Sign-constrained synapses and biased patterns in neural networks

1993 J. Phys. A: Math. Gen. 26 6195

(http://iopscience.iop.org/0305-4470/26/22/020)

J. Phys. A: Math. Gen. 26 (1993) 6195–6203. Printed in the UK

Sign-constrained synapses and biased patterns in neural networks

R Raju Viswanathan

Centre for Artificial Intelligence and Robotics, Raj Bhavan Circle, Bangalore 560001, India

Received 20 November 1992, in final form 16 June 1993

Abstract. The storage of biased patterns is examined in neural networks with sign-constrained synapses. Every neuron has outgoing synapses which are either all inhibitory or all excitatory. For random patterns stored in such networks, it is known that the presence of a discrete gauge symmetry makes the maximal storage capacity independent of the proportion of excitatory neurons to inhibitory neurons. When the stored patterns are biased, however, this discrete gauge symmetry is broken, with the result that the maximal capacity depends on the proportion of excitatory neurons to inhibitory ones. The dependence of the capacity on the fraction of excitatory neurons in the network, f, is calculated using the space of interactions approach. It is found that the storage capacity is maximal at $f=0.5$; this result is true regardless of the particular value of the bias in the stored patterns. The significance of this result in the neurophysiological context is discussed.

1. Introduction

Neural networks as models of associative memory have been considered seriously in recent years as viable models for understanding the organization of memories in biological systems. In addition, a detailed knowledge of the functioning of such networks would prove helpful in the construction of artificial neural networks with a good amount of control over how the information is stored and recovered associatively.

In this paper, we shall consider a class of neural networks which model faithfully one of the observed properties of neurons in human brains, namely, that neurons in brains obey Dale's law. This law simply says that neurons in brains have outgoing synapses which are either all excitatory or all inhibitory. In the context of artificial neural networks, this means that the synaptic couplings w_{ij}, describing the influence of neuron j on neuron i, have the same sign for all i, for fixed j. Such sign-constrained networks have been studied in the context of storing random patterns [1]. By writing down an appropriate constraint in the space of couplings w, one can evaluate the storage capacity of a sign-constrained network. In the case of unbiased patterns, there is a discrete gauge symmetry in the space of couplings which lets one choose the sign associated with every neuron arbitrarily. The capacity is independent of the associated sign distribution of the neurons in the network.

This discrete gauge symmetry is broken, however, when the stored patterns are no longer random, but are biased. In that case, it is reasonable to expect that the number of fixed points of the network depends on the distribution of inhibitory and excitatory neurons present. As we shall demonstrate, this is in fact the case. In the next section, we shall write down an appropriate measure in the space of couplings that will let us

0305-4470/93/226195+09$07.50

determine the maximal capacity of biased patterns in sign-constrained networks. The saddle point evaluation of the entropy in interaction space (in the large N limit) gives a set of equations in the replica symmetric limit that must be solved to evaluate the capacity. The numerical solution of these equations is presented for various bias values of the stored patterns and for different proportions of inhibitory and excitatory neurons in the network.

In the last section, we discuss the significance of our results, which we believe are of some relevance to the organization of neurons in brains, and end with some speculations.

2. Sign-constrained synapses

We shall consider a neural network consisting of N neurons in the large N limit, with $p=\alpha N$ patterns stored as fixed points of the network. The dynamical rule for the updating of the states $s_i=\pm 1$ of neurons in the network is the standard one,

$$s_i(t+1)=\operatorname{sgn}\left(\sum_{j\neq i} w_{ij}s_j(t)\right).$$

It is apparent that the fixed points of this dynamical rule satisfy

$$s_i\sum_{j\neq i} w_{ij}s_j>0.$$

For the storage of p patterns s^μ, $\mu=1,\ldots,p$, a stronger characterization of fixed points is provided by

$$\gamma_i^\mu\equiv\frac{1}{\sqrt{N}}s_i^\mu\sum_{j\neq i} w_{ij}s_j^\mu>\kappa \tag{1}$$

which ensures a finite radius of attraction for $\kappa>0$ [2, 3, 4, 5, 6, 7]. We can choose to normalize the ws by enforcing the 'spherical' constraint

$$\sum_j w_{ij}^2=N$$

for each row of the matrix of couplings. The signs of all the outgoing synapses from neuron j can be constrained to have the same sign $g_j=\pm 1$ by requiring that $g_jw_{ij}>0$, for every i. Furthermore, we shall assume that a fraction of the neurons f are excitatory; this means that

$$\sum_j g_j=N(f-(1-f))=rN \tag{2}$$

with $r\equiv 2f-1$.

