Analytical expressions for neural interactions which improve upon the projector rule

1991 J. Phys. A: Math. Gen. 24 2665

(http://iopscience.iop.org/0305-4470/24/11/031)

J. Phys. A: Math. Gen. 24 (1991) 2665–2675. Printed in the UK

Analytical expressions for neural interactions which improve upon the projector rule

A C C Coolen

Department of Medical Physics and Biophysics, University of Nijmegen, Geert Grooteplein Noord 21, NL-6525 EZ Nijmegen, The Netherlands
and
Utrecht Biophysics Research Institute, University of Utrecht, Princetonplein 5, NL-3584 CC Utrecht, The Netherlands

Received 25 November 1990, in final form 4 March 1991

Abstract. In the regime of linearly independent patterns we derive, using normal forms, analytical expressions for neural interaction which improve upon the projector rule (in the sense that all stability parameters are larger). Optimal interactions are shown to correspond to the unique fixed-point of a nonlinear differential equation, the flow of which can be related to one of our expressions.

1. Introduction

The proposal of Little [1] and Hopfield [2] to model neural networks as Ising spin systems, as well as the work of Amit, Gutfreund and Sompolinsky [3] (who were the first to systematically show this proposal to be extremely fruitful) have focused the attention of many physicists on statistical mechanical studies of neural network models [4]. Since Gardner published her paper on the space of interactions [5] two particular problems have been studied intensively [6]: the problem of calculating upper bounds for the performance of neural networks as pattern storage devices, and (as a consequence) the problem of finding neural interactions which saturate these bounds (optimal interactions). Unfortunately, it has not yet been possible to derive analytical expressions for optimal interactions; one has to resort to iterative procedures, described by learning rules [5, 7–9], in order to find the optimal interaction strengths. Because of the highly nonlinear nature of these learning rules one cannot study analytically the evolution of interactions which they generate. At most one can derive estimates for learning times as well as alternative formulations of the problem [10, 11].

What is required of an interaction matrix is that a (large) number p of given N-bit patterns (microscopic network states) be fixed-point attractors under the dynamics of the corresponding network. An indirect measure of the sizes of the domains of attraction is given by the so-called stability parameters, which are simple functions of both the connection matrix and the bits of the patterns to be stored. This has in fact turned the problem of finding optimal interactions into an optimization problem: finding the matrix for which the smallest of these stability parameters is maximal. Instead of trying to tackle this unsolved problem directly, our strategy will be less ambitious, and we will first try to find out how one can improve upon sub-optimal interaction matrices for which analytical expressions do exist. The projector matrix [12–14] seems to be most powerful of these sub-optimal matrices. This connection matrix was shown to be able to stablize any set of linearly independent patterns. Those generalizations of the projector matrix which have been suggested in the literature [15] were aimed at improving generalization properties; they did not provide an improvement in terms of pattern stability. The restriction to linear independence appears to be very fundamental, in view of the fact that individual neurons can only perform linear separations. It is our belief that, before trying to find solutions in the complicated regime of linearly dependent patterns, one should first work on solid ground (linear independence), in the hope that a solution which significantly improves upon the projector matrix might possibly be extended to the complicated regime.

In this paper we consider only the storage of linearly independent patterns. By introducing a normal form we first derive analytical expressions for all optimal interaction matrices which, in addition, satisfy the constraint that the stability parameters be pattern-independent (which is a property of the projector matrix). We show that all matrices in this class are completely equivalent in terms of stability parameters and that the projector matrix is the only symmetric member. Next, by breaking the pattern symmetry, we construct interaction matrices for which at every site the smallest of the stability parameters is significantly larger than the stability parameter of the projector matrix. Finally we use the normal form to show that, for linearly independent patterns, optimal interactions correspond to the unique fixed-point of a nonlinear differential equation, whose initial flow direction is directly related to one of the proposed matrices.

2. A normal form for neural interactions

Our aim is to find all $N \times N$ matrices J such that for a given set of p N-dimensional vectors $\boldsymbol{\xi}^\mu$ (patterns, or specific microscopic spin states):

$$\gamma_{i\mu} \equiv \left(\sum_{j\neq i} J_{ij}^2\right)^{-1/2} \xi_i^\mu \sum_{j\neq i} J_{ij}\xi_j^\mu > 0 \quad \text{for all } i, \mu \tag{1}$$

where

$$\xi_i^\lambda \in \{-1, 1\} \qquad \lambda = 1, \ldots, p \qquad i, j = 1, \ldots, N.$$

