Non-linear neural networks with external noise

1987 J. Phys. A: Math. Gen. 20 6553

(http://iopscience.iop.org/0305-4470/20/18/051)

J. Phys. A: Math. Gen. **20** (1987) 6553-6560. Printed in the UK

Non-linear neural networks with external noise

J L van Hemmen† and K Rzążewski‡

† Sonderforschungsbereich 123 der Universität Heidelberg, D-6900 Heidelberg, Federal Republic of Germany and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA

‡ Institute of Theoretical Physics, Polish Academy of Sciences, Al Lotników 32/46, PL-02-668 Warsaw, Poland

Received 16 April 1987, in final form 20 July 1987

Abstract. The performance of neural network models with arbitrary non-linearity and Gaussian external noise superimposed on the synaptic efficacies is analysed. The memory function, though surprisingly robust, gradually fades out as the noise level is increased. In the low-noise limit the best performance is at zero temperature. There is a noise range, however, where optimal performance is obtained at a *non*-zero temperature.

Robustness with respect to both input data errors and internal failures is one of the most prominent—and most valuable—characteristics of an associative memory [1-8]. The neurons being modelled [9] by Ising spins $S(i)$, $1 \leqslant i \leqslant N$, the basic idea [1] behind this robustness is that recalling a memory is equivalent to a downhill motion in a free-energy landscape of a Hamiltonian

$$H_N = -\tfrac{1}{2} \sum_{i \neq j} \mathscr{J}_{ij} S(i) S(j) \tag{1}$$

with suitable symmetric couplings $\mathscr{J}_{ij} = \mathscr{J}_{ji}$. The patterns to be stored in the $\mathscr{J}_{ij}$ are specific spin configurations, associated with the 'bottom' of certain free-energy valleys. The domain of attraction of a pattern is the whole free-energy valley.

Locality [10, 11] is an important physiological requirement which implies that $\mathscr{J}_{ij}$ is determined by the local information available to neurons i and j only. Then [8, 12]

$$\mathscr{J}_{ij} = N^{-1} Q(\boldsymbol{\xi}_i; \boldsymbol{\xi}_j) \tag{2}$$

for some *synaptic kernel* $Q(\boldsymbol{x}; \boldsymbol{y}) = Q(\boldsymbol{y}; \boldsymbol{x})$. Here $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$ are vectors with components $\xi_{i\alpha}$ and $\xi_{j\alpha}$, respectively, where $1 \leqslant \alpha \leqslant q$ labels the stored patterns. The $\xi_{i\alpha}$ are independent, identically distributed random variables which take the values ± 1 with equal probability; for the ensuing argument, however, this is not necessary [8, 12, 13]. A model is called non-linear if the introduction of new data requires a *non*-linear operation on Q. The original Hopfield model [1] is linear whereas, for instance, clipped synapses constitute a non-linear model (see below).

Neither a considerable alteration of the input data (the patterns) nor a dilution of the bonds (the synaptic efficacies) noticeably deteriorates the memory function [1, 4, 5, 8]. There is, however, yet another source of errors: external noise, which *modifies* the $\mathscr{J}_{ij}$. In this paper we study the case where

$$\mathscr{J}_{ij} = N^{-1} Q(\boldsymbol{\xi}_i; \boldsymbol{\xi}_j) + \varepsilon b_{ij} \tag{3}$$

for *general* non-linearity, i.e. arbitrary synaptic kernel Q. The b_{ij}, which represent the noise, are taken to be independent identically distributed Gaussians, with mean zero and variance N^{-1}, as in the Sherrington–Kirkpatrick (SK) model [14]. The non-linearity in Q will be treated exactly while the noise will be handled in the replica-symmetric approximation, which for moderate values of ε will prove to be quite accurate.

