On the macroscopic description of recurrent neural network dynamics

1998 J. Phys. A: Math. Gen. 31 4197

(http://iopscience.iop.org/0305-4470/31/18/009)

J. Phys. A: Math. Gen. **31** (1998) 4197–4202. Printed in the UK PII: S0305-4470(98)88469-2

On the macroscopic description of recurrent neural network dynamics

Toshiyuki Tanaka and Shuho Osawa

Department of Electronics and Information Engineering, Graduate School of Engineering, Tokyo Metropolitan University, 1-1, Minami-Osawa, Hachioji, Tokyo 192-0397, Japan

Received 16 October 1997

Abstract. It is expected that the dynamics of recurrent neural networks can be described by a set of macroscopic order parameters. For such descriptions to exist, macroscopic specification of microscopic states of neural networks during the dynamics has to be possible. We study this problem for the case of Coolen–Sherrington (CS) theory, which successfully describes qualitative aspects of the dynamics of the Hopfield-type recurrent neural network. We show that the macroscopic specification is incomplete in CS theory by providing direct evidence that the *equipartitioning* assumption employed by CS theory prevents the theory from describing the dynamics quantitatively.

1. Introduction

Considering a Hopfield-type autoassociative memory that memorizes an extensive number of randomly generated patterns, it is observed that its dynamics becomes independent of the realization of the patterns as the size N of the model becomes sufficiently large. From observations like this, it is expected that the dynamics of such models can be described by a rather simple deterministic time-evolution rule of macroscopic order parameters. For such descriptions to exist, specification of the microscopic states by the macroscopic order parameters during the dynamics has to be possible. At (thermal) equilibrium of the Hopfield model, such a specification is possible because the corresponding Boltzmann–Gibbs microscopic state distribution is described solely in terms of a macroscopic order parameter, the energy. The problem is whether or not, by a suitable choice of macroscopic order parameters, such a specification is possible throughout the dynamics.

We study this problem for the case of Coolen–Sherrington (CS) theory [1, 2]. The reasons for focusing on CS theory are as follows. First, it is complex enough to reproduce qualitative aspects of the dynamics of the Hopfield model [1, 2] and other related models [3]. Second, it is still simple compared with competing [4] or more advanced [5] theories; it allows us to investigate theoretical details and to execute numerical experiments.

2. CS theory

We consider networks of N binary neurons. $s_i \in \{-1, 1\}$ denotes the state of the ith neuron. The microscopic states of the networks are described by the state vector $\boldsymbol{s} = (s_1, \ldots, s_N)$.

0305-4470/98/184197+06$19.50

States of neurons are assumed to be updated asynchronously. The rate $w_i(\boldsymbol{s})$ of the transitions $s_i \to -s_i$ is determined by the 'local field' $h_i(\boldsymbol{s})$ as

$$w_i(\boldsymbol{s}) = \tfrac{1}{2}(1 - s_i \tanh \beta h_i(\boldsymbol{s})) \qquad h_i(\boldsymbol{s}) = \sum_{j \neq i} J_{ij} s_j \tag{1}$$

where J_{ij} is a synaptic weight from neuron j to neuron i and β the so-called inverse temperature. When $\{J_{ij}\}$ are symmetric, i.e. $J_{ij} = J_{ji}$, the dynamics (1) leads to the Boltzmann–Gibbs equilibrium distribution $p(\boldsymbol{s}) \propto \exp(-\beta E(\boldsymbol{s}))$, determined by the energy $E(\boldsymbol{s}) = -\sum_{i<j} J_{ij} s_i s_j$.

For an autoassociative memory memorizing a set of p random patterns $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \ldots, \xi_N^\mu)$ with Hebbian synapses $J_{ij} = (1/N)\sum_\mu \xi_i^\mu \xi_j^\mu$, the overlap m between the state vector $\boldsymbol{s}$ and the pattern to be retrieved (we assume that pattern 1 is nominated for retrieval) is generally used as a macroscopic order parameter: $m = (1/N)\sum_i \xi_i^1 s_i$. It is known [6] that m is insufficient for specification of microscopic states during the dynamics: different microscopic states with the same m value will evolve differently, depending on their history.

The CS theory [1, 2] employs another macroscopic order parameter r in addition to m:

$$r = \frac{1}{\alpha}\sum_{\mu>1}(m^\mu)^2 \tag{2}$$

where m^μ are the overlaps between the state vector $\boldsymbol{s}$ and non-nominated patterns μ ($\neq 1$), and $\alpha = p/N$ is the memory rate. To obtain deterministic flow equations of m and r the CS theory assumes the following:

(i) self-averaging of the flow with respect to randomly chosen patterns;

(ii) probability equipartitioning within the (m, r) subshells: when the macroscopic state is given by m and r, the probability distribution of the corresponding microscopic states can be assumed to be uniform over the (m, r) subshell, with regard to calculation of the flow equations.