In this section we derive a normal form for the solutions of (1), which is also valid in the case of linearly dependent patterns. First we will adopt the following conventions with regard to notation

$$\boldsymbol{\xi}^\mu \equiv (\xi_1^\mu, \ldots, \xi_N^\mu) \in \{-1, 1\}^N$$
$$\boldsymbol{\xi}_i \equiv (\xi_i^1, \ldots, \xi_i^p) \in \{-1, 1\}^p.$$

The $\gamma_{i\mu}$ are the usual stability parameters [5], which are an indirect measure of the stability of pattern μ at site i. Since the problem decouples for the different values of i we can write problem (1) as: find all vectors $\boldsymbol{J}(i)$, so that $a_{i\mu} > 0$ for all (i, μ), where

$$a_{i\mu} \equiv \boldsymbol{J}(i) \cdot \boldsymbol{\eta}^\mu(i) \tag{2}$$

and

$$\boldsymbol{J}(i) \equiv (J_{i1}, \ldots, J_{ii-1}, J_{ii+1}, \ldots, J_{iN}) \in R^{N-1}$$
$$\boldsymbol{\eta}^\mu(i) \equiv \xi_i^\mu(\xi_1^\mu, \ldots, \xi_{i-1}^\mu, \xi_{i+1}^\mu, \ldots, \xi_N^\mu) \in \{-1, 1\}^{N-1}.$$

The relation between $a_{i\mu}$ and $\gamma_{i\mu}$ is given by

$$\gamma_{i\mu} = a_{i\mu}\left(\sum_{j\neq i} J_{ij}^2\right)^{-1/2}.$$

Since (2) is a linear problem, it can be solved easily. If a solution exists, it is given by

$$J_{ij} = (1-\delta_{ij})\frac{1}{N}\sum_{\mu\nu} a_{i\mu}\xi_i^\mu \sum_{n\geq 0}\left(1-\frac{1}{p}C(i)\right)^n_{\mu\nu}\xi_j^\nu + J_{ij}^\perp \tag{3}$$

where:

$$\sum_{j\neq i} J_{ij}^\perp \xi_j^\mu = 0 \qquad a_{i\mu} > 0 \qquad C(i)_{\mu\nu} \equiv \frac{1}{N}\sum_{K\neq i}\xi_K^\mu \xi_K^\nu .$$

All interaction matrices with strictly positive stability parameters are of the above form. However, the converse does not necessarily hold: not all matrices of the above form need to have positive stability parameters. For the standard form (3) we can express the stability parameters in terms of the parameters $a_{i\mu}$

$$\gamma_{i\lambda} = \hat{a}_{i\lambda}\left(\frac{1}{Np}\sum_{\rho\nu}\hat{a}_{i\nu}\xi_i^\nu \sum_{n\geq 0}\left(1-\frac{1}{p}C(i)\right)^n_{\nu\rho}\xi_i^\rho \hat{a}_{i\rho} + \sum_{j\neq i} J_{ij}^{\perp 2}\right)^{-1/2} \tag{4}$$

where

$$\hat{a}_{i\lambda} \equiv \sum_\mu \xi_i^\lambda P(i)_{\lambda\mu} a_{i\mu}\xi_i^\mu \tag{5}$$

and $P(i)$ is the projection on K_i

$$K_i \equiv \langle \xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_N\rangle \subset R^p.$$

It is clear from (4) that adding a perpendicular part $J^\perp$ always reduces the stability parameters, since adding $J^\perp$ leads only to an increase in the magnitude of the matrix elements of J. For this reason optimal interactions must have $J^\perp = 0$:

$$J_{ij} = (1-\delta_{ij})\frac{1}{N}\sum_{\mu\nu} a_{i\mu}\xi_i^\mu \frac{1}{p}\sum_{n\geq 0}\left(1-\frac{1}{p}C(i)\right)^n_{\mu\nu}\xi_j^\nu \tag{6}$$

$$\gamma_{i\lambda} = \hat{a}_{i\lambda}\left(\frac{1}{Np}\sum_{\rho\nu}\hat{a}_{i\nu}\xi_i^\nu \sum_{n\geq 0}\left(1-\frac{1}{p}C(i)\right)^n_{\nu\rho}\xi_i^\rho \hat{a}_{i_\rho}\right)^{-1/2} \tag{7}$$

$$\hat{a}_{i\lambda} \equiv \sum_\mu \xi_i^\lambda P(i)_{\lambda\mu} a_{i\mu}\xi_i^\mu .$$

It is important to emphasize that an eventual choice for the parameters $a_{i\mu}$ may depend on the specific patterns that are to be stored. In fact, this freedom of choosing the parameters as specific functions of the patterns (instead of *a priori* assigning fixed values to the $a_{i\mu}$) will enable us to improve upon the projector matrix.

