Mean field theory for asymmetric neural networks

H. J. Kappen and J. J. Spanjers
SNN University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands
(Received 5 October 1999)

The computation of mean firing rates and correlations is intractable for large neural networks. For symmetric networks one can derive mean field approximations using the Taylor series expansion of the free energy as proposed by Plefka. In asymmetric networks, the concept of free energy is absent. Therefore, it is not immediately obvious how to extend this method to asymmetric networks. In this paper we extend Plefka's approach to asymmetric networks and in fact to arbitrary probability distributions. The method is based on an information geometric argument. The method is illustrated for asymmetric neural networks with sequential dynamics. We compare our approximate analytical results with Monte Carlo simulations for a network of 100 neurons. It is shown that the quality of the approximation for asymmetric networks is as good as for symmetric networks.

PACS number(s): 87.18.Sn, 87.10.+e, 07.05.+Mh

I. INTRODUCTION

Networks of stochastic binary neurons are an abstract computational model for neural information processing. The dynamics of these networks is defined as a Markov process. Under rather mild conditions, this dynamics converges asymptotically to a stationary probability distribution [1]. For symmetrically connected networks, this stationary distribution is a known function of the network parameters, depending on the type of dynamics. For random sequential dynamics one obtains the Boltzmann-Gibbs distribution; for parallel dynamics one obtains the Little model.

Computing the statistics of the stationary distribution, such as mean firing rates and correlations, is intractable. One can, however, use mean field theory to obtain approximate results. For Boltzmann distributions, it is possible to derive mean field theory as a Taylor expansion in the weights of the free energy around a factorized model. When only the first term in the expansion is considered, one obtains the naive mean field equations [2,3]. When one also includes the second order term, one obtains the Thouless-Anderson-Palmer (TAP) equations [4,5]. The TAP correction improves the quality of the approximation, depending on the amount of frustration in the network. This approach is different from the replica mean field approach because it retains a description in terms of the individual neural activities (and correlations) instead of in terms of a small number of order parameters. No quenched averaging is performed. Such a detailed description is useful when one considers learning in neural networks [6,7]

For asymmetric networks, the stationary distribution is not known. In particular, the concept of free energy does not exist. Therefore, it is not immediately clear how to extend the above procedure to asymmetric networks. One can, however, reformulate the Plefka expansion in an information geometric language [8,9]. One considers a manifold of probability distributions, containing a submanifold of factorized distributions. In geometric terms, the mean field or TAP approximations become orthogonal projections onto the factorized submanifold. The advantage of the geometric interpretation is that it can be directly extended to asymmetric networks. The aim of this paper is to present the mean field approximation for asymmetric networks. This approach is essentially identical to approximately solving the dynamical equations for the mean firing rates and correlations. To first order, this approach was used in [10].

This paper is organized as follows. In Sec. II we briefly introduce stochastic neural networks. In Sec. III we introduce the information theoretic description of the mean field approximation, and we apply the method to asymmetric networks with sequential dynamics. We obtain expressions valid to second order in the weights for the mean firing rates and the correlations. In Sec. IV we compare the approximations with Monte Carlo results.

II. STOCHASTIC NEURAL NETWORKS

Consider a network of n binary neurons $s_i = \pm 1$. Each neuron has a bias or threshold θ_i and the activity of neuron j affects neuron i through a synaptic coupling w_{ij}. The dynamics of the network is sequential Glauber dynamics. Define the operator F_i that flips the value of the ith neuron: $s' = F_i s \Leftrightarrow s'_j = s_j, j \neq i, s'_i = -s_i$. At discrete time steps, the network in state s can make a transition to state $s' = F_i s$ with probability

$$T(s'|s) = \frac{1}{n}\sigma(h_i s'_i) \tag{2.1}$$

where $h_i = \Sigma_{j \neq i} w_{ij} s_j + \theta_i$ and $\sigma(x) = \frac{1}{2}[1 + \tanh(x)]$. The probability of remaining in state s is given implicitly by the equality $\Sigma_{s'} T(s'|s) = 1$.

