Storage capacity of a two-layer perceptron with fixed preprocessing in the first layer

1994 J. Phys. A: Math. Gen. 27 1929

(http://iopscience.iop.org/0305-4470/27/6/018)

J. Phys. A: Math. Gen. 27 (1994) 1929–1937. Printed in the UK

Storage capacity of a two-layer perceptron with fixed preprocessing in the first layer

Anthea Bethge, Reimer Kühn and Heinz Horner
Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 19, W-6900 Heidelberg, Germany

Received 25 May 1993

Abstract. It is well known that one can store up to $2N$ uncorrelated patterns in a simple perceptron with N input neurons. Changing the architecture of the net by adding a hidden layer may enhance the storage capacity. Learning in multilayered networks, however, is difficult and slow compared to perceptron learning. In this work a different approach is taken. A large hidden layer with N' neurons is used onto which the patterns are mapped according to a one-to-one code which is fixed beforehand. Only the connections from the hidden layer to the output unit are modified by learning. Here we show how to treat the correlations which are introduced by the coding. We find that the storage capacity of such a net can be made exponentially large. Moreover, our results shed new light on the optimal capacity problem for single-layer perceptrons. We find the optimal capacity to be determined by the dimension of the space spanned by the input patterns, rather than by the size of the input layer.

1. Introduction

Input signals to the brain are processed on various, hierarchical levels. Both convergent and divergent structures exist. Within the visual cortex, for instance, divergent preprocessing of the data is found. Most of the synapses are developed very early during the growth of an organism. Others get changed by learning at a later stage. In this paper we try to integrate both of these features into a model network with a simple architecture. It needs to have at least one hidden layer in order to simulate preprocessing. We choose this layer to be very large, imitating a divergent structure. Random input patterns are mapped onto this layer by a transformation which is fixed. This mapping may be achieved through unsupervised learning in an early stage of development. We are not dealing with learning on this level but rather concentrate on supervised learning on the second level.

Learning in a multilayered network is known to be difficult [1, 2]. Our choice of architecture avoids this problem. The hidden layer and the output unit together form a simple perceptron which has been widely studied. Efficient learning algorithms are known for this problem. Networks of this type may also be of relevance for technical applications. A Gardner-type calculation [3, 4] will yield the storage capacity of the whole network.

The preprocessing of the input patterns will, in general, lead to spatial correlations in the hidden layer, even if the patterns of the input layer were free of any correlations. The problem of storing spatially correlated patterns in a network of perceptron type was recently addressed by Monasson [5] and Lewenstein and Tarkowski [6]. Here we use the general ideas of Monasson to treat spatial correlations due to preprocessing. However, it turns out that the correlation matrices describing such spatial correlations may be singular, rendering the direct approach of Monasson inapplicable in the present case. We propose a

0305-4470/94/061929+09$19.50

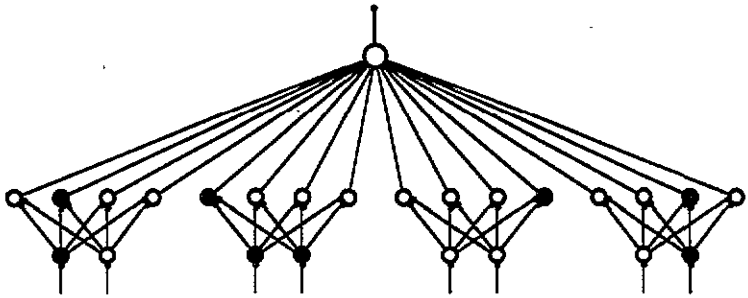

Figure 1. The input layer has N neurons, the hidden layer has $N' = (2^n/n)N$ of them. The connections between input and hidden layer are fixed. An example is given in the next paragraph. The connections between the hidden layer and the output unit are the ones which get changed by learning.

reformulation of Monasson's approach which is capable of dealing with singularly correlated input patterns. They may also be taken as binary, unlike in [5]. For the sake of definiteness we consider a specific coding which exhibits these features. Our results will be seen to also shed new light on the optimal capacity problem for single-layer perceptrons.