3. Linearly independent patterns

In the case of linearly independent patterns $K_i = R^p$ and our results become very simple (for $P(i) = 1$ and thus $\hat{a}_{i\lambda} = a_{i\lambda}$):

$$J_{ij} = (1-\delta_{ij})\frac{1}{N}\sum_{\mu\nu} a_{i\mu}\xi_i^\mu C(i)^{-1}_{\mu\nu}\xi_j^\nu \tag{8}$$

$$\gamma_{i\lambda} = a_{i\lambda}\left(\frac{1}{N}\sum_{\rho\nu} a_{i\nu}\xi_i^\nu C(i)^{-1}_{\nu\rho}\xi_i^\rho a_{i\rho}\right)^{-1/2}. \tag{9}$$

Exploring the space of all matrices with positive stability parameters is now completely equivalent to varying the parameters $a_{i\lambda}$ ($a_{i\lambda} > 0$).

To see how the projector matrix fits into this scheme we now, in addition, impose the constraint $\gamma_{i\lambda} = \gamma_i > 0$. If the stability parameters are required to be independent of the pattern index, the equations become

$$J_{ij} = (1 - \delta_{ij}) a_i \frac{1}{N} (\boldsymbol{\xi}_i \cdot C(i)^{-1} \boldsymbol{\xi}_j) \tag{10}$$

$$\gamma_i = \left(\frac{1}{N} \boldsymbol{\xi}_i \cdot C(i)^{-1} \boldsymbol{\xi}_i \right)^{-1/2}. \tag{11}$$

We define the correlation matrix C as

$$C_{\mu\nu} \equiv \frac{1}{N} \boldsymbol{\xi}^\mu \cdot \boldsymbol{\xi}^\nu = C(i)_{\mu\nu} + \frac{1}{N} \xi_i^\mu \xi_i^\nu.$$

From

$$C(i)^{-1}_{\mu\nu} = C^{-1}_{\mu\nu} + \frac{1}{N} (C(i)^{-1} \boldsymbol{\xi}_i)_\mu (C^{-1} \boldsymbol{\xi}_i)_\nu$$

it follows that

$$\boldsymbol{\xi}_i \cdot C(i)^{-1} \boldsymbol{\xi}_j = \left(1 + \frac{1}{N} \boldsymbol{\xi}_i \cdot C(i)^{-1} \boldsymbol{\xi}_i \right) (\boldsymbol{\xi}_i \cdot C^{-1} \boldsymbol{\xi}_j)$$

so (10) can be written as

$$J_{ij} = (1 - \delta_{ij}) a_i \left(1 + \frac{1}{N} \boldsymbol{\xi}_i \cdot C(i)^{-1} \boldsymbol{\xi}_i \right) \frac{1}{N} \boldsymbol{\xi}_i \cdot C^{-1} \boldsymbol{\xi}_j. \tag{12}$$

By choosing $a_i^{-1} = 1 + 1/N \boldsymbol{\xi}_i \cdot C(i)^{-1} \xi_i$, (12) reduces to the definition of the projector matrix. Apparently (10) describes a family of matrices which can be constructed from the projector matrix by rescaling rows. As a consequence the stability parameters (11) are those which correspond to the projector rule. It is a trivial matter to prove that the projector rule is the only symmetric matrix of the form (10) (disregarding an overall scaling factor).

4. Broken pattern symmetry

We found that if the constraint $\gamma_{i\lambda} = \gamma_i$ is imposed it is not possible to improve upon the projector matrix; the latter appeared to be simply the symmetric member of the family of interactions which are optimal with respect to the constraint $\gamma_{i\lambda} = \gamma_i$. This is consistent with the picture depicted by Abbott and Kepler [16], who imposed the peaked distribution of stabilities as constraints in $N \to \infty$ Gardner calculations and found that near criticality there is no way to improve upon the projector rule. Here we find that this statement is also true for finite N. If the aim is to improve upon the projector matrix, we have to abandon the constraint and allow for stability parameters which are pattern dependent. First we rewrite (9)

$$\gamma_{i\lambda}^{-2} = a_{i\lambda}^{-2} \frac{1}{N} \sum_{\mu\nu} \xi_i^\mu \xi_i^\nu a_{i\mu} a_{i\nu} C(i)^{-1}_{\mu\nu}. \tag{13}$$

The smallest stability $\gamma_{i\rho}$ for a given site i is found by taking the index ρ of the smallest component of the vector $\boldsymbol{a}_i$

$$\forall\lambda:\ \gamma_{i\lambda} \geqslant \gamma_{i\rho} \quad \text{if}\ \forall\lambda:\ a_{i\lambda} \geqslant a_{i\rho}.$$

