J. Phys. I France **5** (1995) 409–432 MARCH **1995**, PAGE 409

Classification
Physics Abstracts
05.45 — 05.40

Increase in Complexity in Random Neural Networks

B. Cessac(*)

Centre d'Etudes et de Recherches de Toulouse, 2 avenue Edouard Belin, BP 4025, 31055 Toulouse Cedex, France

(Received 1 September 1994, revised 18 November 1994, accepted 1 December 1994)

Résumé. — Nous étudions la dynamique d'un réseau de neurones à temps discret, dont les couplages sont asymétriques, aléatoires, et les seuils aléatoires. L'évolution des neurones est donnée à la limite thermodynamique par un jeu d'équations de champ moyen dynamique obtenues via une hypothèse de chaos local. Nous étudions l'évolution de la distance quadratique moyenne entre deux trajectoires, et montrons qu'il existe deux régimes selon la valeur des paramètres de contrôle. Dans le premier (régime statique) deux conditions initiales arbitrairement proches convergent vers le même point fixe, alors que, dans le second (régime chaotique), elles divergent avec une vitesse exponentielle et évoluent vers une distance constante non nulle. La condition critique de transition est obtenue dans un cadre général, mais pour un cas particulier nous retrouvons l'équation de la ligne AT suggérant une forte analogie avec le modèle SK. De plus, la distance quadratique limite en régime chaotique est la même, quelles que soient les conditions initiales, montrant que notre modèle présente une structure ultramétrique. Nous montrons numériquement que cette propriété n'est cependant pas associée à un morcèlement complexe de l'espace des phases comme pour le modèle SK. En outre, nous montrons que le processus d'évolution des neurones est, à la limite thermodynamique, un bruit blanc. Le comportement de notre modèle à la traversée de la ligne AT peut être illustré en étudiant l'entropie de Kolmogorov-Sinaï qui présente une transition brutale à la limite thermodynamique. Elle est nulle en régime statique et infinie en régime chaotique.

Abstract. — We study the dynamics of a discrete time, continuous state neural network with random asymmetric couplings and random thresholds. The evolution of the neurons is given in the thermodynamic limit by a set of dynamic mean-field equations obtained by using a local chaos hypothesis. We study the evolution of the mean quadratic distance between two trajectories, and show there exist two different regimes according to the value of the control parameters. In the first one (static regime) two initially close trajectories evolve to the same fixed point, while, in the second one, (chaotic regime) they diverge with an exponential rate, and evolve to a constant, non zero distance. The critical condition for the transition is obtained in a general frame, but, in a specific case, we recover the equation for the De Almeida-Thouless line suggesting strong analogy with the SK model. Besides, the limit for the quadratic distance is the same for all initial conditions choice, showing that ultrametricity occurs in our model. However, we show numerically that this property is not associated to a complex breaking up of the phase

(*) *Present address*: Universität Bielefeld, Fakultät für Physik, Forschungszentrum BiBos, Postfach 100131, 33501 Bielefeld, Germany.

© Les Editions de Physique 1995

space like in the SK model. Besides, the quenched stochastic process giving the evolution of the neurons is a white noise in the thermodynamic limit. The behaviour of our model when crossing the AT line can be characterized by studying the Kolmogorov-Sinai entropy, which exhibits a sharp transition in the thermodynamic limit. This entropy is zero in the static phase, while it becomes infinite in the chaotic regime.

The emergence of complexity in disordered systems with many degrees of freedom has been the subject of many works for the last decades. In this context, spin-glasses, and more particularly the Sherrington-Kirkpatrick model [1], have been the focus of interest of a large number of studies. Indeed, this model has shown a very rich structure in the low temperature regime, below the De Almeida-Thouless line (AT line) where the replica symmetry has to be broken [2]. In this phase, the model exhibits complex features, such as the existence of many equilibrium states and breaking of ergodicity [3], ultrametricity [4], and lack of self-averaging. However these characteristics are not specific to spin-glasses. Other models such as the Kauffman model [5], or the generalized Hopfield neural network model [6] exhibit similar features.

Neural networks with asymmetric synaptic weights also exhibit a complex behaviour. One of the most striking features, is the existence of a chaotic regime in a given range of values for the gain parameter [7]. However, for finite size, they have a wider diversity of dynamical behaviour. For a large class of neural networks with random asymmetric synaptic weights, in the finite size case, it has been shown in [8, 9] that the generic way leading to chaos by increasing the non-linearity of the transfer function is a quasi-periodicity one. When the gain parameter increases, the system goes from a static to a periodic regime by a Hopf bifurcation. A second Hopf bifurcation then occurs giving rise to a biperiodic regime and the dynamics lives on a T2 torus. Frequency locking occurs on this torus leading to chaos. The intermediate range of parameter values corresponding to the quasi-periodicity route shrinks to zero when the size tends to infinity, leading to a sharp transition from fixed point to chaos in the thermodynamic limit.

Recently, we have numerically shown that the occurrence of chaos in a given class of neural nets is given by the equation of the AT line, suggesting a close relationship between the spin-glass phase for the SK model and the chaotic phase in these neural nets [10]. In this paper we inspect this relationship more deeply. The evolution of the neurons is given in the thermodynamic limit by a set of dynamic mean-field equations. These equations can be obtained by using a local chaos hypothesis [11], whose justification is given in appendix, for *asymmetric* couplings. We first derive these equations and expose some of their consequences, such as the breaking of ergodicity for certain regions in the space of control parameters. Next, we study the evolution of the mean quadratic distance between two trajectories in the thermodynamic limit, and show that there exist two different regimes according to the value of the control parameters. In the first one, corresponding to a static regime, two initially close trajectories evolve to the same fixed point. Hence, they are finally identical with a zero distance. In the second regime, there is a sensitivity to the initial conditions, namely two initially close trajectories diverge with an exponential rate, and evolve to a non zero distance. The critical condition for the transition is obtained in a general frame, but, in the particular case of the model studied in [10], we recover the equation for the AT line. The limit for the quadratic distance is the same for all the choices of initial conditions, showing that *ultrametricity* occurs in this model. However, we numerically show that this property is not associated to a complex breaking up of phase space like in the SK model. Besides, we show that the quenched stochastic process giving the evolution of the neurons is a white noise in the thermodynamic limit.

The behaviour of our model, when crossing the AT line, can be characterized by studying the Kolmogorov-Sinai entropy, which exhibits a sharp transition in the thermodynamic limit. This entropy is zero in the static phase, while it becomes infinite in the chaotic regime. We suggest that there may exist a relationship between this entropy and the exponential rate of growth for the number of TAP solutions in the SK model. In these two cases, the crossing of the AT line corresponds to an increase in complexity.

Model

We consider the following discrete time neural networks([1]), with dynamics:

$$\begin{cases} x_i(t+1) &= f(u_i(t+1)) \\ u_i(t+1) &= \sum_{j=1}^{N} J_{ij} x_j(t) + \theta_i \end{cases} \tag{1}$$

The net is fully connected. The J_{ij}'s are independent, identically distributed random variables with expectation $E(J_{ij}) = \overline{J}/N$ and variance $\mathrm{Var}(J_{ij}) = J^2/N$. They are not symmetric, i.e., $J_{ij} \neq J_{ji}$. The thresholds θ_i are independent, identically distributed *Gaussian* random variables with expectation $E(\theta_i) = \overline{\theta}$ and variance $\mathrm{Var}(\theta_i) = \sigma_\theta^2$. The disorder is quenched, i.e the couplings and thresholds do not change with time. f is an arbitrary (derivable) sigmoidal function of slope g. As an example, it can be $f(x) = \tanh(gx)$ or $f(x) = (1 + \tanh(gx))/2$.

Local Chaos Hypothesis

The system (1) may be viewed as a N-dimensional stochastic process with *quenched* disorder induced by the couplings J_{ij} and the thresholds θ_i. The probability law of the $x_i(t)$'s and $u_i(t)$'s may be obtained by using the *"local chaos hypothesis"* initiated by Amari [11] in the field of neural networks([2]). It is the assumption that, when N is large, the system behaves as if the random variables $x_i(t)$ were independent of each other and of the random variables J_{kl}, $\forall k$, l [12,13]. This hypothesis allows us to state conjectures about the behaviour of large systems of randomly coupled equations.

However, this hypothesis was shown to be false in many models with *symmetric* couplings. For example, applied to the Sherrington-Kirkpatrick model, it leads to the replica symmetric SK solutions and predicts a Gaussian distribution for the local field. Such a result is known to be wrong in the low temperature regime [14]. This is because the mean-field equation for the ith spin obtained under this assumption does not take into account the "reaction term" which is the contribution due to the influence of the ith spin on the others [15]. In the field of neural networks, the same objection can be formulated. For example the rigorous results obtained in [16,17] for extremely diluted models, or in [18,19] for the Little-Hopfield model with parallel dynamics, show that the *feedback* effects due to the *symmetry* of the couplings have a dramatical influence on the distribution of the local field by adding a *non-Gaussian* contribution. Then, from a general point of view, the main defect of the local chaos hypothesis is that it neglects these feedback effects.

([1]) This kind of model is also called the (fully) asymmetric SK model.

([2]) The terminology "local chaos hypothesis" is a little bit confusing in this paper, where we also deal with dynamical chaos. However, by reference to Amari, to the molecular chaos hypothesis of Boltzmann and to what the probabilists call "chaos propagation" we will keep this terminology.

However, we argue in the appendix that the local chaos hypothesis leads to correct results when the couplings are *asymmetric* because *the feeback effects vanish in the thermodynamic limit.* From a mathematical point of view, some results in favour of the local chaos hypothesis have been given by Geman [12], Geman and Hwang [13] and more recently by Guionnet and Ben Arous [20] for *linear* models with asymmetric couplings. Applied to the continuous time neural network of Sompolinsky *et al.* [7], the local chaos hypothesis leads to the *correct* mean-field equations, i.e. the same as those obtained by these authors by using a dynamic mean-field theory [12,21]. We have widely numerically checked this hypothesis for our model in a previous article [9]. Additional evidence for its validity will appear in this paper.