The composition of our paper is as follows. In section 2 we discuss the architecture and our coding scheme, and we demonstrate that it leads to spatial correlations with a singular correlation matrix in the hidden layer. In section 3 we present our reformulation of Monasson's approach which is able to cope with such correlations. In section 4 we calculate the optimal storage capacity of our network. Some technicalities are relegated to an appendix. Section 5 finally contains a brief discussion and an outlook on open problems.

2. The network

2.1. The architecture

We study a net with an architecture of the kind shown in figure 1.

2.2. The code

We consider random input patterns each containing N binary digits in $\{\pm 1\}$. These are mapped onto the N' neurons of the hidden layer in the following way. The input neurons are divided into m disjoint modules of n neurons ($N = m \cdot n$). Each module can represent one of $d = 2^n$ different states which all have the same probability. In the hidden layer these d states are represented by d $\{0, 1\}$-neurons of which only one is on ($= 1$). The $(k + 1)$th neuron in a module of the hidden layer is on, when the original module shows the binary representation of the number k. The coding of the four possible states of an $(n = 2)$-module is also shown in figure 1. The modules can be regarded as receptive fields or filters extracting mutually exclusive features.

2.3. The correlations

This coding leads to patterns of low activity, globally and within every module,

$$\langle \xi_i^\mu \rangle = \frac{1}{2^n} = \frac{1}{d} \,. \tag{1}$$

Here, the $\xi_i^\mu \in \{0, 1\}$ denote the pattern bits as represented in the *hidden* layer, i enumerates the neurons of the hidden layer, and μ the patterns. The angular brackets $\langle\rangle$ designate an average over the distribution of these patterns. The patterns are correlated among each other,

$$\langle \xi_i^\mu \xi_j^\nu \rangle = \frac{1}{d^2} \qquad \text{for } \mu \neq \nu\,. \tag{2}$$

One also gets a *spatial* correlation within each pattern. It can be described by the correlation matrix C with elements

$$C_{ij} = \langle \xi_i^\mu \xi_j^\mu \rangle\,. \tag{3}$$

Given our code, C has the following structure:

$$C = \begin{pmatrix} D & A & \dots & \dots & A \\ A & D & A & \dots & \vdots \\ \vdots & A & \ddots & A & \vdots \\ \vdots & \dots & A & \ddots & A \\ \vdots & \dots & \dots & A & D \end{pmatrix} \tag{4}$$

with

$$D = \frac{1}{d} I_d \qquad A = \frac{1}{d^2} 1_d \tag{5}$$

where I_d is the d-dimensional unit matrix and 1_d the d-dimensional matrix with 1 everywhere. The matrix C has three different eigenvalues. These are (including degeneracies)

$$\begin{aligned} \lambda_1 &= \frac{1}{d}\left(1 + N'\frac{1}{d} - d\frac{1}{d}\right) = \frac{N'}{d^2} && \text{single} \\ \lambda_2 &= \frac{1}{d}\left(1 - d\frac{1}{d}\right) = 0 && ((N'/d) - 1)\text{-fold} \\ \lambda_3 &= \frac{1}{d} && (N' - (N'/d))\text{-fold}\,. \end{aligned} \tag{6}$$

The vanishing eigenvalue λ_2 indicates that the correlation matrix is singular and cannot be inverted.