We first turn to the results, which are formulated most conveniently in terms of a specific model. They are, however, valid in a much wider context. We take Q to represent an inner-product model [12]

$$Q(\boldsymbol{x};\boldsymbol{y}) = \sqrt{q}\,\phi(\boldsymbol{x}\cdot\boldsymbol{y}/\sqrt{q}) \tag{4}$$

and concentrate on a non-linear but technologically highly interesting model, *clipped synapses*, with $\phi(x) = \mathrm{sgn}(x)$. Note that the linear Hopfield model has $\phi(x) = x$.

The case of finitely many patterns already turns out to be quite interesting. Throughout most of this work we therefore take q to be finite, even when $N \to \infty$. If equation (4) holds, then all stored patterns bifurcate from zero [8, 12] (i.e. the paramagnetic phase) at the same temperature $T_c = \Lambda_1$. For clipped synapses [8], $\Lambda_1 \approx (2/\pi)^{1/2} = 0.80$ is proportional to the *largest* eigenvalue of Q. Figure 1 shows the phase diagram. There is a paramagnetic phase (P) where no information can be retrieved since thermal motion destroys every free-energy valley. There is also a spin-glass phase (SG) where again no information can be retrieved since the noise dominates, and there is a retrieval phase (R), where the quality of recovery can be measured by an order parameter $m(\varepsilon, T)$, which is a function of the noise level ε and the temperature T. The closer m is to one, the better the retrieval. For $\varepsilon > \varepsilon_c = (2/\pi)^{1/2}\Lambda_1$, *no* pattern can be retrieved at zero temperature. The critical line R-SG for $\varepsilon_c < \varepsilon < \Lambda_1$ is second order.

Figure 2 shows a plot of $m(\varepsilon, T=0)$. The memory function is very robust. At $T=0$ the error fraction, which is given by $\frac{1}{2}(1-m)$, is less than 0.5% if $\varepsilon \leqslant 0.39\Lambda_1$, so excellent performance is guaranteed for $\varepsilon \lesssim \frac{1}{2}\varepsilon_c$.

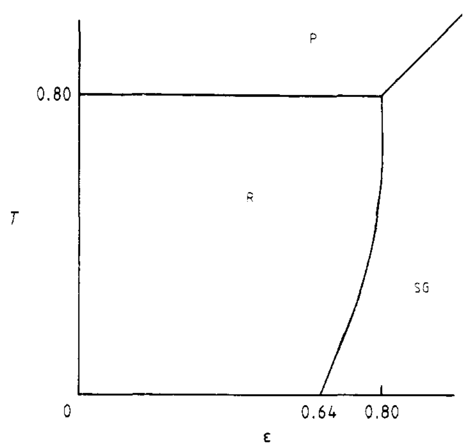

Figure 1. Phase diagram for finitely many patterns stored in clipped synapses with external noise. For high enough temperature only a paramagnetic phase (P) exists and no information can be retrieved. For $T < \Lambda_1 = 0.80$ and $\varepsilon < \varepsilon_c = 0.80\Lambda_1 = 0.64$ there is a stable retrieval phase (R) down to $T = 0$. If the noise level ε is too high ($\varepsilon > \varepsilon_c$), only a spin-glass phase (SG) exists. To the left of ε_c the retrieval phase is stable. The critical line R-SG is second order.

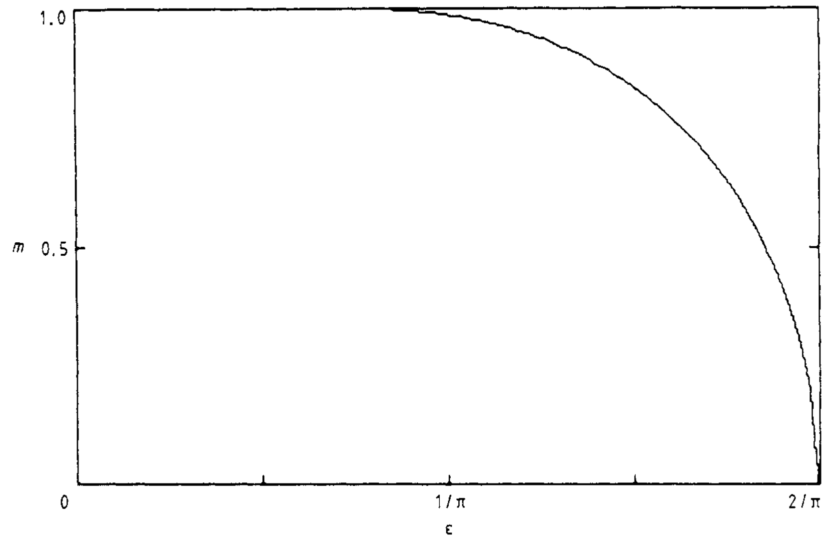

Figure 2. Zero-temperature retrieval quality m as a function of ε. The error fraction is given by $\frac{1}{2}(1-m)$. There is a continuous transition to the spin-glass phase at $\varepsilon=\varepsilon_c$. For clipped synapses, $\varepsilon_c=2/\pi=0.64$.

Fixing $\varepsilon=0.55$ we show m in figure 3 as a function of T. We took ε rather near to $\varepsilon_c=0.64$ (clipped synapses). At $T_c=\Lambda_1$, $m(\varepsilon, T)$ has a square root singularity. But, the maximum of $m(\varepsilon, T)$ is *not* at $T=0$ but at a *finite* non-zero T, which is rather surprising. In figure 4 we show $m(\varepsilon, T)$ as a function of both ε and T. The effect displayed in figure 3 is clearly brought out.

Before turning to the arguments leading to these results we note that the case of extensively many patterns [15] is now relatively well understood and does not produce essentially new results. In the case of the inner-product models (4) there is a *universal* function $F(x)$ which determines the retrieval quality at $T=0$ [15]. It contains a constant C which only depends on the synaptic function ϕ. The very same constant

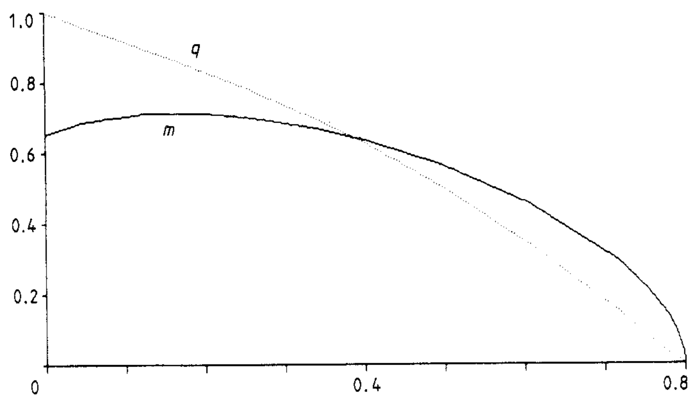

Figure 3. Retrieval quality m (full curve) and spin-glass order parameter q (dotted curve) as a function of the temperature T for $\varepsilon=0.55<\varepsilon_c=0.64$. The temperature varies between zero and $T_c=\Lambda_1$, where for clipped synapses $\Lambda_1=(2/\pi)^{1/2}=0.80$. Note that the maximum of m occurs at *non*-zero temperature.

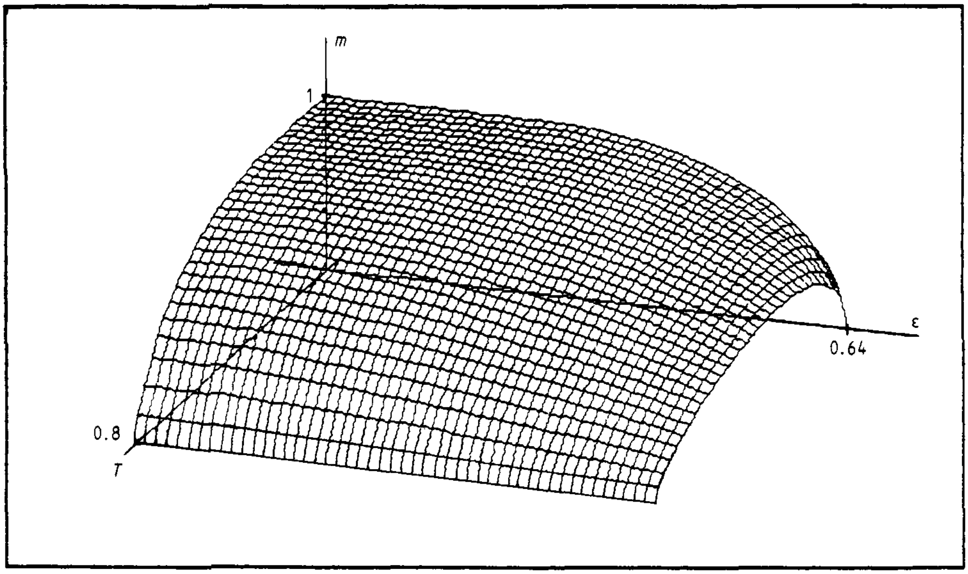

Figure 4. Retrieval quality m as a function of the noise level ε and the temperature T for clipped synapses. In the noise range near $\varepsilon_c = 0.64$ the maximum of m occurs at a non-zero temperature; see figure 3.

C occurs in the equations which govern the finite-temperature behaviour. Adding external noise simply means that we replace C by $(C+\varepsilon^2)$.

For ease of reading we have divided the argument below into several steps.

(*a*) *The replica method.* Let $Z_N = \mathrm{Tr}\exp(-\beta H_N)$ be the partition function, a sum over all Ising spin configurations. Furthermore, let $\phi_N(n) = N^{-1}\ln\langle Z_N^n\rangle$ for positive integer n. The replica method [16] consists of sending N to infinity first, extending the limit $\phi(n)$ to a neighbourhood of $n=0$ and obtaining the free energy $-\beta f(\beta)$ by calculating the derivative $\phi'(0)$. The angular brackets in $\langle Z_N^n\rangle$ denote an average over the Gaussian noise. Performing this average, we obtain

$$\langle Z_N^n\rangle = C_N(n)\,\mathrm{Tr}\exp(-\beta H_N^{(n)}) \tag{5}$$

where $C_N(n)$ is a constant which we drop for the moment, and

$$-\beta H_N^{(n)} = N\left[\tfrac{1}{2}\beta N^{-2}\sum_{i,j} Q(\boldsymbol{\xi}_i;\boldsymbol{\xi}_j)\left(\sum_{\sigma=1}^{n} S_\sigma(i)S_\sigma(j)\right)\right.$$
$$\left. +\tfrac{1}{2}(\beta\varepsilon)^2\sum_{(\sigma,\sigma')}\left(N^{-1}\sum_{i=1}^{N} S_\sigma(i)S_{\sigma'}(i)\right)^2\right]. \tag{6}$$

Here σ and σ' label the n replicas and (σ,σ') ranges through all pairs.

(*b*) *The synaptic kernel.* We now have to handle the non-linearity in Q [8, 12, 13]. Let $\mathcal{O}$ denote the (discrete) set of outcomes $\boldsymbol{x}$ of the random vector $\boldsymbol{\xi}$ and let $p(\boldsymbol{x})$ be their probability. For instance, if the $\xi_{i\alpha}$ are ±1 with equal probability, then $\mathcal{O}=\{-1,1\}^q$ and $p(\boldsymbol{x})=2^{-q}$ for all $\boldsymbol{x}$. We partition the index set $\{1,\ldots,N\}$ into disjoint subsets

$$I(\boldsymbol{x}) = \{i : \boldsymbol{\xi}_i = \boldsymbol{x}\} \tag{7}$$

$x \in \mathcal{C}$, and define order parameters

$$m_\sigma(x) = |I(x)|^{-1} \sum_{i \in I(x)} S_\sigma(i) \tag{8a}$$

$$q_{\sigma\sigma'}(x) = |I(x)|^{-1} \sum_{i \in I(x)} S_\sigma(i) S_{\sigma'}(i). \tag{8b}$$

These allow us to rewrite the Hamiltonian (6)

$$-\beta H_N^{(n)} = N\left[\tfrac{1}{2}\beta \sum_\sigma \sum_{x,y} m_\sigma(x) p(x) Q(x;y) p(y) m_\sigma(y) + \tfrac{1}{2}(\beta\varepsilon)^2 \sum_{(\sigma,\sigma')} \left(\sum_x p(x) q_{\sigma\sigma'}(x)\right)^2\right]. \tag{9}$$

Equation (9) may be summarised in words by stating that $-\beta H_N^{(n)}$ is a function $F(\vec{m}, \vec{q})$ of the *intensive* variables $\vec{m} = (m_\sigma(x);\ x \in \mathcal{C},\ 1 \leqslant \sigma \leqslant n)$ and $\vec{q} = (q_{\sigma\sigma'}(x);\ x \in \mathcal{C},\ 1 \leqslant \sigma < \sigma' \leqslant n)$ times the extensive variable N.

(*c*) *Thermodynamic limit.* As $N \to \infty$, because q is fixed, the size $|I(x)|$ of the set $I(x)$ is $p(x)N$ and thus grows linearly with N. Furthermore, to each x there belongs a group of order parameters $m_\sigma(x)$ and $q_{\sigma\sigma'}(x)$, and for different x these are not directly correlated. More precisely, as stochastic variables they are independent. We may therefore apply a large-deviations argument [8, 12, 13, 17] to each of the $I(x)$ *separately* and then 'glue' the parts together. That is, instead of the Ising spin variables $S_\sigma(i)$ we can introduce new variables $\vec{m}$ and $\vec{q}$ whose common distribution is given by the density

$$\mathcal{D}(\vec{m}, \vec{q}) = \prod_x \mathcal{D}(\vec{m}(x), \vec{q}(x)) = \exp\left[-N\left(\sum_x p(x) c^*(\vec{m}(x), \vec{q}(x))\right)\right] \tag{10}$$

where c^* is the Legendre transform [17] of the convex c function

$$c(\vec{u}, \vec{v}) = \ln \operatorname{Tr} \exp\left(\sum_{\sigma=1}^{n} u_\sigma S_\sigma + \sum_{(\sigma,\sigma')} v_{\sigma\sigma'} S_\sigma S_{\sigma'}\right). \tag{11}$$

Here $\vec{u}$ and $\vec{v}$ are vectors in $\boldsymbol{R}^n$ and $\boldsymbol{R}^{n(n-1)/2}$ respectively. As $N \to \infty$, an integration over $\vec{m}$ and $\vec{q}$ replaces the ordinary trace and $\phi_N(n) = N^{-1} \ln\langle Z_N^n \rangle$ converges to

$$\phi(n) = \lim_{N\to\infty} N^{-1} \ln \int d\vec{m}\, d\vec{q}\, \mathcal{D}(\vec{m}, \vec{q}) \exp\{NF(\vec{m}, \vec{q})\}$$

$$= \sup_{\vec{m},\vec{q}} \left(F(\vec{m}, \vec{q}) - \sum_x p(x) c^*(\vec{m}(x), \vec{q}(x))\right). \tag{12}$$

The supremum in (12) is realised for all those $(\vec{m}, \vec{q})$ which satisfy the fixed-point equation

$$(\vec{m}(x), \vec{q}(x)) = \nabla c(\beta \vec{a}(x), (\beta\varepsilon)^2 \vec{q}(x)) \tag{13}$$

where, for later usage, we define

$$a_\sigma(x) = \sum_y Q(x;y) p(y) m_\sigma(y) \tag{14a}$$

$$q_{\sigma\sigma'}(x) = \sum_y p(y) q_{\sigma\sigma'}(y). \tag{14b}$$

Taking advantage of (13) one can use a simple argument (see § III.A of [7]) to rewrite (12) in the form

$$\phi(n) = \max_{\vec{m},\vec{q}}\Big(-F(\vec{m},\vec{q}) + \sum_x p(x)c(\beta\vec{a}(x), (\beta\varepsilon)^2\vec{q}(x))\Big). \tag{15}$$

One has to choose that solution to the fixed-point equation (13) which maximizes (15). In equations (13)-(15), $\vec{a}(x)$ is a vector with components $a_\sigma(x)$, $1 \leqslant \sigma \leqslant n$ and $\vec{q}(x) = (q_{\sigma\sigma'}(x);\ 1 \leqslant \sigma < \sigma' \leqslant n)$.

(*d*) *Replica symmetry.* Assuming replica symmetry we can drop all the indices σ and σ' from $\vec{m}$ and $\vec{q}$ in equations (9)-(15). This assumption is consistent with the fixed-point equation (13). Performing the 'evident' real variable extension of $\phi(n)$ to a neighbourhood of $n=0$ and including the $C_N(n)$ of equation (5) we then find

$$-\beta f(\beta) = \lim_{n\to 0} n^{-1}\phi(n) = \tfrac{1}{4}(\beta\varepsilon)^2(1-\mathscr{q})^2 - \tfrac{1}{2}\beta J Q(\boldsymbol{m}) + \sum_x p(x)\langle \ln\{2\cosh[\beta(a(x)+\varepsilon\sqrt{\mathscr{q}}z)]\}\rangle \tag{16}$$

where $\boldsymbol{m}$ and $\mathscr{q} = \Sigma_x p(x)q(x)$ satisfy the fixed-point equation

$$\begin{aligned} m(x) &= \langle \tanh[\beta(a(x)+\varepsilon\sqrt{\mathscr{q}}z)]\rangle \\ q(x) &= \langle \tanh^2[\beta(a(x)+\varepsilon\sqrt{\mathscr{q}}z)]\rangle \end{aligned} \tag{17}$$

for all $x \in \mathscr{C}$. In (16) and (17) and throughout the following, angular brackets denote a Gaussian average over a single z with mean zero and variance one. $Q(\boldsymbol{m})$ is the double sum involving Q in (9). By their very definition (8), the $m(x)$ govern the retrieval quality. The $q(x)$ are spin-glass order parameters—exactly as in the SK model, to which (16) and (17) reduce if $\boldsymbol{m}=0$.

(*e*) *Energy and entropy.* The energy $u(\beta)$ is easily obtained through the relation $u(\beta) = (\partial/\partial\beta)(\beta f(\beta))$

$$u(\beta) = -\tfrac{1}{2}Q(\boldsymbol{m}) - \tfrac{1}{2}\beta\varepsilon^2(1-\mathscr{q}^2). \tag{18}$$

The entropy $s(\beta)$ then follows from $f = u - Ts$ or through $s(\beta) = \beta^2(\partial/\partial\beta)f(\beta)$. Both methods give

$$s(\beta) = -[\tfrac{1}{2}\beta\varepsilon(1-\mathscr{q})]^2 + s_0(\beta) \tag{19}$$

where

$$s_0(\beta) = -\beta Q(\boldsymbol{m}) - (\beta\varepsilon)^2\mathscr{q}(1-\mathscr{q}) + \sum_x p(x)\langle \ln\{2\cosh[\beta(a(x)+\varepsilon\sqrt{\mathscr{q}}z)]\}\rangle. \tag{20}$$

(*f*) *Zero-temperature limit.* One can show that $s_0(\beta) \geqslant 0$ and $s_0(\beta) \to 0$ as $\beta\to\infty$. The first term in (19) is negative and converges to a finite *non*-zero limit as $\beta\to\infty$. We now determine this limit.

We first note that $q(x)$ and hence $\mathscr{q}$ converge to one as $\beta\to\infty$. Let $\mathscr{C}(\beta;x) = \beta(1-q(x))$. This expression also converges to a finite limit

$$\lim_{\beta\to\infty} \varepsilon\mathscr{C}(\beta;x) = (2/\pi)^{1/2}\exp[-\tfrac{1}{2}(a(x)/\varepsilon)^2] < 1. \tag{21}$$

The zero-temperature entropy $s(\infty)$ is then given by

$$\lim_{\beta\to\infty} s(\beta) = \lim_{\beta\to\infty} -\tfrac{1}{4}[\beta\varepsilon(1-\mathscr{q})]^2 \equiv -\tfrac{1}{4}\mathscr{C}^2 \tag{22}$$

where, by virtue of (21) and the fact that $q = \Sigma_x p(x) q(x)$,

$$\mathscr{E} = \sum_x p(x)(2/\pi)^{1/2} \exp[-\tfrac{1}{2}(a(x)/\varepsilon)^2]. \tag{23}$$

$\mathscr{E}$ is strictly smaller than $(2/\pi)^{1/2}$ unless $a(x)$ vanishes for all x. In the pure spin-glass phase $\boldsymbol{m} = 0$ and the network has lost its memory completely. Then $\mathscr{E} = (2/\pi)^{1/2}$ and $s(\infty) = -1/2\pi$, which is the value for the replica-symmetric solution to the SK model [14]. We will see shortly that for not too large a noise level (i.e. ε), $s(\infty)$ is orders of magnitude closer to zero than the SK value.

Turning to $m(x)$ in (17) we observe that, as $\beta \to \infty$, $\tanh\{\beta(\ldots)\}$ converges to $\operatorname{sgn}\{\beta(\ldots)\}$ and thus

$$\lim_{\beta \to \infty} m(x) = \operatorname{erf}(a(x)/\sqrt{2}\varepsilon) \tag{24}$$

for all $x \in \mathscr{C}$.

(*g*) *Bifurcation analysis and examples.* For high enough temperature, or small enough β, the *only* solution to the fixed-point equation (17) is $\boldsymbol{m} = 0$ and $\boldsymbol{q} = 0$. As we lower the temperature, a bifurcation occurs as we cross the line $T = \varepsilon$ or $T = \Lambda_1$ where Λ_1 is the largest eigenvalue of the matrix with elements $Q(x; y)p(y)$. See figure 1.

We now have to relate Λ_1 to the physics of the problem. Let us suppose, to simplify the discussion, that Q is of the inner-product form (4) and that $p(x) = 2^{-q}$ for all $x \in \mathscr{C} = \{-1, 1\}^q$. Then one can show [8, 12, 15] the following. (i) The components $v_\rho(x)$ of the 2^q eigenvectors v_ρ of Q may be assumed to have absolute value one. (ii) By the central limit theorem, Λ_1 becomes independent of q as $q \to \infty$. For instance, for clipped synapses with $\phi(x) = \operatorname{sgn}(x)$ we have $\Lambda_1 = (2/\pi)^{1/2} = 0.80$. (iii) Finally, as in the case of clipped synapses, the stored patterns are related to Λ_1 and thus should bifurcate *first*.

Picking a specific pattern, say α, we make the ansatz $m(x) = m v_\alpha(x)$ with $m \geq 0$ in (17). Then $a(x) = m\Lambda_1 v_\alpha(x)$ and (17) reduces to only two coupled equations

$$\begin{aligned} m &= \langle \tanh[\beta(m\Lambda_1 + \varepsilon\sqrt{q}z)] \rangle \\ q &= \langle \tanh^2[\beta(m\Lambda_1 + \varepsilon\sqrt{q}z)] \rangle. \end{aligned} \tag{25}$$

These equations can be solved numerically. For $\Lambda_1 = (2/\pi)^{1/2}$ the result is shown in figures 2–4. The closer m is to one, the better the retrieval.

By virtue of (24) we get at zero temperature

$$m = \operatorname{erf}(m\Lambda_1/\sqrt{2}\varepsilon). \tag{26}$$

Because the error function behaves like the hyperbolic tangent, equation (26) has a non-trivial solution $m \neq 0$ only if

$$\varepsilon < \varepsilon_c = (2/\pi)^{1/2}\Lambda_1 = 0.80\Lambda_1. \tag{27}$$

For $\varepsilon < \varepsilon_c$, the energy of the retrieval state is always lower than that of the spin-glass phase.

In the case of clipped synapses the error fraction $\frac{1}{2}(1-m)$ is less than 0.005 if ε does not exceed $\tilde{\varepsilon} = 0.39\Lambda_1 \approx \frac{1}{2}\varepsilon_c$. For this particular value of ε the zero-temperature entropy $s(\infty)$ is 1.3×10^{-3} times the SK value $-1/2\pi$, making it very close to zero indeed. For $\varepsilon \leq \frac{1}{2}\varepsilon_c$ the effects of replica symmetry breaking can be safely ignored.

In summary, the performance of a non-linear neural network with Gaussian noise superimposed on the synaptic efficacies has been analysed in detail. The non-linearity may be arbitrary and through a large-deviations argument [7, 8, 15, 17] the statistical mechanics could be obtained exactly. For suitable values of the noise strength ε, the optimal performance of the network is obtained at a *non*-zero temperature; see figure 4.

The above results may be compared with the ones presented in a recent paper of Sompolinsky [18]. A closer examination reveals that his method is restricted to inner-product models. The non-linearity is treated approximately as Gaussian noise, which is then mapped onto a (linear) Hopfield model. As we have seen, this kind of approximation is not needed. Furthermore, no attention is paid to the dependence of the retrieval quality upon ε and the temperature T which is, after all, rather surprising.

Acknowledgments

Most of this work was done while the first author was visiting the Institute of Theoretical Physics of the Polish Academy of Sciences in Warsaw (September 1986). Partial support of the Deutsche Forschungsgemeinschaft is gratefully acknowledged. Completion of the project was made possible by his stay at the ITP in Santa Barbara. It is a pleasure to thank R G Palmer for a critical reading of the manuscript. This research was supported in part by the National Science Foundation under Grant No. PHY82-17853, supplemented by funds from the National Aeronautics and Space Administration, at the University of California at Santa Barbara.

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554; 1984 *Proc. Natl Acad. Sci. USA* **81** 3088
[2] Little W A 1974 *Math. Biosci.* **19** 101
Little W A and Shaw G L 1978 *Math. Biosci.* **39** 281
[3] Peretto P 1984 *Biol. Cybern.* **50** 51
[4] Toulouse G, Dehaene S and Changeux J-P 1986 *Proc. Natl Acad. Sci. USA* **83** 1695
[5] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530; 1987 *Ann. Phys., NY* **173** 30
[6] Kinzel W 1985 *Z. Phys.* B **60** 205
[7] van Hemmen J L 1986 *Phys. Rev.* A **34** 3435
[8] van Hemmen J L and Kühn R 1986 *Phys. Rev. Lett.* **57** 913
[9] McCulloch W A and Pitts W 1943 *Math. Biophys.* **5** 115
[10] Hebb D O 1947 *The Organization of Behavior* (New York: Wiley)
[11] Diederich S and Opper M 1987 *Phys. Rev. Lett.* **58** 949
[12] van Hemmen J L, Grensing D, Huber A and Kühn R 1988 *J. Stat. Phys.* **50** to be published
[13] van Hemmen J L, Grensing D, Huber A and Kühn R 1986 *Z. Phys.* B **65** 53
[14] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792
[15] van Hemmen J L 1987 *Phys. Rev.* A **36** 1959
[16] van Hemmen J L and Palmer R G 1979 *J. Phys. A: Math. Gen.* **12** 563
[17] van Hemmen J L 1982 *Phys. Rev. Lett.* **49** 409; 1983 *Heidelberg Coll. on Spin Glasses* (*Lecture Notes in Physics* **192**) (Berlin: Springer) pp 203–33 and appendix
[18] Sompolinsky H 1987 *Heidelberg Coll. on Glassy Dynamics* (*Lecture Notes in Physics* **275**) (Berlin: Springer) pp 485–527