Since (13) is not sensitive to rescaling $|\boldsymbol{a}_i|$ and since (according to (9)) we are only interested in positive $a_{i\rho}$, we can choose a scale where

$$a_{i\rho} = 1 \quad \text{and} \quad \forall\lambda:\ a_{i\lambda} \geqslant 1$$

or

$$a_{i\lambda} = 1 + \varepsilon_{i\lambda} \qquad \varepsilon_{i\lambda} \geqslant 0 \qquad \exists\rho:\ \varepsilon_{i\rho} = 0. \tag{14}$$

We can now write the smallest stability $\gamma_{i,\min} \equiv \gamma_{i\rho}$ (for a given site) as

$$\gamma_{i,\min}^{-2} = \gamma_{i,\text{pro}}^{-2} + \frac{2}{N}\sum_{\mu\nu} \xi_i^\mu \xi_i^\nu C(i)^{-1}_{\mu\nu}\varepsilon_{i\mu} + \frac{1}{N}\sum_{\mu\nu} \xi_i^\mu \xi_i^\nu C(i)^{-1}_{\mu\nu}\varepsilon_{i\mu}\varepsilon_{i\nu} \tag{15}$$

where $\gamma_{i,\text{pro}}$ the stability parameter (11) of the projection matrix. This expression is still completely general. Each choice of $\boldsymbol{\varepsilon}$, in accordance with (14), specifies a model which stabilizes the patterns. Optimal interactions correspond to the optimal choice of $\boldsymbol{\varepsilon}_i$, i.e. the choice for which the right-hand side of (15) is minimal. The pattern-symmetric solutions correspond to taking $\boldsymbol{\varepsilon}_i = \boldsymbol{0}$. We improve upon the pattern-symmetric (projector) solutions if

$$-2\boldsymbol{x}_i \cdot \boldsymbol{\varepsilon}_i + \frac{1}{N}\sum_{\mu\nu} \xi_i^\mu \xi_i^\nu C(i)^{-1}_{\mu\nu}\varepsilon_{i\mu}\varepsilon_{i\nu} < 0 \tag{16}$$

where

$$x_{i\mu} \equiv -\frac{1}{N}\,\xi_i^\mu (C(i)^{-1}\boldsymbol{\xi}_i)_\mu.$$

Since the second term in (16) is always positive, the first term has to be chosen negative. This suggests rather a simple way to break the pattern symmetry

$$\varepsilon_{i\mu} \equiv \varepsilon_i f(x_{i\mu}) \tag{17}$$

where ε_i is positive scaling factor (to be determined later) and f is an arbitrary function which obeys $f(x) > 0$ for $x > 0$ and $f(x) = 0$ for $x < 0$. First we must ensure that our choice (17) is in accordance with (14). By definition, all $\varepsilon_{i\mu}$ are non-negative and there will be at least one $\varepsilon_{i\mu}$ equal to zero, since:

$$\sum_\mu x_{i\mu} = -\frac{1}{N}\,\boldsymbol{\xi}_i \cdot C(i)^{-1}\boldsymbol{\xi}_i < 0$$

(because $C(i)$ is positive definite). Next we must check whether (17) does not simply give the trivial $\boldsymbol{\varepsilon}_i = \boldsymbol{0}$, i.e. we must be sure that not all $x_{i\mu}$ are negative. Suppose, however, that all $x_{i\mu}$ were indeed negative, it would then follow that (16) could have no solutions at all, since both terms of (16) are non-negative. This, in turn, means that $\boldsymbol{\varepsilon}_i = \boldsymbol{0}$ is the optimal choice and that there is no interaction matrix which improves upon the projector rule. We can safely proceed; if improvements upon the projector rule exist (which is obviously the case for randomly drawn patterns [5, 6]), then $\boldsymbol{\varepsilon}_i \neq \boldsymbol{0}$.

Finally we must prove that the inequality (16) is satisfied. To show this we must first find the optimal choice of the overall scaling factor ε_i

$$-2\boldsymbol{x}_i\cdot\boldsymbol{\varepsilon}_i+\frac{1}{N}\sum_{\mu\nu}\xi_i^\mu\xi_i^\nu C(i)^{-1}_{\mu\nu}\varepsilon_{i\mu}\varepsilon_{i\nu}=-2\varepsilon_iA_i+\varepsilon_i^2B_i$$
$$=B_i(\varepsilon_i-A_i/B_i)^2-A_i^2/B_i$$

where

$$A_i\equiv\sum_\mu x_{i\mu}f(x_{i\mu})\tag{18}$$

$$B_i\equiv\frac{1}{N}\sum_{\mu\nu}\xi_i^\mu\xi_i^\nu C(i)^{-1}_{\mu\nu}f(x_{i\mu})f(x_{i\nu})\tag{19}$$