The former assumption is supported by numerical simulations with sufficiently large N. The latter assumption, the *equipartitioning* assumption, will be safe if (m, r) is sufficient for specification of microscopic states. At thermal equilibrium, the equipartitioning condition is a natural consequence of the Boltzmann–Gibbs distribution, which is described in terms of m and r only. It should be noted that assuming equipartitioning is equivalent, in the thermodynamical limit, to employing the so-called maximum entropy principle [7], which states that we are maximally non-committal with respect to information missing from the (m, r) specification.

Under these assumptions the resulting deterministic flow equations are

$$\begin{aligned} \frac{\mathrm{d}m}{\mathrm{d}t} &= \int \mathrm{d}z\, D_{m,r}[z] \tanh[\beta(m+z)] - m \\ \frac{1}{2}\frac{\mathrm{d}r}{\mathrm{d}t} &= \frac{1}{\alpha}\int \mathrm{d}z\, D_{m,r}[z] z \tanh[\beta(m+z)] + 1 - r \end{aligned} \tag{3}$$

where $D_{m,r}[z]$ is the distribution of the 'noise' terms that represent the interference in the local field h_i caused by non-nominated patterns. A replica calculation gives the following result for $D_{m,r}[z]$ within the so-called RS ansatz,

$$\begin{aligned} D_{m,r}[z] = {} & \frac{\mathrm{e}^{-(\Delta+z)^2/2\alpha r}}{2\sqrt{2\pi\alpha r}}\left\{1 - \int Dy \tanh\left[\lambda y\left(\frac{\Delta}{\rho\alpha r}\right)^{1/2} + (\Delta+z)\rho\frac{r_{\mathrm{AGS}}}{r} + \mu\right]\right\} \\ & + \frac{\mathrm{e}^{-(\Delta-z)^2/2\alpha r}}{2\sqrt{2\pi\alpha r}}\left\{1 - \int Dy \tanh\left[\lambda y\left(\frac{\Delta}{\rho\alpha r}\right)^{1/2} + (\Delta-z)\rho\frac{r_{\mathrm{AGS}}}{r} - \mu\right]\right\} \end{aligned} \tag{4}$$

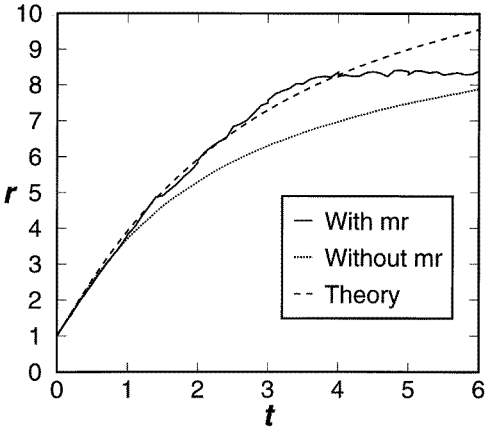

Figure 1. Time evolution of r obtained by simulations with (m, r) annealing (full curve), by ordinary simulations without (m, r) annealing (dotted curve) and by CS theory (dashed curve). $\alpha = 0.1$, $\beta = +\infty$ and $m(t = 0) = 0.3$.

$$\Delta \equiv \rho\alpha(r - r_{\mathrm{AGS}}) \tag{5}$$

$$r_{\mathrm{AGS}} \equiv \frac{\lambda^2}{\rho^2\alpha} \tag{6}$$

where $Dy = (\mathrm{d}y/\sqrt{2\pi})\,\mathrm{e}^{-y^2/2}$ is the Gaussian measure. The parameters $\{q, \lambda, \rho, \mu\}$ are determined from m, r and α by the following CS saddle-point equations:

$$r = \frac{1 - \rho(1-q)^2}{[1-\rho(1-q)]^2} \qquad \lambda = \frac{\rho\sqrt{\alpha q}}{1-\rho(1-q)}$$
$$m = \int Dy \tanh(\lambda y + \mu) \qquad q = \int Dy \tanh^2(\lambda y + \mu). \tag{7}$$

The CS theory successfully describes the observation that the noise distribution is not actually Gaussian. It approximately reproduces the trajectories in the (m, r) plane obtained by numerical simulations. However, it fails to describe the time evolution of m and r, especially when retrieval fails (see, e.g., [8]): simulation results exhibit an overall slowing down compared with the theory. It has been reported that the noise distributions given by the CS theory are quantitatively different from those given by simulations [9].