This probabilistic dynamics is a first order Markov process. When the weights are finite, the dynamics is ergodic and converges to a unique stationary distribution $p(s|w,\theta)$, which is a right eigenvector of T with eigenvalue 1:

$$p = Tp. \tag{2.2}$$

For symmetric networks, p is the Boltzmann distribution. For asymmetric networks the dependence of p on the weights and the thresholds is not known.

From Eq. (2.2), one can derive that the stationary mean firing rates and correlations satisfy

$$\langle s_i \rangle = \langle \tanh[h_i(s)] \rangle, \tag{2.3}$$

$$\langle s_i s_j \rangle = \tfrac{1}{2} \langle s_i \tanh[h_j(s)] \rangle + (i \leftrightarrow j). \tag{2.4}$$

III. INFORMATION THEORY AND MEAN FIELD APPROXIMATION

Let $\mathcal{P} = \{p(s|w,\theta)\}$ be the manifold of all probability distributions that can be obtained by considering different values of w and θ. $\mathcal{P}$ contains a submanifold $\mathcal{M} \subset \mathcal{P}$ of factorized probability distributions. This submanifold is described by

$$\mathcal{M} = \{q(s|\theta,w) \in \mathcal{P} | w = 0\}.$$

$\theta = (\theta_1, \ldots, \theta_n)$ parametrizes the submanifold $\mathcal{M}$, and w parametrizes directions in $\mathcal{P}$ orthogonal to $\mathcal{M}$. Since $q \in \mathcal{M}$ is factorized, we can write the stationary distribution explicitly:

$$q(s|\theta^q) = \prod_i \sigma(\theta_i^q s_i) = \prod_i \tfrac{1}{2}(1 + m_i^q s_i),$$

where $m_i^q = \langle s_i \rangle_q = \tanh(\theta_i^q)$ and $\langle \cdot \rangle_q$ denotes the expectation value with respect to the distribution q.

Consider a network whose weights and thresholds are given by θ and w. This network has a stationary distribution $p(s|\theta,w) \in \mathcal{P}$. We want to find its *mean field approximation*, which we define as the factorized distribution $q \in \mathcal{M}$ that we obtain by minimizing the relative entropy

$$D(p,q) = \sum_s p(s|\theta,w) \ln\left(\frac{p(s|\theta,w)}{q(s|\theta^q)} \right)$$

with respect to the coordinates θ^q of the factorized distribution q [8,9]. We find

$$\frac{dD(p,q)}{d\theta_i^q} = m_i^q - m_i^p = 0, \tag{3.1}$$

with $m_i^p = \langle s_i \rangle_p$. This equation states that the closest factorized model has its first moments equal to the first moments of the target distribution p. This is illustrated in Fig. 1.

We need to solve Eq. (3.1) for $m_i^q = \tanh(\theta_i^q)$. However, we cannot compute m_i^p since we do not know the stationary distribution p. Even if we knew p (for instance, a Boltzmann-Gibbs distribution) it would be of little help, since computation of m_i^p is intractable. In order to proceed, we assume that the distribution p is somehow close to the submanifold $\mathcal{M}$. Define $d\theta_i = \theta_i - \theta_i^q$ and $dw_{ij} = w_{ij} - 0 = w_{ij}$. Expanding $dm_i = m_i^p - m_i^q$ to second order, we obtain

$$0 = dm_i \approx \sum_J \left. \frac{\partial m_i}{\partial \Theta_J} \right|_q d\Theta_J + \frac{1}{2} \sum_{J,K} \left. \frac{\partial^2 m_i}{\partial \Theta_J \partial \Theta_K} \right|_q d\Theta_J d\Theta_K, \tag{3.2}$$

where $\Theta_I = (\theta_i, w_{ij})$ is the vector of all weights and thresholds, and I runs over all relevant indices.