3. Learning correlated patterns

The aim is to calculate the maximum number of correlated patterns which can be stored in the perceptron. Patterns are counted as stored, if the following inequality holds:

$$X_\mu = \zeta^\mu \frac{1}{\sqrt{N'}} \sum_j J_j \xi_j^\mu \geqslant \kappa > 0 \tag{7}$$

where κ is the desired stability of the solution and $\zeta^\mu \in \{\pm 1\}$ the output value. The classification $\xi^\mu \to \zeta^\mu$ is random. These choices yield $\langle X_\mu \rangle = 0$. Otherwise a threshold would have to be introduced. To determine the critical storage capacity, a Gardner-type calculation [3, 4] will be carried out. In order to give a scale to κ, one has to give a scale to the interactions. Here, only interactions satisfying the spherical constraint ($\sum_j J_j^2 = N'$) are considered. Within the space of these interactions one needs to calculate the volume spanned by the couplings solving (7) for a given number of patterns,

$$V = \frac{\int \prod_j \mathrm{d}J_j \prod_\mu \Theta(X_\mu - \kappa)\delta(\sum_j J_j^2 - N')}{\int \prod_j \mathrm{d}J_j\, \delta(\sum_j J_j^2 - N')} . \tag{8}$$

The indices $j(= 1 \ldots N')$ count the neurons and $\mu(= 1 \ldots p)$ the patterns. The δ-distribution fixes the spherical constraint. As $\ln V$ is an extensive variable, one assumes that it is self-averaging. Thus it becomes independent of the special choice of patterns in the limit $N' \to \infty$. Because of the replica identity $\langle \ln V \rangle = \lim_{n\to 0} (1/n) \ln \langle V^n \rangle$ one only needs to calculate $\langle V^n \rangle$. That is, one introduces replicated interactions J_i^σ and correspondingly replicated stabilities X_μ^σ, where the index σ will label the n replica. The averaging is over the correlated patterns ξ^μ,

$$\langle . \rangle = \int \mathrm{d}\xi^\mu p(\{\xi^\mu\})(.) . \tag{9}$$

We can average over the variables X_μ^σ instead, because V depends on the ξ^μ *only* through the X_μ^σ. The first two moments of their joint distribution are

$$\begin{aligned} \langle X_\mu^\sigma \rangle &= 0 \\ \langle X_\mu^\sigma X_\nu^{\sigma'} \rangle &= \delta_{\mu\nu} \frac{1}{N'} \sum_{ij} C_{ij} J_i^\sigma J_j^{\sigma'} \\ &= \delta_{\mu\nu} \frac{1}{N'} \sum_\gamma \lambda_\gamma J_\gamma^\sigma J_\gamma^{\sigma'} = \delta_{\mu\nu} q_{\sigma\sigma'} . \end{aligned} \tag{10}$$

In (10) the λ_γ denote the eigenvalues of C. As C is symmetrical, it can be diagonalized by an orthogonal transformation, $C = U \mathrm{diag}(\lambda_\gamma) U^{-1}$. Orienting the J in the appropriate way, $J_\gamma = \sum_j U_{\gamma j} J_j$, one gets the above expression.

Effecting the ξ^μ average through an average over the X_μ^σ allows us to generalize Monasson's approach to singular correlation matrices. The coding yields a correlation function which is short-ranged. Therefore, relying on the central limit theorem, we can assume a Gaussian distribution for the X_μ^σ, entailing

$$\begin{aligned} p(\{\xi^\mu\}) &= \prod_\mu p(\{X_\mu^\sigma\}_\sigma) \\ &= \prod_\mu \frac{1}{(2\pi)^{(n/2)} \sqrt{\det \mathbf{q}}} \exp\left\{ -\tfrac{1}{2} \mathbf{X} \mathbf{q}^{-1} \mathbf{X} \right\} \end{aligned} \tag{11}$$

because the distribution of the X_μ^σ factorizes in μ. Note that, unlike in [5], the inverse of C is not needed in our formulation to perform the average over the ξ^μ.

