Storing Block-wise Semantically Correlated Patterns in a Perceptron: Results from a Cavity Method *

J.Ó. WINKEL[1] F. GERL[2]
U. KREY[1]

[1] Institut für Physik II der Universität Regensburg,
Universitätsstr. 31, D-93040 F.R.G.

[2] Institut für Theoretische Physik der Universität Göttingen,
Bunsenstr. 9, D-30373 Göttingen F.R.G.

June 16, 1995

Abstract

In this paper we calculate the storage capacity of the perceptron for block-wise semantically correlated patterns. The set of patterns is divided into blocks, each block consisting of n binary patterns. The patterns in a block have the overlap R with each other and patterns in different blocks have overlap 0. Using a cavity method, which was recently developed by F. Gerl [1], we derive a general expression for the storage capacity $\alpha(\kappa, R, n)$ for all possible values of R and n with $-1/(n-1) \leq R \leq 1$. Evaluation leads to enhanced resp. reduced storage capacities for positive resp. negative correlation R. Asymptotically for $n \to \infty$ we find that $\alpha(\kappa = 0, n) \to nR/[2(1-R)\ln(nR\sqrt{1-R})]$, if the logarithm in the denominator is $\gg 1$. We are also able to interpret the various terms in the expression for α in an intuitive way.

*based on the Diploma thesis of J.Ó. Winkel, Regensburg 1995

1 Introduction

Many real-world applications of neural networks involve classification of data. Now, in praxis the data to be classified is usually correlated in some way. Because practical classification problems usually involve recognizing some features in the input patterns, it is often so that input patterns which are in some way similar (i.e. containing the same features) belong to the same class.

An example of such a classification problem would be the recognition of handwritten letters: If each input is a handwritten character, then every handwritten "A" belongs to the same class and every handwritten "B" belongs to the same class, etc. Another task would be to recognize certain object-classes in a photo.

These considerations lead to the following abstract problem: Let a set of $p := n \cdot \mathcal{N}$ patterns be given, which can be divided into $\mathcal{N}$ blocks. Within a block, the patterns are correlated with an overlap R (see below), whereas patterns from different blocks are uncorrelated.

In the following paper, this abstract problem is treated by the methods of statistical mechanics. Of course it would be natural to apply the well-known replica method for the derivation of the learning capacity as a function of the parameters considered, see e.g. [2]; but for general n this would lead to difficult calculations. We therefore used a *cavity method* recently derived by one of the authors, [1], which is simpler, and which is known from [1] to give the same results as the replica method as long as the replica symmetry is not broken. Thus, with the cavity method we are able to treat the problem considered for general values of the parameters $\alpha := p/N$ (where N is the number of input units of the simple perceptron considered), R and n (see above), and the stability parameter κ.

Concerning the input vector $\vec{\xi}$, we assume as usual that the N components ξ_k of $\vec{\xi} := (\xi_1, ..., \xi_N)$ take the values ± 1, and also the possible values of the output unit ξ_0 are ± 1. Input and output are related by the usual perceptron prescription (see below). Concerning the output value, we naturally assume that patterns belonging to the same group have the same output.

2 The Model

2.1 Notations and Definitions

We consider a set of p binary input vectors $\vec{\xi}^\mu$ and p outputs ξ_0^μ. The input vectors are grouped together in blocks, each block containing n input vectors. The input vectors in a block have an overlap R with each other, and all outputs corresponding to a particular block are equal. Introducing the oriented patterns $\vec{\sigma}^\mu := \xi_0^\mu \cdot \vec{\xi}^\mu$, this can be written as:

$$\begin{aligned} \frac{1}{N}\langle \vec{\sigma^\mu} \cdot \vec{\sigma^\nu} \rangle &\equiv R \quad \text{for } \mu \neq \nu \text{ in the same block} \\ \frac{1}{N}\langle \vec{\sigma^\mu} \cdot \vec{\sigma^\nu} \rangle &\equiv 0 \quad \text{for } \mu, \nu \text{ in different blocks}. \end{aligned} \tag{1}$$

From this definition it follows that $-1/(n-1) \leq R \leq 1$. The coupling vector $\vec{J}$ of the perceptron can be written as a linear combination of the oriented patterns, [3]. The coefficients x^μ in this combination are called *embedding strengths*:

$$\vec{J} =: \frac{1}{\sqrt{N}} \sum_{\nu=1}^{p} x^\nu \vec{\sigma}^\nu . \tag{2}$$

Defining the *oriented field* E^μ as:

$$E^\mu := \frac{1}{\sqrt{N}} \vec{J} \cdot \vec{\sigma}^\mu , \tag{3}$$

we can write it in terms of the correlation matrix of the oriented patterns $B^{\mu\nu} := \vec{\sigma}^\nu \cdot \vec{\sigma}^\mu / N$ and define the *cavity field* $\tilde{E}^\mu$ as follows:

$$E^\mu = \sum_\nu x^\nu B^{\mu\nu} = x^\mu + \sum_{\nu(\neq\mu)} x^\nu B^{\mu\nu} =: x^\mu + \tilde{E}^\mu . \tag{4}$$

The importance of the cavity fields lies in their following two properties: Firstly, the embedding strengths can be uniquely determined by the cavity fields and the Kuhn-Tucker conditions (see next section). Secondly, for patterns not explicitly embedded in the couplings (i.e. $x^\mu = 0$), they are identical with the oriented fields E^μ, and thus we can construct their probability distribution using simple statistical arguments (see below).

