

## The signal-to-noise analysis of the Little–Hopfield model revisited

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2004 J. Phys. A: Math. Gen. 37 1951

(<http://iopscience.iop.org/0305-4470/37/6/001>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.65

The article was downloaded on 02/06/2010 at 19:50

Please note that [terms and conditions apply](#).

# The signal-to-noise analysis of the Little–Hopfield model revisited

**D Bollé, J Busquets Blanco and T Verbeiren**

Instituut voor Theoretische Fysica, KU Leuven, B-3001 Leuven, Belgium

E-mail: [desire.bolle@fys.kuleuven.ac.be](mailto:desire.bolle@fys.kuleuven.ac.be), [jordi.busquets@fys.kuleuven.ac.be](mailto:jordi.busquets@fys.kuleuven.ac.be)  
and [toni.verbeiren@fys.kuleuven.ac.be](mailto:toni.verbeiren@fys.kuleuven.ac.be)

Received 11 August 2003

Published 28 January 2004

Online at [stacks.iop.org/JPhysA/37/1951](http://stacks.iop.org/JPhysA/37/1951) (DOI: 10.1088/0305-4470/37/6/001)

## Abstract

Using the generating functional analysis an exact recursion relation is derived for the time evolution of the effective local field of the fully connected Little–Hopfield model. It is shown that, by leaving out the feedback correlations arising from earlier times in this effective dynamics, one precisely finds the recursion relations usually employed in the signal-to-noise approach. The consequences of this approximation as well as the physics behind it are discussed. In particular, it is pointed out why it is hard to notice the effects, especially for model parameters corresponding to retrieval. Numerical simulations confirm these findings. The signal-to-noise analysis is then extended to include all correlations, making it a full theory for dynamics at the level of the generating functional analysis. The results are applied to the frequently employed extremely diluted (a)symmetric architectures and to sequence processing networks.

PACS numbers: 87.18.Sn, 64.60.Cn, 75.10.Hk, 02.50.–r

## 1. Introduction

During the last few years, the treatment of dynamics using generating functional techniques (GFA) has received a lot of attention in the field of statistical mechanics of disordered systems, in particular neural networks (see, e.g., [1–5] and references therein). Such a treatment allows for the exact solution of the dynamics and finds all relevant physical order parameters at any time step via the derivatives of a generating functional [6–9]. An alternative method for studying the dynamics of neural networks is the so called signal-to-noise analysis (SNA) or statistical neurodynamics (see, e.g., [10–13]) where the idea is to start from a splitting of the local field into a signal part originating from the pattern to be retrieved and a noise part arising

from the other patterns. The differences in the existing versions of this approach consist of the different treatments of this noise term, ranging from the assumption that it is Gaussian with various approximations for its variance [10, 11, 14] to a supposedly exact treatment [15, 16].

In two recent papers some comparisons have been made between the two methods for a sequence processing network [4] respectively the fully connected Blume–Emery–Griffiths network [2]. For the first system, surprisingly, the order parameter equations obtained through the exact GFA solution are shown to be completely equivalent to those of statistical neurodynamics, known to be an approximation that assumes Gaussian noise. These theoretical results are verified by computer simulations. We recall that this system contains no feedback correlations. For the second system that is fully connected and, hence, does contain feedback correlations, it has been shown that the results of the GFA and SNA coincide up to the third time step. Some numerical experiments then indicated that they may differ for further time steps, certainly for those parameters of the system corresponding to spin-glass behaviour.

The idea of the present work is to perform a systematic analytical study of the relationship between the two techniques using the fully connected Little–Hopfield model. In order to do so, we take the GFA one step further by deriving a recursion relation for the effective local field. To our knowledge such a recursion relation has not yet been reported in the literature. Furthermore, it is precisely this relation that we use as the basis for studying the correspondence between the two methods.

For the fully connected model we show that the SNA as it has been applied up to now in fact approximates the exact dynamics because it forgets about a part of the correlations. We discuss the physics behind such a short-memory approximation and explain why it leads to very good results in the case of retrieval. Moreover, we show how to apply the SNA correctly leading to a complete equivalence with the GFA. The results obtained are applied to other architectures and to sequence processing networks.

The paper is organized as follows. In section 2, we recall the fully connected Little–Hopfield model and discuss the SNA approach to solve its dynamics. Section 3 shortly reviews the GFA and derives recursion relations for the effective local field in this framework. Section 4 introduces the short-memory approximation and shows how this reduces the results of the GFA to those of the SNA. It also explains the physics behind this approximation and discusses some numerical results. Section 5 presents a scheme of how to apply exactly the SNA. In section 6, we apply our findings to the extremely diluted symmetric and asymmetric architectures and to sequence processing models. Finally, section 7 contains some concluding remarks.

## 2. Signal-to-noise analysis of the Little–Hopfield model

### 2.1. The Little–Hopfield model

By now, the Little–Hopfield model is a standard model for associative memory and can be found in many textbooks (see, e.g., [12]). Consider a system of  $N$  Ising spins  $\sigma_i$ ,  $i = 1, \dots, N$ . We want to store  $p = \alpha N$  patterns  $\xi^\mu = \{\xi_1^\mu, \dots, \xi_N^\mu\}$ ,  $\mu = 1, \dots, p$ , independent and identically distributed with respect to  $i$  and  $\mu$ . The local field in neuron  $i$  is defined by

$$h_i(t) = \sum_{j=1}^N J_{ij} \sigma_j(t) \quad (1)$$

with the couplings given by the Hebb rule

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \quad J_{ii} = 0. \tag{2}$$

All neurons are updated in parallel according to the Glauber dynamics

$$\text{Prob}[\sigma_i(t+1) = s | h_i(t)] = \frac{e^{\beta s h_i(t)}}{2 \cosh(\beta h_i(t))} \quad s = \pm 1 \tag{3}$$

which becomes, in the limit  $\beta = 1/T \rightarrow \infty$ , equivalent to the gain function formulation

$$\sigma_i(t+1) = \text{sgn}(h_i(t)). \tag{4}$$

In general, we write

$$\sigma_i(t+1) = g(h_i(t)). \tag{5}$$

The long time behaviour of the system is governed by the Hamiltonian [17]

$$H = -\frac{1}{\beta} \sum_i \ln[2 \cosh(\beta h_i(t))]. \tag{6}$$

The thermodynamic and retrieval properties are visualized in the  $(T - \alpha)$  phase diagram [18]. There is a transition curve starting at  $T = 0, \alpha = 0.138$  and ending at  $T = 1, \alpha = 0$  beyond which the system does not retrieve any patterns anymore and behaves like a spin-glass. For higher temperatures the system undergoes a transition to the paramagnetic phase.

The parallel dynamics of this model using both the SNA and GFA and, in particular, the comparison between these approaches is the subject of the following sections.

### 2.2. Signal-to-noise analysis

In order to keep this paper self-contained we briefly review the signal-to-noise analysis. We assume that the system has an initial finite overlap with one of the patterns, say the first one, which we call condensed. The other patterns (non-condensed) act as noise, making it harder for the network to retrieve the condensed pattern.