In the space of couplings w_{ij}, the partition function can be written, up to a normalization factor, in the form [2, 3]

$$Z=\int_{-\infty}^{\infty}\left(\prod_{i\neq j}\mathrm{d}w_{ij}\right)\rho[w]\,\mathrm{e}^{-\beta H[w]} \tag{3}$$

where $\rho[w]$ is an appropriate measure in the space of couplings encoding the above constraints, and with the Hamiltonian

$$H[w]=\sum_{i=1}^{N}\sum_{\mu=1}^{p}\theta(\kappa-\gamma_i^{\mu}[w]) \tag{4}$$

measuring the number of sites s which are not fixed points of the above dynamical evolution. In the limit when the inverse 'temperature' goes to infinity, $\beta\to\infty$, the partition function Z is just the fractional volume of zero energy states, $\Omega(0)$; therefore the logarithm of Z in this limit measures the volumetric entropy of the zero energy state, $S=\ln\Omega(0)$. The number of fixed points is maximized when the fractional volume of the zero energy states shrinks to zero.

Without loss of generality, in order to strictly impose Dale's law, we shall assume that the first $N_1=fN$ neurons in the network are constrained to have only excitatory ($g_j=+1$) synapses, while the remaining $N_2=(1-f)N$ neurons have purely inhibitory ($g_j=-1$) ones. The measure $\rho[w]$ which encodes the above constraints can be written in the form.

$$\rho[w]=\prod_i \delta\left(\sum_j w_{ij}^2-N\right)\prod_{j=1}^{N_1}\theta(w_{ij})\prod_{j=N_1+1}^{N}\theta(-w_{ij}). \tag{5}$$

In the $\beta\to\infty$ limit, the term $e^{-\beta H}$ in Z reduces to a produce of step functions

$$e^{-\beta H}\to\prod_i\prod_{\mu=1}^{p}\theta(\gamma_i^{\mu}-\kappa).$$

Since the constraints do not mix rows of the matrix w, the logarithm of the partition function reduces to a sum,

$$S=\sum_{i=1}^{N}\ln\Omega_i \tag{6}$$

with each Ω_i coming from one of the factors in Z. The quenched average entropy is given by averaging this expression over all possible different patterns $\{s_i^{\mu}\}$. We shall assume that all the stored patterns have the same bias m; accordingly, the probability distribution of each s is

$$p(s)=\frac{1+m}{2}\delta_{s,1}+\frac{1-m}{2}\delta_{s,-1}.$$

Introducing integral representations for the step functions and the delta functions in the usual manner, one can evaluate the quenched average above by using the standard replica trick,

$$\langle\langle\ln\Omega_i\rangle\rangle=\lim_{n\to 0}\frac{\langle\langle\Omega_i^n\rangle\rangle-1}{n}.$$

Now we have

$$\langle\langle(\Omega_i)^n\rangle\rangle=\int\left(\prod_{a=1}^{n}\prod_{j=1}^{N}\mathrm{d}w_{ij}^a\right)\prod_{a=1}^{n}\delta\left(\sum_j w_{ij}^{a2}-N\right)$$
$$\times\prod_{a=1}^{n}\left(\prod_{j=1}^{N_1}\theta(w_{ij}^a)\prod_{j=N_1+1}^{N}\theta(-w_{ij}^a)\langle\langle\prod_{\mu=1}^{p}\theta(\gamma_i^{\mu}[w]-\kappa)\rangle\rangle\right). \tag{7}$$

In the usual manner, one can introduce integral representations for the delta and theta functions in the above expression. Specifically, we write

$$\theta(w^a g)=\int_0^\infty \mathrm{d}R^a \int_{-\infty}^{\infty} \frac{\mathrm{d}Z^a}{2\pi} \exp(\mathrm{i}Z^a(w^a g - R^a)) \tag{8}$$

and

$$\delta\left(\sum_j w_{ij}^{a2} - N\right)=\int_{-\infty}^{\infty} \frac{\mathrm{d}E^a}{2\pi} \exp\left(\mathrm{i}E^a\left(\sum_j (w_{ij}^a)^2 - N\right)\right) \tag{9}$$

