Capacity of diluted multi-state neural networks

1994 J. Phys. A: Math. Gen. 27 1151

(http://iopscience.iop.org/0305-4470/27/4/011)

J. Phys. A: Math. Gen. 27 (1994) 1151–1161. Printed in the UK

Capacity of diluted multi-state neural networks

D Bollé† and J van Mourik

Instituut voor Theoretische Fysica een Interdisciplinair Centrum voor Neurale Netwerken, KU Leuven, B-3001 Leuven, Belgium

Received 19 April 1993, in final form 19 November 1993

Abstract. The optimal storage capacity is studied for diluted networks with multi-state neurons and continuous respectively discrete couplings, within the replica-symmetric Gardner approach. The Gardner–Derrida line, the de Almeida–Thouless line and the zero-entropy line are compared and the validity of the replica-symmetric approximation is discussed in detail. The distribution of the synaptic couplings is determined. The results are analysed in terms of the number of states of the neuron, the distribution of the stored patterns, the amount of dilution and the number of discrete values for the couplings.

1. Introduction

This paper is concerned with the storage capacity of diluted multi-state networks based upon the n_S-Ising spin-glass. Both spherical models with continuous synaptic couplings and models with discrete local constraints on the couplings are considered. Within the replica-symmetric Gardner theory in the space of couplings [1, 2] fixed-point equations are discussed for the relevant order parameters. In order to look for optimal storage of the patterns, only annealed dilution [3] is being considered.

For the continuous coupling models an analytic expression is derived for the Gardner–Derrida (GD) optimal capacity α_c as a function of the dilution f. Furthermore the de Almeida–Thouless (AT) condition [4] is examined to study the local stability of the replica-symmetric solution in the full order parameter space. This leads to values for the capacity α in function of f referred to as the AT line. Both lines are compared for different values of the number of neuron states and for different distributions of the stored patterns. Finally the distribution of the coupling coefficients [3] is determined.

For models with discrete couplings there is a third line of α-values of interest, i.e. the line where the quenched entropy, being the logarithm of the number of states available to the system in the space of couplings, vanishes [5–7]. Again a comparison is made between this zero-entropy (ZE) line and the two other lines for different values of the number of neurons, the number of couplings and different distributions of the patterns, to find the optimal α consistent within the approximation used. Also the distribution of the coupling coefficients is discussed.

Some of the underlying motivations behind this work are the study of locally connected architectures, the study of robustness against malfunctioning of some of the neurons and the fact that the dynamics of such networks (in the case of extreme dilution [8, 9] can be completely described analytically. For a more detailed discussion and additional references

† E-mail: FGBDA18@cc1.kuleuven.ac.be .

see [3]. Moreover on a more technical level one would like to get a better understanding of the effects causing replica symmetry breaking (RSB). In this respect it is argued here that for multi-state models with $n_S > 2$ the convexity argument used for $n_S = 2$ models to show that RSB is precluded no longer holds. The same has been seen for graded response neurons [10]. Furthermore it is shown that for any finite degree of dilution RSB already occurs for a loading α lower than the GD optimal α_c. In models with discrete couplings this may already happen for values of α below the ZE values, indicating that in these cases a calculation of the entropy is needed to at least first order in the RSB approximation.

These discussions generalize and extend some of the results on fully connected $n_S = 2$ Hopfield-type networks [12] and n_S-Ising networks [11, 12] with continuous synaptic couplings. For these systems the replica symmetric approximation is found to be stable.

They also extend the analysis for discrete synaptic couplings of fully connected $n_S = 2$ models [2, 5, 7]. There it has been found that the replica symmetric approximation is no longer valid and that $\alpha_{ZE} < \alpha_{AT} < \alpha_{GD}$.

Finally the analysis of this work also extends the treatments of [3] and [6] on diluted Hopfield systems with continuous and binary couplings, respectively.

The rest of this paper is organized as follows. In section 2 the multi-state network models are introduced. In section 3 the replica-symmetric approximation to Gardner's space of interaction approach is discussed in terms of the relevant order parameters characterizing the volume (number of solutions) for models with continuous (discrete) couplings. In section 4 the numerical solutions for the GD-, AT-, and ZE line for the loading capacity α are analysed and compared in terms of the number of states of the neuron, the distribution of the stored patterns, the amount of dilution and the number of discrete values for the couplings. Since the number of possibilities for chosing the different parameters defining the models is extremely high, an exhaustive treatment is not within the aims of the present work. In section 5 the validity of replica symmetry is discussed in detail. Section 6 contains a summary of the most important aspects of this study.