In the appendix we evaluate

$$\langle V^n\rangle = \frac{\int \prod_\mu p(\{X_\mu^\sigma\}_\sigma)\,\mathrm{d}\mathbf{X}_\mu[\int \prod_{\gamma\sigma} \mathrm{d}J_\gamma^\sigma \prod_{\mu\sigma}\theta(X_\mu^\sigma-\kappa)\prod_\sigma \delta(\sum_\gamma (J_\gamma^\sigma)^2 - N')]}{\int \prod_{\gamma\sigma}\mathrm{d}J_\gamma^\sigma\,\delta(\sum_\gamma (J_\gamma^\sigma)^2-N')} . \tag{12}$$

The aim is to write the integral in a form which easily yields the free energy

$$G(\mathbf{q}) = \lim_{N'\to\infty} \frac{1}{N'}\langle \ln V\rangle . \tag{13}$$

The influence of the correlations is given only through the eigenvalues of the correlation matrix. For the limit $N'\to\infty$ we will use the notation introduced by Monasson [5]:

$$\frac{1}{N'}\sum_{\gamma=1}^{N'} f(\lambda_\gamma) \overset{N'\to\infty}{\longrightarrow} \{f(\lambda)\}_\lambda . \tag{14}$$

Assuming replica symmetry

$$\begin{aligned} q_{\sigma\sigma} &= q_0 \quad \forall\sigma \\ q_{\sigma\sigma'} &= q_1 \quad \forall\sigma\neq\sigma' \end{aligned} \tag{15}$$

the free energy we get is

$$\begin{aligned} G(q_1, q_0, \hat{q}_1, \hat{q}_0, \hat{u}) = \alpha \int Dz \ln H\left(\frac{\kappa+\sqrt{q_1}z}{\sqrt{q_0-q_1}}\right) + \mathrm{i}q_0\hat{q}_0 - \mathrm{i}\tfrac{1}{2}q_1\hat{q}_1 + \mathrm{i}\hat{u} + \tfrac{1}{2}\ln\pi \\ - \left\{\tfrac{1}{2}\ln\left(\mathrm{i}[\hat{u}+\lambda(\hat{q}_0-\tfrac{1}{2}\hat{q}_1)]\right)\right\}_\lambda - \left\{\frac{\hat{q}_1\lambda}{4[\hat{u}+\lambda(\hat{q}_0-\frac{1}{2}\hat{q}_1)]}\right\}_\lambda \end{aligned} \tag{16}$$

where we use the convention $H(x)=\int_x^\infty Dt$ with the Gaussian measure Dt.

This is a generalization of the result for the uncorrelated case. Setting all $\lambda_\gamma = 1$ we retreive the free energy for the uncorrelated problem. The correlations which imply $\lambda_\gamma \not\equiv 1$ lead to the modified result.

4. Saddle-point equations and solution

Looking for the extremum of the free energy $G(q_1, q_0, \hat{q}_1, \hat{q}_0, \hat{u})$ we get the five saddle-point equations (A3) shown in the appendix. We can eliminate the three conjugate variables algebraically. The two remaining equations can be simplified for $q_1, q_0 \to q_c$. This is the so-called Derrida–Gardner limit. It is motivated by the idea that the volume of solutions in the space of connections will shrink to a point when the number of patterns to be stored increases. Then $q_1 = q_0$. We call this limiting value q_c. Two equations remain which determine both the critical storage capacity and q_c:

$$\begin{aligned} \alpha_c(\kappa,\{\lambda\}) &= \alpha_p\left(\frac{\kappa}{\sqrt{q_c}}\right)\left\{\frac{\lambda q_c}{[\alpha_p(\kappa/\sqrt{q_c})H(-\kappa/\sqrt{q_c})(\lambda-q_c)+q_c]^2}\right\}_\lambda \\ \alpha_c(\kappa,\{\lambda\}) &= \alpha_p\left(\frac{\kappa}{\sqrt{q_c}}\right)\left\{\frac{\lambda^2}{[\alpha_p(\kappa/\sqrt{q_c})H(-\kappa/\sqrt{q_c})(\lambda-q_c)+q_c]^2}\right\}_\lambda \end{aligned} \tag{17}$$

where $\alpha_p(x)$ is the storage capacity of a perceptron storing uncorrelated patterns with stability x ($\int_{-x}^\infty Dz\,(x+z)^2 = 1/\alpha_p(x)$).