2.2 The Kuhn-Tucker conditions and the cavity method

In this section we briefly sketch the cavity method developed by F. Gerl [1]. The optimal perceptron corresponds to the following quadratic optimization problem [3, 4]:

$$\text{MAXIMIZE} \quad \kappa := \min_{\mu}\{E^{\mu}\} \quad \text{WITH CONSTANT} \quad |\vec{J}|\,. \tag{5}$$

Its solution is uniquely determined by the *Kuhn-Tucker* conditions (KTCs for short) [3, 4]:

$$\text{EITHER} \quad E^{\mu} > \kappa \;\;\wedge\;\; x^{\mu} = 0 \quad \text{OR} \quad E^{\mu} = \kappa \;\;\wedge\;\; x^{\mu} > 0\,. \tag{6}$$

Using the previous definitions and the KTCs, we have the following equation for the norm of the coupling vector:

$$|\vec{J}|^2 \;\;=\;\; \sum_{\nu,\mu} x^{\nu} B^{\nu\mu} x^{\mu} = \sum_{\mu} x^{\mu} E^{\mu} \stackrel{\text{KTC}}{=} \kappa \sum_{\mu} x^{\mu}\,. \tag{7}$$

The last sum is just the expectation value of the embedding strengths times p, the number of patterns. Therefore, if the probability distribution w of x^{μ} is known, setting $|\vec{J}|^2 = N$ and taking the limit $N \to \infty$, we get the following equation:

$$1 = \frac{\kappa}{N} \sum_{\mu} x^{\mu} = \kappa\alpha \int x w(x)\, dx\,, \tag{8}$$

where $\alpha := p/N$ is the desired storage capacity.

This formalism now allows us to set up a simple but powerful cavity scheme in the spirit of Onsager to determine the storage capacity. Here the perceptron couplings play the role of the spin system, and the "interactions" $\vec{J}$ are determined by the KTCs. The recipe now goes as follows (see [1]):

- Assume a perceptron fulfilling the KTCs for a given set of p patterns $\mathcal{P}$. That is, set up the KTCs and determine x^{μ} as a function of the cavity fields, $\forall \vec{\sigma}^{\mu} \in \mathcal{P}$
- Add a new pattern (called *test-pattern*) to $\mathcal{P}$ and assume it to be embedded with a certain embedding strength x^0. Then calculate the updates δx^{μ} for the x^{μ}, i.e. the reaction of the system necessary to restore the KTCs for the old embedded patterns.

- Set up the KTCs for the test-pattern using the *updated* embedding strengths of the stored patterns, i.e. reaction included, ($x^\mu \rightarrow x^\mu + \delta x^\mu$), and express the embedding strengths of the test-pattern as a function of the cavity fields.
- Use eq. (8) to derive the storage capacity $\alpha(\kappa, n, R)$.

3 Cavity method for block-wise correlated patterns

3.1 Embedding strengths of the patterns and the cavity fields

Since the patterns are correlated, we have to treat them block-wise, i.e. we set up the KTCs for each block. Now, the KTCs split the set of patterns into two categories, explicitly ($x^\mu > 0$) and automatically ($x^\mu = 0$) stored patterns. This means that the blocks get split into $(n+1)$ categories, blocks which consist of 0, 1, 2, ...or n explicitly stored patterns. The KTCs for a block containing m explicitly stored patterns (let's call such a block an "m-block") are:

$$\begin{aligned} E^\mu &= x^\mu + R\sum_{\rho=1}^{m-1} x^{\mu+\rho} + \tilde{E}^\mu = \kappa \\ &\vdots \\ E^{\mu+m-1} &= x^{\mu+m-1} + R\sum_{\rho=0}^{m-2} x^{\mu+\rho} + \tilde{E}^{\mu+m-1} = \kappa \,. \end{aligned} \tag{9}$$

To keep the notation as simple as possible we assume that the first m patterns are the explicitly stored ones, whereas the rest is automatically stored. Note that the cavity fields in eqs. (9) are the fields generated when the whole *block* is excluded, i.e. used as test block. The eqs. (9) can be solved for x^μ in terms of the cavity fields:

$$x_m^\mu = \frac{R\sum_{\eta=1}^{m-1}(\tilde{E}^{\mu+\eta} - \kappa) - [(m-2)R + 1]\,(\tilde{E}^\mu - \kappa)}{1 + (m-2)R - (m-1)R^2} \,. \tag{10}$$

The subscript m on x_m^μ means that we are talking about a pattern in an m-block.