2. The model

Consider a network of N neurons taking the n_S different values

$$-1 = s_1 < \cdots < s_k < \cdots < s_{n_S} = +1 . \tag{1}$$

The dynamics of the system is characterized by the following transition probabilities:

$$w[\sigma_i(t+1) = s|h_i(\boldsymbol{\sigma}(t))] = \frac{\exp\{-\beta E_i[s|h_i(\boldsymbol{\sigma}(t))]\}}{\mathrm{Tr}_\sigma \exp\{-\beta E_i[\sigma|h_i(\boldsymbol{\sigma}(t))]\}} \tag{2}$$

with $\boldsymbol{\sigma}(t) = \{\sigma_i(t), i = 1, \dots, N\}$ where $\sigma_i(t)$ is the state of neuron i at time t and with $\beta = \frac{1}{T}$ the inverse temperature. The quantity E_i is given by

$$E_i[\sigma|h_i] = -\tfrac{1}{2}(2h_i\sigma - a\sigma^2) \tag{3}$$

with h_i the local field

$$h_i(\boldsymbol{\sigma}(t)) = \frac{1}{\sqrt{fN}} \sum_{j \neq i} c_{ij} J_{ij} \sigma_j(t) \tag{4}$$

where the J_{ij} denote the synaptic couplings and the c_{ij} take the value 1 when the neurons i and j are connected and 0 otherwise. The fraction of connected neurons is denoted by f.

The term $a\sigma^2$ in (3) favours or discourages high activities, i.e. the highest (in absolute value) neuron states are strengthened ($a < 0$) or suppressed ($a > 0$). For $a < 0$ this dynamics becomes the usual two-state signum dynamics. Hence the storage capacity for random *multi*-state patterns is trivially zero. For $a > 0$ this parameter is a measure of the separation between the plateaus of the gain function [11] and the dynamics can be written in the form

$$\sigma_i(t+1) = s \Leftrightarrow E_i[s|h_i(\sigma(t))] < E_i[\sigma|h_i(\sigma(t))] \qquad \forall\sigma \in \{s_k\} \setminus \{s\} \quad \forall i\,. \tag{5}$$

In the noiseless case ($\beta \to \infty$) the dynamics becomes deterministic: $\sigma_i(t+1)$ is that state out of $\{s_k\}$ that minimizes $E_i[\sigma|h_i(\sigma(t))]$.

We want to store p random patterns $\boldsymbol{\xi}^\mu = \{\xi_i^\mu, i = 1, \ldots, N\}, \mu = 1, \ldots, p$, obeying the following probability distribution

$$P(\xi_i^\mu) = \sum_{k=1}^{n_S} (1 + b_k)/n_S \;\; \delta(\xi_i^\mu - s_k)\,. \tag{6}$$

This distribution is well defined if $b_k \in [-1, n_S - 1]$ and $\sum_{k=1}^{n_S} b_k = 0$. We furthermore suppose that the stored patterns are uncorrelated with $\langle\langle\xi\rangle\rangle = 0$ and $\langle\langle\xi^2\rangle\rangle = l$ (the symbol $\langle\langle\cdots\rangle\rangle$ denotes the average over the distribution of the random patterns). At this point we remark that the parameters b_k allow us to give more weight to the values s_k in the middle of the interval $[-1, +1]$. In this way we are able to mimic a Gaussian distribution of the patterns.

The condition for pattern μ to be a fixed point of the dynamics is given by (5) for $s = \xi_i^\mu, \forall i, \mu$. In order to have a large basin of attraction proportional to κ we have to require [11]

$$h_i(\boldsymbol{\xi}^\mu) \in (O(\xi_i^\mu) + \kappa, B(\xi_i^\mu) - \kappa) \qquad \forall i, \mu \tag{7}$$

with

$$\begin{aligned} O(s_j) &= \begin{cases} -\infty & \text{if } j = 1 \\ \frac{a}{2}(s_{j-1} + s_j) & \text{if } j \neq 1 \end{cases} \\ B(s_j) &= \begin{cases} \frac{a}{2}(s_j + s_{j+1}) & \text{if } j \neq n_S \\ +\infty & \text{if } j = n_S \end{cases} \end{aligned} \tag{8}$$

and κ given by

$$\kappa = \frac{n_S - 1}{2} \min_{\mu,i} \left\{ \min_{\sigma \neq \xi_i^\mu} \left\{ E_i[\sigma|h_i(\boldsymbol{\xi}^\mu)] - E_i[\xi_i^\mu|h_i(\boldsymbol{\xi}^\mu)] \right\} \right\}. \tag{9}$$

In the following we consider models with continuous and models with discrete synaptic couplings J_{ij}.

3. Replica symmetric theory

We apply Gardner's space of interaction approach [1] to study the optimal capacity of these multi-state models with annealed dilution within replica symmetry [13] in function of f, κ, b_k and a. The capacity α of the network is defined as the number of stored patterns divided by the number of neurons, i.e. $\alpha = p/N$.