We first focus on zero temperature. The key idea is to separate the signal from the noise in the local field

$$h_i(t) = \xi_i^1 \frac{1}{N} \sum_j \xi_j^1 \sigma_j(t) + \frac{1}{N} \sum_{\mu \neq 1} \xi_i^\mu \sum_j \xi_j^\mu \sigma_j(t) - \alpha \sigma_i(t) \tag{7}$$

whereby it is technically convenient to include the self-interaction and subtract it again leading to the term  $\alpha \sigma_i(t)$ . Quantities that have a site index, or an index  $N$  are quantities for a finite system. We define the overlap, respectively the residual overlap, by

$$m_N(t) = \frac{1}{N} \sum_j \xi_j^1 \sigma_j(t) \quad r_N^\mu(t) = \frac{1}{\sqrt{N}} \sum_j \xi_j^\mu \sigma_j(t) \quad \mu > 1. \tag{8}$$

The local field (7) then becomes

$$h_i(t) = \xi_i^1 m_N(t) + \frac{1}{\sqrt{N}} \sum_{\mu \neq 1} \xi_i^\mu r_N^\mu(t) - \alpha \sigma_i(t). \tag{9}$$

Our aim is to determine the form of the local field in the thermodynamic limit  $N \rightarrow \infty$ . To the signal term we apply the law of large numbers (LLN)

$$\lim_{N \rightarrow \infty} m_N(t) \stackrel{Pr}{=} m(t) \tag{10}$$

where the convergence is in probability. Taking a closer look at the second term in (9), we could apply the central limit theorem (CLT) if all the terms in the sum would be independent. This, of course, depends on the architecture and it is true, e.g., for asymmetrically extremely diluted models [19, 20]. In general, there are non-trivial correlations between the terms and the different implementations of the SNA mentioned before treat these correlations in a different way.

For the fully connected architecture at hand, the most naive and simple approach [14] is to keep the assumption that all terms in the sum are uncorrelated, and that one can simply apply the CLT theorem to the second term in (9). The local field becomes normally distributed with mean  $\xi_i^1 m(t)$  and variance  $\alpha$ . This approach has then been refined in the theory of statistical neurodynamics [10]. There also one assumes that the noise part of the local field is normally distributed but one calculates explicitly its variance starting from its definition and taking into account part of the correlations between the embedded pattern and neuron states in the dynamics. This results in a more complex structure of the variance of the local field. For more technical details we refer to [10, 12]. Detailed simulations show that the assumption of Gaussian noise is approximately valid and close to reality as long as retrieval is successful. In particular, it succeeds in explaining qualitatively the dynamical behaviour of retrieval in the associative memory model. However, in addition to a different critical capacity at  $T = 0$ , the basin of attraction calculated in this scheme is larger than that obtained by computer simulations [21]. This is attributed to the fact that the correlations of the local fields at successive time steps are neglected. More specifically, the average over the Gaussian distribution of the local field at time  $t + 1$  is taken independently of the average of the neuron state at time  $t$ , which clearly depends on the local field at time  $t - 1$ .

Later on these correlations between the local field at successive time steps (and therefore also correlations between neuron states at different time steps) have been taken into account. In such a treatment one rewrites the noise part of the local field as the sum of two correlated terms to which one can apply the CLT but one keeps the Gaussian assumption throwing away possible non-Gaussian noise. For more details we refer to [11]. This further refined theory gives a better explanation of the dynamics and the basin of attraction. Moreover, the storage capacity resulting from this method is in good agreement with the results from computer simulations.

However, we recall that the *a priori* Gaussian approximation is only valid when the retrieval is successful. An improvement of the approximation is obtained by taking into account explicitly the feedback: the network state at time  $t + 1$ ,  $\sigma(t + 1)$ , depends on its previous states up to  $t - 1$  namely  $\sigma(0), \dots, \sigma(t - 1)$ . One proposal has been to assume two Gaussian peaks with variance calculated from the residual noise and separated by an appropriately chosen distance [22] but here also the correlations between the network states  $\sigma(t), \sigma(t - 1), \sigma(t - 2), \dots$  are only partially taken into account.

More recently, one has studied the distribution of the local field without any *a priori* assumption on the residual noise, trying to take into account all correlations of the different terms of the sum (9) using the insight gained before [15, 16] (see also [13] and references therein).

In the treatment of the correlations between the variables appearing in this expression for the local field, one has to be very careful, because even a small dependence of order  $\mathcal{O}(1/\sqrt{N})$  may give rise to a macroscopic contribution after summation. Therefore, we rewrite the local field

$$h_i(t) = \hat{h}_i^\mu(t) + \frac{1}{\sqrt{N}} \xi_i^\mu r_N^\mu(t) \quad (11)$$

where we have split the part of  $h_i(t)$  that depends strongly on  $\xi_i^\mu$ , such that  $\hat{h}_i^\mu$  only depends

weakly on  $\xi_i^\mu$ . Note that the second term is small for large  $N$ , so

$$\begin{aligned} \sigma_i(t+1) &= g(h_i(t)) \\ &= g\left(\hat{h}_i^\mu(t) + \frac{1}{\sqrt{N}}\xi_i^\mu r_N^\mu(t)\right) \\ &= g(\hat{h}_i^\mu(t)) + \left(\frac{1}{\sqrt{N}}\xi_i^\mu r_N^\mu(t)\right) \frac{\partial g}{\partial h_i} \Big|_{\hat{h}_i^\mu(t)} + \mathcal{O}(N^{-1}). \end{aligned} \tag{12}$$

Using this and the definition of the residual overlap in (8), we get

$$r_N^\mu(t+1) \approx \tilde{r}_N^\mu(t) + \chi_N(t)r_N^\mu(t) \tag{13}$$

where we have defined

$$\tilde{r}_N^\mu(t) = \frac{1}{\sqrt{N}} \sum_i \xi_i^\mu g(\hat{h}_i^\mu(t)) \quad \mu > 1 \quad \chi_N(t) = \frac{1}{N} \sum_i \frac{\partial g}{\partial h_i} \Big|_{\hat{h}_i^\mu(t)}. \tag{14}$$

Now, we take the limit  $N \rightarrow \infty$ . In this limit the density distributions of  $\hat{h}$  and  $h$  are equal since they only differ up to a factor  $1/\sqrt{N}$ . Since  $\hat{h}_i^\mu(t)$  only depends weakly on  $\xi_i^\mu$ , we assume that in the limit we can use the CLT for the first term. To the second term in (13), we apply the LLN arriving at

$$r^\mu(t+1) = \tilde{r}^\mu(t) + \chi(t)r^\mu(t) \tag{15}$$

with  $\tilde{r}^\mu(t)$  being a normally distributed random variable  $\mathcal{N}(0, 1)$  and

$$\chi(t) = \left\langle \left\langle \frac{\partial g}{\partial h} \Big|_{h(t)} \right\rangle_h \right\rangle = \left\langle \int \prod_{s=0}^t dh(s) P(h(0), h(1), \dots, h(t)) \frac{\partial g}{\partial h} \Big|_{h(t)} \right\rangle \tag{16}$$

where the average denoted by  $\langle \cdot \rangle_h$  is over the distribution of the local field with probability density  $P(h(0), h(1), \dots, h(t)) \equiv P(\mathbf{h})$  and where  $\langle \langle \cdot \rangle \rangle$  denotes the average over the initial conditions and the condensed pattern.

Next, in order to find the distribution of the local field in the thermodynamic limit, we start from (7) and use (8) and (13)

$$\begin{aligned} h_i(t+1) &= \xi_i^1 m_N(t+1) + \frac{1}{N} \sum_{\mu \neq 1} g(\hat{h}_i^\mu(t)) + \frac{1}{N} \sum_{\mu \neq 1} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu g(\hat{h}_j^\mu(t)) \\ &\quad + \chi_N(t) \frac{1}{\sqrt{N}} \sum_{\mu \neq 1} \xi_i^\mu r_N^\mu(t) - \alpha \sigma_i(t+1). \end{aligned} \tag{17}$$

For the first term on the rhs we use (10). In the second term, we can replace  $\hat{h}_i^\mu(t)$  by  $h_i(t)$  leading to a contribution  $\alpha \sigma_i(t+1)$ . The third term is a sum of independent random variables and by the CLT it converges to  $\mathcal{N}(0, \alpha)$ . In the fourth term, we employ (9). From this we obtain, omitting the site index  $i$

$$h(t+1) = \xi^1 m(t+1) + \chi(t)[h(t) - \xi^1 m(t) + \alpha \sigma(t)] + \mathcal{N}(0, \alpha). \tag{18}$$

We remark that  $h(0) = \xi^1 m(0) + \mathcal{N}(0, \alpha)$ . In this way we arrive at the recursion relation (18) for the local field and (15) for the residual overlap. We still want a recursion relation for the overlap by using the dynamics (5) starting from (8)

$$m(t) = \langle \langle \xi^1 g(h(t-1)) \rangle_h \rangle. \tag{19}$$

Finally, from (18) it is clear that the local field consists of a discrete part and a normally distributed part

$$h(t) = M(t) + \mathcal{N}(0, \alpha D(t)) \tag{20}$$

where  $M(t)$  can be found by iterating the recursion relation for the local field, i.e.

$$M(t) = \xi^1 m(t) + \alpha \sum_{s=0}^{t-1} \left( \prod_{s'=s}^{t-1} \chi(s') \right) \sigma(s). \tag{21}$$

The variance of the noise in (20) can be calculated by using (15)

$$\text{Var}(r^\mu(t+1)) = D(t+1) = 1 + \chi^2(t)D(t) + 2\chi(t)\text{Cov}(\tilde{r}^\mu(t), r^\mu(t)). \tag{22}$$

Clearly,  $D(0) = 1$ .