The next step is to present a test-block to the perceptron. Again there are $(n+1)$ conceivable such block categories. If the test-block is to become an l-block, the cavity fields of the stored patterns will be shifted according to:

$$\tilde{E}^{\mu}_{\text{new}} = \sum_{\lambda'=1}^{l} x^{\lambda'} B^{\mu\lambda'} + \tilde{E}^{\mu} \,. \tag{11}$$

The labeling is to be understood such that the sum is performed over the patterns in the *test*-block. Replacing $\tilde{E}^{\mu}$ with $\tilde{E}^{\mu}_{\text{new}}$ in eq. (10) the corresponding shift in the embedding strengths can be calculated:

$$\begin{aligned} \delta x^{\mu}_{m,l} &:= x^{\mu}_{\text{new}} - x^{\mu} \\ &= \frac{\sum_{\lambda'=1}^{l} \left[R \sum_{\eta=1}^{m-1} B^{(\mu+\eta)\,\lambda'} - [(m-2)R+1]\, B^{\mu\lambda'} \right] x^{\lambda'}}{1+(m-2)R-(m-1)R^2} . \end{aligned} \tag{12}$$

We now know, how the embedding strengths of the stored patterns react to the test-block. Setting up the KTCs for the test-block (analogous to eq. (9)), using the *updated* embedding strengths ($x^{\mu} \rightarrow x^{\mu} + \delta x^{\mu}$) and solving for x_l^{λ}, the embedding strengths can be written as

$$x_l^{\lambda} = \frac{R \sum^{l}_{\substack{\lambda'=1 \\ (\lambda' \neq \lambda)}} \left(\tilde{E}^{\lambda'} - \kappa \right) - [(l-2)\,R+1] \left(\tilde{E}^{\lambda} - \kappa \right)}{\left(1 - \sum_{\sigma=1}^{n} \alpha_{\sigma}\right) [1+(l-2)R-(l-1)R^2]} \,, \quad \lambda \in \{1, \ldots, l\} \,. \tag{13}$$

This is the embedding strength of the *now stored* test-block. After being stored, the test-block is in no way unique: Rather it is as typical as the other blocks, so eq. (13) describes the embedding of an l-block in general (see [1]).

Comparing eqs. (10) and (13) we notice an extra factor of $G := (1-\sum_{\sigma=1}^{n} \alpha_{\sigma})$ in the denominator. This enhancement factor is necessary to compensate the reaction of the system to the test-block. Here, $\alpha_{\sigma} \cdot N$ is the number of explicitly stored patterns coming from σ-blocks, and thus the sum is the total number of explicitly stored patterns. The α_{σ} can be written as

$$\alpha_{\sigma} := \frac{\sigma p}{nN} P_{\sigma} \,, \tag{14}$$

where P_{σ} is the probability that a given block is a σ-block. In general these probabilities depend on κ, n and R.

Before going on with the next step in the recipe, we need to know something about the cavity fields and when a certain block is an m-block. Using the central-limit theorem, it is easy to show that the cavity fields are Gaussian with the correlation matrix

$$M_{ij} = \delta_{ij} + (1 - \delta_{ij})R\,, \tag{15}$$

and thus their probability distribution is given by

$$\rho_R^n(\tilde{E}^1 = y_1, \ldots, \tilde{E}^n = y_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det M}} \exp(-\frac{1}{2}\vec{y}^{\,\mathrm{t}} M^{-1} \vec{y})\,. \tag{16}$$

Here one should note that $\det M$ is > 0 as long as $-1/(n-1) < R < 1$.

3.2 Integration-regions for the cavity-fields

To find out when a block is an m-block we use a kind of self-consistency argument, assuming the cavity fields to be ordered as follows

$$\tilde{E}_1 < \tilde{E}_2 <, \ldots, < \tilde{E}_n\,. \tag{17}$$

This leads to a very systematic ordering of the embedding strengths: For a particular m-block x_m^m is the *smallest* of the non-vanishing embedding strengths. Furthermore we have that $x_{m-1}^{m-1} > x_m^m$. Very roughly this means for example that if the first three patterns in a block are explicitly stored, then surely the first two patterns must be stored etc.

Our self-consistency argument is that for an m-block there have to be at least m positive embedding strengths. So for a given block we have to choose m such that x_m^m is the smallest positive embedding strength. If we would take $(m-1)$, there would be patterns left unstored, and if we would take $(m+1)$, we are overdoing it. In fig. 1 the corresponding integration regions are shown for $n = 2$.