Mean-Field Equations

The network law is symmetric, i.e., all the weights (resp. all the thresholds) have the same distribution. So if we suppose, without loss of generality, that the initial conditions are identically distributed, all the x_i's will have the same distribution at each time step. Then, by the central limit theorem([3]), the consequence of the local chaos hypothesis is that the $u_i(t)$'s become, in the thermodynamic limit, *independent, identically distributed random processes with a Gaussian distribution.* Let $\mu(t) = \langle u(t) \rangle$ and $\nu(t) = \langle u^2(t) \rangle - \mu^2(t)$ be the mean and variance of the $u_i(t)$'s, in the thermodynamic limit, where $\langle\ \rangle$ denotes the average over the quenched disorder. Knowing $\mu(t)$ and $\nu(t)$, all the moments of the $x_i(t)$'s together with their probability density can be obtained. By using the local chaos hypothesis, one obtains the evolution equations for $\mu(t)$ and $\nu(t)$, namely:

$$\mu(t+1) = \overline{J} m(t) + \overline{\theta} \tag{2a}$$

$$\nu(t+1) = J^2 q(t) + \sigma_\theta^2 \tag{2b}$$

The quantities $m(t)$ and $q(t)$ are respectively the first and the second order moments of the $x_i(t)$'s in the thermodynamic limit, given by:

$$m(t) = \int_{-\infty}^{+\infty} \mathrm{D}h f(\sqrt{\nu(t)}h + \mu(t)) \tag{3a}$$

$$q(t) = \int_{-\infty}^{+\infty} \mathrm{D}h f^2(\sqrt{\nu(t)}h + \mu(t)) \tag{3b}$$

where $\mathrm{D}h = \frac{1}{\sqrt{2\pi}} \exp -\frac{h^2}{2}\, \mathrm{d}h$ is the Gaussian measure.

To characterize the Gaussian process $u_i(t)$ completely, one also needs to compute the evolution equation for the covariance $\Delta(t,t') = \langle u_i(t) u_i(t') \rangle - \langle u_i(t) \rangle \langle u_i(t') \rangle$. By using the local chaos hypothesis one obtains:

$$\Delta(t+1, t'+1) = J^2 C(t,t') + \sigma_\theta^2 \tag{4a}$$

where

$$C(t,t') = \langle x_i(t) x_i(t') \rangle$$
$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{D}h \mathrm{D}h' f\left(\frac{\sqrt{\nu(t)\nu(t') - \Delta^2(t,t')}}{\sqrt{\nu(t')}} h + \frac{\Delta(t,t')}{\sqrt{\nu(t')}} h' + \mu(t) \right) f\left(h' \sqrt{\nu(t')} + \mu(t') \right) \tag{4b}$$

([3]) The u_i's have a finite variance because of the scaling law for the couplings and because each x_i's is bounded.

The set of equations (2), (3a, b), (4a, b) allows us to know the evolution laws of the quenched Gaussian process (1). These equations have been previously obtained by Molgedey *et al.* [22] by using a dynamic mean field theory, for the restricted case $\overline{\theta} = 0$, $\sigma_\theta^2 = 0$, $\overline{J} = 0$, $f(x) = \tanh(gx)$, and in the presence of an external white noise.

Fixed Point Equations

After a certain transient, the dynamics flows towards a stationary state that may be a static fixed point or a more complex attractor [9]. We are mainly interested in this stationary regime. The convergence towards a stationary state for (1) is firstly materialized by the convergence of (2a, b) towards a fixed point. The following fixed point equations give the *instantaneous* distribution of the u_i's (resp. the x_i's) in the stationary regime([4]):

$$\mu = \overline{J}m + \overline{\theta} \tag{5a}$$

$$\nu = J^2 q + \sigma_\theta^2 \tag{5b}$$

where:

$$m = \int_{-\infty}^{+\infty} \mathrm{D}h f\left(h\sqrt{J^2 q + \sigma_\theta^2} + \overline{J}m + \overline{\theta}\right) \tag{6a}$$

$$q = \int_{-\infty}^{+\infty} \mathrm{D}h f^2\left(h\sqrt{J^2 q + \sigma_\theta^2} + \overline{J}m + \overline{\theta}\right) \tag{6b}$$

More general information can be gained by studying the covariance $\Delta(t, t')$ in the stationary regime. This point will be discussed below. First we shall note some interesting features revealed by (5a, b), (6a, b).

Breaking of Ergodicity

The self-consistent equations (5a, b), (6a, b) can have several solutions. This corresponds to the existence of several attractors in the phase space of (1) [9]. This allows a critical manifold to be computed in the (4-dimensional) space of the control parameters g, $\overline{\theta}$, σ_θ^2, $\overline{J}$, (J is a redundant parameter), that divides this space into 2 regions. In the first region the system (1) admits only one attractor, while in the other region several attractors exist simultaneously whose attraction basins divide the phase space([5]). Hence, the crossing of this manifold corresponds to a breaking of (global) ergodicity for (1). It has been computed in [9] for the case $\overline{J} = 0$, $f(x) = (1 + \tanh(gx))/2$, $\sigma_\theta^2 = 0$ and in [23] for the case $\overline{J} = 0$, $f(x) = (1 + \tanh(gx))/2$. In this particular case, the crossing of the critical (two-dimensional) manifold in the space (g, $\overline{\theta}$, σ_θ^2) corresponds to the appearance of a second attractor, corresponding to a *saddle-node* bifurcation for (2a, b).

The general case is more difficult to study. For $\overline{J} \neq 0$, one has to look for all the fixed points of the non-linear, two-dimensional system (3). On the contrary for $\overline{J} = 0$ we only have to seek the fixed point of a one-dimensional recurrence. In this paper, all the *numerical* investigations

([4]) As a consequence of the local chaos hypothesis these equations are the SK replica symmetric solutions. This point is discussed below.

([5]) Notice that this second region may *a priori* be divided into sub-regions, each one corresponding to a different number of attractors. For the moment we have only looked at the case $\overline{J} = 0$, and we have not encountered this situation. Notice however that the solutions of (6a, b) are generically isolated and in a finite number.

will be restricted to the model where $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, $J = 1$ and $\overline{\theta} \geq 0$ (but the theoretical results will be given in the general frame). In this case, there is always only one attractor([6]), whatever the gain parameter value.

Evolution of the Mean Quadratic Distance

We now study the evolution of the mean quadratic distance between two trajectories $\mathbf{u}^1(t) = \{u_i^1(t)\}_{i=1\ldots N}$, $\mathbf{u}^2(t) = \{u_i^2(t)\}_{i=1\ldots N}$ in the stationary regime, and in the thermodynamic limit, i.e the evolution of:

$$d_{12}^2(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left\langle \left[u_i^1(t) - u_i^2(t)\right]^2 \right\rangle \tag{7}$$

This method is similar to that used by Derrida and Pomeau for the Kauffman model [24].

Under the local chaos hypothesis the $u_i^1(t)$'s (resp the $u_i^2(t)$'s) are identically distributed, so it is sufficient to study the distance between one component of each trajectory $u^1(t)$, $u^2(t)$, namely:

$$d_{12}^2(t) = \left\langle \left[u^1(t) - u^2(t)\right]^2 \right\rangle = 2\left[\nu - \Delta_{1,2}(t)\right] \tag{8}$$

where ν is given by (5b) and where $\Delta_{1,2}(t) = \langle u^1(t)u^2(t)\rangle - \langle u^1(t)\rangle \langle u^2(t)\rangle$.

$u^1(t)$, $u^2(t)$ being Gaussian, the equation giving the evolution of $\Delta_{1,2}(t)$ in the stationary regime is:

$$\begin{aligned} &\Delta_{1,2}(t+1) = H(\Delta_{1,2}(t)) = \\ &= J^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{D}h\mathrm{D}h' f\left(\frac{\sqrt{\nu^2 - \Delta_{1,2}^2(t)}}{\sqrt{\nu}} h + \frac{\Delta_{1,2}(t)}{\sqrt{\nu}} h' + \mu \right) f(h'\sqrt{\nu} + \mu) + \sigma_\theta^2 \end{aligned} \tag{9}$$

In the particular case $\sigma_\theta^2 = 0$, $\overline{J} = 0$, $\overline{\theta} = 0$, $f(x) = \tanh(gx)$ one recovers an equation obtained by Molgedey *et al.* [22].

This one-dimensional map admits, for *every* value of the control parameters, the fixed point $\Delta_{1,2} = \nu$; Indeed, $H(\nu) = J^2 q + \sigma_\theta^2 = \nu$. This simply means that two initially identical trajectories remain equal forever in this quenched model.

We expect ν to be a *stable* fixed point for (9) in the static regime for (1). This means that in this regime two arbitrary close trajectories converge to the same fixed point. On the contrary the situation may be reversed in a chaotic regime. Hence, the entry into chaos for (1) may be seen by the destabilization of ν for the mapping (9).

The critical condition for the destabilization is:

$$\frac{\mathrm{d}H}{\mathrm{d}\Delta_{1,2}}(\Delta_{1,2} = \nu) = 1 \tag{10}$$

This derivative can easily be computed by writing $H(\Delta_{1,2})$ as a Fourier transform:

$$H(\Delta_{1,2}) = J^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\mathrm{d}k\mathrm{d}k'}{2\pi 2\pi} \mathrm{e}^{\imath k'\mu} \mathrm{e}^{\imath k\mu} \hat{f}(k)\hat{f}(k') \mathrm{e}^{-[\nu(k^2+k'^2)+2\Delta_{1,2}kk']/2} + \sigma_\theta^2 \tag{11}$$

([6]) For $\overline{\theta} = 0$ the model admits two attractors symmetric with respect to the origin, but one can select one attractor or the other, only by inverting the sign of the initial conditions. For $\overline{\theta} < 0$, there exists a region in phase space where two attractors coexist [9].

where $\hat{f}(k)$ is the Fourier transform of f. One then obtains the critical condition:

$$\frac{dH}{d\Delta_{1,2}}(\Delta_{1,2}=\nu) = J^2 \int_{-\infty}^{+\infty} Dh f'^2 \left(h\sqrt{J^2 q + \sigma_\theta^2} + \overline{J}m + \overline{\theta} \right) = 1 \tag{12}$$

where f' is the derivative of f. This equation is *general* because it depends on *all* the control parameters of our model (namely g, $\overline{J}$, $\overline{\theta}$, σ_θ^2) and because f is arbitrary.

In the particular case $f(x) = (1+\tanh(gx))/2$, $\overline{J} = 0$, equation (12) is identical to the critical condition for the entrance to the dynamical regime (i.e. the destabilization of the fixed points) found in [9]. For $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, we recover the critical condition for the entrance to the dynamical regime found in [10]. In these two papers, we have only given the theoretical value for the *destabilization* of the fixed points ; we have then checked numerically that the destabilization line and the line for the entrance to the chaotic regime become closer and closer when the system size tends to infinity, suggesting that they are identical in the thermodynamic limit. Here we give a proof of this identity.

For $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, (12) is the equation of the AT line, where g plays the role of the inverse temperature β, and $\overline{\theta}$ is an external field. In the general case, equation (10) does not correspond to a line, but rather to a manifold in the space of control parameters. However, by analogy with spin glasses and because our numerical investigations are only made for the model $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, the case $\frac{dH}{d\Delta_{1,2}}(\Delta_{1,2} = \nu) < 1$ will be referred to as "above the AT line". Indeed, g is equivalent to the inverse temperature β and the slope $\frac{dH}{d\Delta_{1,2}}(\Delta_{1,2} = \nu)$ increases with g. The case $\frac{dH}{d\Delta_{1,2}}(\Delta_{1,2} = \nu) > 1$ will be referred to as "below the AT line".