where a is a replica index denoting the ath replica, and takes values from 1 to n. The product over j factorizes into two factors due to the sign constraints on the synapses, each of these factors occurring N_1 and $N-N_1$ times, respectively. In the limit $N\to\infty$, after performing the average over patterns, we are left with an expression of the form

$$\langle\langle \Omega_i^n \rangle\rangle = \int \left(\prod_a \frac{\mathrm{d}E^a}{2\pi}\right) \int \left(\prod_{a<b} \frac{\mathrm{d}q_{ab}\,\mathrm{d}F_{ab}}{2\pi/N}\right) \int \left(\prod_a \frac{\mathrm{d}M^a\,\mathrm{d}H^a}{2\pi/\sqrt{N}}\right)$$
$$\times \exp\left[N\alpha G_1(q_{ab}, M_a) + NG_2(F_{ab}, E^a, H^a) + \mathrm{i}N \sum_{a<b} F_{ab} q_{ab}\right]. \tag{10}$$

We shall write down the expressions for G_1 and G_2 shortly. The N in the exponent comes from the factorization over the js. Here α is the normalized storage capacity p/N. The order parameters q_{ab} and M^a arise from averaging over the biased patterns, and F_{ab} and H^a are additional order parameters which serve to enforce the definitions

$$q_{ab}=\frac{1}{N}\sum_j w_j^a w_j^b \tag{11}$$

and

$$M^a=\frac{1}{\sqrt{N}}\sum_j w_j^a \tag{12}$$

of the order parameters q_{ab} and M^a, respectively, as delta-function constraints. They arise due to inserting unity in the forms

$$\int \mathrm{d}q_{ab} \frac{\mathrm{d}F_{ab}}{2\pi/N} \exp\left(\mathrm{i}F_{ab}\left(q_{ab} - \frac{1}{N}\sum_j w_j^a w_j^b\right)\right)$$

for $a<b$, and

$$\int \mathrm{d}M^a \frac{\mathrm{d}H^a}{2\pi/\sqrt{N}} \exp\left(\mathrm{i}H^a\left(M^a - \frac{1}{\sqrt{N}}\sum_j w_j^a\right)\right)$$

into the expression for $\langle\langle \Omega_i^n \rangle\rangle$.

At this point one can already see that the discrete gauge invariance alluded to earlier is broken. For the case of random ($m=0$) patterns, given a particular realization of the g_is, a change in the sign of one of these, say $g_j\to -g_j$, can be compensated by a corresponding change in the sign of the jth bit, s_j, of all the stored patterns [1]. However, in evaluating the quenched average over the patterns, the s_js are averaged over both possible values ± 1, weighted equally over both. Consequently $\langle\langle \Omega_i^n \rangle\rangle$ for random patterns is independent of the particular realization of the signs of the outgoing synapses

of each neuron $\{g_j\}$. This local gauge invariance is broken, however, when the stored patterns are biased, since the averaging procedure over the spins s_j now involves unequally weighting $s_j=+1$ and $s_j=-1$. One therefore expects, in general, that the entropy, and the storage capacity, would depend on the particular choice of the signs g_j.

To simply the evaluation of the integrals, it helps to assume the replica symmetric forms $F_{ab}=\mathrm{i}F$, $q_{ab}=q$, $M^a=M$, $E^a=\mathrm{i}E$ and $H^a=\mathrm{i}H$ for the various order parameters, for all a and for $b\neq a$. Then $G_1(q, M)$ is found to be [2]

$$G_1(q, M)=n\left[\frac{1+m}{2}\int_{-\infty}^{\infty}\mathrm{D}z\ln H(\tau_-)+\frac{1-m}{2}\int_{-\infty}^{\infty}\mathrm{D}z\ln H(\tau_+)\right] \tag{13}$$

where the quantities $\tau_\pm$ are defined as

$$\tau_\pm=\frac{1}{(1-q)^{1/2}}\left(\sqrt{q}z+\frac{\kappa\pm mM}{(1-m^2)^{1/2}}\right) \tag{14}$$

and $H(\tau)$ is the complementary error function $H(\tau)\equiv\int_\tau^\infty \mathrm{D}z$, with the Gaussian measure $\mathrm{D}z\equiv \mathrm{e}^{-z^2/2}\,\mathrm{d}z/\sqrt{2\pi}$.