3. (m, r) annealing

The CS saddle-point equations (7) are formally quite similar to the AGS saddle-point equations [10]. Based on the formal similarity between them, we formulate a process, which we will call (m, r) annealing, whose thermal equilibrium corresponds to the equipartitioning condition. First, we introduce an external field proportional to the nominated pattern $\boldsymbol{\xi}^1$. The local field $h_i(\boldsymbol{s})$ is now given by

$$h_i(\boldsymbol{s}) = \sum_{j \neq i} J_{ij} s_j + b\xi_i^1 \tag{8}$$

where b is a coefficient representing the magnitude of the external field. One can then show the following proposition.

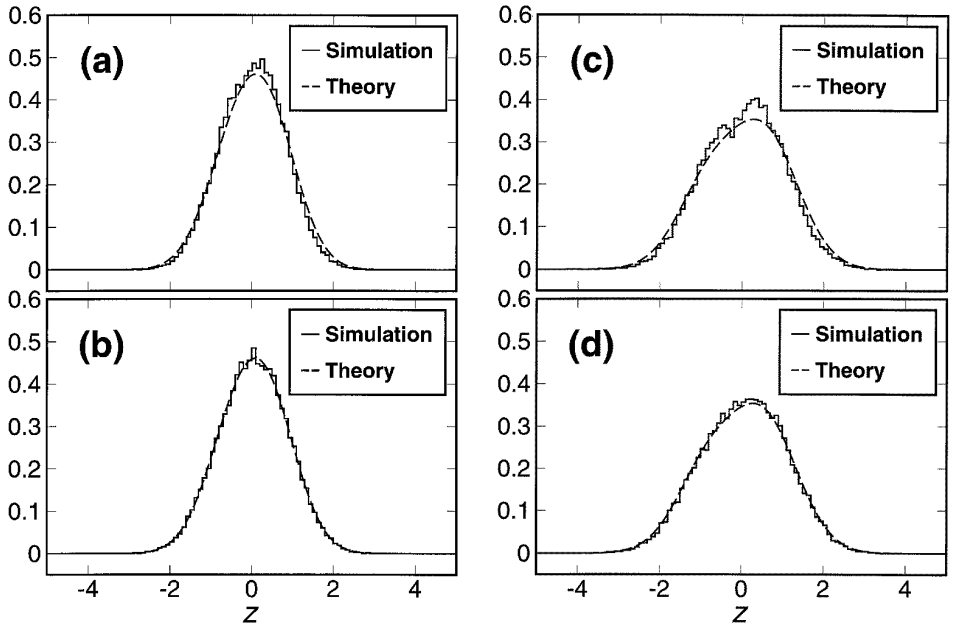

Figure 2. Noise distributions: (a) $t = 1$, before (m, r) annealing; (b) $t = 1$, after (m, r) annealing; (c) $t = 2$, before (m, r) annealing; (d) $t = 2$, after (m, r) annealing. $\beta = +\infty$, $\alpha = 0.1$ and $m(t = 0) = 0.1$.

Proposition. For a given m, r and α, solve the CS saddle-point equations (7) and define b by the identity $\mu = \rho(m + b)$ using the parameters μ and ρ. Then the dynamics, with the local field given by equation (8) and with the inverse temperature given by ρ, in the RS ansatz, has a thermal equilibrium distribution corresponding to the equipartitioning of the (m, r) subshell.

We call the dynamics defined above (m, r) annealing. The proof of the proposition is straightforward; one can derive the CS saddle-point equations, for example, by following the analysis of Amit *et al* [10], or by using the standard procedure of the cavity method [11], which leads to the AGS saddle-point equations. It should be noted that (m, r) annealing actually achieves the equipartitioning distribution in the limit $N \to \infty$ since the energy associated with the (m, r) annealing is written in terms of m and r.

The proposition is regarded as showing a practical method for realizing the equipartitioning condition by the use of (m, r) annealing. A sufficiently large network working in the '(m, r) annealing mode' is seen, in thermal equilibrium, as a Gibbs sampler sampling microscopic states from the (m, r) subshell with uniform probability. If one executes (m, r) annealing in the course of the ordinary dynamics with sufficient frequency, the equipartitioning condition will be satisfied throughout the dynamics and hence we can expect that CS theory becomes exact. Figure 1 shows the time evolution of r obtained by ordinary dynamics, by ordinary dynamics with (m, r) annealing executed with a short interval (averaged over three samples), and by CS theory. The result with (m, r) annealing is in good agreement with CS theory, as expected. It should be noted that the trajectory is in the RSB region for about $t > 4$. Since the consideration in this paper is based on the RS ansatz we cannot expect it to hold in the RSB region. We can see in figure 1 that the agreement becomes worse at a later stage, about $t > 4$, and it can be ascribed to RSB.