(if $\boldsymbol{\varepsilon}_i\neq\boldsymbol{0}$, then both $A_i>0$ and $B_i>0$). The optimal choice for ε_i is found to be $\varepsilon_i=A_i/B_i$, and, furthermore, with this choice for ε_i we always satisfy (16). We can now summarize our results. If connection matrices exist which improve upon the projector matrix, then for any function $f(x)$ which satisfies

$$f(x)>0\quad\text{for } x>0\quad\text{and}\quad f(x)=0\quad\text{for } x<0\tag{20}$$

we can construct the following matrix, for which at each site i the smallest stability parameter is larger than the corresponding stability parameter of the projection matrix:

$$J_{ij}=(1-\delta_{ij})\frac{1}{N}\sum_{\mu\nu}\xi_i^\mu C(i)^{-1}_{\mu\nu}\xi_j^\nu\left(1+f(x_{i\mu})\frac{\Sigma_\rho x_{i\rho}f(x_{i\rho})}{1/N\,\Sigma_{\rho\lambda}\,\xi_i^\rho\xi_i^\lambda C(i)^{-1}_{\lambda\rho}f(x_{i\rho})f(x_{i\lambda})}\right)\tag{21}$$

where

$$x_{i\mu}\equiv-\frac{1}{N}\xi_i^\mu(C^{-1}(i)\boldsymbol{\xi}_i)_\mu$$

$$\forall i,\mu\colon\ \gamma_{i\mu}^{-2}\leqslant\gamma_{i,\text{pro}}^{-2}-\frac{(\Sigma_\rho x_{i\rho}f(x_{i\rho}))^2}{1/N\,\Sigma_{\rho\lambda}\,\xi_i^\rho\xi_i^\lambda C^{-1}(i)_{\lambda\rho}f(x_{i\rho})f(x_{i\lambda})}.\tag{22}$$

For randomly drawn patterns and large networks ($N\to\infty$) the stability parameters of the projector matrix vanish at the critical point $\alpha=1$ (as soon as the patterns become linearly dependent). This means that near criticality the proposed expression (21) will only be a significant improvement if the second term on the right-hand side of (22) diverges in the same way as $\gamma_{i,\text{pro}}^{-2}$. If this turns out to be the case, then (21) is not only an improvement of the projector matrix, but it also belongs to a different universality class [16]. Furthermore one might expect the specific choice of the function $f(x)$ to become irrelevant (in the limit $N\to\infty$) and both (18) and (19) (and thus the amplitude ε_i) to be self-averaging (i.e. functions of the ratio $\alpha=p/N$ only).

Specific examples

As an illustration of the behaviour of (21) we will now make some explicit choices for the function $f(x)$. One of the simplest choices would be $f(x)=x^n\theta(x)$ (n is not necessarily an integer), which would give

$$\gamma_{i\mu}^{-2}\leqslant\gamma_{i,\text{pro}}^{-2}-\frac{(\Sigma_\rho x_{i\rho}^{n+1}\theta(x_{i\rho}))^2}{1/N\,\Sigma_{\rho\lambda}\,\xi_i^\rho\xi_i^\lambda C^{-1}(i)_{\lambda\rho}\theta(x_{i\rho})\theta(x_{i\lambda})x_{i\rho}^n x_{i\lambda}^n}.\tag{23}$$

Because of the step-function it is extremely difficult to average (parts of) this expression over the pattern components, which would be the usual procedure (assuming (23) to be self-averaging for $N \to \infty$). Therefore we have done numerical calculations of the stability parameters (for $n=0$ and $n=1$ and for randomly drawn patterns), which can be compared to the projector matrix results.

Figure 1 shows a histogram of the distribution $\rho(\gamma)$ of stability parameters for $N=200$ and α ranging from 0.6–1.0 (step size α: 0.1). For $\alpha \geq 1$ the inverse of a correlation matrix is replaced by the pseudo-inverse, according to (9). Clearly, for $N=200$ both the $n=0$ and the $n=1$ matrix yield stability parameters which are significantly larger than the corresponding quantities of the projector matrix. Figure 1 also indicates that the transition from the $\alpha<1$ regime to the $\alpha>1$ regime is less dramatic for the $n=0$ and $n=1$ matrices than for the projector matrix. This is emphasized by figure 2, which shows the distributions of the stability parameters near the critical point $\alpha=1$.