The destabilization of ν implies that $\Delta_{1,2}(t)$ converges to another stable fixed point, corresponding to *a non-zero quadratic distance* d^*. As an example, the curve $H(\Delta)$ for the model $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, and $\overline{\theta} = 0.1$ is drawn in Figure 1 in the two situations. This result implies that in this regime two arbitrary close initial conditions diverge with an exponential rate, *corresponding to a chaotic regime.*

Another interesting feature is that all trajectories converge to a stationary state where *the mutual distance between trajectories is always equal to* d^*. This means that the space of trajectories of (1) is *ultrametric* [25]. In the chaotic phase all trajectories evolve towards a stationary, chaotic state where the average distance between two arbitrary trajectory is constant. Explanations of this property are given below.

REMARK. — When several attractors exist, corresponding to different solutions for equations (5, 6), one has to solve (12) *for each attractor.* This case has been studied in [9], for the model $f(x) = (1 + \tanh(gx))/2$, $\overline{J} = 0$. It has been shown that, in the region where two attractors coexist, corresponding to two sets of solutions (m, q), equation (12) has a solution for *only one* of these sets. In this case, that means that only one fixed point leads to a strange attractor by increasing the gain parameter while the other fixed point remains stable forever.

NUMERICAL CHECK. — To check our result numerically we have studied the evolution of the Hamming distance between two configurations, *in the same net*, after a sufficiently long transient $t_0 = 1000$ time steps, insuring that the dynamics lives on the attractor. A strange attractor carries an ergodic measure [26]. Thus we averaged the Hamming distance over time, namely:

$$D_N(\mathbf{u}^1, \mathbf{u}^2) = \frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[u_i^1(t) - u_i^2(t)\right]^2 \tag{13}$$

where $T = 100$, and $N = 500$. $D_N(\mathbf{u}^1, \mathbf{u}^2)$ is drawn *versus* g for various $\overline{\theta}$ values (Figs. 2, 3).

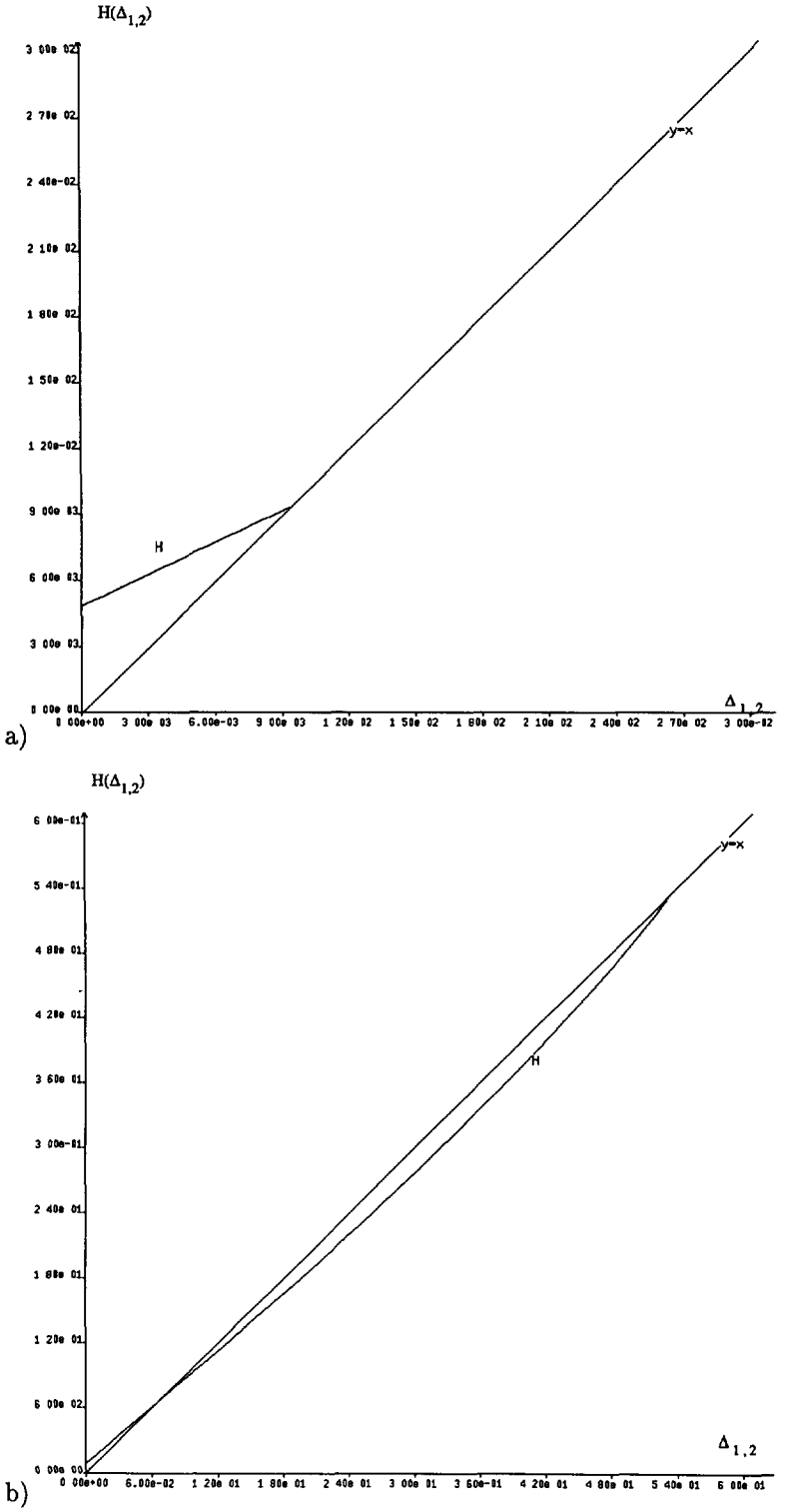


Fig. 1. — Fixed points of the function $H(\Delta_{1,2})$, for $g = 0.7$ (Fig. 1a) and $g = 2$. (Fig. 1b).

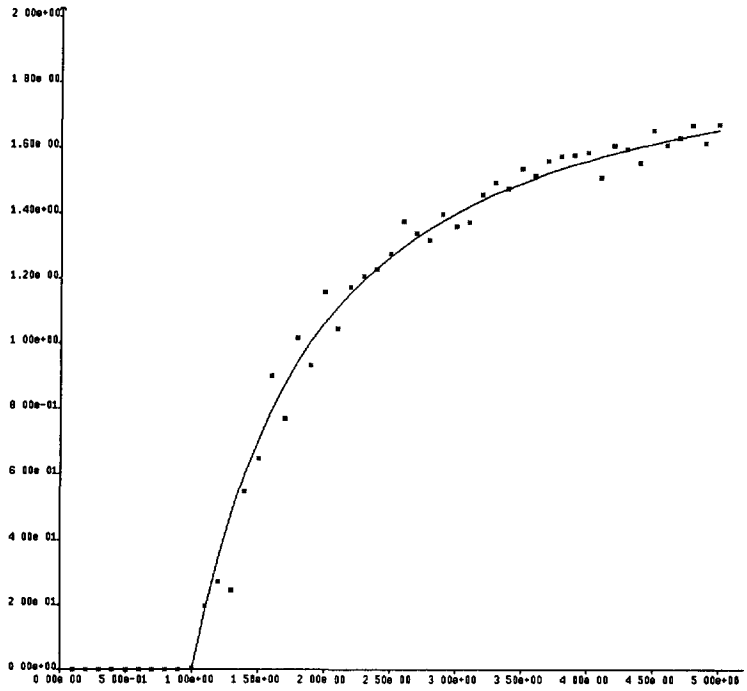


Fig. 2. — Evolution of the average distance $D_N(\mathbf{u}^1, \mathbf{u}^2)$ *versus* g, for $\overline{\theta} = 0$, $N = 500$. The theoretical curve is drawn in full line. The theoretical transition occurs at $g = 1$.

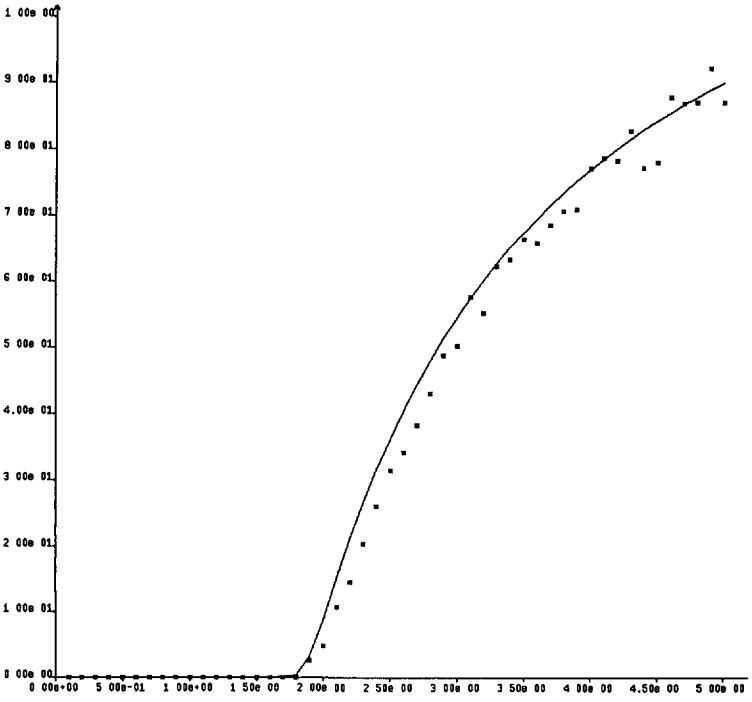


Fig. 3. — Evolution of the average distance $D_N(\mathbf{u}^1, \mathbf{u}^2)$ *versus* g, for $\overline{\theta} = 0.5$, $N = 500$. The theoretical transition occurs at $g = 1.87$.

The numerical results follow the theoretical curve with a good accuracy. The critical transition is very close to the theoretical value predicted by (12).

Nature of the Chaotic Phase

DISTRIBUTION OF THE OVERLAPS. — Equations (6) are identical to the SK equations while (12) is the equation of the AT-line. This suggests a strong analogy between the SK model and our neural net. In particular we expect a close relationship between the spin-glass phase and the chaotic regime of (1). We now inspect this relationship.

Below the AT line, the SK model presents a very rich structure. Many equilibrium states coexist, separated by free-energy barriers whose heights diverge in the thermodynamic limit, and whose number tends to infinity with the system size [27]. As a consequence the ergodicity is broken in the spin-glass phase according to a very complex scheme [28,29]. The distribution of overlaps between the pure states for a given sample is given by:

$$W_J(q) = \lim_{N\to\infty} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \delta(q - Q(t)) \tag{14}$$

$$Q(t) = \frac{1}{N} \sum_{i=1}^{N} u_i^1(t) u_i^2(t) \tag{15}$$

In these formulae $u_i^1(t)$, $u_i^2(t)$ are two distinct spins trajectories and $Q(t)$ is the overlap between these trajectories. The quantity $W_J(q)$ is not self-averaging and fluctuates from sample to sample even in the thermodynamic limit; this is because the weights of the pure states (corresponding to the contribution of a pure state to the Gibbs state) are not self- averaging. This situation is consistent with the replica symmetry breaking scheme of Parisi [28, 29] and the average distribution $W(q) = \langle W_J(q) \rangle$ can be analytically computed under this scheme [4]. It has a complex (and non-Gaussian) shape.

By analogy with spin-glasses, we expect a similar behaviour in our model. One may expect *a priori* a complex breaking up of the phase space for (1), when the model is in the chaotic phase (below the AT line). The analogy between (1) and the SK model is more relevant in the case $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, $J = 1$ and $\overline{\theta} \geq 0$. In this case, our theory predicts that only one attractor exists. This attractor is chaotic and the theory of chaotic systems asserts that the dynamics on a strange attractor is ergodic. Thus, with our theory we do not expect any breaking of ergodicity. In the general case (in particular $\overline{J} \neq 0$) several attractors exist leading to a breaking of ergodicity. However, these attractors are generically in finite number, so the situation is far much simpler than in the SK model. This raises a first opposition between the expected analogy with the SK model and our predictions based on the local chaos hypothesis.

Another straightforward contradiction can be seen on equations (6a, b). These equations are the SK solutions. In the SK model, these equations, corresponding to replica symmetry, are unstable below the AT line. On the other hand, the local chaos hypothesis implies the replica symmetry([7]). Thus we are faced with the following alternatives: either our mean-field equations are incorrect, or our model has a structure different from the SK model.

To decide what is the correct alternative, we use a numerical method developed by Young [31] for spin-glasses. For $t > t_0 = 1000$ time steps, in the case $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, $\overline{\theta} \geq 0$, we calculate the probability distribution $W(q) = \langle W_J(q) \rangle$ where $W_J(q)$ is given by (14).

([7]) The opposition between the local chaos hypothesis and the replica symmetry breaking below the AT line has been previously noticed by Crisanti *et al.* [30].

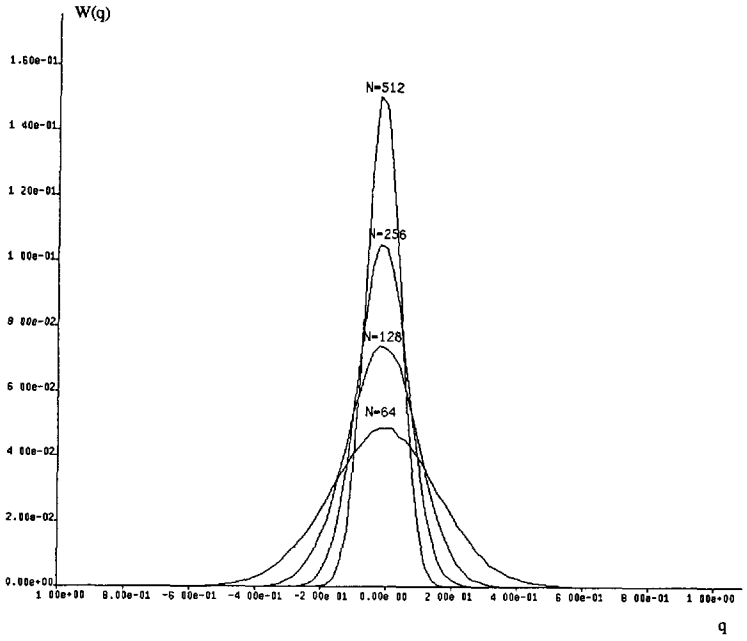


Fig. 4. — Histogram of $W(q)$ for $N = 64,\ 128,\ 256,\ 512$ and $\overline{\theta} = 0$.

If replica symmetry breaking occurs in the chaotic regime, we expect $W(q)$ to have a complex shape. On the contrary, the replica symmetric solutions (6a, b) (and (9)) predict that $W(q)$ is in the thermodynamic limit a Dirac distribution located at $q = q_{12} = \Delta^*_{1,2} + m^2$, where $\Delta^*_{1,2}$ is the stable fixed point for (9) and m is given by (6a). We numerically look for the shape of $W(q)$, when the size N increases in the case $g = 2.5$, $\overline{\theta} = 0$, i.e., under the same conditions as Young's experiments ($1/g = 0.4$). In all the experiments $T = 10000$ while the sample average is made over 20 nets. For these parameter values, our mean-field theory predicts $\Delta^*_{1,2} = 0$ and $m = 0$. Hence, $W(q)$ may have, for finite size, a Gaussian shape centered at zero with a variance decreasing proportionally to $1/N$. We observe such a behaviour. We have computed $W(q)$ for $N = 64,\ 128,\ 256,\ 512$ (Fig. 4). The evolution of the variance with N is shown Figure 5. We have a decrease of this variance with a power law $N^{-\nu}$, and with a computed exponent $\nu = 1.104 \pm 0.041$ in good agreement with our prediction.

These experiments show that *no replica symmetry breaking occur in our model, below the AT line.* They also answer the question of self-averaging raised in [10]. It has been argued in this paper that the characteristics of the ergodic measure carried by the strange attractor in (1) are expected to depend widely on the dynamics of a particular net. Therefore, this measure might depend on the *peculiar realization of the couplings* J_{ij}, and so it might be *sample-dependent, even in the thermodynamic limit.* The quenched randomness generates a family of random measures carried by random strange attractors in the chaotic regime. In this case, a sample averaging might be different from a quenched averaging. Hence (1) is expected not to be self-averaging below the AT-line. On the contrary, the local chaos hypothesis implies self-averaging (by the law of large numbers).

Our experiments show that the crossing of the AT line in (1) does not lead to this loss of self-averaging. This is a very surprising result, because it implies that the strange attractors are almost surely identical (i.e. carry the same ergodic measure) in the thermodynamic limit. This behaviour will be clarified in the following section.

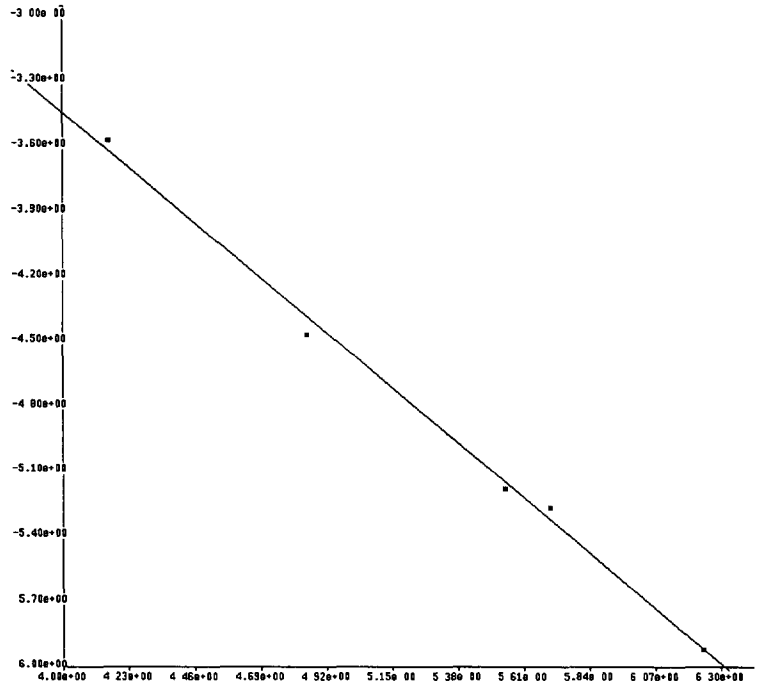


Fig. 5. — Plot of the logarithm for the computed width of $W(q)$ *versus* $\log(N)$ for $N = 64, 128, 256, 300, 512$. In full line the regression line is drawn.

NATURE OF THE GAUSSIAN PROCESS. — These experiments show that our neural net has simpler properties than the SK model below the AT line. To further characterize the dynamics of this model, we study the evolution of the covariance $\Delta(t,t')$, given by (4a, b). However, this is not a simple task. For example the covariance $\Delta(t+1,t)$ between two successive time steps is given by:

$$\Delta(t+1,t) =$$
$$= \sigma_\theta^2 + J^2 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{D}h\mathrm{D}h' \left(h\frac{\sqrt{\nu(t)\nu(t-1) - \Delta^2(t,t-1)}}{\sqrt{\nu(t-1)}} + \frac{h'\Delta(t,t-1)}{\sqrt{\nu(t-1)}} + \mu(t) \right)$$
$$\times f\left(h'\sqrt{\nu(t-1)} + \mu(t-1)\right) \tag{16}$$

To obtain the dynamics for $\Delta(t+1,t)$ one has to iterate simultaneously (2a), (2b), (16). However, the Gaussian process evolves towards a stationary regime where $\nu(t) = \nu$, $m(t) = m$, $\Delta(t,t') = \Delta(t-t')$, with:

$$\Delta(t-t') = H(\Delta(t-t')) =$$
$$= J^2 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{D}h\mathrm{D}h' f\left(h\frac{\sqrt{\nu^2 - \Delta^2(t-t')}}{\sqrt{\nu}} + \frac{h'\Delta(t-t')}{\sqrt{\nu}} + \mu \right) f\left(h'\sqrt{\nu} + \mu\right) + \sigma_\theta^2 \tag{17}$$

where H is the same function as that giving the evolution of the mean quadratic distance between two trajectories (Eq. (9)). Then, $\Delta(t,t')$ evolves towards $\Delta(t-t') = \Delta^*$, which is a *stable fixed point of* H.

According to the position of the system in the control parameter space, with respect to the AT line, one has two different situations. Above the AT line the stable fixed point is $\Delta^* = \nu$. Then each neuron state is a Gaussian process with mean μ and covariance ν. The covariance matrix for the n time step vector $(u_i(1), \ldots, u_i(n))$ has the form: $\begin{pmatrix} \nu & \ldots & \nu \\ \vdots & & \vdots \\ \nu & \ldots & \nu \end{pmatrix}$. Such a Gaussian process is identical to the following one:

$$Y_t : \begin{cases} Y_0 = X \\ Y_{t+1} = Y_t \end{cases} \tag{18}$$

where X is a $N(\mu, \nu)$ Gaussian random variable. Namely, this is a process where almost all trajectories are constants, corresponding to the static regime for (1). The components of the fixed points are distributed according to a $N(\mu, \nu)$ distribution. When there are several fixed points, corresponding to several solutions of (5a, b) (μ_1, ν_1), (μ_2, ν_2), each fixed point is characterized by its own Gaussian distribution $N(\mu_1, \nu_1)$, $N(\mu_2, \nu_2)$, ...

Below the AT line, the stable fixed point is $\Delta^* < \nu$. Then, the covariance matrix for the Gaussian vector $(u_i(1), \ldots, u_i(n))$ becomes:

$$\begin{pmatrix} \nu, \Delta^*, \ldots, \Delta^* \\ \Delta^*, \nu, \ldots, \Delta^* \\ \ldots \nu \ldots \\ \Delta^*, \ldots, \Delta^*, \nu \end{pmatrix} = \begin{pmatrix} \nu - \Delta^*, 0, \ldots, 0 \\ 0, \nu - \Delta^*, 0 \ldots, 0 \\ 0 \ldots \nu - \Delta^* \ldots, 0 \\ 0, \ldots, \nu - \Delta^* \end{pmatrix} + \begin{pmatrix} \Delta^* \ldots \ldots, \Delta^* \\ \Delta^* \ldots \ldots, \Delta^* \\ \ldots \\ \Delta^* \ldots \ldots, \Delta^* \end{pmatrix} \tag{19}$$

The resulting process Z_t is:

$$Z_t : \begin{cases} Z_0 = X \\ Z_{t+1} = Z_t + B_t \end{cases} \tag{20}$$

where B_t is a centered *white noise* of variance $\nu - \Delta^*$, while X is a $N(\mu, \Delta^*)$ Gaussian random variable. This means that Z_t is the superposition of a process with almost all constant trajectories and of a centered *white noise*. The trajectory of a given neuron oscillates stochastically around a fixed value X; for a sample of neurons the X values spread over a Gaussian distribution with mean μ and variance Δ^*. For $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, $\overline{\theta} = 0$, Δ^* is equal to 0 and the trajectory of a given neuron oscillates stochastically around zero [7, 22].

This result calls for several remarks. First, it explains the property of ultrametricity observed above. It is the usual ultrametricity observed over a set of identically distributed random variables ([32] page 51). In fact, to any ultrametric organization appearing in an infinite dimensional space, under certain general conditions, it is possible to assign a stochastic process [25]. The one associated to our model is given by (20).

As a consequence, we find that the "chaotic" regime obtained in the thermodynamic limit is rather simple. In fact all the properties observed here are implied by the local chaos hypothesis, which, in turn is correct only for *asymmetric couplings*. Indeed, we show in the appendix that, for asymmetric couplings, there is no feedback effects in the thermodynamic limit. In this case, the state of the neuron i at time $t + 1$ depends only on the local field at site i and at time t (see Eq. (9) of the Appendix). The loss of feedback effects implies that the system has no memory of its past, and is a Markovian process.

So an important conclusion of this paper is that *all the properties* observed in our model (Gaussian distribution for the local field, replica symmetry, self averaging, Markovian evolution) are due to the *asymmetry* of the couplings. On the contrary, the situation is expected to be drastically different for partially symmetric couplings. In this case, the evolution of the neurons and the stationary state reached after the transients depend on the whole history of the network (see Eq. (9) of the Appendix). This stationary state is then expected to be sample dependent. Obviously the local chaos hypothesis fails in this case.

Maximal Lyapunov Exponent

Our mean-field theory allows the average maximal Lyapunov exponent to be computed, in the thermodynamic limit. Let $u_i^1(t)$, $u_i^2(t)$ be two initially arbitrary close trajectories. The average maximal Lyapunov exponent is:

$$\langle\lambda\rangle = \left\langle \lim_{t\to\infty} \lim_{|u_i^1(0)-u_i^2(0)|\to 0} \frac{1}{2t} \log \frac{[u_i^1(t)-u_i^2(t)]^2}{[u_i^1(0)-u_i^2(0)]^2} \right\rangle \tag{21}$$

where the initial conditions have to be taken in the stationary regime. It is not obvious to compute this quantity. A more straightforward computation can be done on:

$$\langle\omega\rangle = \lim_{t\to\infty} \lim_{d_{12}(0)\to 0} \frac{1}{2t} \log\left(\frac{d_{12}^2(t)}{d_{12}^2(0)}\right) \tag{22}$$

where $d_{12}^2(t) = \left\langle [u_i^1(t)-u_i^2(t)]^2 \right\rangle$ is the mean quadratic distance at time t in the *stationary regime*. Indeed:

$$\begin{aligned}
\langle\omega\rangle &= \lim_{t\to\infty} \lim_{\Delta_{12}(0)\to\nu} \frac{1}{2t} \log\left(\frac{[\nu-\Delta_{1,2}(t)]}{[\nu-\Delta_{1,2}(0)]}\right) \\
&= \lim_{t\to\infty} \lim_{\Delta_{12}(0)\to\nu} \frac{1}{2t} \log\left(\frac{[H^t(\nu)-H^t(\Delta_{1,2}(o))]}{[\nu-\Delta_{1,2}(0)]}\right) \\
&= \lim_{t\to\infty} \frac{1}{2t} \log\left(\frac{dH^t}{d\Delta_{1,2}}\right)_{\Delta_{1,2}=\nu}
\end{aligned}$$

where H is given by (9) and where H^t is H o .. o H, t time. By the chain rule and because ν is a fixed point for H one has $\left(\frac{dH^t}{d\Delta_{1,2}}\right)_{\Delta_{1,2}=\nu} = \left(\frac{dH}{d\Delta_{1,2}}(\nu)\right)^t$. Then:

$$\langle\omega\rangle = \frac{1}{2} \log\left(\frac{dH}{d\Delta_{1,2}}\right)_{\Delta_{1,2}=\nu} \tag{23}$$

$\langle\omega\rangle$ is interpreted by Molgedey *et al.* [22] as the average maximal Lyapunov exponent. But it is not *a priori* evident that $\langle\lambda\rangle$ and $\langle\omega\rangle$ are equal. Indeed, even if one admits that the limits and the expectation commute in (21), one only has, by Jensen inequality:

$$\langle\lambda\rangle \leq \langle\omega\rangle \tag{24}$$

In fact, in a general model, the equality has no reason to hold, in particular if the model is not self-averaging. However, the particular nature of the quenched stochastic process in the

thermodynamic limit (Eqs. (18, 20)) allows for some simplifications. First, this process is ergodic, so the time average is almost surely equal to the expectation, i.e.:

$$\langle\lambda\rangle \overset{a.s}{=} \lim_{t\to\infty} \lim_{|u_i^1(0)-u_i^2(0)|\to 0} \frac{1}{2t} \log \frac{[u_i^1(t)-u_i^2(t)]^2}{[u_i^1(0)-u_i^2(0)]^2} \tag{25}$$

where $\overset{a.s}{=}$ signify almost-surely equal, according to the probability induced by the disorder (joint probability for the couplings and thresholds). Second, the self-averaging property for d_{12} (see previous Sect.) implies that its value obtained for one sample is almost-surely equal to its expectation over the disorder in the thermodynamic limit. Namely:

$$\langle\lambda\rangle \overset{a.s}{=} \lim_{t\to\infty} \lim_{\langle(u_i^1(0)-u_i^2(0))^2\rangle\to 0} \frac{1}{2t} \log \frac{\langle[u_i^1(t)-u_i^2(t)]^2\rangle}{\langle[u_i^1(0)-u_i^2(0)]^2\rangle} = \omega \tag{26}$$

Then, the equality between $\langle\omega\rangle$ and $\langle\lambda\rangle$ occurs only because of the particular behaviour of our model. The mean maximal Lyapunov exponent is given by:

$$\langle\lambda\rangle = \frac{1}{2}\log\left(J^2\int_{-\infty}^{+\infty} \mathrm{D}h f'^2\left(\sqrt{J^2q+\sigma_\theta^2}+\overline{J}m+\overline{\theta}\right)\right) \tag{27}$$

Above the AT line (static regime) $\langle\lambda\rangle < 0$ while below the AT line (chaotic regime) it is positive.

Kolmogorov-Sinai Entropy

All these results have been obtained in the thermodynamic limit. We have shown that a sharp transition occurs when crossing the critical value (10) where the system goes from a static regime to fully developed chaos (white noise). However, for finite sized systems the situation is different. The chaos occurs *gradually* by a cascade of transitions, as the gain parameter is increased. The system goes generically from a fixed point to a limit cycle, then to a T2 torus, and finally to chaos (Figs. 6a, b, c). It may be interesting to understand the mechanism of transition from the finite sized case with "classical" chaos to the infinite dimensional case, where the chaos becomes white noise.

The route to chaos observed in Figures 6a, b, c, corresponds in fact to a gradual growing of complexity. We also expect this complexity to increase with the non-linearities inside the chaotic regime; this is revealed by the loss of apparent structure for the strange attractor when g increases (Fig. 6d) (see also Fig. 7). There are several ways to measure the complexity (Lyapunov exponents, fractal dimensions or Kolmogorov-Sinai entropy). A good candidate to measure it in the chaotic regime is the sum of positive Lyapunov exponents $\lambda_i \quad \sum_{\lambda_i>0} \lambda_i$. Indeed, this quantity is closely related to the Kolmogorov-Sinai entropy $h(\rho)$ (where ρ is an ergodic measure carried by the strange attractor [26]) by [33]:

$$h(\rho) \leq \sum_{\lambda_i>0} \lambda_i \tag{28}$$

The equality occurs for a particular class of ergodic measure (Sinai-Ruelle-Bowen measures) living on axiom A strange attractors [26], but it is very difficult to prove that a system is axiom A. From the Ruelle-Takens-Newhouse theorem we just know that in any (C^2) neighbourhood of a T^3 torus exists an axiom A strange attractor [34]. So, because in our model

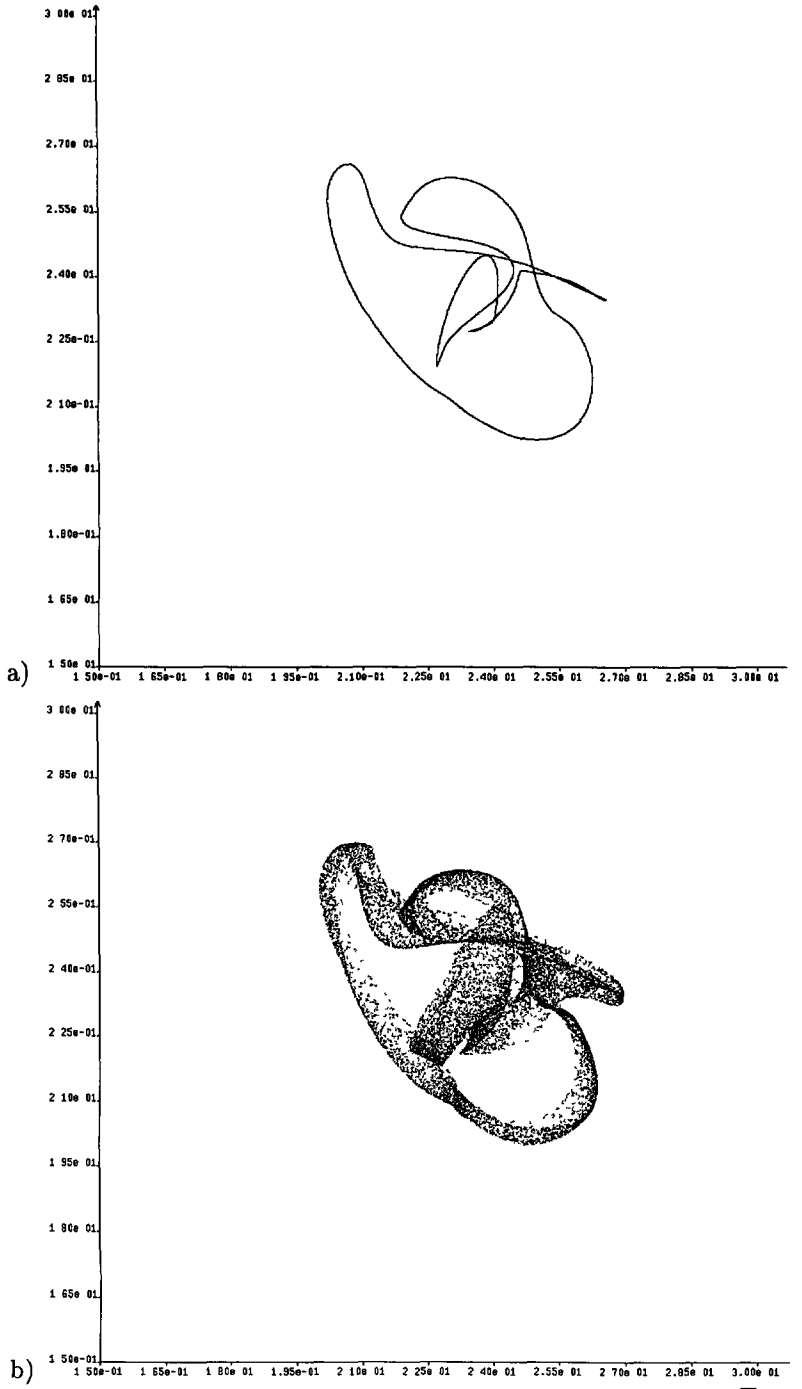


Fig. 6. — Route to chaos for a neural net with $N = 100$, $f(x) = \tanh(gx)$, $\overline{J} = 0$, $\sigma_\theta^2 = 0$, and $\overline{\theta} = 0.1$, when increasing the gain parameter g. We have drawn $m_N(t+1)$ *versus* $m_N(t)$ where $m_N(t)) = \frac{1}{N}\sum_{i=1}^{N} x_i(t)$. This gives a projection of the attractor in two dimensions. Fig. 6a) limit cycle ($g = 1.45$); Fig. 6b) T2 torus ($g = 1.457$); Fig. 6c) Chaos ($g = 1.47$); Fig. 6d) Developed chaos ($g = 1.8$).

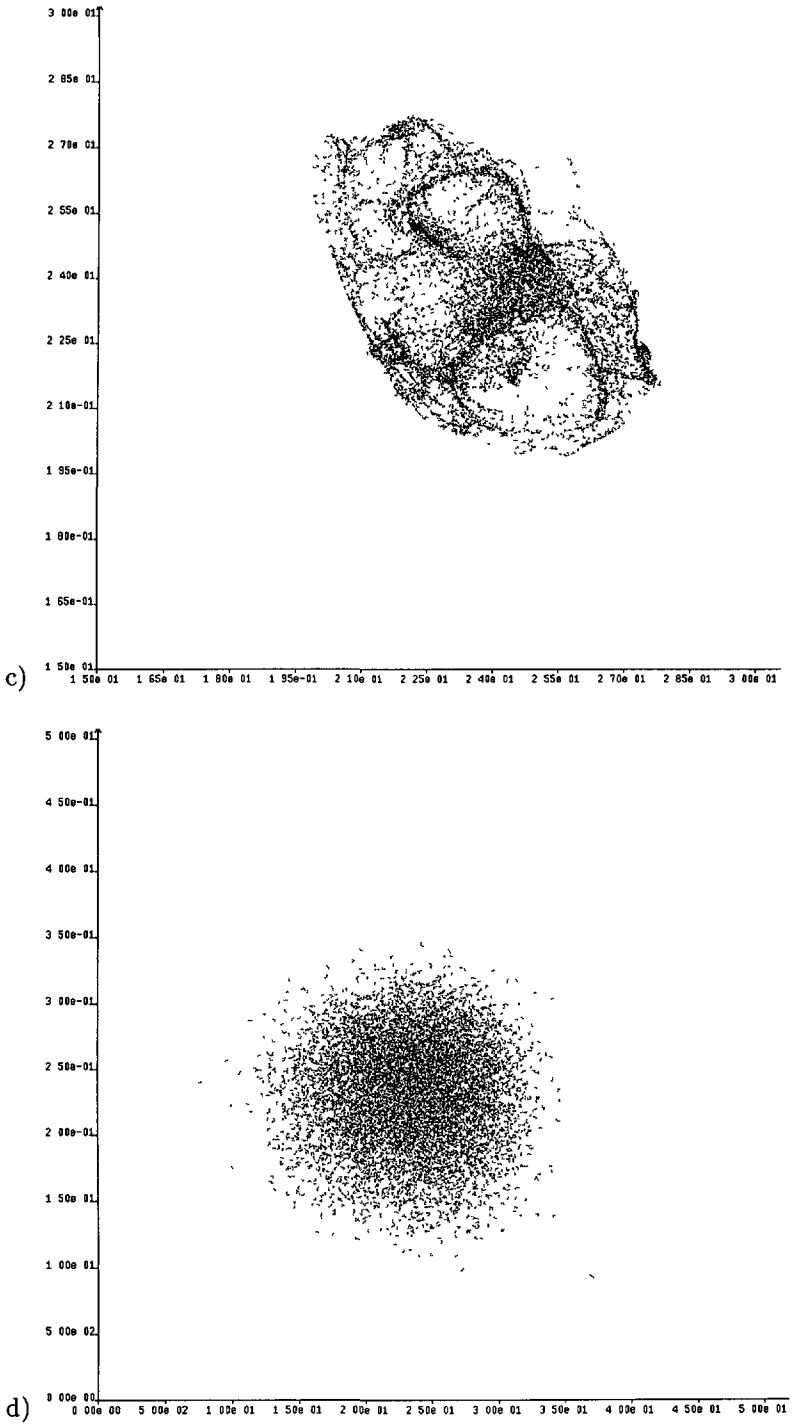


Fig. 6. — continued.

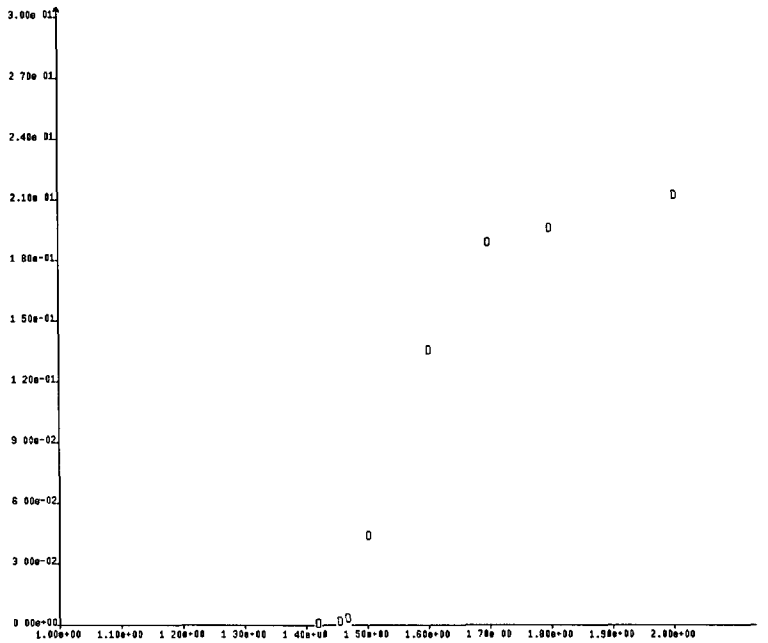


Fig. 7. — Sum of the positive Lyapunov exponent *versus* g for the previous net.

the strange attractors are generated by a quasi-periodicity route, it is reasonnable to suppose that our attractors are axiom A. By the way, for the following, we assume that equality holds in our model.

This quantity is always finite for the finite size model; in particular the number of positive Lyapunov exponents on the attractor is lower than the dimension of phase space. We have drawn the sum of positive Lyapunov exponents in Figure 7, for the net whose dynamical evolution is represented in Figures 6a, b, c, d. The Lyapunov spectrum has been computed by using the algorithm of Eckmann *et al.* [35], with a 30000 time step long trajectory. As expected, it is zero before chaos, and increases sharply after the transition, before saturation. However, we must emphasize that it is difficult to extend the numerical investigations for high gain. Indeed, when the gain g increases, the attractor dimension increases, and so the embedding dimension needed to compute the Lyapunov spectrum. In this case more and more trajectory points are needed and the time of computation becomes prohibitively long.

In the thermodynamic limit (1) becomes a set of independent, identically distributed, monodimensional stochastic processes. All these processes have the same maximal Lyapunov exponent $\langle\lambda\rangle$ given by (27). Above the AT line $\langle\lambda\rangle \leq 0$, then, the Kolmogorov-Sinai entropy is zero. On the contrary, below the AT line $\langle\lambda\rangle$ is positive and the Kolmogorov-Sinai entropy becomes infinite. Hence, this entropy exhibits a sharp transition in the thermodynamic limit. Returning to the finite size case, we expect it to increase more and more accurately with the non-linearity parameter g, when the size increases. Indeed, when the size increases the transition from fixed point to chaos is faster and faster, leading to a sharp transition in the thermodynamic limit. This situation can be analyzed by looking at the Jacobian matrix of the fixed point [8], at least for centered couplings. The distribution of eigenvalues of this random matrix converges in the thermodynamic limit to a uniform density in the complex plane, distributed over a disk with

a radius increasing with g [36]. The destabilization of the fixed point occurs when at least one eigenvalue crosses the unity disk. For finite size the eigenvalues cross two by two (they are complex conjugate for a Hopf bifurcation), leading to a gradual transition from fixed point to chaos. In the thermodynamic limit an *infinite* number of eigenvalues cross simultaneously, leading to an infinite codimension bifurcation, and then to a sharp increase in complexity. The chaos emerging from this bifurcation is infinite-dimensional.

It remains now to understand the relation between this transition and that occurring in the SK model when crossing the AT line. Indeed, we have seen that our chaotic phase is very different from the spin-glass phase. In our model, below the AT line the stationary regime is an ergodic white noise. In the SK model, the dynamics is relaxational but a multiplicity of equilibrium states exists, leading to a complex breaking of ergodicity. However, in the two cases the crossing of the AT line corresponds to an *increase in complexity.* On one hand the Kolmogorov-Sinai entropy in our model grows proportionally to $N\langle\lambda\rangle$ below the AT line, for large size, because (1) consists of N nearly independent, identically distributed monodimensional stochastic processes. On the other hand, in the SK model the complexity is associated to the large number of pure equilibrium states. These pure states are solutions of the Thouless, Anderson, Palmer (TAP) equations [15]. The average number $\overline{N_s}$ of solutions for the TAP equations grows exponentially with the size N, i.e. with an exponent $N\alpha(T) = \log(\overline{N_s})$ computed by Bray and Moore [27]. Above the AT line $\alpha(T) = 0$ while it is positive below the AT line. Hence, a natural measure of the complexity for the SK model is the quantity $N\alpha(T)$([8])

Discussion

In this paper we have fully characterized the dynamics of a discrete time, continuous time neural network with random asymmetric couplings and random thresholds. We have shown, that it exhibits a chaotic transition, given by a general equation, reducing in a specific case to the De Almeida-Thouless line. Even if the chaotic phase is very different from the spin-glass phase in the SK model, the transition corresponds in the two models to a sharp increase in complexity.

However, the situation observed in this paper depends dramatically on the couplings asymmetry. If this asymmetry is broken, the local chaos hypothesis fails, and the mean-field equations will be far more complicated (see Eq. (9) of the Appendix). It would then be interesting to explore the situation from totally symmetric couplings to totally asymmetric ones, by varying continuously the parameter k of asymmetry (see Appendix). For symmetric couplings the model (1) is expected to admit a simple dynamics with trivial attractors (fixed points or period 2 cycles) whatever the gain value ; for sufficiently low gain, it admits only one fixed point, while a multiplicity of trivial attractors exist in the high gain regime, organized into a structure that could be similar to the spin-glass phase. On the contrary, for total asymmetry the organization of attractors is simpler, but the dynamics is more complicated. The intermediate situation is then expected to be very rich. In fact it is not difficult to show that (1) admits only one stable fixed point for sufficiently low gain, whatever is the k value (for sufficiently low gain, f is a contraction). Then, the increase in g is always expected to lead to an increasing complexity. It would be interesting to know whether the critical transition is always given by an equation like (12).

The analogy with spin glasses suggests that the SK model and (1) belong to the same family. It would be worth to know whether this family incorporates other models and what are the reasons for this analogy.

([8]) In fact a more interesting quantity would be log(average number of pure states).

Acknowledgments

I am very grateful to D. Hansel, J. Bellissard, M. Mézard, J.P. Nadal for stimulating and illuminating conversations. I also thank M. Benaïm, M. Samuelides, M. Quoy, B. Doyon for helpful discussions. I would especially like to thank V. Zagrebnov for helpful comments and constructive criticism. I am grateful to "programme cognisciences, CNRS", for financial support.

Appendix A

In this Appendix we give a heuristical argument for the local chaos hypothesis. It is based on the *cavity approach* previously used by Mézard *et al.* [37] for the SK model. For the sake of simplicity, we work with centered couplings, namely $\overline{J} = 0$; in this case the J_{ij}'s are proportional to $1/\sqrt{N}$. Besides, the couplings will have a uniform law; their variance being J^2/N this means that $J_{ij} \in \left[-J\sqrt{3}/\sqrt{N}; J\sqrt{3}/\sqrt{N}\right]$, $\forall i,\ j$. Without loss of generality we will take $f(x) = \tanh(gx)$.

We consider a system of N neurons of internal state $\{u_i(t)\}_{i=1,\ldots,N}$ evolving according to equation (1). Then, we add, at time t_0, a new neuron u_0, $(N+1)$ couplings J_{0j} and a threshold θ_0. Each neuron now feels a new influence due to the added neuron. Consequently their trajectories become new functions $\{\tilde{u}_i(t+1)\}_{i=0,\ldots,N}$. We want to study the evolution of the influence of this new spin when we iterate the dynamics.

The evolution of this system of $N+1$ neurons is now given by:

$$\begin{cases} \tilde{x}_i(t+1) &= f(\tilde{u}_i(t+1)) \\ \tilde{u}_i(t+1) &= \displaystyle\sum_{j=1}^{N} J_{ij}\tilde{x}_j(t) + \theta_i + J_{i0}\tilde{x}_0(t) \end{cases} \tag{A.1}$$

Because of the scaling law for the couplings we have:

$$\begin{aligned} \tilde{x}_i(t+1) &= f\left(\sum_{j=1}^{N} J_{ij}\tilde{x}_j(t) + \theta_i\right) + J_{i0}\tilde{x}_0(t) f'\left(\sum_{j=1}^{N} J_{ij}\tilde{x}_j(t) + \theta_i\right) \\ &\quad + \sum_{n=2}^{+\infty} \frac{J_{i0}^n \tilde{x}_0^n(t)}{n!} f^{(n)}\left(\sum_{j=1}^{N} J_{ij}\tilde{x}_j(t) + \theta_i\right) \end{aligned} \tag{A.2}$$

where f' is the derivative of f (and is proportional to g) and where $f^{(n)}$ is the n-th derivative of f.

At time t_0, $\tilde{x}_i(t_0) = x_i(t_0)$, $i = 1 \ldots N$, then:

$$\tilde{x}_i(t_0+1) = x_i(t_0+1) + J_{i0}\tilde{x}_0(t_0) f'(u_i(t_0+1)) + R_i(t_0+1); \quad i = 0 \ldots N \tag{A.3}$$

where:

$$R_i(t_0+1) = \sum_{n=2}^{+\infty} \frac{J_{i0}^n \tilde{x}_0^n(t_0)}{n!} f^{(n)}(u_i(t_0+1))$$

Here $x_i(t_0+1) = f(u_i(t_0+1)$ and $u_i(t_0+1) = \displaystyle\sum_{j=1}^{N} J_{ij}x_j(t_0) + \theta_i$ is the local field at site i, at time t_0+1, resulting from the influence of all neurons *but the 0 neuron* (cavity field). It follows, that $u_i(t_0+1)$ is independent of $\tilde{x}_0(t_0)$ and of the J_{i0}'s.

The total local field at site $i \neq 0$, at time $t_0 + 2$ is:

$$\tilde{u}_i(t_0+2) = \sum_{j=1}^{N} J_{ij}x_j(t_0+1) + \theta_i + J_{i0}\tilde{x}_0(t_0+1) + \sum_{j=1}^{N} J_{ij}J_{j0}\tilde{x}_0(t_0)f'(u_j(t_0+1)) + \sum_{j=1}^{N} J_{ij}\sum_{n=2}^{+\infty}\frac{J_{j0}^n\tilde{x}_0^n(t_0)}{n!}f^{(n)}(u_j(t_0+1)) \quad (A.4)$$

We are now interested in the behaviour of each term containing $\tilde{x}_0$ in the thermodynamic limit. Because of the scaling law for the couplings the term $J_{i0}\tilde{x}_0(t_0+1)$ tends to zero.

Let us now have a look at the second term. The derivative $f'(u_j(t_0+1))$ is positive and is bounded by a positive quantity g. Besides, the J_{i0}'s are independent of the J_{ij}'s and of the $f'(u_j(t_0+1))$'s, $j \neq 0$. Then,

$$\left\langle \left(\sum_{j=1}^{N} J_{ij}J_{j0}f'(u_j(t_0+1)) \right)^2 \right\rangle =$$

$$= \left\langle \sum_{j=1}^{N} J_{ij}^2 J_{j0}^2 f'^2(u_j(t_0+1)) \right\rangle + 2\left\langle \sum_{\substack{j,k=1 \\ j<k}}^{N} J_{ij}J_{ik}J_{j0}J_{k0}f'(u_j(t_0+1))f'(u_k(t_0+1)) \right\rangle$$

$$\leq g^2 \sum_{j=1}^{N}\langle J_{ij}^2\rangle\langle J_{j0}^2\rangle + 2\sum_{\substack{j,k=1 \\ j<k}}^{N}\langle J_{j0}\rangle\langle J_{k0}\rangle\langle J_{ij}J_{ik}f'(u_j(t_0+1))f'(u_k(t_0+1))\rangle \quad (A.5)$$

The first term converges to zero when $N \to \infty$ while the second is always zero for centered couplings. Then, $\sum_{j=1}^{N} J_{ij}J_{j0}f'(u_j(t_0+1))$ tends to zero (in L^2 sense). In fact, by the same argument it is possible to show L^p convergence for arbitrary $p \geq 2$.

The third term is much more complicated to control. Let $M_n = \sup_x \left| \frac{\mathrm{d}^n}{\mathrm{d}x^n}\tanh(x) \right|$ or, in other words, $\sup_x \left| f^{(n)}(x) \right| = g^n M_n$. Each M_n's is lower than $n!$. Then, the series $\sum_{n=2}^{+\infty} \frac{M_n y^n}{n!}$ converges, and is bounded by $\frac{y^2}{1-y}$, as soon as $|y| < 1$. Then, $\sum_{n=2}^{+\infty} \frac{J_{j0}^n\tilde{x}_0^n(t0)}{n!}f^{(n)}(u_j(t_0+1))$ converges absolutely provided that $g\,|J_{j0}|\,|\tilde{x}_0(t_0)| < 1$.

Now $|\tilde{x}_0(t_0)| \geq 1$, while $|J_{j0}| \leq J\sqrt{3}/\sqrt{N}$ by hypothesis. Then, for each g, there exists $N_g =$ integer part of $3g^2J^2$, such that, if $N > N_g$, the series $\sum_{n=2}^{+\infty} \frac{J_{j0}^n\tilde{x}_0^n(t_0)}{n!}f^{(n)}(u_j(t_0+1))$ converges absolutely. Let:

$$S = \lim_{N\to\infty} \sum_{j=1}^{N} J_{ij} \sum_{n=2}^{+\infty} \frac{J_{j0}^n\tilde{x}_0^n(t_0)}{n!}f^{(n)}(u_j(t_0+1))$$

We have:

$$\begin{aligned} |S| &\leq \lim_{N\to\infty} \sum_{j=1}^{N} |J_{ij}| \sum_{n=2}^{+\infty} |J_{j0}^n| \, |\tilde{x}_0^n(t_0)| \, \frac{g^n M_n}{n!} \\ &\leq \lim_{N\to\infty} \sum_{j=1}^{N} |J_{ij}| \, \frac{g^2 \, |J_{j0}^2|}{1 - g \, |J_{j0}| \, |\tilde{x}_0(t_0)|} \\ &\leq \lim_{N\to\infty} \sum_{j=1}^{N} \frac{\sqrt{3}^3 g^2 J^3}{\sqrt{N}^3} \frac{1}{1 - g\dfrac{\sqrt{3}J}{\sqrt{N}} \, |\tilde{x}_0(t_0)|} = 0 \end{aligned}$$

Then:

$$\tilde{u}_i(t_0+2) = \sum_{j=1}^{N} J_{ij} x_j(t_0+1) + \theta_i + O\left(\frac{1}{\sqrt{N}}\right) = u_i(t_0+2) + O\left(\frac{1}{\sqrt{N}}\right); \quad i = 1 \ldots N$$

where $O\left(1/\sqrt{N}\right)$ is the vanishing contribution of the three terms examined above. Then in the thermodynamic limit:

$$\tilde{x}_i(t_0+2) = x_i(t_0+2); \; i = 1 \ldots N \tag{A.6}$$

Now we can iterate the reasoning at all times showing that neurons $i = 1 \ldots N$ evolve independently of the neuron 0. Extending this argument by adding an arbitrary number of neurons, it follows that in the thermodynamic limit all neurons are independent.

Looking now at the neuron 0 we have, at time $t_0 + 2$:

$$\begin{aligned} \tilde{u}_0(t_0+2) &= \sum_{j=1}^{N} J_{0j}[x_j(t_0+1) + J_{j0}\tilde{x}_0(t_0) f'(u_j(t_0+1)) + R_j(t_0+1)] \\ &\quad + \theta_0 + J_{00}\tilde{x}_0(t_0+1) \end{aligned} \tag{A.7}$$

The $x_j(t_0+1)$'s are independent of the J_{0j}'s by construction and of each other by the previous argument. So, by the central limit theorem, the term $\sum_{j=1}^{N} J_{0j} x_j(t_0+1)$ becomes, in the thermodynamic limit a Gaussian process η.

The contribution of the term in containing $R_j(t_0+1)$ converges to zero by the same arguments as above. However, the second term $\sum_{j=1}^{N} J_{j0} J_{0j} f'(u_j(t_0+1))$ leads to a *non-Gaussian* contribution in the thermodynamic limit. In fact, it converges to zero *only if the couplings are asymmetrics.*

Let us consider the more general situation, where the couplings are partially symmetric, i.e. $J_{ij} = J_{ij}^{\mathrm{s}} + k J_{ij}^{\mathrm{as}}$, where the J_{ij}^{s}'s are symmetric ($J_{ij}^{\mathrm{s}} = J_{ji}^{\mathrm{s}}$), the J_{ij}^{as}'s are antisymmetric ($J_{ij}^{\mathrm{as}} = -J_{ji}^{\mathrm{as}}$) and where the J_{ij}^{s}'s and the J_{ij}^{as}'s have the same expectation $\langle J_{ij}^{\mathrm{s}} \rangle = \langle J_{ij}^{\mathrm{as}} \rangle = 0$ and variance $\mathrm{Var}\,(J_{ij}^{\mathrm{s}})) = \mathrm{Var}\,(J_{ij}^{\mathrm{s}})) = \dfrac{J^2}{N} \dfrac{1}{1+k^2}$. We also assume that they are independent. The parameter k controls the asymmetry of the couplings; for $k = 0$ the couplings are symmetric while for $k = 1$ we recover the case studied in this paper. The correlation between J_{ij} and J_{ji} is given by:

$$\mathrm{Cov}(J_{ij}, J_{ji}) = \langle J_{ij} J_{ji} \rangle = \frac{1-k^2}{1+k^2} \cdot \frac{J^2}{N} \tag{A.8}$$

Returning to (7) we see, that the term $\sum_{j=1}^{N} J_{j0} J_{0j} f'(u_j(t_0+1))$ leads, in thermodynamic limit, to a contribution proportional to $J^2 \frac{1-k^2}{1+k^2}$, that vanishes *only in the asymmetric case* $k = 1$. In fact $\tilde{u}_0(t)$ becomes a stochastic process:

$$\tilde{u}_0(t_0+2) = \eta(t_0+1) + J^2 \frac{1-k^2}{1+k^2} G(t_0+1)\tilde{x}_0(t_0) + \theta_0 \qquad \text{(A.9)}$$

where η is a Gaussian process and $G(t_0+1)$ is the average linear response function to the cavity field $u(t_0+1)$ i.e. $G(t_0+1) = \langle f'(u(t_0+1)) \rangle$. Equation (9) is similar to that established by Crisanti and Sompolinsky for a system of continuous spins [38] with continuous time dynamics. We see that a feedback term $J^2 \frac{1-k^2}{1+k^2} G(t_0+1)\tilde{x}_0(t_0)$ appears in the case $k \neq 1$. This term corresponds to the fact that the neuron 0 at time t_0 acts on the neuron i at time t_0+1; these neurons act in turn on the neuron 0 at time t_0+2. Then the neuron 0 at time t_0+2 feels its own influence at time t_0. This influence disappears in the thermodynamic limit only for asymmetric couplings. In the other cases the dynamics is drastically different from the asymmetric case([9]) because of the presence of the feedback, non-Markovian term $J^2 \frac{1-k^2}{1+k^2} G(t_0+1)\tilde{x}_0(t_0)$. This term implies that the evolution of each neuron depends on the whole history of the network. This implies the failure of the local chaos hypothesis. This might also lead to a loss of self averaging.

References

[1] Sherrington D., Kirkpatrick S., *Phys. Rev. Lett.* **35** (1975) 1792-1795; Kirkpatrick S., Sherrington D., *Phys. Rev. B* **11** (1978) 4384-4403.

[2] De Almeida J.R.L, Thouless D.J., *J. Phys. A: Math. Gen.* **11** (1978) 983-990.

[3] Mackenzie N.D., Young A.P., *Phys. Rev. Lett.* **49** (1982) 301-304.

[4] Mézard M., Parisi G., Sourlas N., Toulouse G., Virasoro M., *J. Phys.* **45** (1984) 843-854.

[5] Derrida B., Dynamics of Automata , Spin Glasses and Neural Network Models, in "Nonlinear Evolution And Chaotic Phenomena" (Plenum Publishing Corporation, 1988) and references therein.

[6] Amit D., Gutfreund H., Sompolinsky H., *Phys. Rev. A* **32** (1985) 1007-101β; *Phys. Rev. Let.* **55** 1530-1533.

[7] Sompolinsky H., Crisanti A., Sommers H.-J., *Phys. Rev. Lett.* **61** (1988) 259-262.

[8] Doyon B., Cessac B., Quoy M., Samuelides M., *Int. J. Bifurcation Chaos* **3** (1993) 279-291.

[9] Cessac B., Doyon B., Quoy M., Samuelides M., *Physica D* **74** (1994) 24-44.

[10] Cessac B., *Europhys. Lett.* **26** (1994) 577-582.

[11] Amari S., *IEEE Trans. Syst. Man. Cyb.* **SMC-2** (1972) 643-657.

[12] Geman S., *SIAM J. Appl. Math.* **42** (1982) 695-703.

[13] Geman S., Hwang C.R.Z., *Wahr. verw. Gebiete* **60** (1982) 291-314.

[14] Thomsen M., Thorpe M. F., T.C. Choy, D. Sherrington, H.J. Sommers, *Phys. Rev. B.* **33** (1986) 1931-1947.

[15] Thouless D. J., Anderson P. W., Palmer R.J., *Philos. Mag.* **35** (1977) 593-601.

([9]) The role of asymmetry in random neural nets has been previously emphasized by Krauth *et al.* [39].

[16] Patrick A. E., Zagrebnov V. A., *J. Phys. A.* **23** (1990) L1323-1329.

[17] Patrick A. E., Zagrebnov V. A., *J. Phys. A.* **25** (1992) 1009-1011.

[18] Patrick A. E., Zagrebnov V. A., *J. stat. Phys.* **63** (1991) 59-71.

[19] Patrick A. E., Zagrebnov V. A., *J. Phys. A.* **24** (1991) 3413-3426.

[20] Guionnet A.- Ben Arous G., (1994) to be published in *Prob. Th. and Rel. Fields.*

[21] Cessac B., Ph.D. thesis (1994).

[22] Molgedey L., Schuchhardt J., Schuster H. G. , *Phys. Rev. Lett.* **69** (1992) 3717-3719.

[23] Quoy M., Cessac B., Doyon B., Samuelides M., Dynamique de réseaux de neurones à temps discret, 7èmes journées Neurosciences et sciences de l'ingénieur, NSI 94, Chamonix (1994).

[24] Derrida B.- Pomeau Y, *Europhys. Lett.* **1** (1986) 45-49.

[25] Rammal R., Toulouse G., Virasoro M., *Rev. Mod. Phys.* **58** (1986) 765-788.

[26] Eckmann J.P, Ruelle D., *Rev. Mod. Phys.* **57** (1985) 617-655.

[27] Bray A.J., Moore M.A., *J. Phys. C* **13** (1980) L469-476.

[28] Parisi G.,*J. Phys. A* **13** (1980) 1101-1112.

[29] Parisi G., *Phys. Rev. Lett.* **50** (1983) 1946-1948.

[30] Crisanti A., Sommers H.-J., Sompolinsky H., Chaos in Neural Networks : chaotic solutions, Preprint (1990).

[31] Young A.P., *Phys. Rev. Lett.* **51** (1983) 1206-1209.

[32] Mézard M., Parisi G., Virasoro M. A., Spin glass theory and beyond (World scientific, Singapore, 1987).

[33] Ruelle D., *Bol. Soc. Bras. Mat.* **9** (1978) 83.

[34] Newhouse S., Ruelle D., Takens F., *Commun. Math. Phys.* **64** (1978) 35.

[35] Eckmann J.P., Kamphorst S. O., Ruelle D., Ciliberto S., *Phys. Rev. A.* **34** (1986) 4971-4979.

[36] Girko V.L., *Theory Prob. Its Appl. (USSR)* **29** (1985) 694-706.

[37] Mézard M., Parisi G., Virasoro M. A., *Europhys. Lett.* **1** (1986) 77-82.

[38] Crisanti A., Sompolinsky H., *Phys. Rev. A* **36** (1987) 4922-4939.

[39] Krauth W., Nadal J.P., Mézard M., *J. Phys. A* **21** (1988) 2995-3011.