This is Monasson's result [5]. We confirm it also for the case in which some of the eigenvalues λ_γ of the correlation matrix C vanish.

4.1. Evaluating the result

Assuming that $q_c \neq 0$ and $\alpha_p(\kappa/\sqrt{q_c}) \neq 1/H(-\kappa/\sqrt{q_c})$, and separating zero from non-zero eigenvalues of C in the evaluation of (17), we get

$$\alpha_c(\kappa, \{\lambda\}) = \alpha_p\left(\frac{\kappa}{\sqrt{q_c}}\right)\left\{\frac{\lambda q_c}{[\alpha_p(\kappa/\sqrt{q_c})H\left(-\kappa/\sqrt{q_c}\right)(\lambda - q_c) + q_c]^2}\right\}_{\lambda\neq 0} \tag{18}$$

where the symbol $\{\cdots\}_{\lambda\neq 0}$ denotes an average as in (14), except that only non-zero eigenvalues are taken to contribute to the sum. For the other equation

$$\alpha_c(\kappa, \{\lambda\}) = \alpha_p\left(\frac{\kappa}{\sqrt{q_c}}\right)\left\{\frac{\lambda^2}{[\alpha_p(\kappa/\sqrt{q_c})H\left(-\kappa/\sqrt{q_c}\right)(\lambda - q_c) + q_c]^2}\right\}_{\lambda\neq 0}. \tag{19}$$

Setting both of them equal, we can calculate q_c. Using (6), and taking the limit $N' \to \infty$, we obtain

$$q_c = \lambda_3 = \frac{1}{d} \tag{20}$$

independently of the stability κ. For the storage capacity we find

$$\alpha_c(\kappa) \equiv \frac{p_{\max}}{N'} = \alpha_p(\kappa\sqrt{d})\left(1 - \frac{1}{d}\right). \tag{21}$$

This result may be interpreted on at least two levels.

First, it gives the optimal storage capacity of a simple perceptron for the storage of singularly correlated structured patterns. A reduction is observed if compared with the optimal capacity $\alpha_p(\kappa)$ of a simple perceptron storing uncorrelated patterns with the same stability. The reduction is twofold. On the one hand, $\alpha_c(\kappa)$ is expressed in terms of α_p, at a stability that is renormalized by a factor of $\sqrt{d} = 1/\sqrt{\lambda_3}$. This leads to a decrease of the number of storable patterns as $\alpha_p(x)$ is a decreasing function of x. On the other hand there is the factor $(1-(1/d))$ which is equal (as $N' \to \infty$) to the fraction of non-zero eigenvalues of C. This may be interpreted in the sense that the rank of C defines the dimension of the space spanned by the input patterns, and thus an *effective* size of the perceptron's input layer (here, the hidden layer of a two-layer machine). It is this effective size $N'(1 - 1/d)$ of the input layer which determines the perceptron's optimal capacity. The reduction of the storage capacity due to the first factor becomes more pronounced, the one due to the second factor less pronounced as d increases. To conclude these considerations, the correlations influence the storage capacity not only by implying $\lambda_\gamma \not\equiv 1$. There is also an effect due to the singularity of the correlation matrix; this effect is *quantitative* and determined by the (reduced) rank of C.

Second, coming back to the original problem, we observe that (21) also yields the number of patterns that can be stored in our two-layer network. The number of storable patterns normalized with respect to the size of the input layer is

$$\tilde{\alpha}_c(\kappa) \equiv \frac{p_{\max}}{N} = \alpha_p(\kappa\sqrt{d})\,\frac{(d-1)}{\log_2 d}. \tag{22}$$