From now on we assume, for convenience, an equidistant distribution for the values s_k of the neurons reading

$$s_k = -1 + \frac{2(k-1)}{n_S - 1} \qquad k = 1, \ldots, n_S\,. \tag{10}$$

Furthermore we will use the shorthand notation

$$B = \frac{B(\xi) - \kappa}{\sqrt{l}} \qquad O = \frac{O(\xi) + \kappa}{\sqrt{l}} \qquad \mathrm{D}z = \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right). \tag{11}$$

3.1. Continuous couplings

To fix the degree of dilution and to normalize the non-zero connections we have to impose the following global constraints

$$\sum_{i \neq j} c_{ij} = fN \qquad \sum_{i \neq j} c_{ij} J_{ij}^2 = fN. \tag{12}$$

Focusing our attention on a specific neuron i we start from the following volume in the space of couplings

$$\begin{aligned} V = \frac{1}{\mathcal{N}} \prod_j \mathrm{Tr}'_{\{c_{ij} J_{ij}\}} \prod_{\mu=1}^{\alpha N} \Big(\theta(h_i^\mu - O) - \theta(h_i^\mu - B) \Big) \\ \times \delta\Big(\sum_j c_{ij} J_{ij}^2 - fN \Big) \delta\Big(\sum_j c_{ij} - fN \Big) \end{aligned} \tag{13}$$

$$\mathcal{N} = \prod_j \mathrm{Tr}'_{\{c_{ij} J_{ij}\}} \delta\Big(\sum_j c_{ij} J_{ij}^2 - fN \Big) \delta\Big(\sum_j c_{ij} - fN \Big)$$

where

$$\mathrm{Tr}'_{\{c,J\}} \cdots = \int \mathrm{d}c \cdots \delta(c) + \int \mathrm{d}c\, \delta(c-1) \int \mathrm{d}J \cdots. \tag{14}$$

This restricted trace is introduced to get the relevant volume for the couplings determined by the constraints imposed by the stored patterns. In this way we avoid adding a constant volume due to the integration over the couplings for which $c_{ij} = 0$ [3]. Following [1] we use the replica method to evaluate $v = \lim_{N\to\infty} N^{-1} \langle\langle \ln V \rangle\rangle$. Assuming that replica symmetry is unbroken we get the result

$$v = \alpha G_1 + G_2 + \frac{f}{2}(\psi + \hat{q}_0 - \hat{q} q_0) \tag{15}$$

with

$$G_1 = \int \mathrm{D}z \left\langle\!\!\left\langle \ln\left\{ \frac{1}{2}[\mathrm{erf}(f_B) - \mathrm{erf}(f_O)] \right\} \right\rangle\!\!\right\rangle \tag{16}$$

$$G_2 = \int \mathrm{D}z \, \ln\left\{ 1 + \sqrt{\frac{2\pi}{\hat{q}_0}} \exp\left(\frac{\sqrt{\hat{q}} z^2}{2\hat{q}_0} - \frac{\psi}{2} \right) \right\} \tag{17}$$

where

$$f_r = \frac{r + z\sqrt{1 - q_0}}{\sqrt{2 q_0}} \qquad g_r = r + \frac{z}{\sqrt{1 - q_0}} \qquad r = B, O. \tag{18}$$

The four order parameters $\psi, \hat{q}_0, \hat{q}$ and q_0 appearing in these expressions have the following meaning: $q_0 = 1 - q$ where q is the overlap between two solutions in the coupling space, ψ expresses the first condition in (12) and $\hat{q}_0$ and $\hat{q}$ are conjugate variables. For these order parameters fixed-point equations can be derived in the normal way.

For general values of α and f these fixed-point equations have to be solved numerically. To study the optimal capacity of the network according to the GD criterion, one considers the limit $q_0 \to 0$. This can be done analytically in a manner anologous to [3], leading to

$$\alpha_c(f) = (f + 2u\exp(-u^2)/\sqrt{\pi})\alpha_c(f=1) \tag{19}$$

$$\alpha_c^{-1}(f=1) = \left\langle\!\!\left\langle \int_{-\infty}^{-B} \mathrm{D}z\,(z+B)^2 + \int_{-O}^{+\infty} \mathrm{D}z\,(z+O)^2 \right\rangle\!\!\right\rangle. \tag{20}$$

where u and f are related through $f = \mathrm{erfc}(u)$. For a detailed discussion of the fully connected case, $f = 1$, we refer the reader to [11].

Another value of interest of α is given by the requirement of stability of the replica-symmetric approximation we have used, i.e. the AT criterion. To check this stability it is sufficient to examine if the sign of the product of the eigenvalues of the tranverse fluctuations, the so called replicon eigenvalue λ_R [4], is negative.