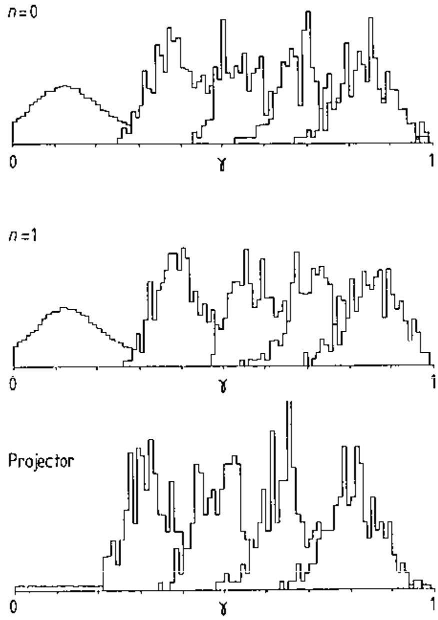

Figure 1. Distribution $\rho(\gamma)$ of stabilities for $N=200$ and randomly drawn patterns. From left to right: $\alpha=1.0$, $\alpha=0.9$, $\alpha=0.8$, $\alpha=0.7$, $\alpha=0.6$.

Figure 2 shows that above the critical point $\alpha=1$ all matrices fail (since all γ's were required to be positive). The distribution $\rho(\gamma)$ of the projector matrix becomes practically flat as soon as the patterns become linearly dependent; for the $n=0$ and $n=1$ matrices the collapse is less dramatic. However, one cannot yet conclude that the latter matrices do not belong to the universality class of the projector rule; such a conclusion could only be drawn if a detailed study were to be made of the N-dependence of the distributions $\rho(\gamma)$. The proposed constructions are still far from being the optimal connection matrix: this can be seen if one compares the results shown in figures 1 and 2 with the $N \to \infty$ smallest stability $\gamma_{\min}(\alpha)$ of the optimal matrix, as given by Gardner's [5] expression:

$$\int_{-\gamma_{\min}(\alpha)}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}z^2)(\gamma_{\min}(\alpha)+z)^2 = \alpha^{-1}.$$

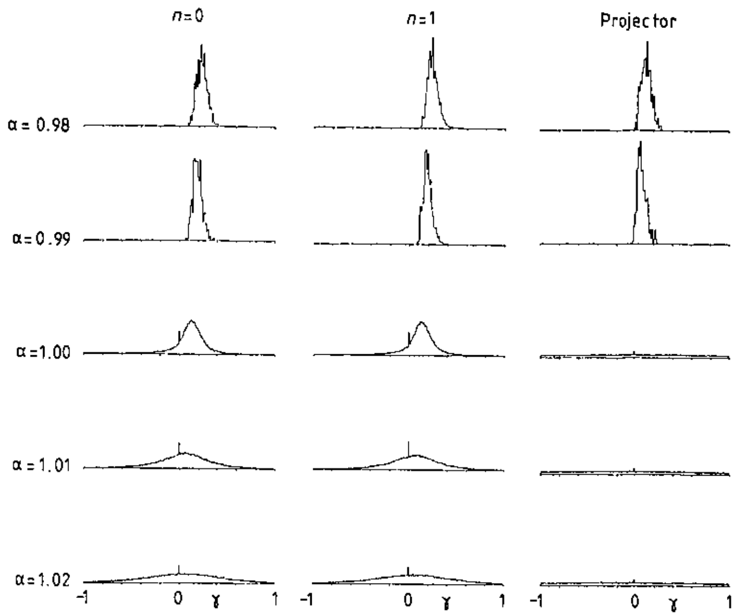

Figure 2. Distribution $\rho(\gamma)$ of stabilities for $N=200$ and randomly drawn patterns near the critical point $\alpha=1$.

This expression predicts $\gamma_{\min}(0.6)\approx 0.875$, $\gamma_{\min}(0.7)\approx 0.747$, $\gamma_{\min}(0.8)\approx 0.640$, $\gamma_{\min}(0.9)\approx 0.549$ and $\gamma_{\min}(1.0)\approx 0.471$.

Finally we studied the behaviour of the average stability $\bar{\gamma}$ (the average over all patterns and all sites). In figure 3 we have drawn the average of this quantity $\bar{\gamma}$ over two $N=200$ trials for $\alpha\in[0,1]$. For $\alpha<0.5$ the difference between the results of the three models is negligible. For $\alpha>0.5$, however, the $n=0$ and $n=1$ curves are again