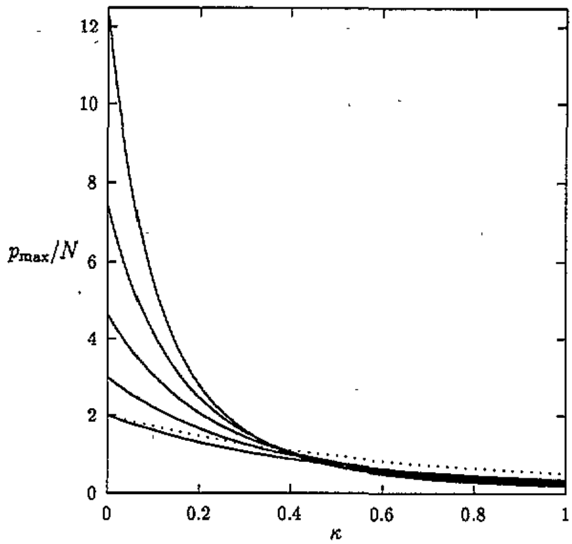

Figure 2. Optimal capacity of our feed-forward architecture as a function of the stability κ, measured relative to the size N of the input layer. The full lines correspond to $d = 2^n$, where, from top to bottom, n decreases from 5 to 1. For comparison, the broken curve shows Gardner's result, $\alpha_p(\kappa)$, for a simple perceptron storing uncorrelated patterns.

This result may well be much larger than $\alpha_p(\kappa)$ as can be seen in figure 2.

At any $\kappa > 0$, however, the storage capacity $\tilde{\alpha}_c(\kappa)$ becomes lower than that for the storage of uncorrelated patterns in a simple perceptron if d is sufficiently large, since $\alpha_p(\kappa)/\tilde{\alpha}_c(\kappa) \sim \alpha_p(\kappa)[\kappa^2 d/(d-1)] \log_2 d > 1$, when $d \to \infty$.

If we set $n = cN$ with $c < 1$, the number of storable patterns at stability $\kappa = 0$ scales exponentially with the size N of the input layer,

$$p_{max} = \frac{2}{c}(2^{cN} - 1) \,. \tag{23}$$

5. Discussion

We have found an easy way to enhance the storage capacity of a simple perceptron when the number of input neurons is fixed. We changed the architecture by introducing a large hidden layer. *It seems possible that the storage ability of the brain may profit from this* type of enhancement by using divergent preprocessing.

We have shown how to calculate the storage capacity of a network with one hidden layer onto which random patterns are mapped according to a fixed code. We have found that for small stabilities the storage capacity may be significantly larger than that of the network without the hidden layer. The enhancement factor depends on the preprocessing, i.e. on our type of coding and within this code on the size of the modules. We noticed that preprocessing leads to correlated, often singularly correlated patterns in the hidden layer. We used a reformulation of Monasson's approach which enabled us to handle singular correlation matrices. We have found that Monasson's result (17) also remains true for singularly correlated patterns. Furthermore we evaluated the result for our type of coding. An analytical expression for the number of storable patterns with finite stability is given by (21) or (22). We found that the correlation matrix influences the storage capacity by implying that its eigenvalues λ_γ need not be equal to 1 as in the uncorrelated case. If C is singular, another modification is given by the reduced fraction of its non-zero eigenvalues. Thus there is a quantitative effect due to the singularity of the preprocessing considered.

If the size of the modules is proportional to the number of the input neurons, the storage capacity scales exponentially with the size of the input layer.

Our approach may also be applied to other types of preprocessing, such as generated by a feed-forward structure with random interactions, which may be taken from $\{\pm 1\}$ or from some larger set. Such kinds of preprocessing may, in fact, be closer to biological data. As yet, we have not investigated them in quantitative detail, however.

It should be noted that our results are stable with respect to replica symmetry breaking. This is because the set of couplings satisfying (7) is convex *and* connected on the surface of the hypersphere defined by the spherical constraint.

Finally, our analysis also naturally applies to simple perceptrons storing spatially correlated patterns. We find the optimal capacity (at $\kappa = 0$) of such systems to be bounded by twice the dimension of the space spanned by the input patterns, i.e. by 2 rank(C), which may be smaller than twice the size of the input layer.