3.2. Discrete couplings

The number of discrete values of the J_{ij} is denoted by n_J. We assume for convenience an equidistant distribution of these values between -1 and $+1$.

We start from the number of solutions Ω replacing the fractional volume

$$\Omega = \mathrm{Tr}'_{\{c_{ij}J_{ij}\}} \prod_{\mu} (\theta(h_i^\mu - O) - \theta(h_i^\mu - B))\, \delta\left(\sum_j c_{ij} - fN\right) \tag{21}$$

where

$$\mathrm{Tr}'_{\{c,J\}} \cdots = \int \mathrm{d}c \cdots \delta(c) + \int \mathrm{d}c\, \delta(c-1)\ \mathrm{Tr}_J \cdots . \tag{22}$$

Again the restricted trace is introduced in order to avoid adding an entropy to couplings for which $c_{ij} = 0$ (compare (14)). Furthermore the value 0 for the couplings J is excluded in order to avoid double counting of solutions.

We then evaluate the quenched entropy, $S_{\mathrm{qu}} = \lim_{N\to\infty} N^{-1}\langle\!\langle \ln \Omega \rangle\!\rangle$, over the different sets of memories using again the replica method. We obtain

$$S_{\mathrm{qu}} = \alpha G_1 + G_2 + (f/2)(\psi + \hat{q}_0 Q - \hat{q} q_0) \tag{23}$$

with G_1 formally given by (16) and

$$G_2 = \int \mathrm{D}z\, \ln\left\{1 + \mathrm{Tr}_J \exp\left(\sqrt{\hat{q}}zJ - \hat{q}_0\frac{J^2}{2} - \frac{\psi}{2}\right)\right\} \tag{24}$$

where

$$f_r = \frac{r + z\sqrt{Q-q_0}}{\sqrt{2q_0}} \qquad g_r = r + \frac{zQ}{\sqrt{Q-q_0}} \qquad r = B, O\,. \tag{25}$$

The five order parameters $\hat{q}, \hat{q}_0, \psi, q_0$ and Q now describe the following: $q_0 = Q - q$ where q is the overlap between two solutions in the coupling space and Q is the self-overlap of a solution; the others have the same meaning as before. Again, fixed-point equations can be derived.

To study the optimal capacity there are now three values of interest [7]. Besides the GD- and AT criteria, the ZE criterion is also relevant. We remark that, in contrast with the spherical coupling models no closed expression for α_c is found in the GD limit. Furthermore, the order parameters have to be rescaled in order to keep them finite ($\hat{q}' = q_0^2\hat{q}$, $\psi' = q_0\psi$, $\hat{q}_0{}' = q_0\hat{q}_0$).

Numerical solutions for these three values will be analysed and compared in section 4 leading to an upper limit for the optimal storage capacity within replica symmetry.

3.3. Distribution of couplings

It is interesting to write down the distribution of the couplings following [3]. For continuous and discrete couplings for which the corresponding $c_{ij} \neq 0$ one gets, respectively

$$\rho(J) = \frac{1}{f} \int \mathrm{D}z \, \frac{\exp(\sqrt{\hat{q}}zJ - \hat{q}_0 J^2/2 - \psi/2)}{1 + \sqrt{(2\pi/\hat{q}_0)} \exp(\hat{q}z^2/\hat{q}_0 - \psi/2)} \tag{26}$$

$$\rho(J) = \frac{1}{f} \int \mathrm{D}z \, \frac{\mathrm{Tr}_{J'} \exp(\sqrt{\hat{q}}zJ' - \hat{q}_0 J^2/2 - \psi/2)\delta(J - J')}{1 + \mathrm{Tr}_{J'} \exp(\sqrt{\hat{q}}zJ' - \hat{q}_0 J'^2/2 - \psi/2)} \,. \tag{27}$$

In the GD limit the expression (26) formally simplifies to the formula obtained in [3] for the $n_S = 2$ diluted network. For discrete couplings the GD limit of (27) is given by

$$\rho(J) = \frac{1}{f} \mathrm{Tr}_{J'} \, \delta(J - J') \int_{I_{J'}} \mathrm{D}z \, \theta\left(\sqrt{\hat{q}'}zJ' - \frac{\hat{q}_0{}' J'^2}{2} - \frac{\psi'}{2}\right) \tag{28}$$

where

$$I_{J'} = \left\{ z \in \mathbb{R} \middle| \left(z - \frac{J'\hat{q}_0{}'}{\sqrt{\hat{q}'}}\right)^2 = \min_J \left[\left(z - \frac{J\hat{q}_0{}'}{\sqrt{\hat{q}'}}\right)^2\right] \right\}. \tag{29}$$

These distributions of the couplings J_{ij} are explicitly dependent upon f. The dependence upon n_S, a, κ and b_k only occurs through the fixed-point values of the order parameters.

4. Results for the optimal capacity

4.1. Continuous couplings

Analysing the general behaviour of the fixed-point equations we find that for fully connected models, i.e. $f = 1$, the GD solution may be stable or unstable with respect to replica symmetry, depending on the gain parameter a, the number of spin states n_S and the distribution of the patterns b_k. In fact, for $n_S > 2$, the replicon eigenvalue λ_R is negative for a large enough, while λ_R may become positive for a small. For the stability parameter $\kappa = 0$ a straightforward calculation leads to the identity

$$\mathrm{sign}(\lambda_R) = \mathrm{sign}(\partial\alpha_c/\partial a) \,. \tag{30}$$

This has also been seen in graded-response networks for any input–output relation that is monotonically non-decreasing (or non-increasing) [10]. This implies the existence of an optimal a_{opt}, that can be 0 or finite, for which α_c is maximal and λ_R is vanishing. It also means that for $a < a_{\mathrm{opt}}$ replica symmetry is broken and that for $a > a_{\mathrm{opt}}$ replica symmetry is valid. For all cases with equidistant spins and $b_k = 0$ that we have examined we find that $a_{\mathrm{opt}} = 0$.

For $\kappa \neq 0$ equation (30) is to be replaced by the following:

$$\text{at } \alpha_c(a_{\mathrm{opt}}) \qquad \text{such that } \frac{\partial\alpha_c}{\partial a} = 0 : \lambda_R < 0 \,. \tag{31}$$

In general, increasing κ lowers α_c and enhances the stability against RSB. We remark that for given κ, a is bounded from below (recall (9)). This behaviour is illustrated in figure 1 for a $n_S = 3$ model.

Finally we find that in the GD limit the solution is unstable against RSB for all diluted models, i.e. for any $f \neq 1$. This can be seen in figure 2 where the GD line and the AT line are plotted as a function of f for the $n_S = 2$ model with $\kappa = 0$ and for a $n_S = 3$ model

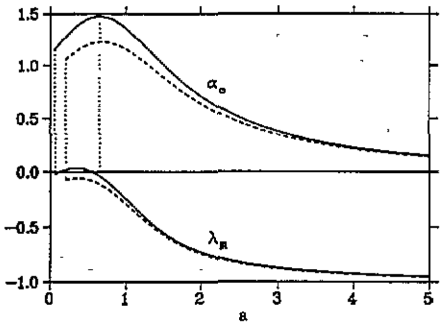

Figure 1. The storage capacity α_c and the replicon eigenvalue λ_R as a function of the gain parameter a for the continuous $n_S = 3$ model with $b_k = \{-0.5, 1, -0.5\}$ and $\kappa = 0.03$ (solid curve) respectively $\kappa = 0.1$ (dashed curve).

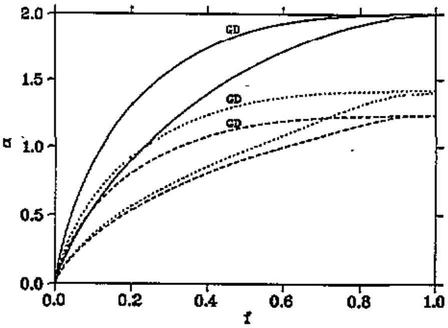

Figure 2. The GD and AT line for the storage capacity α as a function of the dilution f for the continuous $n_S = 2$ model with $\kappa = 0$ (solid curves) and for the $n_S = 3$ model with $b_k = \{-0.5, 1, -0.5\}$ and $a = 0.23$, $\kappa = 0$ (dotted curves) and $\kappa_c = 0.0488$ (dashed curves).

for two values of κ. The results on the $n_S = 2$ model (solid lines) confirm the GD line first calculated in [3]. It further demonstrates that the GD limit is stable at $f = 1$. For the $n_S = 3$ model this limit is unstable at this point for small κ. There exist a critical κ_c depending on a (in the case shown $\kappa_c = 0.0488$ for $a = 0.23$), above which the replica symmetric solution is locally stable at $f = 1$.

In the case of extreme dilution, $f \to 0$ ($u \to \infty$) , we see that both α_{GD} and α_{AT} behave as $\alpha \sim -f \ln f$ for all n_S. The former can be shown analytically using (19)–(20).

4.2. Discrete couplings

For any finite number of coupling values n_J the solution for the capacity on the GD line is unstable with respect to RSB and the quenched entropy (23) is negative.

In the fully connected $n_S = 2$ model we find the result of [7] (see also [5]) that the ZE criterion gives the optimal storage capacity. For fully connected $n_S > 2$ models we find that the ZE solution is stable with respect to RSB, for all the combinations of the parameters n_J, a, κ and b_k we have examined. If we impose an additional constraint on these models by fixing the self-overlap Q to a value between $\min\{J^2\}$ and $\max\{J^2\}$ this is no longer true in general. In figure 3 we show a situation where for increasing n_J the ZE solution approaches the GD solution while the AT solution does not. This stresses the role of the spherical constraint in the occurrence of RSB, as we will discuss in further detail in the next section.

In the case of dilution, $f \neq 1$, we find that there are models in which the ZE solution is unstable with respect to RSB even for $n_S = 2$ and no additional constraint on Q, in contrast with the $f = 1$ case. For increasing n_J the ZE solution will eventually approach the GD solution while the AT solution not necessarily does so. In figures 4–6 we present some of the different types of behaviour that we have observed for the GD, AT and ZE line as a function of f.

In the case of binary neurons we see in figure 4 that for $n_J = 2$ the ZE line is the lowest within the whole range of f, for higher values of n_J and decreasing f it stays the lowest until a certain value of the latter where it crosses the ZE line. Furthermore for increasing n_J the ZE line approaches the GD line while the AT line approaches the GD line only in $f = 1$

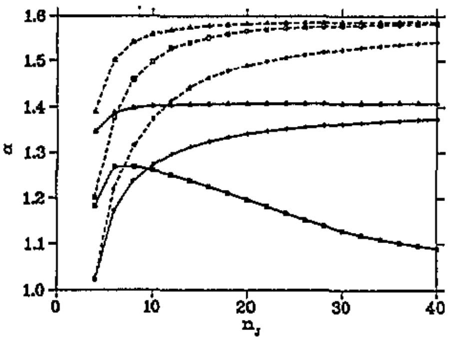

Figure 3. The GD, AT and ZE values for the storage capacity α as a function of n_J for the discrete $n_S = 3$ model with $a = 0.3$, $\kappa = 0$ and $b_k = \{-0.5, 1, -0.5\}$ (triangles, squares and diamonds on the dashed curves). The points on the solid curves are results with the self-overlap Q fixed to be 0.5.

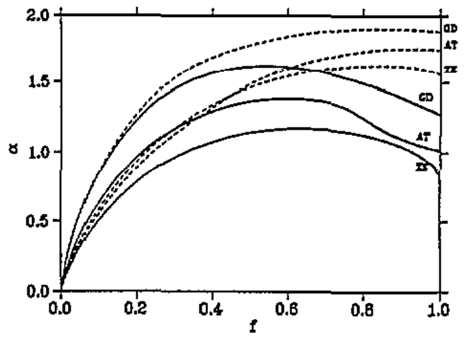

Figure 4. The GD, AT and ZE line for the storage capacity α as a function of the dilution f for the $n_S = 2$ model with $n_J = 2$ (solid curves) and $n_J = 6$ (dashed curves) both with $\kappa = 0$.

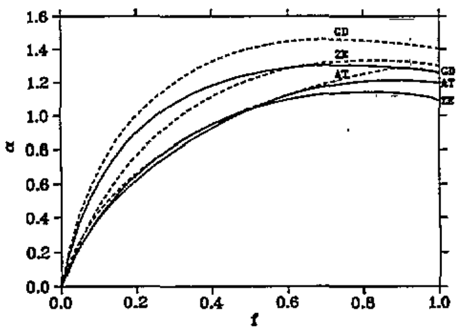

Figure 5. As figure 4, but for the $n_J = 6, n_S = 3$ model with $\kappa = 0, a = 0.3, b_k = \{0, 0, 0\}$ and $Q = 0.5$ (solid lines) and for the $n_J = 12, n_S = 3$ model with $\kappa = 0, a = 0.3, b_k = \{-0.5, 1, -0.5\}$ and $Q = 0.5$ (dashed lines).

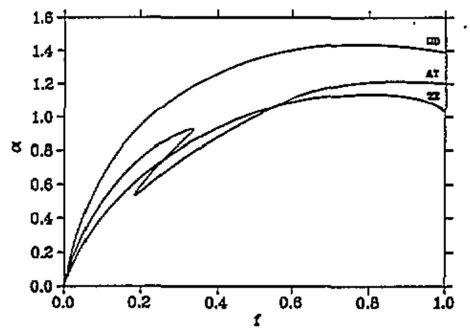

Figure 6. As figure 4, but for the $n_J = 4, n_S = 3$ model with $a = 0.3, \kappa = 0, b_k = \{-0.5, 1, -0.5\}$.

such that the crossing point of the AT and the ZE line eventually shifts to $f \simeq 1$.