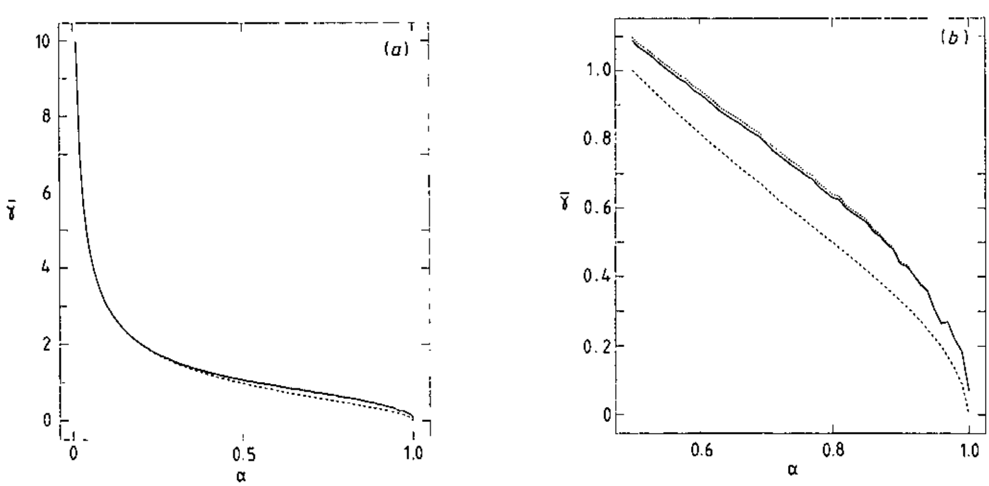

Figure 3. Average stability $\bar{\gamma}$ for $N=200$ as a function of α (averaged over two trials of randomly drawn patterns and over all site- and pattern-indices) for $n=1$ matrix (upper curve), $n=0$ matrix (middle curve) and projector rule (lower curve). (a) $\alpha\in[0,1]$; (b) close-up near the critical point.

clearly above the projector matrix curve (the $n=1$ average stability being highest). Nevertheless, in all three cases $\bar{\gamma}\to 0$ for $\alpha\to 1$.

6. Optimal symmetry breaking

According to (15), in the regime of linearly independent patterns optimal interactions are found by solving the optimization problem:

$$Q(\boldsymbol{\varepsilon})\equiv(1,\ldots,1)\cdot A\boldsymbol{\varepsilon}+\tfrac{1}{2}\boldsymbol{\varepsilon}\cdot A\boldsymbol{\varepsilon}\quad\text{is minimal}\tag{24}$$
$$\varepsilon_\mu\geqslant 0\quad\text{for all }\mu,\qquad \exists\rho:\ \varepsilon_\rho=0.$$

All local minima of Q are solutions of

$$\forall\rho:\left(\varepsilon_\rho=0\quad\text{and}\quad\sum_\lambda A_{\rho\lambda}(\varepsilon_\lambda+1)\geqslant 0\right)\ \text{or}\ \left(\varepsilon_\rho>0\quad\text{and}\quad\sum_\lambda A_{\rho\lambda}(\varepsilon_\lambda+1)=0\right).\tag{25}$$

There can only be one local minimum, for if we assume that at least two such minima $\boldsymbol{\varepsilon}'$ and $\boldsymbol{\varepsilon}''$ exist, then we can construct a path $\boldsymbol{\varepsilon}(\lambda)$ from $\boldsymbol{\varepsilon}'$ to $\boldsymbol{\varepsilon}''$ (the allowed region is convex)

$$\boldsymbol{\varepsilon}(\lambda)=\boldsymbol{\varepsilon}'+\lambda(\boldsymbol{\varepsilon}''-\boldsymbol{\varepsilon}')\qquad \lambda\in[0,1]$$
$$Q(\lambda)=Q(0)+\lambda(\boldsymbol{\varepsilon}''-\boldsymbol{\varepsilon}')\cdot A((1,\ldots,1)+\boldsymbol{\varepsilon}')+\tfrac{1}{2}\lambda^2(\boldsymbol{\varepsilon}''-\boldsymbol{\varepsilon}')A(\varepsilon''-\varepsilon').$$

It is clear that $Q(\lambda)$ can never have a local maximum, which it would have had if both $\lambda=0$ and $\lambda=1$ were to correspond to local minima of Q. We can conclude that there can only be one local minimum, so there is only one solution of (25). The corresponding value of Q would be $Q(\varepsilon)\equiv\frac{1}{2}\boldsymbol{\varepsilon}\cdot A(1,\ldots,1)$. We will now denote by S the set of all indices μ for which $\varepsilon_\mu=0$. If we write A_S for the $(p-|S|)\times(p-|S|)$ matrix which can be constructed from A by taking only entries corresponding to indices which are not in the set S, we find

$$\begin{aligned}&\mu\notin S:\ \varepsilon_\mu=-\sum_{\nu\notin S}(A_S^{-1})_{\mu\nu}\left(\sum_\lambda A_{\nu\lambda}\right)\\&\mu\in S:\ \varepsilon_\mu=0\ \text{(by definition)}\end{aligned}\tag{26}$$

(note that A being positive definite implies that A_S is invertible). We can now write our problem as follows: find the unique index set S for which (26) is the solution of (25), or

$$\rho\in S:\ \sum_\lambda A_{\rho\lambda}\geqslant\sum_{\lambda\nu\notin S}A_{\rho\lambda}(A_S^{-1})_{\lambda\nu}\left(\sum_\mu A_{\nu\mu}\right)$$
$$\rho\notin S:\ \sum_{\nu\notin S}(A_S^{-1})_{\rho\nu}\left(\sum_\lambda A_{\nu\lambda}\right)<0.$$