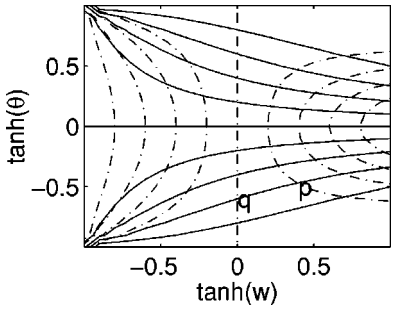

FIG. 1. Manifold of probability distributions $\mathcal{P}$ is computed for a Boltzmann distribution on two variables $p(s_1,s_2|w,\theta) = \exp[ws_1s_2 + \theta(s_1+s_2)]/Z$. Solid lines are lines of constant $\langle s_1 \rangle = \langle s_2 \rangle$. Broken lines are lines of constant $\langle s_1 s_2 \rangle$. Both (w,θ) and $(\langle s_1 \rangle, \langle s_1 s_2 \rangle)$ are coordinate systems of $\mathcal{P}$. $\mathcal{M}$ is given by the line $w = 0$. For any $p \in \mathcal{P}$, the closest $q \in \mathcal{M}$ satisfies $\langle s \rangle_q = \langle s \rangle_p$.

Equations (3.2) are the mean field equations (including TAP corrections). They can be applied to any manifold of probability distributions that contains a submanifold of tractable distributions (such as factorized distributions) and for which the evaluation of the derivatives at q is tractable.[1] For Boltzmann distributions this derivation is essentially identical to the approach introduced in [5].

The computation of the derivatives of m_i with respect to θ, w at the factorized point q can be obtained from Eq. (2.3). The computation of the derivatives is tedious but straightforward. It is presented in the Appendix. The result is

$$m_i = \tanh\left(\sum_j w_{ij} m_j + \theta_i - m_i \sum_j w_{ij}^2 (1 - m_j^2) \right), \tag{3.3}$$

where $m_i = m_i^p = m_i^q$ because of Eq. (3.1).

Note that this result is identical to the TAP equations for symmetric networks. This is somewhat surprising if one tries to understand this equation from a cavity type of argument. The cavity argument is that in computing the mean field equation for neuron i, the mean firing rates of all neurons $j \neq i$ are subject to a polarization $\delta m_j = \chi_{jj} w_{ji} m_i$ which shifts their firing rates to $m_j - \delta m_j$. Here, $\chi_{ij} = \partial m_j / \partial \theta_i$. From the linear response theory one obtains that $\chi_{jj} = 1 - m_j^2$. Substituting this altered value of m_j in the naive mean field equation then gives the TAP equation. When one applies this argument to the asymmetric network, one obtains a TAP term of the form $-m_i \Sigma_j w_{ij} w_{ji} (1 - m_j^2)$, in disagreement with Eq. (3.3). The paradox is resolved when one observes that the linear response relation does not hold for the asymmetric network due to the absence of equilibrium.

If one wants to learn the parameters w_{ij} and θ_i from data, one needs also an approximate expression for the correlations. Due to the absence of equilibrium this expression cannot be obtained from the linear response theorem, as was done in [6] for the Boltzmann machine. Instead, one must compute the correlations in a similar perturbative manner to the mean firing rates. Although we will not pursue the issues

[1] This is not the case, for instance, for directed graphs, when the number of parents is large. In that case additional approximations must be made [11].

of learning in this paper, we give the approximate expressions for the correlations. We will address learning in a separate publication.

Unlike the mean firing rates, the stationary correlations depend on the type of dynamics. We restrict ourselves to sequential dynamics and equal time correlations. When we expand $\chi_{ij}=\langle s_i s_j\rangle-\langle s_i\rangle\langle s_j\rangle$ around the factorized solution $\chi^q_{ij}=0$, we obtain

$$\chi_{ij}=\tfrac{1}{2}(1-m_i^2)(1-m_j^2)\times\left(w_{ij}+\sum_{k\neq i} w_{jk}w^s_{ik}(1-m_k^2)+2m_i m_j (w_{ji})^2\right)+(i\leftrightarrow j), \tag{3.4}$$

where w^s_{ij} denotes the symmetric part of w_{ij}. The derivation is given in the Appendix.

IV. NUMERICAL RESULTS

To evaluate the quality of our mean field approximations, we compare them to results of Monte Carlo simulations. We consider networks of $n=100$ neurons. We choose $w^0_{ij}, i\neq j$, randomly and independently from a normal distribution with mean zero and variance $1/\sqrt{n}$. We consider two different types of weights: symmetric weights $w^0_{ij}=w^0_{ji}$ and asymmetric weights, where w^0_{ij} and w^0_{ji} are drawn independently. We consider two types of thresholds: $\theta^0_i=0$ and θ^0_i chosen randomly and independently from a normal distribution with mean zero and variance 1. Since the approximation is expected to deteriorate with increasing weight size, we consider networks with $(w_{ij},\theta_i)=\beta(w^0_{ij},\theta^0_i)$ and vary $0\leqslant\beta\leqslant 1$.

We use Monte Carlo simulation to estimate the mean firing rates $\langle s_i\rangle$ and correlations χ_{ij}. The states are generated using sequential Glauber dynamics. To minimize the initialization (burn in) effect, we start the network in a random state and do not include the first t_0 iterations. We compute the average over the subsequent τ states:

$$\langle s_i\rangle^{\mathrm{MC}}=\frac{1}{\tau}\sum_{t=t_0}^{t=t_0+\tau} s_i(t), \tag{4.1}$$

$$\chi^{\mathrm{MC}}_{ij}=\frac{1}{\tau}\sum_{t=t_0}^{t=t_0+\tau} s_i(t)s_j(t)-\langle s_i\rangle^{\mathrm{MC}}\langle s_j\rangle^{\mathrm{MC}}. \tag{4.2}$$

The results are rather dependent on a proper choice of t_0 and τ. We obtained stable results by choosing $t_0=10^5 n$ and $\tau=10^6 n$. These values are rather large, but necessary to get results accurate enough to compute the small χ_{ij}'s. (The χ_{ij}'s are small because to lowest order $\chi_{ij}\propto w_{ij}\propto 1/\sqrt{n}$.)

From Eq. (3.3) we compute the mean field approximation of the mean firing rates. In order to assess the importance of the second order (TAP) contribution, we also compute these approximate values taking only the terms of $O(w)$ into account. In Fig. 2, we show the root mean square (RMS) values of the mean firing rates as a function of β for the Monte Carlo solution (MC), the mean field solution (MF) solution, and the TAP solution. The statistical errors in the Monte Carlo results for m_i are of the order $\delta m_i\approx 0.002$. In addition, we show the RMS values of the difference between the MF and MC solutions and between the TAP and MC solutions. We conclude that the second order approximation is significantly better than the first order approximation when $\beta<1$, for both symmetric and asymmetric networks.

The results for the correlations are presented in Fig. 3. We compute the TAP values for the mean firing rates and insert these in Eq. (3.4). With these values of m, we consider separately the $O(w)$ approximation and the $O(w^2)$ approximations of Eq. (3.4). The statistical errors in the Monte Carlo results for χ_{ij} are very small due to the large sampling times. For instance, at $\beta=0.5$ they are of the order of $\delta\chi_{ij}\approx 0.002$. We conclude that the second order approximation of the correlations gives a small, but statistically significant, improvement over the first order approximation when $\beta<0.5$, for both symmetric and asymmetric networks.

V. DISCUSSION

We have derived a mean field theory for asymmetric networks including $O(w^2)$ (''TAP-like'') corrections. Surprisingly, these equations are identical to the well-known equations for symmetric networks. In addition, we have derived an approximation for the correlations which is valid to $O(w^2)$. Numerical results show that the mean field results are equally accurate for symmetric and asymmetric networks.

It is easy to show that Eqs. (2.3) also hold for parallel dynamics. Therefore, Eqs. (3.3) also describe the approximate mean firing rates for parallel dynamics. The time-delayed correlations are given by $\langle s_i(t+1)s_j(t)\rangle=\langle s_j\tanh(h_i)\rangle$, which is identical to the unsymmetrized version of Eqs. (2.4). Therefore, the unsymmetrized version of Eqs. (3.4) describes the time-delayed correlations to $O(w^2)$. The equal-time correlations are given by $\langle s_i s_j\rangle=\langle\tanh(h_i)\tanh(h_j)\rangle$ and are not related to any of the results of this paper, but can be expanded using the same method.

ACKNOWLEDGMENT

This research was funded in part by the Dutch Technology Foundation (STW).

APPENDIX

In this Appendix we present the main steps to deriving the TAP equations, Eqs. (3.3), and the equal-time correlations, Eqs. (3.4).

1. TAP equations

We start with the computation of the derivatives in Eq. (3.2). From Eq. (2.3) we obtain

$$\left.\frac{\partial\langle s_i\rangle}{\partial\theta_j}\right|_q=\sum_s\left.\frac{\partial p(s)}{\partial\theta_j}\right|_q\tanh(\theta^q_i)+q(s)(1-m_i^2)\delta_{ij}=(1-m_i^2)\delta_{ij}.$$

The first term is zero because of the normalization $\Sigma_s p(s)=1$ and $m_i=m^q_i$. Similarly,

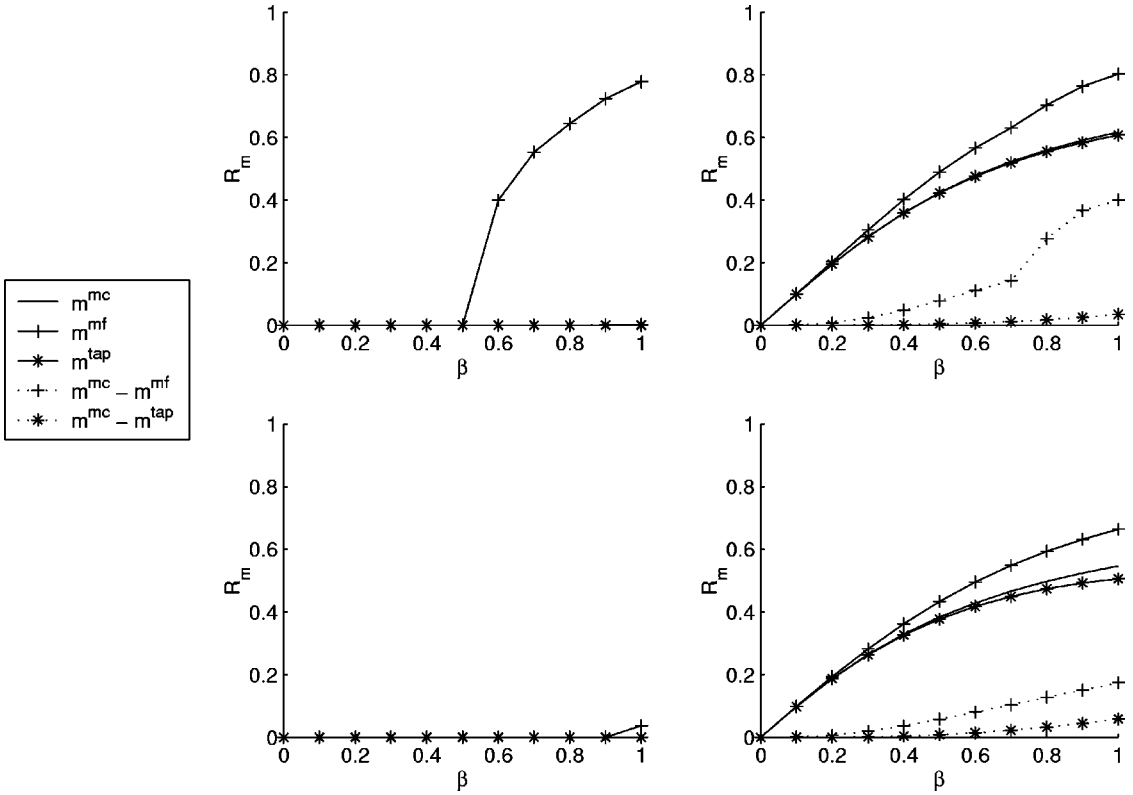

FIG. 2. Mean firing rates as a function of the strength of the connections for sequential dynamics, $n=100$. RMS values $R_m^2=(1/n)\Sigma_i^n m_i^2$ of Monte Carlo results (—), first order approximation (+–), and second order approximation (⋆—). In addition, RMS values of the difference between the first order approximation and the MC value (+···) and the difference between the second order approximation and the MC value (⋆···). Top row, symmetric connections ($w_{ij}=w_{ji}$). Bottom row, asymmetric connections (w_{ij} and w_{ji} are drawn independently). Left column $\theta_i=0$, right column θ_i random. In the top left figure, both the TAP results and the Monte Carlo results give $m_i=0$ due to symmetry. Therefore, the errors in the TAP results are zero. The mean field solution breaks at $\beta=0.5$ and the errors in the mean field results ($m^{\mathrm{MC}}-m^{\mathrm{MF}}$) equal the mean field results (m^{MF}).

$$\left.\frac{\partial\langle s_i\rangle}{\partial w_{jk}}\right|_q=(1-m_i^2)\delta_{ij}m_k.$$

Inserting in Eq. (3.2), we obtain to lowest order

$$0=dm_i=(1-m_i^2)\left(d\theta_i+\sum_j m_j dw_{ij}\right)+O(d\Theta^2). \tag{A1}$$

Using $d\theta_i=\theta_i-\theta_i^q$ and $dw_{ij}=w_{ij}$, this is equivalent to

$$m_i=\tanh\left(\sum_j w_{ij}m_j+\theta_i\right).$$

In a similar way one computes the second order derivatives:

$$\sum_{jk}\left.\frac{\partial^2\langle s_i\rangle}{\partial\theta_j\partial\theta_k}\right|_q d\theta_j d\theta_k=-2m_i(1-m_i^2)(d\theta_i)^2,$$

$$\sum_{jkl}\left.\frac{\partial^2\langle s_i\rangle}{\partial\theta_j\partial w_{kl}}\right|_q d\theta_j dw_{kl}=(1-m_i^2)\sum_j[(1-m_j^2)\times d\theta_j-2m_i m_j d\theta_i]dw_{ij},$$

$$\sum_{jklm}\left.\frac{\partial^2\langle s_i\rangle}{\partial w_{jk}\partial w_{lm}}\right|_q dw_{jk}dw_{lm}=(1-m_i^2)\sum_{jk}[(1-m_k^2)m_j dw_{kj}dw_{ik}+(1-m_j^2)m_k dw_{jk}dw_{ij}-2m_i\langle s_j s_k\rangle dw_{ij}dw_{ik}].$$

Substituting these into Eq. (3.2) we obtain

$$0=dm_i=(1-m_i^2)\left(A_i-m_iA_i^2+\sum_j(1-m_j^2)w_{ij}A_j-m_i\sum_j w_{ij}^2(1-m_j^2)\right)+O(d\Theta^3), \tag{A2}$$

where we have defined $A_i=d\theta_i+\Sigma_j dw_{ij}m_j$. Since $A_i=0+O(d\Theta^2)$, because of Eq. (A1), we obtain

$$A_i=m_i\sum_j w_{ij}^2(1-m_j^2)+O(d\Theta^3),$$

which is equivalent to Eq. (3.3).

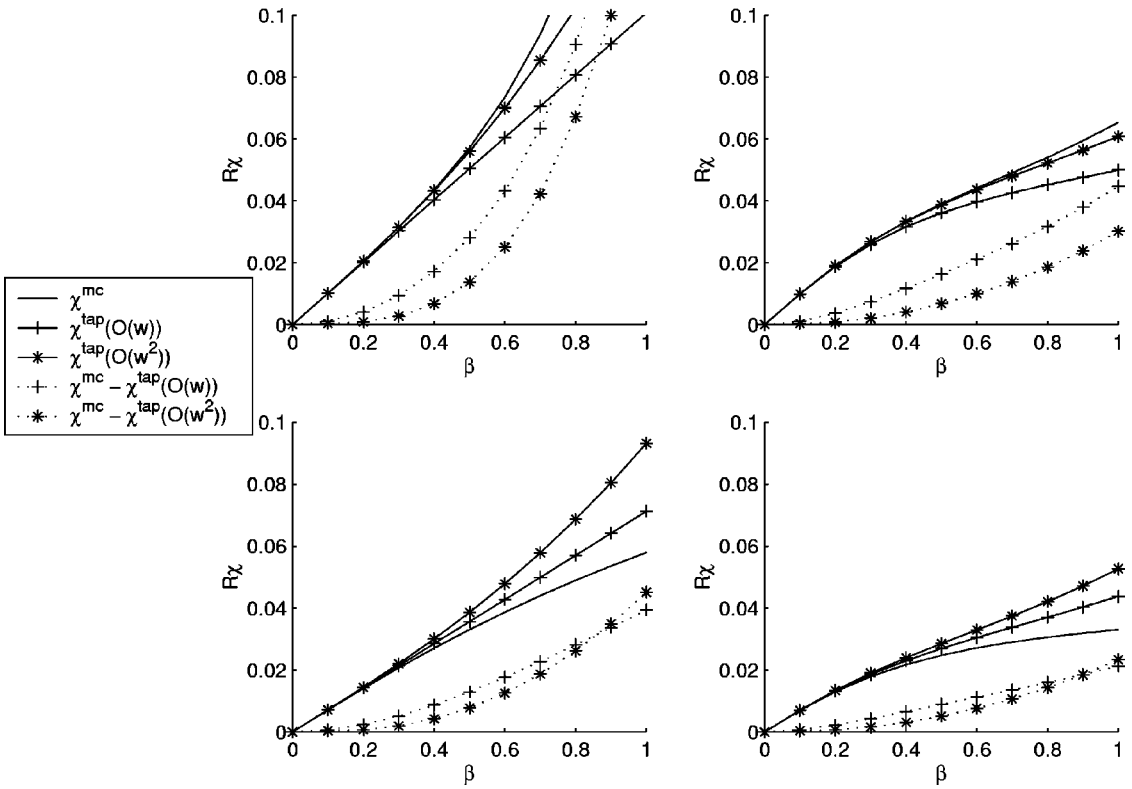

FIG. 3. Correlations as a function of the strength of the connections for sequential dynamics, $n=100$. RMS values $R\chi^2=[2/n(n-1)]\Sigma_{i>j}^n\chi_{ij}^2$ of Monte Carlo results (—), first order approximation (+—), and second order approximation (⋆—). In addition, RMS values of the difference between the first order approximation and the MC value (+ ⋯) and the difference between the second order approximation and the MC value (⋆ ⋯). Top row, symmetric connections ($w_{ij}=w_{ji}$). Bottom row, asymmetric connections (w_{ij} and w_{ji} are drawn independently). Left column $\theta_i=0$, right column θ_i random.

2. Correlations

From Eqs. (2.3) and (2.4) we obtain

$$2\chi_{ab}=\sum_s p(s)s_a\{\tanh[h_b(s)]-m_b\}+(a\leftrightarrow b).$$

From this equation, we directly compute the derivatives of χ_{ab} with respect to θ_i and w_{ij}:

$$2\frac{\partial\chi_{ab}}{\partial\theta_i}\bigg|_q=0,\quad 2\frac{\partial\chi_{ab}}{\partial w_{ij}}\bigg|_q=(1-m_a^2)(1-m_b^2)\delta_{ia}\delta_{jb}+(a\leftrightarrow b),\quad 2\frac{\partial^2\chi_{ab}}{\partial\theta_i\partial\theta_j}\bigg|_q=0,$$

$$2\frac{\partial^2\chi_{ab}}{\partial\theta_i\partial w_{kl}}\bigg|_q=(1-m_b^2)\left[\delta_{bk}\left(\frac{\partial\langle s_as_l\rangle}{\partial\theta_i}-m_l\frac{\partial\langle s_a\rangle}{\partial\theta_i}\right)-2\delta_{ib}\delta_{kb}m_b(\langle s_as_l\rangle-m_am_l)-m_a(1-m_l^2)\delta_{kb}\delta_{il}\right]+(a\leftrightarrow b),$$

$$2\frac{\partial^2\chi_{ab}}{\partial w_{ij}\partial w_{kl}}\bigg|_q=(1-m_b^2)\left[\delta_{bk}\left(\frac{\partial\langle s_as_l\rangle}{\partial w_{ij}}-m_l\frac{\partial\langle s_a\rangle}{\partial w_{ij}}\right)+\delta_{bi}\left(\frac{\partial\langle s_as_j\rangle}{\partial w_{kl}}-m_j\frac{\partial\langle s_a\rangle}{\partial w_{kl}}\right)-2\delta_{ib}\delta_{kb}m_b(\langle s_as_js_l\rangle-m_a\langle s_js_l\rangle)\right.$$
$$\left.-m_am_j(1-m_l^2)\delta_{kb}\delta_{il}-m_am_l(1-m_j^2)\delta_{ib}\delta_{kj}\right]+(a\leftrightarrow b).$$

Thus,

$$
\begin{aligned}
\chi^p_{ab} &= \sum_{ij} \frac{\partial \chi_{ab}}{\partial w_{ij}} dw_{ij} + \sum_{ijkl} \frac{1}{2} \frac{\partial^2 \chi_{ab}}{\partial w_{ij} \partial w_{kl}} dw_{ij} dw_{kl} + \sum_{ikl} \frac{\partial^2 \chi_{ab}}{\partial \theta_i \partial w_{kl}} d\theta_i dw_{kl} \\
&= \sum_{ij} \frac{\partial \chi_{ab}}{\partial w_{ij}} dw_{ij} + \sum_{ijkl} \left(\frac{1}{2} \frac{\partial^2 \chi_{ab}}{\partial w_{ij} \partial w_{kl}} - m_j \frac{\partial^2 \chi_{ab}}{\partial \theta_i \partial w_{kl}} \right) dw_{ij} dw_{kl} \\
&= \frac{1}{2}(1-m_b^2) \Bigg[(1-m_a^2) w_{ab} + \sum_{ijl} dw_{ij} dw_{bl} \Bigg(\frac{\partial \langle s_a s_l \rangle}{\partial w_{ij}} - m_j \frac{\partial \langle s_a s_l \rangle}{\partial \theta_i} - m_b \delta_{ib} (\langle s_a s_j s_l \rangle \\
&\quad - m_a \langle s_j s_l \rangle) \\
&\quad + 2 m_j m_b \delta_{ib} (\langle s_a s_l \rangle - m_a m_l) \Bigg) \Bigg] + (a \leftrightarrow b) \\
&= \frac{1}{2}(1-m_a^2)(1-m_b^2) \Bigg(w_{ab} + \sum_{l \neq a} dw_{bl} dw^s_{al} (1-m_l^2) + 2 m_a m_b (dw_{ba})^2 \Bigg) + (a \leftrightarrow b).
\end{aligned}
$$

This is identical to Eq. (3.4).

[1] G.R. Grimmett and D.R. Stirzaker, *Probability and Random Processes* (Clarendon Press, Oxford, 1992).

[2] R. Glauber, J. Math. Phys. **4**, 294 (1963).

[3] M. Suzuki and R. Kubo, J. Phys. Soc. Jpn. **24**, 51 (1968).

[4] D.J. Thouless, P.W. Anderson, and R.G. Palmer, Philos. Mag. **35**, 593 (1977).

[5] T. Plefka, J. Phys. A **15**, 1971 (1982).

[6] H.J. Kappen and F.B. Rodríguez, Neural Comput. **10**, 1137 (1998).

[7] H.J. Kappen and F.B. Rodríguez, in *Advances in Neural Information Proceessing Systems No. 11*, edited by M. S. Kearns, S. A. Solla, and D. A. Cohn (MIT Press, Cambridge, MA, 1999), pp. 280–286.

[8] S.-I. Amari, IEEE Trans. Neural Netw. **3**, 260 (1992).

[9] T. Tanaka, in *Advances in Neural Information Processing Systems No. 11*, pp. 351–357.

[10] I. Ginzburg and H. Sompolinsky, Phys. Rev. E **50**, 3171 (1994).

[11] L.K. Saul, T. Jaakkola, and M.I. Jordan, J. Artif. Intell. Res. **4**, 61 (1996).