We still have to write the probability density of the local field used to define  $\chi(t)$  in (16). The evolution equation tells us that  $\sigma(t)$  can be replaced by  $g(h(t-1))$ , such that the second term of  $M(t)$  in (21) is the sum of step functions of correlated variables. These variables are also correlated with the Gaussian part of the local field through the dynamics. Therefore, the local field can be seen as a transformation of a set of correlated Gaussian variables  $x(s)$ ,  $s = 0, 1, \dots, t-2, t$  which we choose to normalize. These correlations are given by (recall equation (20))

$$W(s, s') \equiv \frac{E[(h(s) - M(s))(h(s') - M(s'))]}{\sqrt{\alpha D(s)}\sqrt{\alpha D(s')}} \tag{23}$$

with  $E$  denoting the expectation value. They can be calculated explicitly when working out the recursion relations (18) and (22). For example, knowing that  $h(1)$  and  $h(0)$  are correlated due to the dependence of  $\sigma(1)$  and  $\sigma(0)$  we find from relation (18) for  $h(1)$

$$W(0, 1) = \frac{\langle\langle \sigma(0) \int \mathcal{D}z g(\xi^1 m^1(0) + \sqrt{\alpha}z) \rangle\rangle + \chi(0)}{[1 + \chi^2(0) + 2\chi(0)\langle\langle \sigma(0) \int \mathcal{D}z g(\xi^1 m^1(0) + \sqrt{\alpha}z) \rangle\rangle]^{1/2}} \tag{24}$$

where

$$\chi(0) = \frac{1}{\sqrt{\alpha}} \left\langle\left\langle \int \mathcal{D}z z g(\xi^1 m^1(0) + \sqrt{\alpha}z) \right\rangle\right\rangle \tag{25}$$

with  $\langle\langle \cdot \rangle\rangle$  denoting the average over  $\xi$  and over the initial configuration and where  $\mathcal{D}z$  is the Gaussian measure with variable  $z$ . For more details we refer to [24]. These correlations for different time steps define a correlation matrix  $\mathbf{W}$  such that we arrive at the following expression for the probability density of the local field

$$P(\mathbf{h}) = \int \prod_{s=0}^{t-2} dx(s) dx(t) \frac{1}{\sqrt{\det(2\pi \mathbf{W})}} \exp \left( -\frac{1}{2} \sum_{\substack{s,s'=0 \\ s,s' \neq t-1}}^t x(s) \mathbf{W}^{-1}(s, s') x(s') \right) \times \delta(h(t) - M(t) - \sqrt{\alpha D(t)}x(t). \tag{26}$$

This concludes the SNA treatment of the Little–Hopfield model at zero temperature. The above equations form a recursive scheme in order to calculate the dynamical properties of the system up to an arbitrary time step. The practical difficulty which remains, certainly after a few time steps is the explicit calculation of the correlations.

It is possible to extend the method to arbitrary temperatures by introducing auxiliary thermal fields to express the stochastic dynamics within the gain function formulation of the deterministic dynamics [23]

$$\sigma(t+1) = g(h(t) + T\gamma(t)) \tag{27}$$

where the  $\gamma(t)$  are independent and identically distributed with probability density

$$P(\gamma(t)) = \frac{1}{2}(1 - \tanh^2(\gamma(t))). \tag{28}$$

One then averages the zero temperature results over the auxiliary field  $\gamma(t)$ . This alters the equations in a non-trivial way but such that the idea of the derivation can be completely retained. At this point we remark that in the derivation of the local field distribution one has to replace  $\sigma(t)$  by  $g(h(t - 1))$  with  $h(t - 1)$  given by (20). For arbitrary temperatures one has to replace  $\sigma(t)$  by  $g(h(t - 1) + T\gamma(t - 1))$  and the average over the  $\gamma(t)$  then enters at the same level as the average over the noise. For more details we refer to the literature [23].

2.3. *Explicit results for the first four time steps*

In order to compare with the results of the GFA method to be explained in the following section it is useful to recall the SNA results for the first few time steps as found in [24] (and references therein). We only write down explicitly the expressions for the overlap  $m(t)$  at zero temperature

$$m(1) = \left\langle\left\langle \xi \int \mathcal{D}z g(\xi m(0) + \sqrt{\alpha}z) \right\rangle\right\rangle \tag{29}$$

$$m(2) = \left\langle\left\langle \xi \int \mathcal{D}z g(\xi m(1) + \alpha\chi(0)\sigma(0) + \sqrt{\alpha D(1)}z) \right\rangle\right\rangle \tag{30}$$

$$m(3) = \left\langle\left\langle \xi \int \mathcal{D}W^{0,2}(x, y) g(\xi m(2) + \alpha\chi(1)[g(\xi m(0) + \sqrt{\alpha}x) + \chi(0)\sigma(0)] + \sqrt{\alpha D(2)}y) \right\rangle\right\rangle \tag{31}$$

$$m(4) = \left\langle\left\langle \xi \int \mathcal{D}W^{0,1,3}(x, y, z) g[\xi m(3) + \alpha\chi(2)g(\xi m(1) + \alpha\chi(0)\sigma(0) + \sqrt{\alpha D(1)}y) + \alpha\chi(2)\chi(1)g(\xi m(0) + \sqrt{\alpha}x) + \alpha\chi(2)\chi(1)\chi(0)\sigma(0) + \sqrt{\alpha D(3)}z] \right\rangle\right\rangle. \tag{32}$$

Here  $\mathcal{D}W^{0,1,\dots,t-2,t}(x_1, \dots, x_t)$  is the multidimensional Gaussian measure with correlation matrix  $W$  after  $t$  time steps, as it appears in (26). For example,

$$\mathcal{D}W^{0,2}(x, y) = \frac{dx dy}{2\pi\sqrt{1 - W(0, 2)^2}} \exp\left(-\frac{x^2 - 2W(0, 2)xy + y^2}{2(1 - W(0, 2)^2)}\right). \tag{33}$$

We also omitted the superscript 1 in  $\xi$ , indicating the condensed pattern. For the explicit form of the expressions for the variance of the residual overlap, the function  $\chi$  and the correlations we refer to the literature [24].