3.3 The probabilities P_m

Having expressed the embedding strengths in terms of the cavity fields (13) and knowing how the cavity fields are distributed, we can write down an equation for P_m. Starting with $m = n$ we simply integrate the distribution $\rho_R^n(y)$ over the region which corresponds to the criterion

$x_n^n > 0$. We then work our way down, remembering to subtract P_n, $P_{(n-1)}$,..., $P_{(m+1)}$. This can be written as

$$P_m = n!\,\mathcal{F} \int\limits_{x_m^m > 0} \rho_R^n(\tilde{E}'_1 \dots \tilde{E}'_n)\, d\tilde{E}'_1 \dots d\tilde{E}'_n - \sum_{i=m+1}^{n} P_i\,, \qquad (18)$$

where we have introduced the field-ordering-operator $\mathcal{F}$, which simply orders the integration variables according to eq. (17). As already mentioned, the corresponding integration regions for the case $n = 2$ are depicted in fig. 1. Results for the probabilities P_m will be presented in section 5.

3.4 An expression for $\alpha(\kappa, n, R)$

Now it is easy to see that eq. (8) can be written as

$$\frac{1}{\alpha} = \frac{1}{n} \sum_i i P_i + \kappa \int x' w(x)\, dx\,. \qquad (19)$$

where x' is defined as the embedding strength in eq. (13) but *without* the factor $G = 1 - \sum \alpha_\sigma$. Writing the embedding strengths in terms of the cavity fields, integrating them over the distribution in eq. (16) and inserting everything into eq. (19) the storage capacity is given by

$$\frac{1}{\alpha} = \frac{1}{n} \sum_i i P_i + \kappa \sum_{i=1}^{n} i!\,\mathcal{F} \int \left[\textstyle\sum_{l=1}^{i} x_l'^i\right] \rho_R^n(\tilde{E}'_1 \dots \tilde{E}'_n)\, d\tilde{E}'_1 \dots d\tilde{E}'_n\,. \quad (20)$$

The integration bounds are to be taken as in eq. (18).

The terms on the r.h.s. of eq. (20) can be interpreted in a rather intuitive way: The first term on the right side of eq. (20) has to do with the degrees of freedom, i.e. at $\kappa = 0$ we have $N = (p/n) \sum_i i P_i$, which means that the degrees of freedom are equal to the number of explicitly stored patterns in the system. If we want to store the patterns with a positive stability, $\kappa > 0$, we have to "pay more" than with just the degrees of freedom, i.e. with the second term on the right side of eq. (20).

4 Asymptotic behaviour for $n \to \infty$

For $\kappa = 0$, α is determined by

$$\frac{1}{\alpha} = \frac{1}{n} \sum_{m=1}^{n} m P_m \,. \tag{21}$$

so we only need to evaluate P_m. For very large n, the complicated criteria for the integration regions, see section 3.2 and eq. (13), can be approximated by

$$x_m^m \to 0 \text{ for } \mathrm{n} \to \infty \;\Rightarrow\; [(\mathrm{m}-1)\mathrm{R}+1]\,\tilde{\mathrm{E}}_\mathrm{m} = \mathrm{R}\sum_{\mathrm{i}=1}^{\mathrm{m}} \tilde{\mathrm{E}}_\mathrm{i}\,. \tag{22}$$

Additionally for each block the cavity fields are Gaussian with variance $(1-R)$ and a certain mean z, see below. For the different blocks, z is also Gaussian with variance R and zero mean. The sum in eq. (21) can now be approximated by an integral. Inserting these assumptions into eq. (22) and doing the math, we arrive at the formula:

$$\alpha \overset{n\to\infty}{\to} \frac{nR}{2(1-R)\ln(\frac{nR\sqrt{1-R}}{\sqrt{2\pi}})}\,. \tag{23}$$

This approximation applies as long as $\ln(nR\sqrt{1-R}/\sqrt{2\pi}) \gg 1$.

5 Results

For $n = 2$, the multiple integral in eq. (20) can easily be evaluated analytically, and if one is patient one might also be able to do it for $n = 3$ analytically. But due to the rather complicated integration bounds in eqs. (18) and (20) this becomes harder as n grows. As we are not very patient, we use a Monte-Carlo integration method for $n > 2$.

5.1 Storage capacity for $n = 2$ and $n = 3$

First we calculate the probabilities P_m using eq. (18). For $n = 2$, the criteria for the integration regions are (see fig. 1 for $\kappa = 0$):

$$x_2^2 > 0 \;\Rightarrow\; (\tilde{E}^2 - \kappa) < R \cdot (\tilde{E}^1 - \kappa) \tag{24}$$

$$x_1^1 > 0 \;\Rightarrow\; \tilde{E}^1 < \kappa\,. \tag{25}$$

The regions in fig. 1 have a very simple interpretation: Assuming we want to store a pair of patterns, we simply calculate their cavity fields (they are identical to the fields produced by patterns *before* they are embedded). If the fields are in region III, both patterns have to be embedded, whereas if the fields are in region II, only one of the patterns has to be explicitly stored. There is a part of region II where *both* fields are negative but still we only have to embed *one* of the patterns (i.e. the one with field $\tilde{E}_1$). Due to the correlations, the other pattern (the one with field $\tilde{E}_2$) will be dragged along as the first one is embedded.