The function G_2 is determined by the constraints in the measure in the space of couplings, and is given by

$$\mathrm{e}^{NG_2}=(L_+)^{N_1}(L_-)^{N-N_1} \tag{15}$$

where

$$L_\pm=\int_0^\infty\left(\prod_a \mathrm{d}R^a\right)\int_{-\infty}^{\infty}\left(\prod_a\frac{\mathrm{d}Z^a}{2\pi}\right)\int_{-\infty}^{\infty}\left(\prod_a \mathrm{d}w^a\right)\exp\left(\sum_{a<b}Fw^aw^b\right)$$
$$\times\exp\left(-\sum_a E[(w^a)^2-1]+\mathrm{i}\sum_a w^a(g_\pm Z^a-H)-\mathrm{i}\sum_a Z^aR^a\right) \tag{16}$$

where $g_\pm=\pm 1$. Changing variables to $Z^a=g^a-H^a$, the integral over the Z^a's in each of the factors above can be done and leaves us with the following integral over the R^a's:

$$\int_0^\infty\left(\prod_{a=1}^{n}\mathrm{d}R^a\right)\int_{-\infty}^{\infty}\mathrm{D}z\exp\left(-\frac{2E+F}{2}\sum_a(R^a)^2\right)$$
$$\times\exp\left(-z\sqrt{F}\sum_a R^a+gH\sum_a R^a+E\right). \tag{17}$$

This integral factorizes into separate integrals over each of the R^a's. By changing variables to

$$t=R(2E+F)^{1/2}-\frac{\sqrt{F}z+gH}{(2E+F)^{1/2}}$$

we obtain

$$\mathrm{e}^{NG_2}=\int_{-\infty}^{\infty}\mathrm{D}z(G_+(z))^{nN_1}\int_{-\infty}^{\infty}\mathrm{D}y\,(G_-(y))^{nN_2} \tag{18}$$

where

$$G_{\pm}(z)=e^{E}\frac{\sqrt{2\pi}}{\sqrt{2E+F}}\exp\left(\frac{1}{2}\frac{(\sqrt{F}z\pm H)^2}{2E+F}\right)\int_{-t_{\pm}}^{\infty}Dt. \tag{19}$$

Here $t_{\pm}\equiv(z\sqrt{F}\pm H)/(2E+F)^{1/2}$.

The final expression for G_2 in the limit of small n is

$$G_2=n\left[E+\frac{1}{2}\ln\left(\frac{2\pi}{2E+F}\right)+\int_{-\infty}^{\infty}Dz(f\log P_{+}+(1-f)\log P_{-})\right]. \tag{20}$$

The functions $P_{+}(z)$ and $P_{-}(z)$ here are defined by

$$P_{+}(z)=\exp\left(\frac{1}{2}\frac{(z\sqrt{F}+H)^2}{2E+F}\right)\int_{-t_{+}}^{\infty}Dt$$

and

$$P_{-}(z)=\exp\left(\frac{1}{2}\frac{(z\sqrt{F}-H)^2}{2E+F}\right)\int_{-t_{-}}^{\infty}Dt.$$

We note that for the case of unbiased patterns, the order parameter H does not exist, and the above expression for G_2 reduces exactly to that given in [1] as it should.

In the $N\rightarrow\infty$ limit, the expression (10) for the quantity $\langle\langle\Omega_i^n\rangle\rangle$ can be evaluated in the saddle point approximation, when the argument of the exponent there takes its minimum value. Defining

$$G\equiv\alpha G_1(q,M)+G_2(F,E,H)+\frac{n}{2}qF$$

the saddle point equations for G are

$$\frac{\partial G}{\partial q}=\frac{\partial G}{\partial F}=\frac{\partial G}{\partial E}=\frac{\partial G}{\partial M}=0 \tag{21}$$

and

$$\frac{\partial G}{\partial H}=0. \tag{22}$$

The replica method, in the limit $n\rightarrow 0$, then yields the following expression for the entropy S:

$$\frac{S}{N^2}=\alpha\int_{-\infty}^{\infty}Dz\left(\frac{1+m}{2}\ln H(\tau_{-})+\frac{1-m}{2}\ln H(\tau_{+})\right)+E+\frac{qF}{2}$$
$$+\frac{1}{2}\ln\left(\frac{2\pi}{2E+F}\right)+\int_{-\infty}^{\infty}Dz\,(f\log P_{+}+(1-f)\log P_{-}). \tag{23}$$