**3. Generating functional approach to the Little–Hopfield model**

3.1. *The effective local field*

The idea of the GFA approach to study dynamics [3, 6, 7] is to look at the probability to find a certain microscopic path in time. The basic tool to study the statistics of these paths is the generating functional

$$Z[\psi] = \sum_{\sigma(0), \dots, \sigma(t)} P[\sigma(0), \dots, \sigma(t)] \prod_{i=1}^N \prod_{s=0}^t e^{-i\psi_i(s)\sigma_i(s)} \tag{34}$$



with  $P[\sigma(0), \dots, \sigma(t)]$  being the probability to have a certain path in phase space

$$P[\sigma(0), \dots, \sigma(t)] = P[\sigma(0)] \prod_{s=0}^{t-1} W[\sigma(s+1)|\sigma(s)] \quad (35)$$

and  $W[\sigma|\sigma']$  the transition probabilities from  $\sigma'$  to  $\sigma$ . For parallel Glauber dynamics, we have

$$W[\sigma(s+1)|\sigma(s)] = \prod_i \frac{e^{\beta\sigma_i(s+1)h_i(s)}}{2 \cosh(\beta h_i(s))} \Bigg|_{h_i(s)=\sum_j J_{ij}\sigma_j(s)+\theta_i(s)} \quad (36)$$

where we have introduced a time-dependent external field  $\theta_i(s)$  in order to define a response function.

One can find all the relevant order parameters, i.e., the overlap  $m(t)$ , the correlation function  $C(t, t')$  and the response function  $G(t, t')$ , by calculating appropriate derivatives of the above functional and letting  $\psi = \{\psi_i\}$  tend to zero afterwards

$$m(t) = i \lim_{\psi \rightarrow 0} \frac{1}{N} \sum_i \xi_i \frac{\delta Z[\psi]}{\delta \psi_i(t)} \quad (37)$$

$$G(t, t') = i \lim_{\psi \rightarrow 0} \frac{1}{N} \sum_i \frac{\delta^2 Z[\psi]}{\delta \psi_i(t) \delta \theta_i(t')} \quad (38)$$

$$C(t, t') = - \lim_{\psi \rightarrow 0} \frac{1}{N} \sum_i \frac{\delta^2 Z[\psi]}{\delta \psi_i(t) \delta \psi_i(t')}. \quad (39)$$

In the thermodynamic limit one expects the physics of the problem to be independent of the quenched disorder and, therefore, one is interested in derivatives of  $\overline{Z[\psi]}$  with the overline denoting the average over this disorder, i.e., all pattern realizations. This results in an effective single spin local field given by

$$h(t) = \xi m(t) + \alpha \sum_{s=0}^{t-1} R(t, s) \sigma(s) + \sqrt{\alpha} \eta(t) \quad (40)$$

with  $\eta(t)$  being temporally correlated noise with zero mean and correlation matrix

$$D = (I - G)^{-1} C (I - G^\dagger)^{-1} \quad (41)$$

and the retarded self-interaction

$$R = (I - G)^{-1} G. \quad (42)$$

We refer to [3] for further details. The order parameters defined above can be written as

$$m(t) = \langle\langle \xi \sigma(t) \rangle_\star \rangle \quad (43)$$

$$C(s, s') = \langle\langle \sigma(s) \sigma(s') \rangle_\star \rangle \quad (44)$$

$$G(s, s') = \left\langle\left\langle \left\langle \frac{\partial \sigma(s)}{\partial \theta(s')} \right\rangle_\star \right\rangle \right\rangle. \quad (45)$$

The average over the effective path measure is given by

$$\langle f \rangle_\star = \text{Tr}_{\{\sigma(1), \dots, \sigma(t)\}} \int d\eta P(\eta) P(\sigma|\eta) f \quad (46)$$

where  $d\eta = \prod_s d\eta(s)$  and with

$$P(\sigma|\eta) = \prod_{s=0}^{t-1} \frac{\exp(\beta\sigma(s+1)h(s))}{2 \cosh(\beta h(s))} \Big|_{h(s)=\xi m(s)+\alpha \sum_{s'} R(s,s')\sigma(s')+\sqrt{\alpha}\eta(s)} \quad (47)$$

$$P(\eta) = \frac{1}{\sqrt{\det(2\pi D)}} \exp\left(-\frac{1}{2} \sum_{s,s'=0}^{t-1} \eta(s) D^{-1}(s,s') \eta(s')\right). \quad (48)$$

The average denoted by the double brackets is again (as in the SNA) over the condensed pattern and initial conditions. We remark that  $G(s,s') = 0$  for  $s \leq s'$  and  $D(s,s') = D(s',s)$ , and that for all  $s < t$

$$G(t,s) = \beta \langle\langle \sigma(t)[\sigma(s+1) - \tanh(h(s))] \rangle\rangle_* \quad (49)$$

where  $h(s)$  is given by (40).

### 3.2. An alternative stochastic description

The GFA at arbitrary temperatures starts from the transition probabilities (36) following from the Glauber dynamics (3). In order to make the comparison with the SNA later on we want to introduce an alternative description starting from the deterministic dynamics as in (27) and (28).

Starting from (47) and (48), taking the limit of zero temperature and introducing the auxiliary thermal field  $\gamma(t)$  both formulations are equivalent. Indeed, suppose that we want to calculate the effective path average of a general function  $f$  using the alternative description. Writing out the expression for this average (the average over the condensed pattern and initial conditions is not relevant here)

$$\begin{aligned} \int d\gamma P(\gamma) \left( \lim_{\beta \rightarrow \infty} \langle f(\sigma) \rangle_* \right)_{h \rightarrow h+T\gamma} &= \text{Tr}_{\{\sigma(1), \dots, \sigma(t)\}} \int d\gamma dh d\eta P(\eta) \\ &\times \prod_{s=0}^{t-1} \left[ \delta \left( h(s) - \xi m(s) - \alpha \sum_{s'} R(s,s') \sigma(s') - \sqrt{\alpha} \eta(s) - T\gamma(s) \right) \right. \\ &\times \left. \frac{1}{2} (1 - \tanh^2(\gamma(s))) \Theta(\sigma(s+1)h(s)) \right] f(\sigma) \end{aligned} \quad (50)$$

where  $\gamma$  is the thermal field. Since the latter only occurs inside the  $\delta$ -function and the probability density, we can evaluate the integral over  $\gamma$ . Then the integration over the local field becomes

$$\prod_{s=0}^{t-1} \int dh(s) \Theta(\sigma(s+1)h(s)) \left( 1 - \tanh^2 \left( h(s) - \xi m(s) - \alpha \sum_{s'} R(s,s') \sigma(s') - \sqrt{\alpha} \eta(s) \right) \right). \quad (51)$$

This integral can be evaluated, and yields

$$\prod_{s=0}^{t-1} \frac{\exp(\beta\sigma(s+1)h(s))}{2 \cosh(\beta h(s))} \Big|_{h(s)=\xi m(s)+\alpha \sum_{s'} R(s,s')\sigma(s')+\sqrt{\alpha}\eta(s)} \quad (52)$$

which is exactly the same as (47) showing that we can use both representations for the dynamics.

Using this alternative description for the thermal dynamics it is straightforward to write the distribution of the local field as

$$P(h) = \int d\gamma d\eta P(\eta) P(\gamma) \delta(h(t) - \xi m(t) - \alpha \sum_{s=0}^{t-1} R(t, s) \sigma(s) - \sqrt{\alpha} \eta(t) - T \gamma(t)) |_{\sigma, h} \quad (53)$$

where we denote by  $|_{\sigma, h}$  the substitutions  $\sigma(s) = g(h(s) - 1)$  and  $h(s) = \xi m(s) + \alpha \sum_{s'} R(s, s') \sigma(s') + \sqrt{\alpha} \eta(s) + T \gamma(s)$  for all  $s \leq t$ . The above discussion makes it trivial to perform the zero temperature limit in the GFA and also shows that it is enough to look at the relation between the two methods at zero temperature in order to be able to compare them in the whole phase diagram, because the way to extend the methods to finite temperature is completely equivalent.

### 3.3. Recursive scheme

The aim of this subsection is to derive recursion relations for the local field and also for the noise appearing in the local field starting from the GFA. In this way we want to gain more insight into the equations at hand and make a detailed comparison between the SNA and GFA.