Since there is only one local minimum of Q we can also find this minimum as the fixed point of the differential equation which describes constrained gradient descent on the surface Q

$$\frac{\mathrm{d}}{\mathrm{d}t}\varepsilon_\mu=-(A[\boldsymbol{\varepsilon}+(1,\ldots,1)])_\mu(\theta^+(\varepsilon_\mu)+(1-\theta^+(\varepsilon_\mu))\theta^+(-(A[\boldsymbol{\varepsilon}+(1,\ldots,1)]_\mu)))\tag{27}$$

where

$$\theta^+(x) = 0 \qquad \text{for } x \leqslant 0$$

$$\theta^+(x) = 1 \qquad \text{for } x > 0.$$

Close inspection reveals that the $n = 1$ matrix, defined in section 4, can in fact be obtained from (27) by choosing as an estimate of the fixed point a vector proportional to $(d/dt)\boldsymbol{\varepsilon}|_{t=0}$ (taking $\boldsymbol{\varepsilon}(0) = \mathbf{0}$). Unfortunately we were not able to find an analytical expression for the fixed-point of (27).

The optimization problem (24) is equivalent to the one formulated by Opper [10]; however, the physical meaning of the solution is different. In [10] the trivial critical point $\boldsymbol{\varepsilon} = \mathbf{0}$ corresponds to the matrix $J_{ij} = 0$, whereas in (24) the trivial critical point corresponds to the projector matrix.

7. Discussion

Using a normal form we have derived analytical expressions for neural interaction matrices which improve upon the projector rule, restricting ourselves to the regime of linearly independent patterns. We have numerically compared the outcome of these matrices with the outcome of the projector rule with respect to the distribution of the stability parameters and the average stability (for $N = 200$ and randomly drawn patterns). The results indicate that the improvement obtained is most significant near the critical point $\alpha = 1$. Finally we have used the normal form to show that optimal interactions correspond to the unique fixed-point of a nonlinear differential equation, of which the $t = 0$ flow direction is directly related to one of the proposed matrices.

The proposed constructions may serve as a first step towards finding an analytical expression for optimal interactions. A next step could be either to further elaborate our scheme for breaking the pattern symmetry in the regime of linearly independent patterns, or to try to find an extension of the proposed constructions in order to handle linearly dependent patterns. It would also be interesting to see how our matrices deal with biased patterns and to find out, in the limit of extreme bias, how they compare with powerful models like the one formulated by Willshaw [17–19].

Acknowledgment

The author would like to thank Theo Ruijgrok for many pleasant and stimulating discussions on this subject.

References

[1] Little W A 1975 *Math. Biosci.* **19** 101
[2] Hopfield J J 1982 *Proc. Natl. Acad. Sci. USA* **79** 2554
[3] Amit D J, Gutfreund H and Somplinsky H 1985 *Phys. Rev.* A **32** 1007; 1985 *Phys. Rev. Lett.* **55** 1530
[4] Amit D J 1989 *Modelling Brain Function: The World of Attractor Neural Networks* (Cambridge: Cambridge University Press)
[5] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[6] Abbott L F 1990 *Network* **1** 105
[7] Diederich S and Opper M 1988 *Phys. Rev. Lett.* **58** 949

[8] Krauth W and Mezard M 1987 *J. Phys. A: Math. Gen.* **20** L745
[9] Anlauf J K and Biehl M 1989 *Europhys. Lett.* **10** 687
[10] Opper M 1988 *Phys. Rev.* A **38** 3824
[11] Kinzel W 1990 *Statistical Mechanics of Neural Networks* ed L Garrido (Berlin: Springer)
[12] Kohonen T 1988 *Self Organisation and Associative Memory* (Berlin: Springer)
[13] Personnaz L, Guyon I and Dreyfus G 1986 *Phys. Rev.* A **34** 4217
[14] Kanter I and Sompolinsky H 1986 *Phys. Rev.* A **35** 380
[15] Vallet F, Cailton J-G and Refregier Ph 1989 *Europhys. Lett.* **9** 315
[16] Abbott L F and Kepler T B 1989 *J. Phys. A: Math. Gen.* **22** 2031
[17] Willshaw D J, Buneman O P and Longuet-Higgins H C 1969 *Nature* **222** 960
[18] Nadal J P and Toulouse G 1989 *Network* **1** 61
[19] Golomb D, Rubin N and Somplinsky H 1989 *Phys. Rev.* A **41** 1843