The maximum storage capacity is reached when this volumetric entropy diverges to negative infinity as the available volume in interaction space shrinks to zero. Correspondingly, $q\rightarrow 1$. We shall assume that, in this limit, $F/(2E+F)\rightarrow\infty$ and $H\rightarrow\infty$, with $(H/\sqrt{F})\rightarrow k$, where k is a finite constant. This assumption will be justified by the

consistency of our solution. In this limit, the saddle point relation (22) then gives the following equation for k:

$$k-f\int_{-\infty}^{-k}\mathrm{D}z(z+k)+(1-f)\int_{-\infty}^{k}\mathrm{D}z(z-k)=0. \tag{24}$$

The order parameter M is determined from the condition $\partial G/\partial M=0$ which gives the equation

$$(1+m)\int_{\frac{mM-\kappa}{(1-m^2)^{1/2}}}^{\infty}\mathrm{D}z\left(z+\frac{\kappa-mM}{(1-m^2)^{1/2}}\right)=(1-m)\int_{\frac{-mM-\kappa}{(1-m^2)^{1/2}}}^{\infty}\mathrm{D}z\left(z+\frac{\kappa+mM}{(1-m^2)^{1/2}}\right). \tag{25}$$

Correspondingly, the saddle point equations $\partial G/\partial E=\partial G/\partial F=0$ yield the equations

$$(2E+F)=\frac{1}{(1-q)}(1-A-kB) \tag{26}$$

and

$$F=\frac{(1-A-kB)^2}{(1+k^2-A)}\frac{1}{(1-q)^2} \tag{27}$$

for E and F, to leading order in $(1-q)^{-1}$. Here we have defined the quantities

$$A=f\int_{-\infty}^{-k}\mathrm{D}z(z+k)^2+(1-f)\int_{-\infty}^{k}\mathrm{D}z(z-k)^2 \tag{28}$$

and

$$B=-f\int_{-\infty}^{-k}\mathrm{D}z(z+k)+(1-f)\int_{-\infty}^{k}\mathrm{D}z(z-k). \tag{29}$$

The maximum storage capacity is then determined by the saddle point condition $\partial G/\partial q=0$, which gives

$$\alpha_{\max}\left[\frac{1+m}{2}\int_{s_-}^{\infty}\mathrm{D}z(z-s_-)^2+\frac{1-m}{2}\int_{s_+}^{\infty}\mathrm{D}z(z-s_+)^2\right]=\frac{(1-A-kB)^2}{(1+k^2-A)} \tag{30}$$

where we have defined

$$s_-=\frac{mM-\kappa}{(1-m^2)^{1/2}}$$

and

$$s_+=-\frac{mM+\kappa}{(1-m^2)^{1/2}}.$$

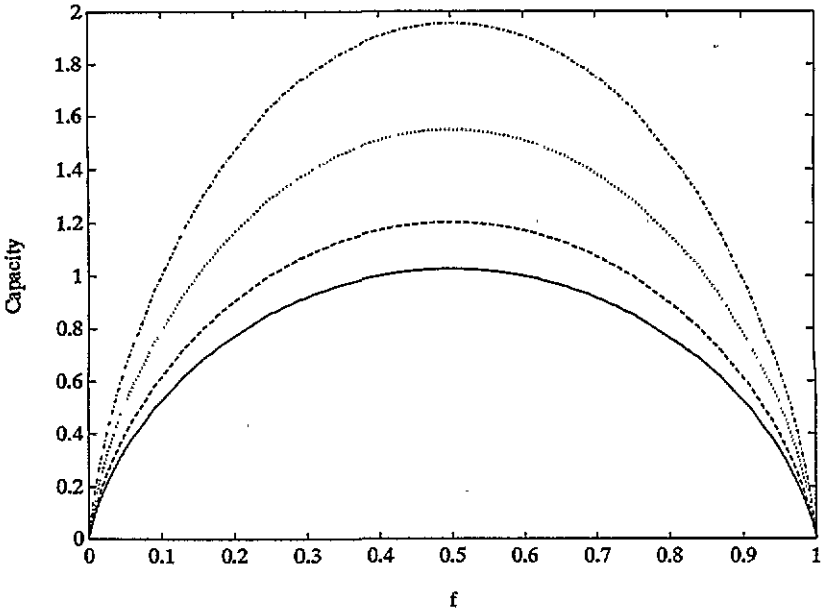

Figure 1. The storage capacity as a function of f for $\kappa=0$ and $m=0.2$, 0.5, 0.7 and 0.8; the upper curves are for the larger values of m.