From expression (40) for the effective local field, it is not immediately obvious how  $h(t)$  depends on  $h(s)$ ,  $s < t$ . First, we want an expression for the retarded self-interaction  $\mathbf{R}$  as a function of the previous time step(s). To this end we write the response matrix at time  $(t + 1)$  in the following way:

$$\mathbf{G}_{t+1} = \begin{pmatrix} \mathbf{G}_t & 0 \\ \mathbf{g}_{t+1} & 0 \end{pmatrix} \quad (54)$$

where  $\mathbf{g}_{t+1}$  is the following vector of dimension  $t + 1$ :

$$\mathbf{g}_{t+1} = (G(t + 1, 0), G(t + 1, 1), \dots, G(t + 1, t)) \quad (55)$$

and  $\mathbf{G}_t$  is the response matrix at time  $t$ . It is clear that adding a time step adds a row and a column to  $\mathbf{G}_t$ . From the decomposition of the response matrix, we calculate  $\mathbf{R}_{t+1}$  recalling (42) and using the inverse of a partitioned matrix (see, e.g., [25] p 18). We obtain

$$\mathbf{R}_{t+1} = \begin{pmatrix} \mathbf{R}_t & 0 \\ \mathbf{g}_{t+1}(\mathbf{I} + \mathbf{R}_t) & 0 \end{pmatrix}. \quad (56)$$

Plugging this result into (40) we have

$$h(t) = \xi m(t) + \alpha \sum_{s'=0}^{t-1} \mathbf{g}_t(\mathbf{I} + \mathbf{R}_{t-1})(s') \sigma(s') + \sqrt{\alpha} \eta(t) \quad (57)$$

$$= \xi m(t) + \alpha \sum_{s=0}^{t-1} G(t, s) \left[ \sigma(s) + \sum_{s'=0}^{t-1} R_{t-1}(s, s') \sigma(s') \right] + \sqrt{\alpha} \eta(t) \quad (58)$$

$$= \xi m(t) + \sum_{s=0}^{t-1} G(t, s) [h(s) - \xi m(s) + \alpha \sigma(s) - \sqrt{\alpha} \eta(s)] + \sqrt{\alpha} \eta(t) \quad (59)$$

$$= \xi m(t) + \sum_{s=0}^{t-1} G(t, s) [h(s) - \xi m(s) + \alpha \sigma(s)] + \sqrt{\alpha} \sum_{s=0}^t (\mathbf{I} - \mathbf{G})(t, s) \eta(s). \quad (60)$$

This is one of the main results of this section and will be used as the basis of the comparison between the GFA and SNA approaches. It leads naturally to the definition of a modified noise variable  $\phi$

$$\phi(t) = \sum_{s=0}^t (\mathbf{I} - \mathbf{G})(t, s)\eta(s) \quad \langle \phi(s)\phi(s') \rangle = C(s, s') \quad (61)$$

and it follows that the covariance matrix of the noises,  $\mathbf{B}$ , is given by

$$B(s, s') = \langle \eta(s)\phi(s') \rangle = \sum_{s''=0}^t (\mathbf{I} - \mathbf{G}^\dagger)(s'', s') \langle \eta(s)\eta(s'') \rangle = [(\mathbf{I} - \mathbf{G})^{-1}\mathbf{C}](s, s'). \quad (62)$$

The correlation matrix of this transformed noise variable is clearly simpler than the original one and is in fact the correlation matrix of the spins.

Next, we derive recursion relations for the noise in a similar way. Analogous to (54) we write

$$\mathbf{C}_{t+1} = \begin{pmatrix} \mathbf{C}_t & \mathbf{c}_{t+1}^\dagger \\ \mathbf{c}_{t+1} & 1 \end{pmatrix} \quad (63)$$

where  $\mathbf{c}_{t+1}$  is the following vector of dimension  $t + 1$ :

$$\mathbf{c}_{t+1} = (C(t + 1, 0), C(t + 1, 1), \dots, C(t + 1, t)). \quad (64)$$

One then finds recalling (41) that

$$\mathbf{D}_{t+1} = \begin{pmatrix} \mathbf{D}_t & \mathbf{D}_t \mathbf{g}_t^\dagger + (\mathbf{I} - \mathbf{G}_t)^{-1} \mathbf{c}_{t+1}^\dagger \\ \mathbf{g}_t \mathbf{D}_t + \mathbf{c}_{t+1} (\mathbf{I} - \mathbf{G}_t^\dagger)^{-1} & 1 + \mathbf{g}_t \mathbf{D}_t \mathbf{g}_t^\dagger + 2\mathbf{c}_{t+1} (\mathbf{I} - \mathbf{G}_t)^{-1} \mathbf{g}_t^\dagger \end{pmatrix}. \quad (65)$$

Again, going one time step further implies adding one row and column to the matrix  $\mathbf{D}_t$ . From (65) we find a relation for the variance of the noise at time  $t + 1$

$$D(t + 1, t + 1) = 1 + \sum_{s,s'=0}^t G(t + 1, s)D(s, s')G(t + 1, s') + 2 \sum_{s=0}^t G(t + 1, s)B(s, t + 1) \quad (66)$$

as a function of the variance of the noise at time  $s < t$  and other quantities. Apparently, the right-hand side of this equation does still depend on  $t + 1$  and, therefore, does not seem to be really a recursion relation. However, as we will show, the quantities on the right-hand side can be calculated without having full information about time  $t + 1$ , e.g.,  $G(t + 1, s)$  can be obtained without knowing  $D(t + 1, t + 1)$ .

## 4. SNA versus GFA

### 4.1. Short-memory approximation

Comparing the expressions for the effective local field (60) and (18) we note that the first one contains a sum over all time steps up to  $t - 1$  while the second one only contains  $t - 1$  itself. Therefore, we introduce the following approximation:

$$\sum_{s=0}^{t-1} G(t, s)h(s) \approx \hat{X}(t - 1)\hat{h}(t - 1) \equiv \sum_{s=0}^{t-1} \hat{G}(t, s)\hat{h}(s) \quad (67)$$

where we have defined  $\hat{X}$ ,  $\hat{h}$  and  $\hat{G}$  in this way. The approximated matrix  $\hat{G}$  now has a simple form

$$\hat{G} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \hat{X}(0) & 0 & \dots & 0 & 0 \\ 0 & \hat{X}(1) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & 0 & 0 \\ 0 & 0 & 0 & \hat{X}(t-1) & 0 \end{pmatrix}. \quad (68)$$

This approximation reduces the expression for the local field (57) to

$$\hat{h}(t+1) = \xi \hat{m}(t+1) + \hat{X}(t)[\hat{h}(t) - \xi \hat{m}(t) + \alpha \hat{\sigma}(t)] + \sqrt{\alpha} \hat{\phi}(t+1). \quad (69)$$

Furthermore, the modified noise equation (61) simplifies to

$$\hat{\phi}(t) = \hat{\eta}(t) - \hat{X}(t-1)\hat{\eta}(t-1) \quad (70)$$

and the variance of the noise itself (66) becomes

$$\hat{D}(t+1, t+1) = 1 + \hat{X}^2(t)\hat{D}(t, t) + 2\hat{X}(t)\hat{B}(t, t+1). \quad (71)$$

Another consequence of this approximation is that the discrete part in the local field (20) can be written down or, in other words, the retarded self-interaction  $\mathbf{R}$  can be calculated explicitly to be

$$R(t, s) = \prod_{s'=s}^{t-1} \hat{X}(s'). \quad (72)$$

Comparing the recursion relations (18) and (69) for the local field, those for the residual overlap, (15) and (70), and those for the variance of the residual overlaps, (22) and (71), we find that they are formally the same by taking

$$\hat{X}(t) = \chi(t) \quad \hat{\phi}(t) = \tilde{r}^\mu(t-1) \quad \hat{\eta}(t) = r^\mu(t). \quad (73)$$