In figure 5 we present the $\alpha - f$ diagram for some $n_S = 3$ models where the self-overlap Q is fixed to 0.5. A completely different behaviour is found for the ordering of the AT and ZE solutions depending on the values of n_J and b_k. We note (solid curves) that $\alpha_{ZE} < \alpha_{AT}$ for $\kappa = 0$ and $f \sim 0$ or $f \sim 1$, while for intermediate f this inequality is reversed. For the other case (dashed curves) $\alpha_{ZE} < \alpha_{AT}$ only for $f \sim 0$. Finally figure 6 shows the results of a similar model without fixing Q. There is a range for f where 3 solutions for the AT line are found. This range shifts to small f with increasing stability parameter κ. For all models we have looked at the ordering $\alpha_{ZE} < \alpha_{AT} < \alpha_{GD}$ remains valid in the neighbourhood of $f \sim 0$.

Finally we have checked numerically that all α_{GD}, α_{AT} and α_{ZE} curves show a $-f \ln f$ behaviour in the limit of extreme dilution $f \to 0$.

4.3. *Distribution of couplings*

Concerning the distribution of the couplings for the continuous models we find that for $\alpha = 0$ it is a Gaussian with mean zero and variance one. For increasing α it converges, in the GD limit, towards the distribution also found in [3] i.e. a Gaussian with mean zero and variance $1/c_f$ with a gap in the middle. The width of this gap is only dependent upon f and grows with increasing f. Furthermore this limiting distribution is independent of the parameters n_S, a, κ and b_k. At intermediate $\alpha < \alpha_{GD}$ the distribution is dependent upon these parameters, but only through the fixed-point values of the order parameters.

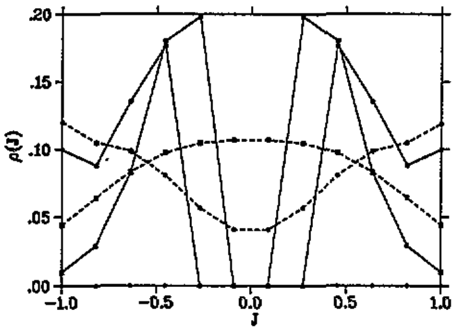

Figure 7. The distribution of the couplings for the discrete $n_J = 12, n_S = 3$ model with $f = 0.5, a = 0.3, \kappa = 0$ and $b_k = \{-0.5, 1, -0.5\}$. The squares indicates the results with no constraints on Q, the circles with $Q = 0.5$. The dashed curves are for $\alpha = 0.862$ (Q free) and $\alpha = 1.15$ ($Q = 0.5$), the solid curves for the limiting values $\alpha_c = 1.47$ (Q free) and $\alpha_c = 1.404$ ($Q = 0.5$).

For the models with discrete couplings figure 7 shows the height of the peaks in (27) and (28). Also the enveloping curves are presented (solid and dashed curves). In the GD limit this enveloping curve has a gap, such as in the case of continuous coupling models. The probability of having small coupling values vanishes for the models with and without the extra spherical constraint on Q. For models with Q fixed, the gap is broader (the two inner J values are not present). The vanishing of these coupling values is not observed for all models, an example being models with low dilution and a small number of n_J. The probability of having extreme coupling values may become very small, especially for models without a constraint on Q (see the squares on the solid curves in the figure) but it never vanishes. For low α the distribution is mostly centred around the origin (e.g. the squares on the dashed curves). For intermediate $\alpha < \alpha_{GD}$ a dip around the origin already develops (e.g. the circles on the dashed curves).

We end with the remark that the distribution of the local fields [14] is independent of the amount of dilution f and takes the same form for both continuous and discrete coupling models as the one given in [12].

5. Discussion

In this section we give a qualitative discussion of the validity of the replica-symmetric approximation used in the foregoing calculations. In particular we want to look at the popular argument stating that the convexity and connectedness of the solution space precludes RSB. This argument has been seen to break down in fully connected graded-response networks with continuous couplings and monotone non-decreasing input–output relations due to the spherical constraint ([10]).

For binary neurons, $n_S = 2$, each of the stored pattern divides the $(N-1)$-dimensional space of couplings for each neuron i by a hyperplane into two regions leading to, respectively the desired output or the wrong output for the stored pattern.

For multi-state neurons, $n_S > 2$, a stored pattern may divide the space of couplings by two parallel hyperplanes (for the intermediate spin-states $|s_k| < 1$) or one hyperplane (for the extremal spin-states $|s_k| = 1$). For the intermediate spin-states the coupling vector has to lie between the two hyperplanes, for the extremal spin-states it has to lie on the correct side of the hyperplane.

All these planes lie at a distance of order $1/\sqrt{N}$ from the origin, or in the origin for $\kappa = 0$ and $n_S = 2$, when one takes the length of the coupling vector of order 1, i.e. f in the continuous coupling models (see (12)) and fQ in the discrete ones. The part of the solution space where the coupling vector has to lie to give the correct output is called the version space.

For $n_S = 2$ the version space has a cone-like shape, with the top in the origin for $\kappa = 0$ or at a distance of order $1/\sqrt{N}$ of the origin for $\kappa \neq 0$. So this space and the intersection with any sphere with a radius of order 1 are connected and convex. For $n_S > 2$ the version space no longer has a cone-like shape but it is still convex and connected, while the intersection with any sphere with a radius of order 1, however, not necessarily is. The latter depends on the ratio of the intermediate and extremal spin-states described by the parameter b_k and on the parameters a and κ.

For the fully connected $n_S = 2$ models with continuous couplings this implies that, until the moment where the version space shrinks to zero (i.e. the GD point), no RSB occurs. For this type of model with $n_S > 2$ it tells us that RSB is not precluded. In the case of dilution replica symmetry is broken at the GD point. This can be understood in the following way. The permitted solutions have to lie in the intersection of $(1 - f)(N - 1)$ of the $(N - 1)$ hyperplanes indicating a zero value of the couplings. In this case the permitted volume is no longer connected and hence RSB is not precluded.

For the models with discrete couplings there are n_J parallel hyperplanes for each coupling indicating the permitted values for that coupling. This gives a discrete set (the corners of a hypercube for $n_J = 2$) of possible solutions in the coupling space. So the solution space is again not connected and RSB is also not precluded here. If we dilute these models or if we impose an additional constraint on these models with $n_S > 2$ by fixing the self-overlap Q, we observe that the version space may become disconnected (AT) already before there are no more points of the discrete set of solutions in it (ZE). This indicates the role of the spherical constraint in the occurrence of RSB.

6. Summary

In this paper we have analysed the capacity of diluted multi-state network models with continuous and discrete couplings, respectively. Within replica symmetry we have studied in some detail the different criteria to determine the optimal storage capacity : the Gardner–Derrida criterion taking the overlap between two solutions in the coupling space to be maximal, the de Almeida–Thouless criterion checking the local stability of the replica symmetric solution in the full space of solutions and, in the case of discrete coupling models the zero-entropy criterion determining the vanishing of the replica-symmetric quenched entropy.

The most important results for all combinations of the relevant parameters we have examined are the following.

(i) All α curves show a $-f \ln f$ behaviour in the limit of extreme dilution $f \to 0$

(ii) If the distribution of the couplings shows a gap the replicon eigenvalue becomes $+\infty$ in the GD limit and hence replica symmetry is broken. For spherical couplings this occurs for any degree of dilution $f < 1$, for discrete couplings it always happens, i.e. for any f.

(iii) For continuous couplings and $b_k = 0$, α_{AT} and α_{GD} always coincide at $f = 1$.

(iv) For discrete couplings the ordering $\alpha_{\mathrm{ZE}} < \alpha_{\mathrm{AT}} < \alpha_{\mathrm{GD}}$ is always valid at $f \sim 0$. If there is no extra spherical constraint this ordering $\alpha_{\mathrm{ZE}} < \alpha_{\mathrm{AT}} < \alpha_{\mathrm{GD}}$ also stays valid at $f \sim 1$.

Thus for discrete couplings the vanishing of the replica-symmetric quenched entropy is, in general, no longer a good criteron to determine the optimal capacity. Therefore a study of replica symmetry breaking effects is needed.

Acknowledgments

This work has been supported in part by the Research Fund of the KU Leuven (grant OT/91/13). The authors are indebted to M Bouten and R Kühn for stimulating discussions. One of them (DB) would like to thank the Belgian National Fund for Scientific Research for financial support as a Research Director.

References

[1] Gardner E 1987 *Europhys. Lett.* 4 481; 1988 *J. Phys. A: Math. Gen.* **21** 257
[2] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
[3] Bouten M, Engel A, Komoda A and Serneels R 1990 *J. Phys. A: Math. Gen.* **23** 4643
[4] de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
[5] Krauth W and Mézard M 1989 *J. Physique* **50** 3057
[6] Bouten M, Komoda A and Serneels R 1990 *J. Phys. A: Math. Gen.* **23** 2605
[7] Gutfreund H and Stein Y 1990 *J. Phys. A: Math. Gen.* **23** 2613
[8] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* 4 167
[9] Bollé D, Vinck B and Zagrebnov V A 1993 *J. Stat. Phys.* **70** 1099
[10] Bollé D, Kühn R and van Mourik J 1993 *J. Phys. A: Math. Gen.* **26** 3149
[11] Bollé D, Dupont P and van Mourik J 1991 *Europhys. Lett.* **15** 893
[12] Mertens S, Köhler H M and Bös S 1991 *J. Phys. A: Math. Gen.* **24** 4941
[13] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[14] Kepler T B and Abbott L F 1988 *J. Physique* **49** 1657