Appendix

To evaluate

$$\langle V^n\rangle = \frac{\int \prod_\mu p(\{X_\mu^\sigma\}_\sigma)\, \mathrm{d}\mathbf{X}_\mu [\int \prod_{\gamma\sigma} \mathrm{d}J_\gamma^\sigma \prod_{\mu\sigma} \theta(X_\mu^\sigma - \kappa) \prod_\sigma \delta(\sum_\gamma (J_\gamma^\sigma)^2 - N')]}{\int \prod_{\gamma\sigma} \mathrm{d}J_\gamma^\sigma\, \delta(\sum_\gamma (J_\gamma^\sigma)^2 - N')} \tag{12}$$

the following steps are taken. We interchange the integrals, see that the X_μ^σ integrals factorize in μ and note that the Θ functions introduce nothing but a lower bound κ to each X_μ^σ integral. Furthermore, we introduce conjugate variables $\hat{q}_{\sigma\sigma'}$ for the $q_{\sigma\sigma'}$ and $\hat{u}_\sigma$ for the Fourier integral of the δ-distribution. The calculations are carried through as in the uncorrelated case. We get

$$\begin{aligned}\langle V^n\rangle = \int \prod_{\sigma\leqslant\sigma'} \frac{\mathrm{d}q_{\sigma\sigma'}\,\mathrm{d}\hat{q}_{\sigma\sigma'}}{2\pi/N'} \int \prod_\sigma \frac{\mathrm{d}\hat{u}_\sigma}{2\pi} \exp\left[\alpha N' \ln \int_\kappa^\infty \prod_\sigma \mathrm{d}X^\sigma\, p(X^\sigma)\right] \\ \times \exp\left[\mathrm{i}\sum_{\sigma\leqslant\sigma'} \hat{q}_{\sigma\sigma'} N' q_{\sigma\sigma'} + \mathrm{i}\sum_\sigma \hat{u}_\sigma N'\right] \\ \times \int \prod_{\gamma\sigma} \mathrm{d}J_\gamma^\sigma \exp\left[-\mathrm{i}\sum_{\sigma\leqslant\sigma'} \hat{q}_{\sigma\sigma'} \sum_\gamma \lambda_\gamma J_\gamma^\sigma J_\gamma^{\sigma'} - \mathrm{i}\sum_\sigma \hat{u}_\sigma \sum_\gamma (J_\gamma^\sigma)^2\right] \\ \times \left\{\int \frac{\mathrm{d}\hat{u}}{2\pi} \int \prod_\gamma \mathrm{d}J_\gamma \exp\left[\mathrm{i}\hat{u}\ (N' - \sum_\gamma J_\gamma^2)\right]\right\}^{-n}.\end{aligned} \tag{A1}$$

The only part of the integral which is affected by the correlations is

$$I(\{\lambda_\gamma\}) = \exp\left[N' \frac{1}{N'} \sum_\gamma \left\{\ln \int \prod_\sigma \mathrm{d}J_\gamma^\sigma \exp\left[-\mathrm{i}\sum_{\sigma\leqslant\sigma'} \hat{q}_{\sigma\sigma'} \lambda_\gamma J_\gamma^\sigma J_\gamma^{\sigma'} - \mathrm{i}\sum_\sigma \hat{u}_\sigma J_\sigma^2\right]\right\}\right]. \tag{A2}$$

The influence of the correlations is given only through the eigenvalues of the correlation matrix. For the limit $N' \to \infty$ we will use the convention

$$\frac{1}{N'}\sum_{\gamma=1}^{N'} f(\lambda_\gamma) \stackrel{N'\to\infty}{\longrightarrow} \{f(\lambda)\}_\lambda . \tag{14}$$

At the saddle point the conjugate variables are imaginary. The eigenvalues equal to zero don't cause any trouble, because the $\hat{u}$ terms don't vanish and thus provide for the convergence of the integral. The integrals can be further evaluated when replica symmetry is assumed,

$$\begin{aligned} q_{\sigma\sigma} &= q_0 \quad \forall\sigma \\ q_{\sigma\sigma'} &= q_1 \quad \forall\sigma \neq \sigma' . \end{aligned} \tag{15}$$