Next, we would like to know what is the physics behind this approximation. In fact, all solutions to (67) can be called  $k$ th order short-memory approximations because they approximate the feedback

$$\sum_{s=0}^{t-1} G(t, s)h(s) \sim \left( \sum_{\substack{s=t-k \\ s \geq 0}}^{t-1} G(t, s) \right) h(t-1). \quad (74)$$

These approximations take into account that responses, in general, decrease very fast as a function of time. As we will show, the first-order approximation obtained for  $k = 1$  and simply called the short-memory approximation, corresponds to the SNA

$$\sum_{s=0}^{t-1} G(t, s)h(s) \sim G(t, t-1)h(t-1) \quad \chi(t) = G(t+1, t). \quad (75)$$

We remark that each of these approximations has a different  $\chi(t)$  so that we can use the latter in order to distinguish between them. We calculate the  $\chi(t)$  corresponding to  $k = 1$ , starting from the alternative stochastic description of the GFA (forgetting about the average over the

initial conditions and condensed pattern). We find

$$\begin{aligned}
 G(t + 1, t) &= \int d\eta d\gamma P(\eta) P(\gamma) \left. \frac{\partial \sigma(t + 1)}{\partial h(t)} \right|_{\sigma, h} \\
 &= \int d\eta d\gamma P(\eta) P(\gamma) \left. \frac{\partial g(h(t))}{\partial h(t)} \right|_{\sigma, h} \\
 &= \int \left( \prod_{s=0}^t \frac{d\eta(s) d\gamma(s)}{\det(2\pi \mathbf{D})} \right) \exp \left( -\frac{1}{2} \sum_{s, s'}^t \eta(s) \mathbf{D}^{-1}(s, s') \eta(s') \right) P(\gamma) \left. \frac{\partial g(h(t))}{\partial h(t)} \right|_{\sigma, h}.
 \end{aligned} \tag{76}$$

Recalling the distribution of the local field at time  $t$  (53), and taking the temperature to be zero (which means vanishing  $\gamma$ ), this expression already resembles the analogous expression (26) in the SNA approach. Furthermore, one sees that  $\eta(t - 1)$  does not occur in the integrand of this expression. As a consequence, one can show that the matrix  $\mathbf{D}$  can be replaced by

$$\bar{\mathbf{D}}(s, s') = \langle \eta(s) \eta(s') \rangle_{\star} \quad s, s' \neq t - 1 \tag{77}$$

or in more detail

$$\bar{\mathbf{D}} = \begin{pmatrix} D(0, 0) & D(0, 1) & \dots & D(0, t - 2) & D(0, t) \\ D(1, 0) & D(1, 1) & \dots & D(1, t - 2) & D(1, t) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D(t - 2, 0) & D(t - 2, 1) & \dots & D(t - 2, t - 2) & D(t - 2, t) \\ D(t, 0) & D(t, 1) & \dots & D(t, t - 2) & D(t, t) \end{pmatrix}. \tag{78}$$

One can then verify that  $G(t + 1, t)$ , for zero temperature, is exactly the same as the expression for  $\chi(t)$  in the SNA approach, namely

$$G(t + 1, t) = \int \left( \prod_{\substack{s=0 \\ s \neq t-1}}^t \frac{d\eta(s)}{\det(2\pi \bar{\mathbf{D}})} \right) \exp \left( -\frac{1}{2} \sum_{\substack{s, s' \\ s, s' \neq t-1}}^t \eta(s) \bar{\mathbf{D}}^{-1}(s, s') \eta(s') \right) \left. \frac{\partial g(h(t))}{\partial h(t)} \right|_{\sigma, h}. \tag{79}$$

In conclusion, compared with the exact GFA method we find that the SNA approach is a short-memory approximation, in which the response from earlier times is not taken into account.

#### 4.2. Discussion

The origin of the approximation inherent in the SNA approach as discussed above lies in the treatment of the residual noise  $\tilde{r}_N^\mu(t)$  (recall (14)). We have assumed that after taking out the term  $N^{-1/2} \xi_i^\mu r_N^\mu(t)$  of the local field,  $\xi_i^\mu$  and  $\hat{h}_i^\mu(t)$  are only weakly correlated, and therefore we have applied the CLT to find that  $\tilde{r}_N^\mu(t)$  converges to a normal distribution. Comparing with the GFA it appears that we took out only part of the correlations between  $\sigma_i(t)$  and  $\xi_i^\mu$ , those coming from the previous time step. In a fully connected system however, there are not only feedback loops of length 1 and 2, but of arbitrary length, although they may be less probable [26].

Therefore, one would expect the SNA method to approximate the dynamics already from the second time step onwards. However, we can show that  $\mathbf{G}(t, t - 2)$  is zero such that the second time step is still exact. Moreover, some of the order parameters in the third time step,

e.g. the overlap, only involve noises up to the second time step, so that they are also correct for the third time step.

In order to show that  $G(t, t - 2) = 0$  we proceed as follows (recall (49))

$$\begin{aligned} G(t, t - 2) &= \frac{\partial \langle \sigma(t) \rangle_{\star}}{\partial \theta(t - 2)} \\ &= \beta \langle \sigma(t) [\sigma(t - 1) - \tanh(\beta h(t - 2))] \rangle_{\star}. \end{aligned} \quad (80)$$

Considering the effective path average the sum over  $\sigma(t)$  can be done explicitly

$$\sum_{\sigma(t)=\pm 1} \sigma(t) \frac{e^{\beta \sigma(t) h(t-1)}}{2 \cosh(\beta h(t-1))} = \tanh(\beta h(t-1)). \quad (81)$$

Moreover, the expression for the distribution only contains summations over  $\sigma(0)$  up to  $\sigma(t-1)$  and there is only one term left that contains  $\sigma(t-1)$  so

$$\sum_{\sigma(t-1)=\pm 1} \sigma(t-1) \frac{e^{\beta \sigma(t-1) h(t-2)}}{2 \cosh(\beta h(t-2))} = \tanh(\beta h(t-2)). \quad (82)$$

Note that further sums over spins cannot be done since  $h(t-1)$  contains the spins at times  $s \leq t-2$ . This shows that  $G(t, t-2) = 0$ .

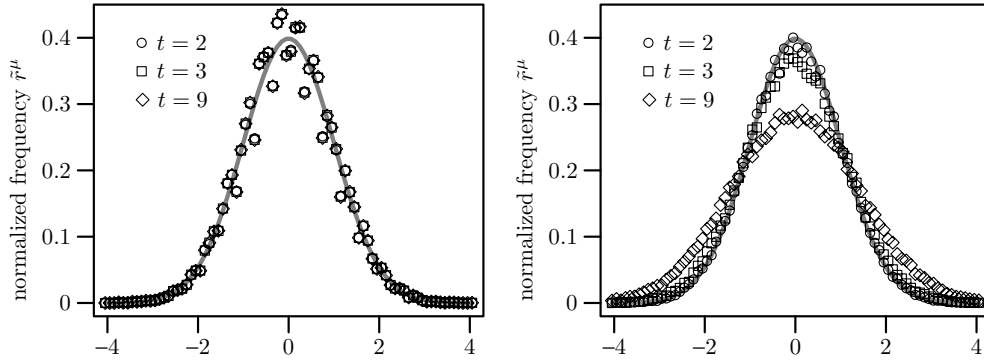
#### 4.3. Accuracy in retrieval

First, for  $\alpha = 0$ , one can easily verify that the GFA analysis yields

$$\mathbf{G} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \chi(0) & 0 & \dots & 0 & 0 \\ 0 & \chi(1) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & 0 & 0 \\ 0 & 0 & 0 & \chi(t-1) & 0 \end{pmatrix} \quad (83)$$

implying that the short-memory approximation, i.e., the SNA is exact. The reason is that the local field at time  $t$  does not depend on previous times because the terms in the retarded self-interaction and the noise are proportional to  $\alpha$ . The only correlation that remains in the system comes from the dependence of the spins on the local field at the previous time step. Consequently, we expect the SNA to be a very good approximation to the exact dynamics for small loading capacities. Some additional reasons to do so is that for small loading capacities, the convergence time to the attractor is very small, only a few time steps, and the SNA is exact up to time step 3. Moreover, to write down the evolution equations for the order parameters such as the overlap one performs averages over the noise. Therefore, not surprisingly, the SNA results for the first few time steps coincide with numerical simulations as has been reported in the literature before. Only discrepancies of the order  $\mathcal{O}(10^{-3})$  that are of the same magnitude as the finite size effects in the simulations, and hence not conclusive, have been observed (see, e.g., [13] and references therein).