In the special-case of $\kappa = 0$ the probabilities P_2 and P_1 are (eq. (18)):

$$P_2(\kappa = 0) = \frac{1}{2\pi}\arccos(R) \quad \text{and} \quad P_1(\kappa = 0) = \frac{1}{2}\,. \tag{26}$$

According to eq. (20) the storage capacity for $\kappa = 0$ is then given by

$$\frac{1}{\alpha(\kappa = 0)} = \frac{1}{4} + \frac{1}{2\pi}\arccos(R)\,. \tag{27}$$

From this equation it follows that for $|R| < 1$ the slope of α is given by:

$$\frac{\partial \alpha}{\partial R} = -\alpha^2 \frac{\partial 1/\alpha}{\partial R} = \frac{\alpha^2}{2\pi\sqrt{(1+R)(1-R)}}\,. \tag{28}$$

For positive κ the storage capacity is

$$\begin{aligned} \frac{1}{\alpha} \quad &= \quad \left(\frac{P_1}{2} + P_2\right) + \kappa \int (\kappa - x)\rho_R^2(x,y)\Big|_{II} dx\, dy \\ &\quad + \kappa \int \frac{2\kappa - (x+y)}{1+R}\rho_R^2(x,y)\Big|_{III} dx\, dy\,, \end{aligned} \tag{29}$$

where the indexes II and III refer to the regions in fig. 1.

For $n = 3$ the corresponding formulae are omitted, since then we preferred the Monte-Carlo integration (see above). The results for $n = 2$ and $n = 3$ are plotted in fig. 2.

5.2 Results for higher n and positive R

In fig. 3, α is plotted as a function of n for $\kappa = 0$ and three values of R. We see that α is almost linear in n which is confirmed by eq. (23).

In order to compare $\alpha(\kappa = 0, n, R)$ for different block-sizes n we define a scaled parameter α_{S}:

$$\alpha_{\mathrm{S}} := 2 + \frac{\alpha - 2}{n - 1}\,, \tag{30}$$

which leads to $\alpha_{\mathrm{S}}(\kappa = 0, R = 0, n) = 2$ and $\alpha_{\mathrm{S}}(\kappa = 0, R = 1, n) = 4$ for all n. In fig. 4, α_{S} is plotted as a function of R for several different values of n. As n grows α_{S} becomes smaller. Comparing with eq. (23) we see that this is consistent with the logarithmic factor in the denominator.

It is also interesting to see *how* the patterns are stored in the perceptron, so let us look at the probabilities P_m. In fig. 5, P_m is plotted as a function of α. First we notice that for very small α (i.e. for very high κ) almost every block is a 3-block, as it should, because if we want high stability, we cannot rely on the correlations alone to drag the patterns into the stability region. Secondly there seems to be no strict ordering of the P_m. The only thing we can say is that as we lower the stability (rise α) the P_m are ordered, but only for very low α.

In fig. 6, the P_m are plotted as functions of R at $\kappa = 0$ for four values of n. At $R = 0$ the patterns are uncorrelated and the probabilities are given by:

$$P_i(R = 0, \kappa = 0) = \frac{1}{2^n}\frac{n!}{i!(n-i)!}\,. \tag{31}$$

For $R = 1$ we find $P_0 = P_1 = 1/2$. At this point the patterns in each block are identical, so if we store one pattern in a block we have stored them all. A further point about these curves is the (numerical) fact that for a given R, the P_m with odd (even) indexes both add up to 1/2.

Interestingly, for $R = 1/2$ and given n, the present model is isomorphous to the Q-state Potts-perceptron with $Q = n + 1$, see [1]. This is due to similar internal correlations in the structure of the Potts-model, which give rise to an isomorphic learning algorithm.

For the problem of letter recognition, the case of $R < 0$ would be unnatural, so we did not put much emphasis on it for $n > 3$, but for $n = 2$ and $n = 3$ we considered the whole range $-1/(n-1) < R < 1$ (For the lower limit, the handwritten letters would be completely anticorrelated). If the limit $R \to -1/(n-1)$ is taken with $\kappa > 0$, then α converges to 0, but with decreasing κ the slope becomes steeper and steeper, in agreement with eq. (28). If we set $\kappa = 0$ and let

$R \to -1/(n-1)$, α approaches a finite value. From eq. (27) we have $\alpha\,(n=2, \kappa=0, R \to -1) = 4/3$, and with increasing n α decreases further (compare figs. 2a,b), but remains ≥ 1. Of course $\alpha(R=0) = 2$ and $\alpha(R=1) = 2n$ for $\kappa = 0$. At $R = 1$ the slope is infinite (comp. eq. (23) and fig. 4).