The free energy we get is

$$\begin{aligned} G(q_1, q_0, \hat{q}_1, \hat{q}_0, \hat{u}) = \alpha \int Dz\, \ln H\left(\frac{\kappa + \sqrt{q_1}z}{\sqrt{q_0 - q_1}}\right) + \mathrm{i}q_0\hat{q}_0 - \mathrm{i}\tfrac{1}{2}q_1\hat{q}_1 + \mathrm{i}\hat{u} + \tfrac{1}{2}\ln\pi \\ - \left\{\tfrac{1}{2}\ln(\mathrm{i}[\hat{u} + \lambda(\hat{q}_0 - \tfrac{1}{2}\hat{q}_1)])\right\}_\lambda - \left\{\frac{\hat{q}_1\lambda}{4[\hat{u} + \lambda(\hat{q}_0 - \frac{1}{2}\hat{q}_1)]}\right\}_\lambda \end{aligned} \tag{16}$$

where we use the abbreviation $H(x) = \int_x^\infty Dt$ with the Gaussian measure Dt. This yields the following saddle-point equations:

$$\begin{aligned} \frac{\partial G}{\partial \hat{q}_1} &= 0 \Rightarrow q_1 = \left\{\frac{\mathrm{i}\hat{q}_1\lambda^2}{4[\hat{u} + \lambda(\hat{q}_0 - \frac{1}{2}\hat{q}_1)]^2}\right\}_\lambda \\ \frac{\partial G}{\partial \hat{q}_0} &= 0 \Rightarrow 2\mathrm{i}(q_0 - q_1) = \left\{\frac{\lambda}{[\hat{u} + \lambda(\hat{q}_0 - \frac{1}{2}\hat{q}_1)]}\right\}_\lambda \\ \frac{\partial G}{\partial \hat{u}} &= 0 \Rightarrow -\mathrm{i} + \frac{1}{2}\left\{\frac{1}{[\hat{u} + \lambda(\hat{q}_0 - \frac{1}{2}\hat{q}_1)]}\right\}_\lambda = \frac{\hat{q}_1}{4}\left\{\frac{\lambda}{[\hat{u} + \lambda(\hat{q}_0 - \frac{1}{2}\hat{q}_1)]^2}\right\}_\lambda \\ \frac{\partial G}{\partial q_0} &= 0 \Rightarrow \frac{\alpha}{2}\int Dz \frac{\exp(-\frac{1}{2}u^2)}{\sqrt{2\pi}H(u)}\,\frac{u}{(q_0 - q_1)} = -\mathrm{i}\hat{q}_0 \\ \frac{\partial G}{\partial q_1} &= 0 \Rightarrow \frac{\alpha}{2}\int Dz \frac{\exp(-\frac{1}{2}u^2)}{\sqrt{2\pi}H(u)}\,\frac{uq_0 + \kappa\sqrt{q_0 - q_1}}{q_1(q_0 - q_1)} = \mathrm{i}\frac{1}{2}\hat{q}_1 \end{aligned} \tag{A3}$$

where

$$u = \frac{\kappa + \sqrt{q_1}z}{\sqrt{q_0 - q_1}} .$$

References

[1] Mitchison G J and Durbin R M 1989 *Biol. Cybern.* **60** 345

[2] Barkai E, Hansel D and Sompolinsky H 1992 *Phys. Rev.* A **45** 4146
Engel A, Köhler H M, Tschepke F, Vollmayr H and Zippelius A 1992 *Phys. Rev.* A **45** 7590

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

[4] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257

[5] Monasson R 1992 *J. Phys. A: Math. Gen.* **25** 3701

[6] Lewenstein M and Tarkowski W 1992 *Phys. Rev.* A **46** 2139