Analytically, of course, one notes the difference from the fourth time step onwards. Starting from the exact GFA approach and writing the expression in a convenient form for



**Figure 1.** Finite size simulations of the residual overlap,  $\tilde{r}^\mu(t)$ , for different time steps compared with the theoretical results using the SNA (full curve) for  $T = 0.1$ ,  $m_0 = 0.6$ . Left:  $\alpha = 0.06$  and  $N = 3000$ ; right:  $\alpha = 0.16$  and  $N = 2000$ , both averaged over 200 samples.

comparison, we get at zero temperature

$$\begin{aligned}
 m(4) = & \left\langle \left\langle \xi \int \mathcal{D}W^{0,1,3}(x, y, z) g[\xi m(3) + \alpha \tilde{\chi}(2)g(\xi m(1) + \alpha \chi(0)\sigma(0) + \sqrt{\alpha D(1)}y) \right. \right. \\
 & + \alpha \tilde{\chi}(2)\tilde{\chi}(1)g(\xi m(0) + \sqrt{\alpha x}) + \alpha \tilde{\chi}(2)\tilde{\chi}(1)\tilde{\chi}(0)\sigma(0) \\
 & \left. \left. + \alpha \underline{G(3, 0)\sigma(0)} + \sqrt{\alpha D(3)}z \right] \right\rangle \right\rangle \quad (84)
 \end{aligned}$$

with  $\tilde{\chi}(t) = G(t + 1, t)$ . The only difference with the SNA result (32) is in the term that is underlined. It is present due to the fact that beyond  $t = 3$ ,  $\mathbf{R}$  is not simply given by (72). We will show numerically in the next subsection that this difference is small (e.g., of the order of 0.3% up to 3% for  $T = 0.1$ ,  $\alpha = 0.1$ ,  $m_0 = 0.4$  respectively  $m_0 = 0.2$ ). It will be interesting to see how this difference behaves for further time steps.

#### 4.4. Numerical results

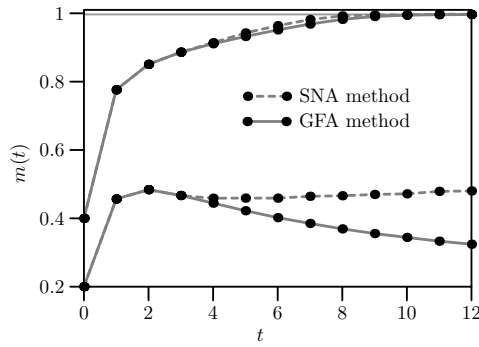
Now that we know precisely the origin of the approximation inherent in the SNA method as usually applied, we can check numerically how accurate it is for retrieval and also for spin-glass behaviour, although from the point of view of neural networks one is not primarily interested in the latter.

We first want to compare the limiting normal distribution of  $\tilde{r}_N^\mu$  in the SNA approach with finite size simulations for different time steps. This is done in figure 1 for two representative values of the capacity, i.e.  $\alpha = 0.06$  which lies in the retrieval phase and  $\alpha = 0.16$  which is above the critical capacity and thus within the spin-glass phase.

We conclude that in the retrieval region ( $\alpha_c < 0.135$  for  $T = 0.1$ ) the simulation results coincide quite well with the limiting SNA distribution, while in the spin-glass region, the results for  $t = 3$  start to divert systematically. This is consistent with the results obtained in the fully connected Blume–Emery–Griffiths neural network [2]. Note that the distribution of  $\tilde{r}^\mu$  starts to divert for  $t = 3$ , while  $m(3)$  is still exact as discussed before.

We also want to compare the evolution of the order parameters. As a typical result we show in figure 2 the overlap  $m(t)$  as a function of time using the method in [27], precisely in order to avoid finite size effects. We see that the SNA results coincide with those of the GFA up to the third time step, as shown analytically. For further time steps the results are





**Figure 2.** The overlap order parameter as a function of time for the SNA versus the GFA approach for  $\alpha = 0.1, T = 0.1$  for retrieval behaviour  $m_0 = 0.4$  and spin-glass behaviour  $m_0 = 0.2$ . The thin line represents the stationary limit for retrieval.

strongly dependent upon the parameters of the network determining its behaviour: retrieval or spin-glass behaviour. The parameters chosen in figure 2 are  $\alpha = 0.1, T = 0.1$ , and  $m_0 = 0.4$  respectively  $m_0 = 0.2$ . For the first choice (the two upper curves) the system evolves to the retrieval attractor and one observes that there is only a marginal difference between the SNA and GFA method. This confirms the previous observations made in the literature ([13] and references therein). For the second choice (the two lower curves) the initial overlap is too small such that we are outside the basin of attraction for retrieval and, hence, we do not evolve to the attractor (spin-glass behaviour). Here the SNA method, as used in the literature, does not give good results for further time steps. We remark that we have also compared these results with finite size simulations. These simulations are not indicated in the figure because they lie almost exactly on the GFA lines.

### 5. The SNA revisited

In this section, we show how to treat the feedback correlations in the SNA approach correctly. Looking at the definition of the residual noise at time  $t$ , i.e.,

$$r_N^\mu(t) = \frac{1}{\sqrt{N}} \sum_i \xi_i^\mu \sigma_i(t) \quad \mu > 1 \tag{85}$$

we want to know how the dependence between spins and patterns evolves in the course of the dynamics.

We focus on a spin at a given site  $i$  and time  $t$ . It clearly depends on  $\xi_i^\mu$  due to the spin flip dynamics, i.e.

$$\sigma_i(s) = \text{sgn} \left[ \frac{1}{\sqrt{N}} \xi_i^\mu r_N^\mu(s-1) + \frac{1}{N} \sum_{v \neq \mu, j} \xi_i^v \xi_j^v \sigma_j(s-1) \right] \tag{86}$$

for a given time  $s < t$ . In the course of the time evolution of the system, correlations such as this may build up and influence  $\sigma_i(t)$  due to feedback loops. Therefore, at every time step, we extract those correlations from the local field obtaining

$$\tilde{h}_i^\mu(s) = h_i(s) + \delta h_i^\mu(s). \tag{87}$$

It is clear from (86) that

$$\delta h_i^\mu(s) = -\frac{1}{N} \sum_j \xi_i^\mu \xi_j^\mu \sigma_j(s) = -\frac{1}{\sqrt{N}} \xi_i^\mu r_N^\mu(s). \tag{88}$$

We assume that all other possible dependences on  $\xi_i^\mu$  lead to no contributions in the thermodynamic limit because of the mean field character of the system.