We have solved the above equations numerically for $\kappa=0$ and for various values of m. These solutions for the capacity as a function of f, the fraction of excitatory neurons, are shown in figure 1 for four different values of the pattern bias m.

All the curves have a maximum when f, the fraction of excitatory neurons in the network, is one half; this is independent of the value of the pattern bias m. It can be directly seen from the symmetry of (24) and (28), and the anti-symmetry of (29), under $k\rightarrow -k$, simultaneously with an interchange of f and $(1-f)$, that $f=\frac{1}{2}$ is an extremum, since (30) for the storage capacity possesses this symmetry as well.

For unbiased patterns, the maximal storage capacity α is 1 for networks with sign-constrained neurons [1]. For patterns with small bias values, we find that the capacity is somewhat less than one. The maximum capacity of sign-constrained networks therefore initially decreases as the bias takes on a non-zero value; this is in contrast to the case of unconstrained networks, where the storage capacity increases continuously with an increase in the bias of the stored patterns. Owing to the invariance of the expression for the entropy under $m\rightarrow -m$, the capacity depends only on the magnitude of the bias m.

3. Discussion

We see that sign-constrained networks can store biased patterns in an optimal manner when excitatory and inhibitory neurons are present in equal numbers. This is true for any non-zero value of the bias. For bias values large in magnitude, one would, at first sight, expect to find a maximal capacity when most of the neurons in a sign-constrained network are excitatory; this would easily ensure that for most neurons i, the stability factor $s_i \Sigma w_{ij} s_j$ is a large positive number, since for large bias magnitudes the spins s

are either mostly positive or mostly negative (depending on whether the bias is positive or negative, respectively). However, a large positive value for this product would mean that these fixed points are very strong attractors, with large basins of attraction, since it is known that larger values of the stability parameter correspond to larger basins of attraction [2–7]. Correspondingly, it is also well known that larger stability parameters in general yield smaller storage capacities. This heuristic argument shows that networks with predominantly ferromagnetic couplings when used to store highly (positively or negatively) biased patterns would actually have a relatively smaller number of attractors with large radii of attraction, and this is why the storage capacity is maximized not for values of f close to 1, but rather for an intermediate value of f. One expects values of f close to one-half to permit the storage of a larger number of attractors with smaller, but non-zero, radii of attraction. It is not unreasonable to expect that the maximal capacity of patterns with any non-zero bias value in sign-constrained networks also corresponds to storing an optimal number of attractors with non-zero radii of attraction.

In the neurophysiological context, a possible advantage of the presence of sign-constrained neurons in brains is therefore that such networks, when used to store biased memories, might be able to store a larger number of them as attractors (rather than just fixed points) with a finite attraction radius when f in these networks is close to one-half. However, we have obtained our results for the case when the threshold potentials of all the neurons in the network are zero. A more realistic study would have to relax this assumption, especially since it is known that the capacity of correlated memories can be enhanced by choosing thresholds appropriately [8].

It is also significant that the value $f=0.5$ is optimal for all non-zero bias values; this would permit the simultaneous optimal storage of memories with varying bias values. This is relevant considering that a large proportion of the memories biological systems store does in fact possess a significant amount of correlation. It is likely that new memories in sign-constrained networks can be stored by changing the magnitude of a relatively smaller number of synaptic strengths, without changing their signs; this would be advantageous biochemically as well. It would be interesting to investigate such issues by means of numerical simulations. Further analytical work along these directions is clearly also necessary before more definitive statements can be made.

References

[1] Amit D J, Campbell C and Wong K Y M 1989 *J. Phys. A: Math. Gen.* **22** 4687
[2] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[3] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
[4] Krauth W, Nadal J-P and Mezard M 1989 *J. Phys. A: Math. Gen.* **21** 2995
[5] Kepler T B and Abbott L F 1988 *J. Physique* **49** 1657
[6] Krauth W, Mézard M and Nadal J.-P. 1988 *Complex Systems* **2** 387
[7] Krauth W and Mézard M 1987 *J. Phys. A: Math. Gen.* **20** L745
[8] Krauth W and Mézard M 1989 *J. Physique* **50** 3056