Next, we investigate the noise distributions. One can expect that, when the equipartitioning condition is actually satisfied by the use of (m, r) annealing, the noise distributions should be identical with those given by CS theory. Figures 2(a) and (c) show the noise distributions obtained by ordinary dynamics (full curves) and calculated by CS theory (equation (4)) using the corresponding (m, r) values (dashed curves), with $\beta = +\infty$, $\alpha = 0.1$ and $m(t = 0) = 0.1$, at $t = 1$ and 2, respectively. Figures 2(b) and (d) show the noise distributions with the same conditions as in (a) and (c), respectively, but *after* executing (m, r) annealing. These results show that (m, r) annealing improves the fit between the numerical experiments and the theory.

Figure 3 shows the time evolution of m and r in numerical simulations, where (m, r) annealing is executed once at $t = 2$ in order to demonstrate how (m, r) annealing affects the dynamics. Execution of (m, r) annealing at $t = 2$ changes the dynamics in such a manner that it somewhat recovers from the slow down, approaching the time evolution curves predicted by CS theory.

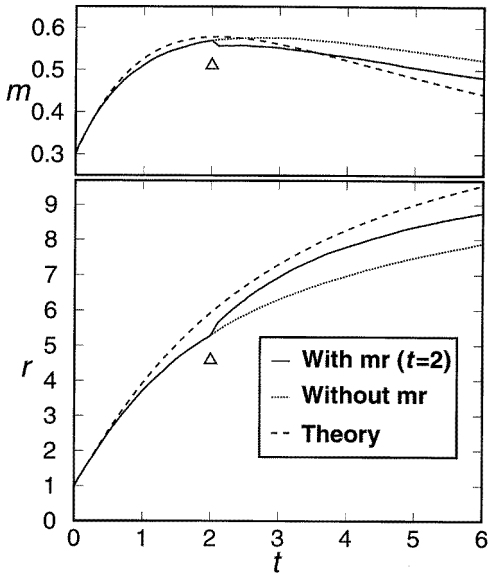

Figure 3. Time evolution of m and r in simulations where (m, r) annealing is executed at $t = 2$ (full curve), of ordinary dynamics (dotted curve), and of CS theory (dashed curve). $\alpha = 0.1$, $\beta = +\infty$, $m(t = 0) = 0.3$ and $r(t = 0) = 1$.

4. Discussion

From these results one can infer that CS theory is indeed exact if the equipartitioning assumption holds; conversely, deviations between the numerical simulation of ordinary dynamics and CS theory, observed in time evolution [8] and in noise distributions [9], are quantitatively explained by the failure of the equipartitioning assumption. This indicates that the dynamics of the model actually selects microscopic states which are not physically *typical* within the (m, r) subshell determined by the current values of m and r.

The failure of the simplest one-parameter theory, based on the overlap m, is explained, in view of CS theory, as follows. For a given m, almost all the microscopic states belonging to the m subshell belong to the (m, r) subshell with $r = 1$ (i.e. the states with $r = 1$ dominate

overwhelmingly in number within the m subshell), whereas the dynamics actually chooses microscopic states with $r > 1$ which are *not typical* within the m subshell. By analogy, the incompleteness of CS theory indicates that the dynamics actually chooses microscopic states that are not typical within the (m, r) subshell. This also indicates that the set of macroscopic order parameters m and r is insufficient for the specification of microscopic states during the dynamics. More advanced or alternative theory is certainly required in order to specify such non-typical microscopic states appropriately by macroscopic quantities, if such specification is possible.

Acknowledgments

TT would like to thank M Yamana, H Nishimori and M Okada for stimulating discussions.

References

[1] Coolen A C C and Sherrington D 1993 *Phys. Rev. Lett.* **71** 3886
[2] Coolen A C C and Sherrington D 1994 *Phys. Rev.* E **49** 1921
Coolen A C C and Sherrington D 1994 *Phys. Rev.* E **49** 5906 (Erratum)
[3] Laughton S N and Coolen A C C 1995 *Phys. Rev.* E **51** 2581
[4] Horner H, Bormann D, Frick M, Kinzelbach H and Schmidt A 1989 *Z. Phys.* B **76** 383
[5] Laughton S N, Coolen A C C and Sherrington D 1996 *J. Phys. A: Math. Gen.* **29** 763
[6] Amari S and Maginu K 1988 *Neural Networks* **1** 63
[7] Jaynes E T 1957 *Phys. Rev.* **106** 620
[8] Coolen A C C, Laughton S N and Sherrington D 1996 *Advances in Neural Information Processing Systems* vol 8, ed D S Touretzky, M C Mozer and M E Hasselmo (Cambridge, MA: MIT Press) p 253
[9] Ozeki T and Nishimori H 1994 *J. Phys. A: Math. Gen.* **27** 7061
[10] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30
[11] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)