The quantity  $\delta h_i^\mu(s)$  tends to zero as  $1/\sqrt{N}$  in the thermodynamic limit, and can therefore be regarded as a perturbation of the field  $h_i(s)$ . We consider this perturbation for all time steps  $s < t$  and find

$$\tilde{\sigma}_i^\mu(t) = \sigma_i(t) + \sum_{s=0}^{t-1} \left. \frac{\partial \sigma_i(t)}{\partial \tilde{h}_i^\mu(s)} \right|_{\delta h_i^\mu(s)=0} \delta h_i^\mu(s) + \mathcal{O}(1/N) \tag{89}$$

where  $\tilde{\sigma}_i^\mu(t)$  is now independent of  $\xi_i^\mu$ . We rewrite the derivatives by introducing an external field  $\theta_i(s)$ ,

$$\left. \frac{\partial \sigma_i(t)}{\partial \delta h_i^\mu(s)} \right|_{\delta h_i^\mu(s)=0} = \left. \frac{\partial \sigma_i(t)}{\partial \theta_i(s)} \right|_{\theta_i(s)=0} \tag{90}$$

Inserting (88) in (89) and using (90) yields

$$\sigma_i(t) = \tilde{\sigma}_i^\mu(t) + \frac{\xi_i^\mu}{\sqrt{N}} \sum_s \left. \frac{\partial \sigma_i(t)}{\partial \theta_i(s)} \right|_{\theta_i(s)=0} r_N^\mu(s) + \mathcal{O}(1/N). \tag{91}$$

We multiply both sides by  $\xi_i^\mu/\sqrt{N}$  and sum over  $i$ , so that we obtain

$$r_N^\mu(t) = \frac{1}{\sqrt{N}} \sum_i \xi_i^\mu \tilde{\sigma}_i^\mu(t) + \sum_s \frac{1}{N} \sum_i \left. \frac{\partial \sigma_i(t)}{\partial \theta_i(s)} \right|_{\theta_i(s)=0} r_N^\mu(s) + \mathcal{O}(1/\sqrt{N}) \tag{92}$$

by using the definition of  $r_N^\mu(t)$ . We now take the limit  $N \rightarrow \infty$ . By construction, the first term on the rhs is a sum of independent terms and thus converges to a normal distribution with zero mean and variance 1. To the second term we apply the LLN. We arrive at

$$r^\mu(t) = \tilde{r}^\mu(t) + \sum_s G(t, s) r^\mu(s). \tag{93}$$

This recursion relation for  $r^\mu(s)$  corresponds to (61) by using the substitution  $r^\mu$  for  $\eta$  and  $\tilde{r}^\mu$  for  $\phi$  as in (73). Finally, we insert this relation into expression (18) of the local field at time  $t + 1$  leading to, after some algebra, the expression for the exact local field as in equation (57) for the GFA. This shows that we have found all feedback correlations.

### 6. Other architectures

In this section, we briefly discuss the application of the SNA approach to other architectures.

For the extremely diluted symmetric architecture one has found an exact solution for the dynamics up to time step 3 by using a probabilistic method analogous to the SNA [28] and comparing it with the generating functional approach [29]. In this case, the effective local field, starting from the GFA is given by

$$h(t) = \xi m(t) + \alpha \sum_{s=0}^{t-1} R(t, s) \sigma(s) + \sqrt{\alpha} \eta(t) \tag{94}$$

with  $\eta(t)$  being a temporally correlated noise with zero mean and correlation matrix  $D$ . Here,

$$D = C \quad R = G. \quad (95)$$

It is clear that, besides the simplification in both the correlations and retarded self-interaction, feedback correlations of arbitrary length survive and make the dynamics as hard to solve as that of the fully connected model. Hence, a completely analogous discussion as the one before can be made in this case.

For the extremely diluted asymmetric architecture the effective local field in the GFA approach is given by (94), where now

$$D = C \quad R = 0. \quad (96)$$

Hence, the retarded self-interaction is zero, which shows that the local field at time  $t$  does not directly depend on the spins at previous times but only indirectly via the noise. This is precisely the reason why for  $\alpha = 0$ , the short-memory approximation gives the exact answer (see above). It amounts to having a response function of the form (83). Therefore, the SNA describes the correct dynamics. The reason is that for asymmetric dilution the probability to have a loop of finite length tends to zero in the thermodynamic limit [19, 20].

For sequence processing networks the GFA effective local field is given by (94) with

$$D = \sum_{n \geq 0} (G^\dagger)^n C (G)^n \quad R = 0 \quad (97)$$

and the situation is analogous to that for the asymmetrically diluted model in the sense that the retarded self-interaction is zero. Again we have a response function of the form (83) and, hence, the short-memory approximation is exact. This is consistent with and further explains the results in [4] where it has been shown through explicit calculation that the order parameter equations obtained through the GFA are equivalent to those of statistical neurodynamics.

## 7. Concluding remarks

In this paper, we have revisited the signal-to-noise approach for solving the dynamics of the fully connected Little–Hopfield model by comparing it with the exact generating functional analysis. In order to do so we have derived a recursion relation for the effective local field in the generating functional approach. We have shown that the signal-to-noise analysis is a short-memory approximation that is exact up to the third time step. For further time steps, it stays very accurate in the retrieval region but not in the spin-glass region. These results are confirmed by numerical simulations. The application of these methods to other architectures and to sequence processing models has also been discussed.

## Acknowledgments

We would like to thank A C C Coolen and G M Shim for informative discussions. This work has been supported in part by the Fund of Scientific Research, Flanders-Belgium.

## References

- [1] Galla T, Coolen A C C and Sherrington D 2003 *J. Phys. A: Math. Gen.* **36** 11159
- [2] Bollé D, Busquets Blanco J, Shim G M and Verbeiren T 2003 *Physica A* **331** 319
- [3] Coolen A C C 2001 *Handbook of Biological Physics* vol 4, ed F Moss and S Gielen (Amsterdam: Elsevier) p 597
- [4] Kawamura M and Okada M 2002 *J. Phys. A: Math. Gen.* **35** 253

- [5] Theumann W K 2003 *Physica A* **328** 1
- [6] Siggia E D, Martin P C and Rose H A 1973 *Phys. Rev. A* **8** 423
- [7] De Dominicis C 1978 *J. Phys. A: Math. Gen.* **11** 4913
- [8] Rieger H, Schreckenberg M and Zittartz J 1989 *Z. Phys. B* **74** 527
- [9] Horner H, Bormann D, Frick M, Kinzelbach H and Schmidt A 1989 *Z. Phys. B* **76** 381
- [10] Amari S and Maginu K 1988 *Neural Netw.* **1** 63
- [11] Okada M 1995 *Neural Netw.* **8** 833
- [12] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing* (Oxford: Oxford University Press)
- [13] Bollé D 2003 *Preprint cond-mat/0307104 (Advances in Condensed Matter and Statistical Mechanics* ed E Korutcheva and R Cuerno (Commack, NY: Nova Science Publishers) at press)
- [14] Kinzel W 1985 *Z. Phys. B* **60** 205
- [15] Patrick A E and Zagrebnov V A 1991 *J. Phys. A: Math. Gen.* **24** 3413
- [16] Patrick A E and Zagrebnov V A 1991 *J. Stat. Phys.* **63** 59
- [17] Peretto P 1984 *Biol. Cybern.* **50** 51
- [18] Fontanari J F and Köberle R 1988 *J. Physique* **49** 13
- [19] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167
- [20] Kree R and Zippelius A 1991 *Models of Neural Networks* ed E Domany, J L van Hemmen and K Schulten (Berlin: Springer)
- [21] Nishimori H and Ozeki T 1993 *J. Phys. A: Math. Gen.* **26** 859
- [22] Henkel R D and Opper M 1990 *Europhys. Lett.* **11** 403
- [23] Patrick A E and Zagrebnov V A 1990 *J. Physique* **51** 1129
- [24] Bollé D, Jongen G and Shim G M 1998 *J. Stat. Phys.* **91** 125
- [25] Horn R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)
- [26] Barkai E, Kanter I and Sompolinsky H 1990 *Phys. Rev. A* **41** 590
- [27] Eissfeller H and Opper M 1992 *Phys. Rev. Lett.* **68** 2094
- [28] Patrick A E and Zagrebnov V A 1992 *J. Phys A: Math. Gen.* **25** 1009
- [29] Watkin T L H and Sherrington D 1991 *J. Phys A: Math. Gen.* **24** 5427