Self-consistent signal-to-noise analysis and its application to analogue neural networks with asymmetric connections

1992 J. Phys. A: Math. Gen. 25 L375

(http://iopscience.iop.org/0305-4470/25/7/017)

J. Phys. A: Math. Gen. 25 (1992) L375-L381. Printed in the UK

LETTER TO THE EDITOR

Self-consistent signal-to-noise analysis and its application to analogue neural networks with asymmetric connections

M Shiino† and T Fukai‡

† Department of Applied Physics, Tokyo Institute of Technology, Ohokayama, Meguroku, Tokyo, Japan

‡ Department of Electronics, Tokai University, Kitakaname 1117, Hiratsuka-shi, Kanagawa, Japan

Received 29 October 1991

Abstract. A new systematic method is proposed for the analysis of the storage capacity of analogue neural networks with general input-output relations. It is based on the self-consistent signal-to-noise analysis in which renormalization of the signal part in the local field is properly performed. A remarkable feature of the present recipe, which in the case of symmetric analogue networks yields the same result as obtained by replica calculations, is the capability of dealing with asymmetric networks of analogue neurons. The theory is applied for the asymmetric network in which each neuron is loaded with biased patterns while some neurons are assigned only to extend inhibitory synaptic couplings free of learning.

The performance of Ising spin type neural networks with symmetric connections as a content addressable memory has been extensively investigated by means of replica symmetric calculations [1, 2] in spin glass theory. Very little is known, however, about the network performance of analogue neural networks with asymmetric connections which have a relevance with physiological nervous systems. Asymmetry as well as analogue property of neurons has so far defied the use of statistical mechanical approach. It will thus be of vital importance to develop a systematic method to cope with network systems with such properties. We have recently reported a successful application of replica calculations to obtaining the storage capacity of the analogue neural networks with symmetric couplings and a sigmoid type input-output relation [3, 4].

In the present paper we develop a self-consistent method based on signal-to-noise analysis [5-8] for evaluating the storage capacity allowing small retrieval errors for analogue neural networks with general input-output relations, and show that it can be applied to certain types of networks with asymmetric connections. Our systematic method, which in the case of symmetric analogue networks yields the same result as that of our recent replicas symmetric analysis [3], differs from the known recipe [9-11]. The latter, which involves the naive treatment of signal-to-noise analysis, seemingly gives the same result as obtained by replica calculations for the stochastic networks of Ising spin type [1, 2]. That naive treatment, however, is considered to be only an approximation in the case of feedback type neural networks like the fully connected models of Hopfield type even if the result itself is not wrong in some cases, since no distinction has been made between the analogue and the stochastic Ising spin type networks. The problem of the Onsager reaction term in the so-called TAP (Thouless, Anderson and Palmer) equation [12, 13] should be taken seriously, since that term has been found to play a key role in discriminating those two kinds of networks [3]. Cannot any signal-to-noise type analysis go beyond an approximate description for what is obtained from the replica symmetric calculations? The answer is affirmative. The characteristic feature of the present self-consistent signal-to-noise analysis is that the separation of the signal part from the noise one in the local field of each neuron is performed legitimately and systematically for analogue neural networks with deterministic updating, which explicitly define the TAP equation without the Onsager reaction term.

We begin by formulating an asymmetric neural network of N analogue neurons to show how the self-consistent signal-to-noise analysis applies to networks with asymmetric connections. Neurons which have a graded response given by a real-valued function $F(u)$ of membrane potential u are assumed to be connected with each other through asymmetric synaptic couplings J_{ij} $(j \neq i)$ of the form

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} (\xi_i^{(\mu)} - a)(\xi_j^{(\mu)} - a)\left(\frac{1+\eta_j}{2}\right) - \frac{1}{N} J \left(\frac{1-\eta_j}{2}\right) \tag{1}$$

where $\{\xi_i^{(\mu)}\}$ $(\mu = 1, \ldots, p,\ i = 1, \ldots, N)$ represent $p(=\alpha N)$ sets of biased random patterns for memories which are specified by independent identical distribution $P_r(\xi_i^{(\mu)})$ with mean a:

$$P_r(\xi_i^{(\mu)}) = \frac{1+a}{2}\,\delta(\xi_i^{(\mu)} - 1) + \frac{1-a}{2}\,\delta(\xi_i^{(\mu)} + 1).$$

Here we assumed that only the synaptic connections extended from wN neurons specified by $\eta_j = 1$ are subject to the Hebb learning rule whereas those from the other $(1-w)N$ neurons with $\eta_j = -1$ are inhibitory $(J > 0)$ and free of learning. We also note that w controls the degree of asymmetric dilution in the fully connected model, with $w = 1$ and $w = 0$ making the system symmetric.

The analogue network dynamics describing the conservation law of currents flowing through the membranes of neurons can usually be written, in terms of membrane potential u's as

$$\frac{\mathrm{d}}{\mathrm{d}t} u_i = -u_i + \sum_{j \neq i} J_{ij} F(u_j) + I \qquad i = 1, \ldots, N$$

where I is an external current. Since the weak asymmetry in the present scheme of couplings (1) can be assumed still to ensure the existence of fixed-point type attractors for the above dynamics, we are allowed to be concerned with equilibrium states of the network. They are determined by

$$x_i = F(h_i) \qquad i = 1, \ldots, N \tag{2}$$

with x_i representing output $F(u_i)$ and h_i local field: $h_i = \sum_{j \neq i} J_{ij} x_j + I$.

Defining modified overlaps

$$m^{(0)} = \frac{1}{N} \sum_j \frac{1-\eta_j}{2} x_j \tag{3a}$$

$$m^{(\mu)} = \frac{1}{N} \sum_j (\xi_j^{(\mu)} - a) \frac{1+\eta_j}{2} x_j \qquad \mu = 1, \ldots, p \tag{3b}$$

as order-parameters for the present system, we proceed to deal with retrieval solutions to equation (2) in which $m^{(1)}=O(1)$ and $m^{(\mu)}=O(1/\sqrt{N})$ for $\mu\geqslant 2$. This corresponds to the assumption that pattern $\{\xi^{(1)}\}$ is chosen for the condensed one in the Amit-Gutfreund-Sompolinsky theory [1]. Note that the usually defined overlap

$$g^{(1)}\equiv\frac{1}{N}\sum_j \xi_j^{(1)}x_j$$

is given by

$$g^{(1)}=\frac{m^{(1)}}{w}+\frac{am^{(0)}}{1-w}.$$

Expressing the local field h_i in terms of the overlaps (3), one obtains

$$h_i=\sum_\mu(\xi_i^{(\mu)}-a)m^{(\mu)}-Jm^{(0)}+I-\alpha(1-a^2)\frac{1+\eta_i}{2}F(h_i) \tag{4}$$

which will be solved to yield $x_i=F(h_i)=\hat{F}(\sum_\mu(\xi_i^{(\mu)}-a)m^{(\mu)}-Jm^{(0)}+I;\ \eta_i)$.

The self-consistent signal-to-noise analysis has its basis in the systematic splitting of the h_i into signal and noise parts. To properly extract noise part, which originates from the sum involving $m^{(\mu)}(\mu\geqslant 2)$, we compute $m^{(\mu)}(\mu\geqslant 2)$ by expanding $\hat{F}$ in the RHS of (3*b*) up to first order in $(\xi_i^{(\mu)}-a)m^{(\mu)}$:

$$\begin{aligned}m^{(\mu)}=\frac{1}{N}\sum_j(\xi_j^{(\mu)}-a)\frac{1+\eta_j}{2}\bigg\{\hat{F}\bigg(\sum_{\nu\neq\mu}(\xi_j^{(\nu)}-a)m^{(\nu)}-Jm^{(0)}+I;\ \eta_j\bigg)\\+(\xi_j^{(\mu)}-a)m^{(\mu)}\hat{F}'\bigg(\sum_{\nu\neq\mu}(\xi_j^{(\nu)}-a)m^{(\nu)}-Jm^{(0)}+I;\ \eta_j\bigg)\bigg\}\qquad\mu\geqslant 2.\end{aligned} \tag{5}$$

Noting that the argument of $\hat{F}'$ does not contain $\xi^{(\mu)}$ and that $1/N\sum_j\ldots$ can be replaced by average $\langle\ldots\rangle$ over ξ's and η, we obtain from equation (5)

$$m^{(\mu)}=\frac{1}{K}\bigg\{\frac{1}{2N}\sum_j(\xi_j^{(\mu)}-a)(1+\eta_j)\hat{F}\bigg(\sum_{\nu\neq\mu}(j);\ \eta_j\bigg)\bigg\}\qquad\mu\geqslant 2 \tag{6}$$

with

$$K=1-(1-a^2)w\bigg\langle\hat{F}'\bigg(\sum_{\nu\neq\mu}(j);\ \eta_j=1\bigg)\bigg\rangle \tag{7}$$

where $\hat{F}(\sum_{\nu\neq\mu}(j);\ \eta_j)$ is the shorthand notation for $\hat{F}(\sum_{\nu\neq\mu}(\xi_j^{(\nu)}-a)m^{(\nu)}-Jm^{(0)}+I;\ \eta_j)$. Keeping in mind that the $j=i$ term in the sum in equation (6) should contribute to the signal part in the local field, we obtain from equations (4) and (6)

$$\begin{aligned}h_i&=(\xi_i^{(1)}-a)m^{(1)}-Jm^{(0)}+I+\sum_{\mu\geqslant 2}\frac{1}{2KN}(\xi_i^{(\mu)}-a)^2(1+\eta_i)\hat{F}\bigg(\sum_{\nu\neq\mu}(i);\ \eta_i\bigg)+\bar{z}\\&\quad-\alpha(1-a^2)\frac{1+\eta_i}{2}F(h_i)\\&=(\xi_i^{(1)}-a)m^{(1)}-Jm^{(0)}+I+\frac{\alpha(1-a^2)(1-K)}{2K}(1+\eta_i)F(h_i)+\bar{z}\end{aligned} \tag{8}$$

with

$$\bar{z} = \sum_{\mu \geq 2} (\xi_i^{(\mu)} - a) \frac{1}{2NK} \sum_{j \neq i} (\xi_j^{(\mu)} - a)(1+\eta_j) \hat{F}\left(\sum_{\nu \neq \mu} (j);\ \eta_j \right). \tag{9}$$

Here, in deriving equation (8) we safely replaced $\hat{F}(\sum_{\nu \neq \mu} (i);\ \eta_i)$ by $\hat{F}(\sum_\nu (i);\ \eta_i) = F(h_i)$. Noting that $\bar{z}$ is a sum of almost uncorrelated random variables, with $\langle \bar{z} \rangle = 0$ and

$$\begin{aligned} \langle \bar{z}^2 \rangle &= \frac{1}{(2NK)^2} \sum_{\mu \geq 2} \sum_{j \neq i} \langle (\xi_i^{(\mu)} - a)^2 (\xi_j^{(\mu)} - a)^2 \rangle \left\langle (\eta_j + 1)^2 \hat{F}\left(\sum_{\nu \neq \mu} (j);\ \eta_j \right)^2 \right\rangle \\ &= \frac{\alpha w (1-a^2)^2}{K^2} \left\langle \hat{F}\left(\sum_\nu (j);\ \eta = 1 \right)^2 \right\rangle \end{aligned} \tag{10}$$

we will be allowed to claim that $\bar{z}$ represents the noise obeying the Gaussian distribution with mean 0 and variance $\sigma^2 = \langle \bar{z}^2 \rangle$. It then follows that the local field h_i in turn is distributed according to a non-Gaussian distribution, owing to the appearance of the term involving $F(h_i)$ itself in the RHS of equation (8). From the renormalized expression (8) for h, output $Y_\eta \equiv F(h)$ turns out to be implicitly determined as a function of $\bar{z}$ by

$$\begin{aligned} Y_\eta = F\bigg(&(\xi^{(1)} - a) m^{(1)} - J m^{(0)} + I \\ &+ \frac{\alpha (1-a^2)(1-K)}{2K} (1+\eta) Y_\eta + \bar{z} \bigg) \qquad \eta = +, -. \end{aligned} \tag{11}$$

Noting that the site summation $1/N \sum_j \ldots$ in the expressions for $m^{(0)}$ and $m^{(1)}$ as well as the average $\langle \ldots \rangle$ in equations (7) and (10) can be written in terms of the average with respect to ξ and noise $\bar{z}$ obeying

$$D(\bar{z}) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-\bar{z}^2}{2\sigma^2}$$

we have

$$m^{(1)} = w \langle (\xi - a) Y_+ \rangle_{\xi, \bar{z}} \qquad m^{(0)} = (1-w) \langle Y_- \rangle_{\xi, \bar{z}} \qquad \sigma^2 = \frac{(1-a^2)^2 \alpha w}{K^2} \langle Y_+^2 \rangle_{\xi, \bar{z}}$$

and

$$K = 1 - (1-a^2) w \left\langle \frac{\mathrm{d}}{\mathrm{d}\bar{z}} Y_+ \right\rangle_{\xi, \bar{z}}.$$

In writing the last equation, we have also noted the following requirement. In order to ensure that the whole process of separating signal from noise in the local field be performed self-consistently, the procedure for obtaining $m^{(\mu)} (\mu \geq 2)$ should be based on use of the renormalized $F(h_i)$, i.e. Y_η rather than $\hat{F}$ in equation (5).

With the change of variables,

$$q = \frac{K^2 \sigma^2}{(1-a^2)^2 \alpha w} \qquad \sqrt{\alpha r} = \sigma,\ U = \frac{1-K}{(1-a^2) w} \qquad \text{and} \qquad \frac{\bar{z}}{\sigma} = z$$

we arrive at a more familiar form of equations, which read

$$m^{(1)} = w \left\langle\!\!\left\langle \int \mathrm{d}z \frac{\exp(-z^2/2)}{\sqrt{2\pi}} (\xi - a) Y_+(z) \right\rangle\!\!\right\rangle \tag{12}$$

$$m^{(0)}=(1-w)\left\langle\!\!\left\langle\int \mathrm{d}z\,\frac{\exp(-z^2/2)}{\sqrt{2\pi}}\,Y_-(z)\right\rangle\!\!\right\rangle \tag{13}$$

$$q=\left\langle\!\!\left\langle\int \mathrm{d}z\,\frac{\exp(-z^2/2)}{\sqrt{2\pi}}\,Y_+(z)^2\right\rangle\!\!\right\rangle \tag{14}$$

$$U\sqrt{\alpha r}=\left\langle\!\!\left\langle\int \mathrm{d}z\,\frac{\exp(-z^2/2)}{\sqrt{2\pi}}\,zY_+(z)\right\rangle\!\!\right\rangle \tag{15}$$

$$Y_\eta(z)=F((\xi-a)m^{(1)}-Jm^{(0)}+I+\Gamma Y_\eta(z)+\sqrt{\alpha r}z) \qquad \eta=+,- \tag{16}$$

with

$$q=\frac{\{1-(1-a^2)wU\}^2 r}{(1-a^2)^2 w} \qquad \text{and} \qquad \Gamma=\frac{\alpha(1-a^2)^2 wU}{1-(1-a^2)wU}\left(\frac{1+\eta}{2}\right)$$

where $\langle\!\langle\ \rangle\!\rangle$ denotes the average over ξ. We note that when self-coupling J_{ii} of the form (equation (1)) is taken into account for local field h_i, Γ should read

$$\Gamma=\frac{\alpha(1-a^2)}{1-(1-a^2)wU}\left(\frac{1+\eta}{2}\right).$$

Setting $F(u)=\tanh\beta u$ (β: analogue gain), $w=1$ (symmetric network) and $a=0$, we see that the result previously obtained by replica calculations is recovered [3, 14]. The presence of ΓY_η in the argument of F in equation (16), which makes the Y be determined only implicitly, was found to give rise to a slight increase in the storage capacity of the analogue networks [3, 14], compared with that of the Ising spin network with inverse temperature β. That additional term arises as a consequence of the absence of the Onsager reaction term for the analogue networks. In other words, the term ΓY_η, which has resulted from the renormalization of signal part (see equation (8)) turns out to be the negative of the Onsager reaction term for the corresponding Ising spin network. In this sense, the present method elucidates the structure of the Onsager reaction term itself of the Ising spin networks. It is then worth noting that the self-consistent signal-to-noise analysis developed in our study is also available for the stochastic networks of Ising type. In fact, as soon as one applies the present analysis to the TAP equation with the Onsager reaction term for the stochastic Ising network [13], one can easily obtain exactly the same result as that of Amit, Gutfreund and Sompolinsky [1] owing to the compensation of the above-mentioned additional term by the presence of the Onsager reaction one.

We finally evaluate the storage capacity as well as the phase diagram of the present asymmetric network in the case where $F(u)=\tanh\beta u$ and $I=0$. We are interested in exploring the effect of decreasing w on the memory retrieval performance together with the existence of spin-glass phase in the asymmetric system. The phase transition line between $m^{(1)}=m^{(0)}=q=0$ (para) and $m^{(1)}=m^{(0)}=0$, $q\neq 0$ (spin-glass) is easily determined from equations (14)–(16) to be

$$\beta_G=\frac{1}{(1-a^2)(w+2\sqrt{\alpha w})}.$$

Before displaying the result of the present analysis for the storage capacity, we note that the condition for retrieval with no errors in the limit $\beta\to\infty$, i.e.

$$\mathrm{sgn}\left(\sum_{j\neq i} J_{ij}\xi_j^{(1)}\right)=\xi_i^{(1)}$$

can be obtained, based on probability theory 6, 7, as

$$\frac{P}{N} \leqslant \frac{\{(1-|a|)(1-a^2)w - J(1-w)|a|\}^2}{(1-a^2)^2 w 2 \ln N} \left(\equiv C(a, w, J)\frac{1}{2\ln N}\right)$$

$$w > w_{\mathrm{th}} \equiv \frac{J|a|}{(1-|a|)(1-a^2)+J|a|}.$$

The threshold arises from the fact that the naive signal part for the local field given by the argument of the sgn function vanishes at $w = w_{\mathrm{th}}$.

The storage capacity near saturation is determined by equations (12)-(16) as the limit of existence of the retrieval state ($m^{(1)} \neq 0$). We solved equations (12)-(16) numerically to obtain the phase diagram showing the storage capacity in the $\alpha - \beta^{-1}$ plane, an example of which is given in figure 1. The effect of dilution with decreasing w on the storage capacity in the limit $\beta \to \infty$ is also shown in figure 2 for several values of J (solid curves). For comparison, the estimate of storage capacity inferred from probability theory $\alpha_c^{\mathrm{prob}} = C(a, w, J)\alpha_c^*$ with α_c^* chosen such that $(1-|a|)^2\alpha_c^* = \alpha(a, w=1)^{15}$ is plotted in this figure (broken curves). It may be interesting to observe that there exist no thresholds in w for the onset of the retrieval state within the context of the self-consistent signal-to-noise analysis, in contrast to the suggestion from probability theory. Computer simulations on the networks with large N ($N = 3000 \sim 5000$), indeed, support the result of the self-consistent signal-to-noise analysis for the storage capacity of the saturation limit even below the threshold suggested by probability theory.

To conclude, we have presented a heuristic description of the self-consistent signal-to-noise analysis to evaluate the storage capacity of the analogue neural networks. The new recipe proves to be very useful in that not only can it be applied to a certain class of asymmetric networks including the present case but also the calculation involved is quite elementary as well as of perspective unlike the replica method. Dealing with analogue networks with such sparse encoding as taken up by Buhmann *et al* [16] and

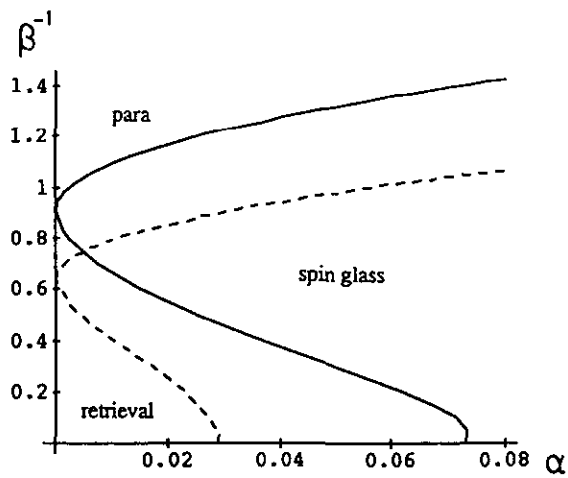

Figure 1. The phase boundaries limiting the regions of retrieval state and spin-glass one for the present neural networks without self-coupling in the case of symmetric coupling ($w = 1$: the solid curve) and asymmetric coupling ($w = 0.7$, $J = 1$: the broken curve) for $a = 0.3$. We see that the weak asymmetry of the present type does not change the qualitative feature of the phase diagram.

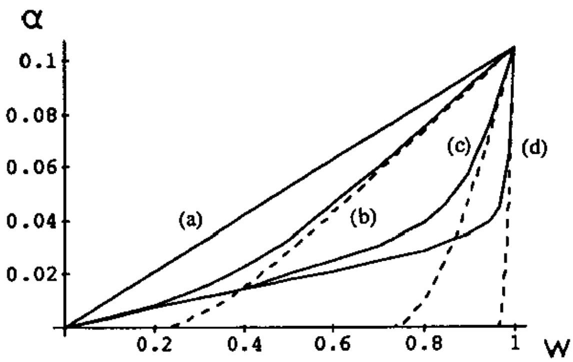

Figure 2. Effect of the asymmetric dilution of the synaptic couplings with learning rule on the storage capacity α in the fully connected model. Plots of α against w in the limit $\beta \to \infty$ are given for $J=0$ (a), 1 (b), 10 (c) and 100 (d). The bias of the stored patterns is set to $a=0.2$. For comparison, w dependence of the storage capacity with no retrieval errors suggested by probability theory is also displayed (broken curves): $\alpha_c^{\rm prob}=C(a,w,J)\alpha_c^*$, where α_c^* is chosen so as to normalize its values at the symmetric limit $w=1$ (see text). Although the behaviours of α with decreasing w around $w=1$ are seen to agree with each other for each value of $J>0$, remarkable differences start to appear with further decreasing of w owing to the existence of the thresholds for the storage capacity from probability theory.

Tsodyks and Feigel'man [17] will be straightforward, say by taking $F(u)=1/2(1+\tanh\beta u)$, since the input-output relation $F(u)$ is free of a constraint such as being odd. A further study, however, will be required to discuss the appearance of the states with the replica symmetry breaking [1] from the view point of the validity of the self-averaging property used in the present study. Details as well as applications to the networks with various types of couplings and $F(u)$ will be studied elsewhere.

References

[1] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30
[2] Amit D J 1989 *Modeling Brain Function* (Cambridge: Cambridge University Press)
[3] Shiino M and Fukai T 1990 *J. Phys. A: Math. Gen.* **23** L1009
[4] Fukai T and Shiino M 1990 *Phys. Rev.* A **42** 7459
[5] Amari S and Maginu K 1988 *Neural Networks* **1** 63
[6] Weisbuch G and Fogelmann-Soulie F 1985 *J. Physique Lett.* **46** L623
[7] McEliece R J, Posner E C, Rodemich E R and Venkatesh S S 1987 *IEEE Trans. Inform. Theor.* **IT-33** 461
[8] Domany E, Kinzel W and Meir R 1989 *J. Phys. A: Math. Gen.* **22** 2081
[9] Peretto P 1988 *J. Physique* **49** 711
[10] Geszti T 1990 *Physical Models of Neural Networks* (Singapore: World Scientific)
[11] Marcus C M, Waugh F R and Westervelt R M 1990 *Phys. Rev.* A **41** 3355
[12] Thouless D J, Anderson P W and Palmer R G 1977 *Phil. Mag.* **35** 593
[13] Mezard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[14] Kuhn R, Bos S and van Hemmen J L 1991 *Phys. Rev.* A **43** 2084
[15] Amit D J, Gutfreund H and Sompolinsky H 1987 *Phys. Rev.* A **35** 2293
[16] Buhmann J, Divko R and Schulten K 1989 *Phys. Rev.* A **39** 2689
[17] Tsodyks M V and Feigel'man M V 1988 *Europhys. Lett.* **6** 101