6 Conclusions

Using a cavity method, which was recently developed by F. Gerl, [1], we have treated the problem of recognition (or classification) of patterns which are "semantically" correlated in the sense that the set consists of $\mathcal{N} = p/n$ groups, such that the n patterns belonging to the same groups have a finite correlation or anti-correlation described by the overlap R ranging from $-1/(n-1)$ to 1, whereas patterns belonging to different groups are assumed to be uncorrelated. Naturally we have assumed that patterns belonging to the same group get the same classification label, as it would be the case in the recognition of handwritten letters written by n different writers, where $\mathcal{N}$ would be the alphabeth length considered. We have been able to solve the problem of the calculation of the learning capacity $\alpha = p/N$ as a function of the stability parameter κ, the correlation parameter R, and the group size n analytically, in such a way that "only" certain multi-dimensional integrals had to be evaluated. For $n = 2$, the integrations could be performed analytically, and for $n \geq 3$ by Monte-Carlo integration. For $R > 0$ (< 0), respectively, an enhancement (reduction) of the capacity was found. When approaching the *upper* limit of R, i.e. for $R \to 1$, the critical capacities are approached with an infinite slope. On the other hand, as long as the *lower* limit $R = -1/(n-1)$ is approached with a small, but fixed value of $\kappa > 0$, one always has $\alpha \to 0$, although with decreasing κ the slope becomes steeper and steeper, whereas for $\kappa = 0$, α converges to a finite value (≥ 1) as R approaches the lower limit: Thus the order of limits $\kappa \to 0$ and $R \to -1/(n-1)$ is important. For $n \to \infty$, or more precisely $\ln(nR\sqrt{1-R}) \to \infty$, we also derived the asymptotic behaviour of α, with the simple result given in eq. (23).

Finally we would like to stress that for the present calculation of α_c the replica symmetry would not be broken. According to [1] this implies that in the present context our cavity approach is rigorous.

When presenting this work and additional results at the Heidelberg workshop on Neural Networks, May 10-12, 1995, we learned that a similar problem has been studied with a different approach by Lopez *et al.* (1995), who also have submitted meanwhile their results for publication, [5]. According to these authors, the limit of α for $\kappa = 0$ and $R \to -1/(n-1)$ is simply $2n/(2n-1)$, corresponding to the fact that in this limit there are $N_c := p \cdot (n-1)/n$ additional constraints such that the critical capacity is given by $p = 2N - N_c$ instead of $p = 2N$. This is in agreement with our results.

Acknowledgments

We would like to thank H. Heyszenau and B. Schottky for useful discussions. One of us (J. Ó. W.) would like to thank the DAAD for its friendly support.

References

[1] Gerl F and Krey U 1994 *J. Phys. A, Math. Gen.* **27** 7353

[2] Gardner E 1987 *Europhys. Lett.* **4** 481

[3] Anlauf J K and Biehl M 1989 *Europhys. Lett.* **10** 687

[4] Fletcher R 1988 *Practical Methods of Optimization* (New York: Wiley)

[5] Lopez B, Schröder M and Opper M, preprint

Figure Captions

Fig. 1: Integration regions for $n = 2$. Because of the field ordering (see text) we only have to consider the half-plane $\tilde{E}^2 > \tilde{E}^1$. The regions I, II and III correspond to 0-, 1-, and 2-blocks, respectively.

Fig. 2: Storage capacity $\alpha(\kappa = 0, R)$ for block-sizes $n = 2$ and $n = 3$. The asteria are the results of simulations done with the well-known AdaTron-algorithm [3].

Fig. 3: α as a function of n at $\kappa = 0$ for three values of R.

Fig. 4: The scaled storage capacity α_S as a function of R at $\kappa = 0$ for different values of n. **Fig. 5**: The probabilities P_m as a function of α for $R = 0.3$ and $R = 0.9$.

Fig. 6: The probabilities P_m as a function of R at $\kappa = 0$ for $n = 2$, $n = 3$, $n = 4$ and $n = 5$. Note that the R-axis in I ranges from -1 to 1.

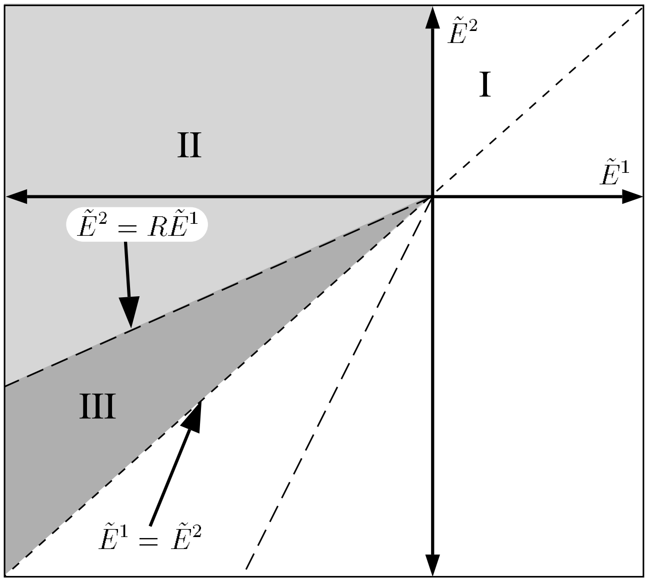

Figure 1:

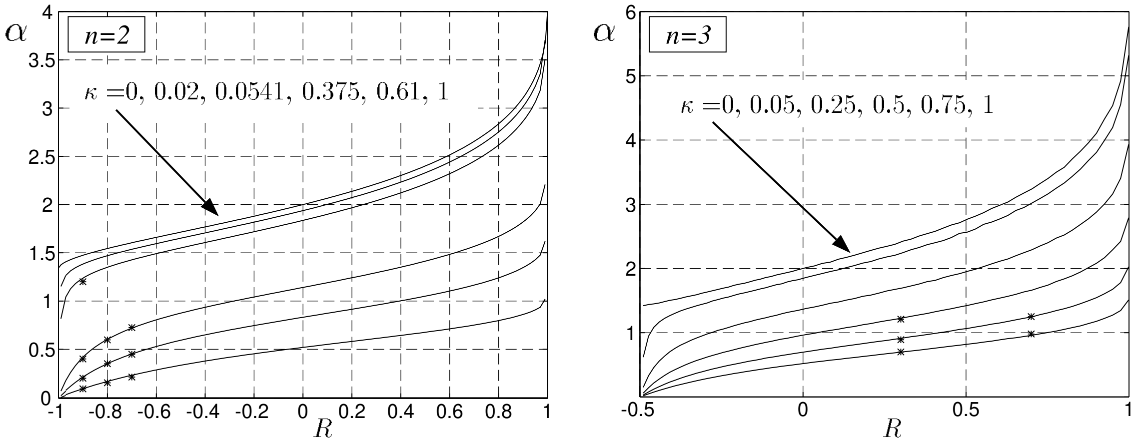

Figure 2:

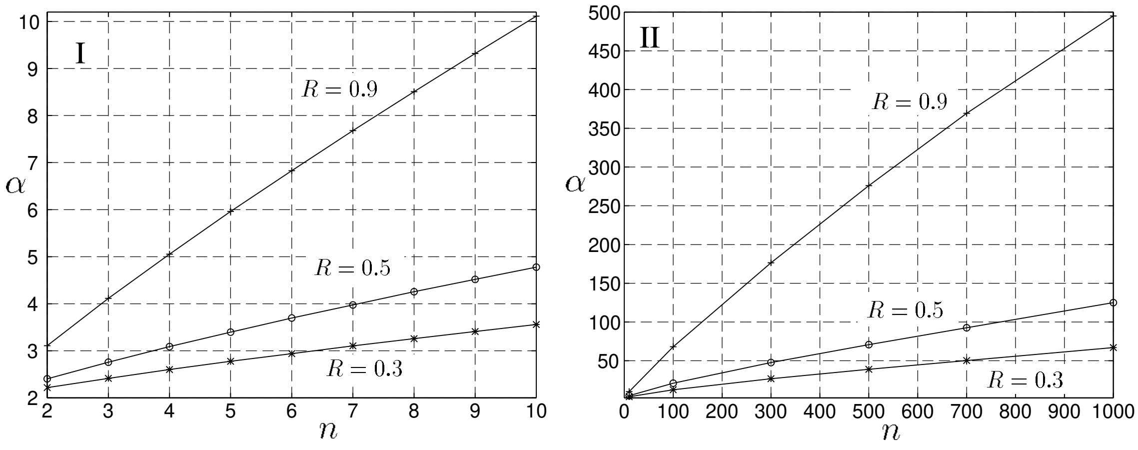

Figure 3:

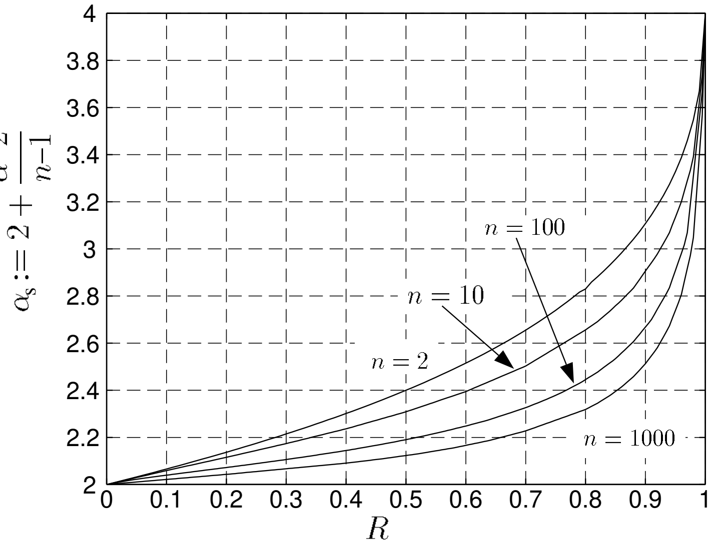

Figure 4:

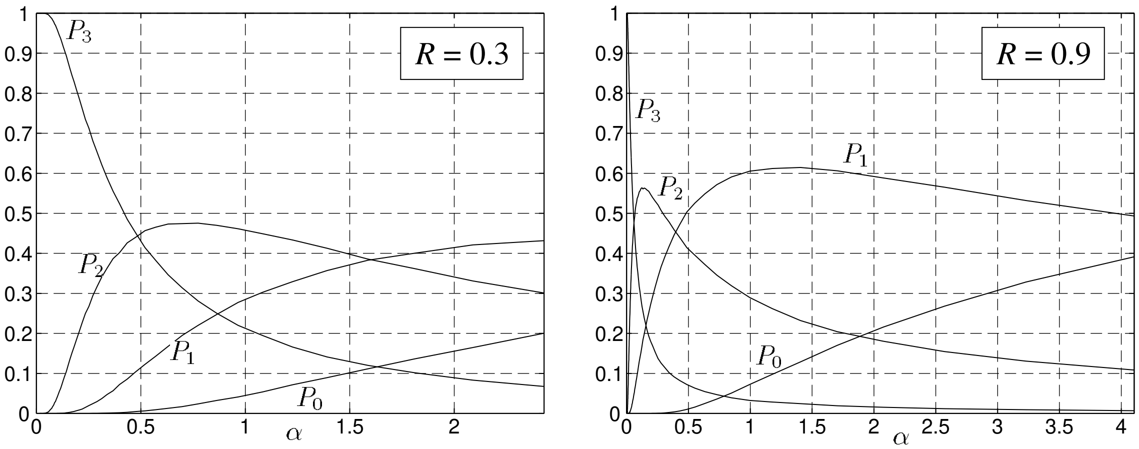

Figure 5:

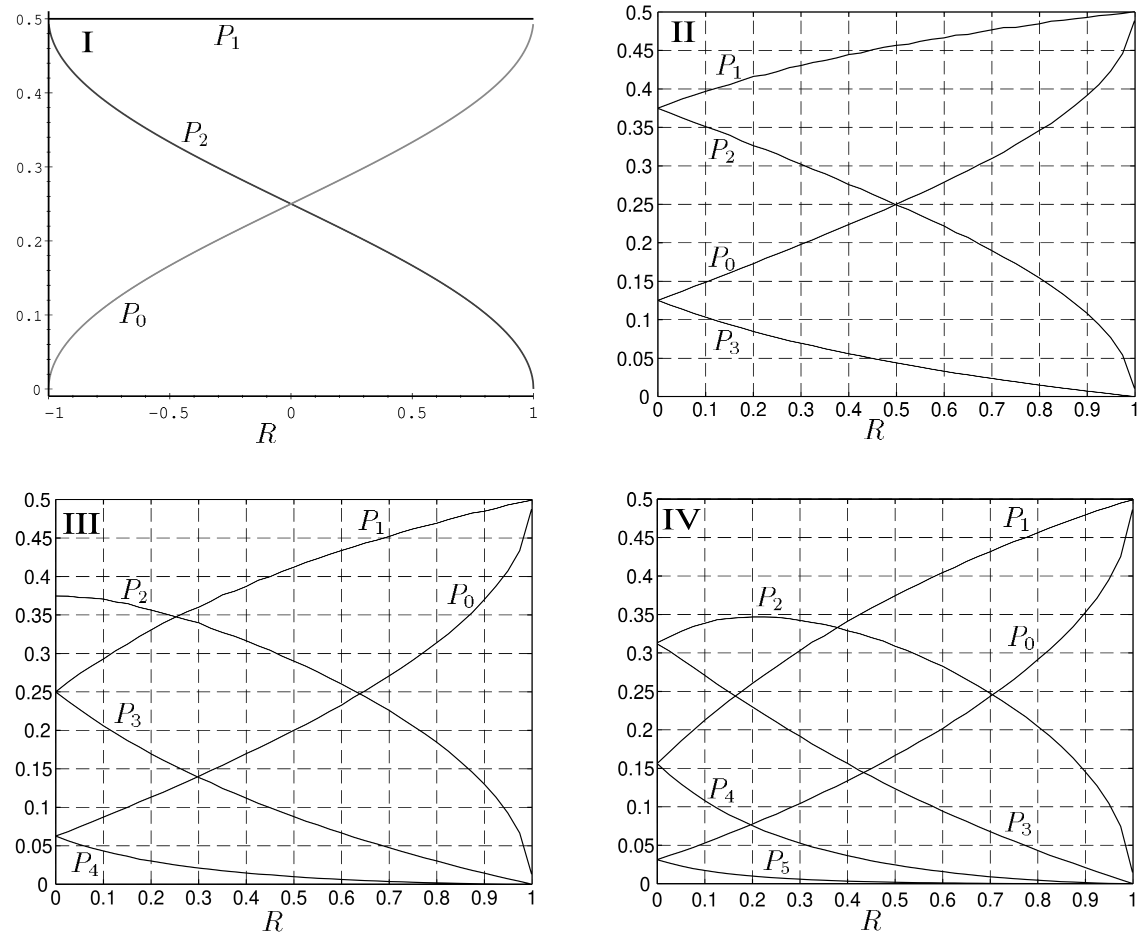

Figure 6: