta} = -2\tau + \frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} Dz \frac{\exp(-\Xi_z^2)}{\operatorname{erfc} \Xi_z} \quad (26)$$

$$\eta(Q - q) = 1 - \frac{1}{\sqrt{\pi\theta}} \int_{-\infty}^{\infty} Dz z \frac{\exp(-\Xi_z^2)}{\operatorname{erfc} \Xi_z} \quad (27)$$

with Ξ_z given by equation (18). In general, these equations can be solved numerically only. In figures 1 and 2 we present the dependence of $Q - q$ and q , respectively, on α for different values of m . For $\alpha = 0$ we find $Q - q = q = 1$, $\theta = (1 + m^2)/2$, and $\eta = 0$, while τ diverges like $1/\sqrt{2\alpha}$. According to the physical meaning of the order parameters given in equations (20) and (21), the difference $Q - q$ measures the average variance of the zero-energy instance entries: the larger this difference, the larger the dispersion of the instance entries. Interestingly, for fixed $m > 0$ this variance reaches its maximum for $\alpha > 0$. The divergence of the order parameters Q and q (and of their difference as well) for $m \rightarrow \infty$ and $\alpha \neq 0$ is expected, since in order for an extremely unbalanced partition to become a perfect partition there must exist some very large entries to compensate for the much larger number of entries in one of the sets. Moreover, we observe from figure 1 that for fixed m there is a value of $\alpha = \alpha_c$ at which the overlap between two zero-energy instances q equals its maximal value Q . This result signals the shrinking of the zero-energy instance subspace to instances differing from a microscopic number of entries a_j only. Besides, it gives the limit of existence of the solutions with zero-energy: for $\alpha > \alpha_c$ there are no zero-energy instances. Taking the limit $q \rightarrow Q$ in the saddle-point equations (23)–(27) yields

$$\alpha_c = 1 - \int_{-\Delta}^{\infty} Dz \quad (28)$$

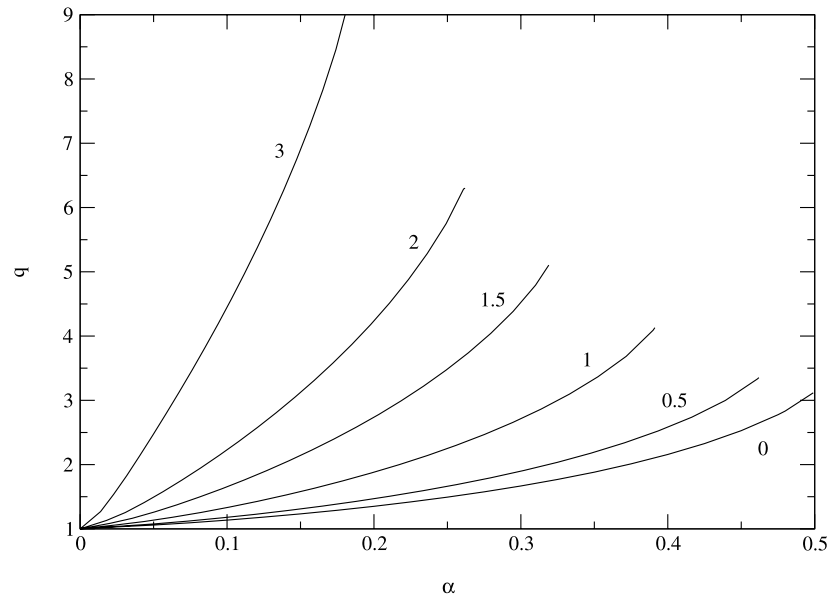


Figure 2. Average overlap q between two different zero-energy instances as a function of α for $m = 0, 0.5, 1, 1.5, 2$ and 3 . The curves end at $\alpha = \alpha_c$.

where

$$\Delta = \left(\frac{\alpha_c m^4}{Q_c + m^2} \right)^{1/2} \quad (29)$$

is the solution of

$$\frac{\Delta}{m^2} = \int_{-\Delta}^{\infty} Dz (z + \Delta) - \Delta. \quad (30)$$

Here Q_c denotes the order parameter Q evaluated at α_c . For $m = 0$ we can solve these equations analytically: we find that Δ vanishes like $m^2/\sqrt{2\pi}$ and so $\alpha_c = \frac{1}{2}$ and $Q_c = \pi$. In figures 3 and 4 we show α_c and Q_c , respectively, as functions of m . The dependence of α_c on m corroborates the statement made in section 1 that perfect random partitions exist only if the difference between the cardinalities of the two sets scales as $mN^{1/2}$, since α_c vanishes very rapidly with increasing m .

2.2. Stability analysis

The condition for local stability of the replica-symmetric saddle point is given by [7]

$$\alpha \gamma_1 \gamma_2 \leq 1 \quad (31)$$

where γ_1 and γ_2 are the transverse eigenvalues [18] of the matrices of second derivatives of G_1 and G_2 with respect to $q_{\rho\delta}$ and $\hat{q}_{\rho\delta}$, respectively, evaluated at the replica-symmetric saddle-point. After some algebra we find that condition (31) reduces to

$$\alpha [\eta(Q - q)]^{-2} \int_{-\infty}^{\infty} Dz (\bar{a}^2 - \bar{a}^2)^2 \leq 1 \quad (32)$$

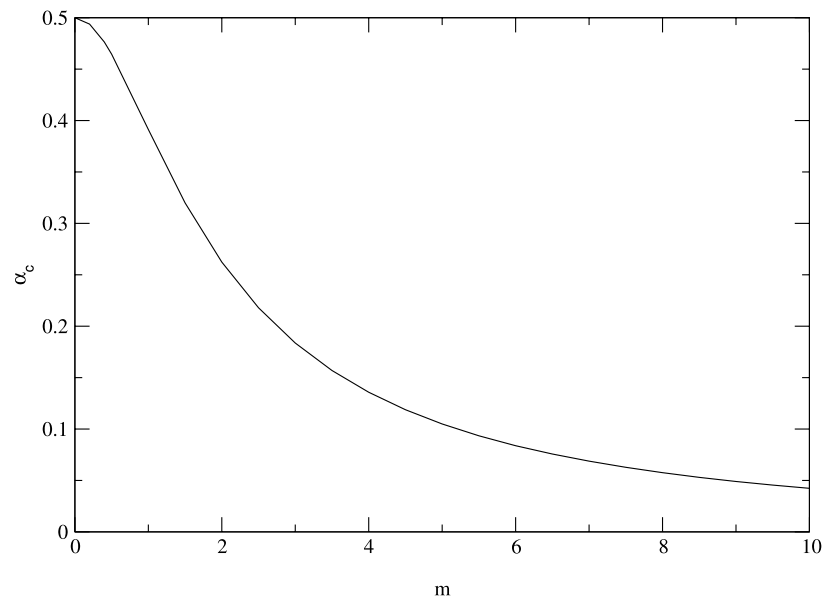


Figure 3. Instance independent upper bound to the number of perfect random partitions α_c as a function of the parameter m which measures the unbalance of the partitions. For balanced partitions ($m = 0$) we find $\alpha_c = \frac{1}{2}$.

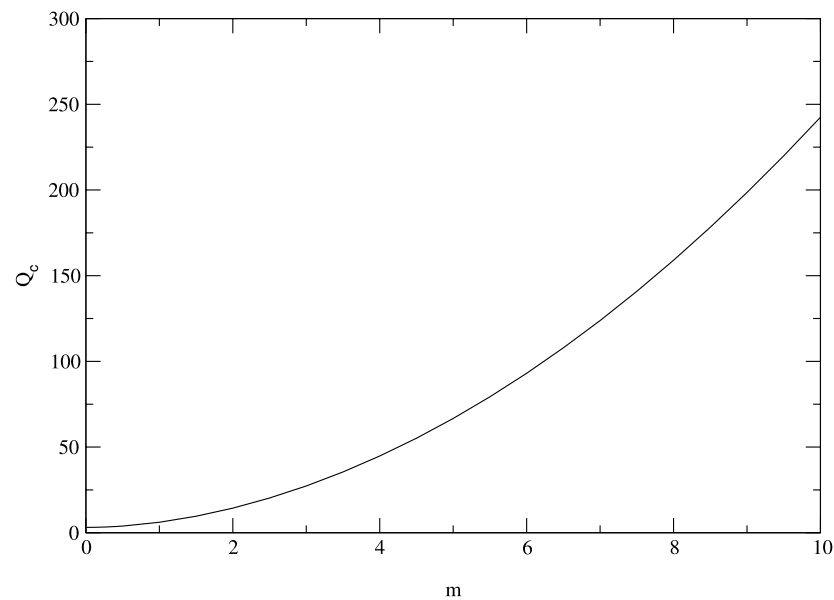


Figure 4. Average overlap of a zero-energy instance with itself calculated at α_c as a function of m . For balanced partitions ($m = 0$) we find $Q_c = \pi$.

where

$$\overline{a^n} = \frac{\int_0^\infty da a^n \exp(-\frac{1}{2}a^2 - a\sqrt{2}\Xi_z)}{\int_0^\infty da \exp(-\frac{1}{2}a^2 - a\sqrt{2}\Xi_z)}. \quad (33)$$

Taking the limit $q \rightarrow Q$ we can easily show that

$$\eta(Q - q) \rightarrow \alpha_c \tag{34}$$

and

$$\int_{-\infty}^{\infty} Dz \left(\overline{a^2} - \bar{a}^2 \right)^2 \rightarrow \alpha_c \tag{35}$$

with α_c given by equation (28), so that the left-hand side of equation (32) equals 1 at $\alpha = \alpha_c$. In addition, we have verified numerically that this stability condition is always satisfied for $\alpha < \alpha_c$.

3. Probability distribution of entries

The traditional probabilistic approach to study optimization problems introduces a probability distribution over the space of instances. The main objection to this approach is that one rarely knows what probability distribution is realistic. In the NPP, for instance, it is usually assumed that the entries a_k are statistically independent random variables distributed uniformly in the unit interval [9–11]. In this section we calculate analytically the distribution of probability that a certain entry, say a_k , of a zero-energy instance assumes the value a , defined by

$$\begin{aligned} \mathcal{P}_k(a) &= \lim_{\beta \rightarrow \infty} \langle \langle \delta(a_k - a) \rangle \rangle_T \\ &= \lim_{\beta \rightarrow \infty} \left\langle \frac{1}{Z} \int_0^{\infty} \prod_j da_j \delta \left(R - \frac{1}{N} \sum_j a_j \right) \delta(a_k - a) e^{-\beta \mathcal{H}(a)} \right\rangle \end{aligned} \tag{36}$$

where Z and \mathcal{H} are given by equations (7) and (3), respectively. As all entries are equivalent we can write $\mathcal{P}_k(a) = \mathcal{P}(a) \forall k$. Hence to evaluate equation (36) we introduce the auxiliary energy

$$\mathcal{H}_{\text{aux}}(a) = \mathcal{H}(a) + h \sum_k \delta(a_k - a) \tag{37}$$

so that

$$\mathcal{P}(a) = - \lim_{\beta \rightarrow \infty} \frac{1}{N\beta} \left. \frac{\partial \langle \ln Z_{\text{aux}} \rangle}{\partial h} \right|_{h=0} \tag{38}$$

where Z_{aux} is the partition function (7) with \mathcal{H} replaced by \mathcal{H}_{aux} . Of course, we note that the entries (a_1, \dots, a_N) are not statistically independent and their joint probability distribution is simply the Gibbs probability distribution

$$\mathcal{W}(a) = \frac{1}{Z} \exp[-\beta \mathcal{H}(a)]. \tag{39}$$

As expected, equation (36) is recovered by integrating this joint distribution over a_j for all $j \neq k$ and then setting $a_k = a$. Using equation (38) the calculations needed to evaluate $\mathcal{P}(a)$ become analogous to those used in the evaluation of the free-energy density (10). Within the replica-symmetric framework and in the limit $\beta \rightarrow \infty$ with $q < Q$ the final result is

$$\mathcal{P}(a) = \sqrt{\frac{2\eta}{\pi R^2}} \int_{-\infty}^{\infty} Dz \frac{1}{\text{erfc } \Xi_z} \exp \left[-\frac{\eta}{2R^2} a^2 - a \Xi_z \left(\frac{2\eta}{R^2} \right)^{1/2} - \Xi_z^2 \right] \tag{40}$$

for $a \geq 0$. Setting $\alpha = 0$ it reduces to

$$\mathcal{P}(a) = \frac{1}{R} \exp\left(-\frac{a}{R}\right) \quad a \geq 0. \tag{41}$$

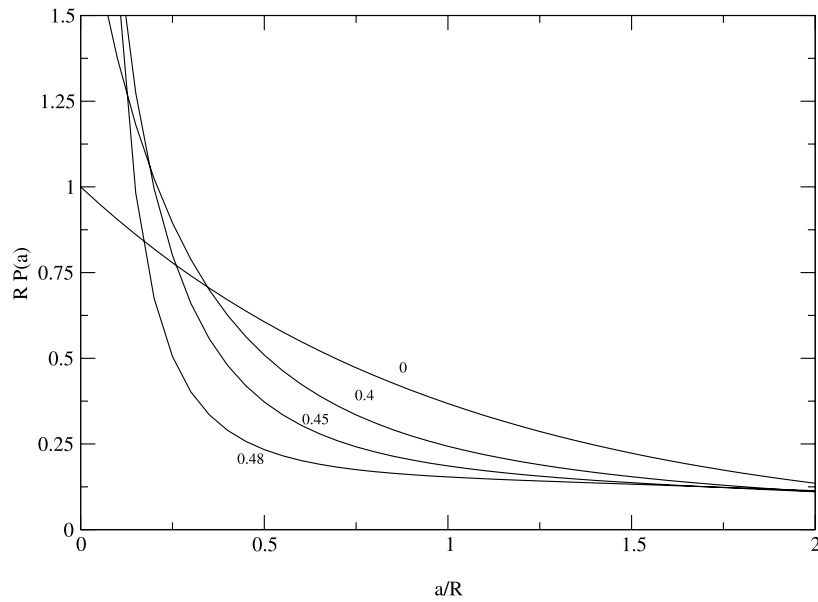


Figure 5. Probability distribution of the zero-energy instance entries for balanced random partitions ($m = 0$) and $\alpha = 0, 0.4, 0.45$ and 0.48 .

To handle a possible singularity in the limit $q \rightarrow Q$ it is more convenient to consider instead the cumulative distribution function defined by

$$\begin{aligned} \mathcal{C}(a) &= \int_0^a da' \mathcal{P}(a') \\ &= 1 - \int_{-\infty}^{\infty} Dz \frac{\operatorname{erfc}[\Xi_z + a(\eta/2R^2)^{1/2}]}{\operatorname{erfc} \Xi_z}. \end{aligned} \quad (42)$$

Taking the limit $q \rightarrow Q$ yields

$$\mathcal{C}_c(a) = 1 - \frac{1}{2} \operatorname{erfc} \left[\frac{\Delta}{\sqrt{2}} \left(1 + \frac{a}{Rm^2} \right) \right] \quad (43)$$

where α_c and Δ are given by equations (28) and (30), respectively. The interesting feature of this distribution is that $\mathcal{C}_c(0)$ is non-zero, thus indicating that the probability distribution (40) evaluated at $\alpha = \alpha_c$ has a delta peak in $a = 0$. Explicitly,

$$\mathcal{P}_c(a) = \mathcal{C}_c(0) \delta(a) + \frac{d\mathcal{C}_c}{da} \quad a \geq 0 \quad (44)$$

which for $m = 0$ reduces to

$$\mathcal{P}_c(a) = \frac{1}{2} \delta(a) + \frac{1}{\sqrt{4\pi R^2}} \exp\left(-\frac{a^2}{4R^2}\right) \quad a \geq 0. \quad (45)$$

In figure 5 we show the probability distribution function $\mathcal{P}(a)$ for $m = 0$ and several values of α . For $\alpha < 0.3$ the distribution is similar to the exponential distribution (41) obtained for $\alpha = 0$. We note that the delta peak at $a = 0$ appears only for $\alpha = \alpha_c$.

4. Conclusion

While the traditional approach of computer science to the validation of combinatorial search algorithms focuses almost exclusively on the instance space (e.g. the *worst-case* analysis is basically a search for instances that give the poorest performance of the algorithm under study [1]), the statistical mechanics approach has concentrated mainly on the configuration space, with the instances being drawn from arbitrary probability distributions [2]. Building on the work of Gardner on neural networks [6, 7], we illustrate in this paper the usefulness of equilibrium statistical mechanics tools to investigate the statistical properties of the instance space as well. For the optimization problem we have considered, namely, the number partitioning problem, we have searched the instance space for the best (easiest) instances to show that there is a maximum number of uncorrelated perfect partitions, $\alpha_c(m)N$ (see figure 3). In particular, for balanced partitions ($m = 0$) we find $\alpha_c(0) = \frac{1}{2}$. Clearly, this result yields an upper bound to the number of perfect random partitions that can be found for any arbitrary instance. Moreover, in the case of unbalanced random partitions, we have shown that zero-energy instances exist only if the cardinalities difference $\sum_i s_i$ scales like $mN^{1/2}$. These results have obvious relevance to cryptography [14] since they yield upper bounds to the number of uncorrelated coded binary messages (i.e. the perfect random partitions) that can be encoded in an arbitrary instance. Of particular interest in this context is the probability distribution of the zero-energy instance entries (see figure 5) since its shape may give a clue to the number of encoded patterns (α) in the instance.

To conclude, some remarks on the relevance of our main result, namely, the upper bound $\alpha_c(m)N$, to non-random partitions are in order. We begin by noting that it is more difficult to find solutions (i.e. zero-energy instances) \mathbf{a} to the equations $\mathbf{a} \cdot \mathbf{s}^l = 0 \forall l$, in the case that the partitions \mathbf{s}^l are random than in the case of correlated partitions. Hence since for $\alpha \leq \alpha_c(m)$ we can guarantee the existence of those solutions for random partitions, they must exist for correlated partitions as well. This leads to the conclusion that $\alpha_c(m)N$ actually yields a *lower* bound to the maximum number of perfect partitions that can be found for any arbitrary instance. Of course, this bound becomes exact in the case of random partitions. Therefore, $\alpha_c(m)N$ can be viewed as a general and robust characteristic of the number partitioning problem, in the sense that it does not depend on arbitrary assumptions about the instance realizations. As in the neural networks case, the instance space analysis proposed in this paper can be extended to virtually all optimization problems.

Acknowledgments

We thank Pablo Moscato for illuminating conversations. The work of JFF was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). FFF is supported by FAPESP.

References

- [1] Garey M R and Johnson D S 1979 *Computers and Intractability: a Guide to the Theory of NP-Completeness* (San Francisco, CA: Freeman)
- [2] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
- [3] Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
- [4] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30
- [5] Hopfield J J and Tank D W 1985 *Biol. Cybern.* **52** 141
- [6] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
- [7] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271

- [8] Watkin T L H, Rau A and Biehl M 1993 *Rev. Mod. Phys.* **65** 499
- [9] Karmakar N, Karp R M, Lueker G S and Odlyzko A M 1986 *J. Appl. Prob.* **23** 626
- [10] Johnson D S, Aragon C R, McGeoch L A and Schevon C 1991 *Oper. Res.* **39** 378
- [11] Ferreira F F and Fontanari J F 1998 *J. Phys. A: Math. Gen.* **31** 3417
- [12] Mertens S 1998 *Phys. Rev. Lett.* **81** 4281
- [13] Ferreira F F and Fontanari J F 1999 *Physica A* **269** 58
- [14] Shamir A 1979 *Proc. 11th Ann. ACM Symp. on Theory of Computing. Association for Computing Machinery (New York)* pp 118–29
- [15] Brunel N and Zecchina R 1994 *Phys. Rev. E* **49** R1823
- [16] Diederich S and Oppen M 1989 *Phys. Rev. A* **39** 4333
- [17] Biscari P and Parisi G 1995 *J. Phys. A: Math. Gen.* **28** 4697
- [18] Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983