Finite-size effects in separable recurrent neural networks

1998 J. Phys. A: Math. Gen. 31 6615

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

J. Phys. A: Math. Gen. **31** (1998) 6615–6634. Printed in the UK PII: S0305-4470(98)92032-7

Finite-size effects in separable recurrent neural networks

A Castellanos†, A C C Coolen‡ and L Viana§

† CICESE, Física de Materiales, A. Postal 2681, Ensenada 22800, BC, Mexico and Dept. de Física, Universidad de Sonora, A. Postal 1626, Hermosillo 83000, Son., Mexico

‡ Department of Mathematics, King's College, University of London, Strand, London WC2R 2LS, UK

§ CCMC, UNAM, A. Postal 2681, 22800 Ensenada, BC, Mexico

Received 24 February 1998, in final form 28 April 1998

Abstract. We perform a systematic analytical study of finite-size effects in separable recurrent neural network models with sequential dynamics, away from saturation. We find two types of finite-size effects: thermal fluctuations, and disorder-induced 'frozen' corrections to the mean-field laws. The finite-size effects are described by equations that correspond to a time-dependent Ornstein–Uhlenbeck process. We show how the theory can be used to understand and quantify various finite-size phenomena in recurrent neural networks, with and without detailed balance.

1. Introduction

Infinite-range spin models of recurrent neural networks, with information stored in the values of the interaction strengths of pairs of spins, have been studied intensively with statistical mechanical tools following [1, 2]. The first wave of such studies involved mainly equilibrium analyses, and was consequently restricted to models obeying detailed balance. Away from the saturation regime (where small numbers of patterns are stored) such models can be solved with standard mean-field techniques and display standard mean-field behaviour. In contrast, in the saturation regime (where an extensive number of patterns are stored) tools from spin-glass theory are required (replica theory), and non-trivial phases occur. The second wave of studies employed tools from non-equilibrium statistical mechanics. Here restriction to detailed balance models is irrelevant. However, in view of the highly non-trivial nature of the glassy non-ergodic dynamics of models close to saturation, most dynamical studies have been restricted to recurrent neural networks with only small numbers of patterns stored. For an overview of the relevant literature see textbooks such as [3] or reviews such as [4].

Despite the fact that finite-size effects have been reported regularly in literature, and that they are known to persist even for system sizes up to $N \approx 10^5$ [5], it appears that systematic studies of finite-size effects in recurrent networks which go beyond pilot studies such as [6] have not yet been performed. The purpose of this paper is to carry out a comprehensive analysis of finite-size effects (in first non-trivial order in the system size) for a reasonably general class of recurrent neural network models, where the interaction matrix has a separable structure. This class contains detailed balance models, as well as models without detailed balance. Away from saturation, finite-size effects in these systems take the form of mainly thermal fluctuations of order $\mathcal{O}(N^{-1/2})$ around mean-field trajectories for dynamical order parameters, as well as disorder-induced 'frozen' corrections to the mean-field laws. Close to saturation even the $N \to \infty$ dynamics cannot be solved in explicit form

0305-4470/98/316615+20$19.50

(describing transients is found to necessitate approximations in all of the present approaches; in the path integral formalism [7, 8] as well as in dynamical replica theory [9]). Therefore in the latter regime the development of a finite-size theory would be premature.

We study the evolution of finite recurrent neural network models away from saturation and with Glauber-type (stochastic) neuronal dynamics. We expand the Kramers–Moyal expansion for the system's natural dynamic order parameters on finite timescales, and calculate the statistical properties of finite-size effects to first non trivial order in $1/N$. The finite-size effects turn out to be governed by a time-dependent Ornstein–Uhlenbeck process. Our theory is used to analyse the dependence of finite-size effects on detailed balance, scaling properties of fluctuations close to phase boundaries, and escape processes in critical models which are driven purely by finite size effects. Comparison with extensive numerical simulations confirms the theoretical predictions in all cases.

2. Derivation of macroscopic laws

2.1. Model definitions and simple relations

We consider a system composed of a large, but finite, number N of interconnected neurons, modelled as Ising spins $\sigma_i \in \{-1, 1\}$. The vector $\boldsymbol{\sigma}(t) = (\sigma_1(t), \ldots, \sigma_N(t))$ defines the state of the system at time t. The dynamics of the system is defined by a master equation for the microscopic probability distribution $p_t(\boldsymbol{\sigma})$:

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(\boldsymbol{\sigma}) = \sum_i \{w_i(F_i\boldsymbol{\sigma}) p_t(F_i\boldsymbol{\sigma}) - w_i(\boldsymbol{\sigma}) p_t(\boldsymbol{\sigma})\} \tag{1}$$

$$w_i(\boldsymbol{\sigma}) = \tfrac{1}{2}[1 - \sigma_i \tanh(\beta h_i(\boldsymbol{\sigma}))] \tag{2}$$

where $w_i(t)$ defines the rate of the single-spin transitions $\sigma_i(t) \to -\sigma_i(t)$, $\beta = T^{-1}$ (the inverse temperature) controls the stochasticity in the dynamics, and F_i is an operator that flips the ith spin, i.e. $F_i f(\sigma_1, \ldots, \sigma_N) = f(\sigma_1, \ldots, -\sigma_i, \ldots, \sigma_N)$. The local field $h_i(\boldsymbol{\sigma})$ is given by the usual linear expression

$$h_i(\boldsymbol{\sigma}) = \sum_j J_{ij}\sigma_j + \theta_i \tag{3}$$

where J_{ij} is the strength of the synaptic connection from neuron j to neuron i, and θ_i is a response threshold. The interactions J_{ij} are assumed to result from a learning process involving a finite number p of randomly chosen binary patterns $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \ldots, \xi_N^\mu) \in \{-1, 1\}^N$, with $\mu = 1, \ldots, p$. We restrict ourselves to situations where the interactions have a separable form (see e.g. [10–12]) and introduce a parameter $\Delta \in \{0, 1\}$ to control whether or not self-interactions J_{ii} will be allowed:

$$J_{ij} = [1 - \Delta\delta_{ij}]\frac{1}{N}\sum_{\mu\nu=1}^{p} \xi_i^\mu A_{\mu\nu}\xi_j^\nu \qquad \begin{cases} \Delta = 0: & J_{ii} \neq 0 \\ \Delta = 1: & J_{ii} = 0\ . \end{cases} \tag{4}$$

Given the process (1), (2), we can define averages over the microscopic ensemble in the usual way, and find simple relations for the temporal derivatives of such averages:

$$\langle f(\boldsymbol{\sigma})\rangle_t \equiv \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma}) f(\boldsymbol{\sigma}) \qquad \frac{\mathrm{d}}{\mathrm{d}t}\langle f(\boldsymbol{\sigma})\rangle_t = \left\langle \sum_i w_i(\boldsymbol{\sigma})[f(F_i\boldsymbol{\sigma}) - f(\boldsymbol{\sigma})] \right\rangle_t . \tag{5}$$

In particular, application to $f(\boldsymbol{\sigma}) = \sigma_k$ gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\sigma_k\rangle_t = \langle\tanh(\beta h_k(\boldsymbol{\sigma}))\rangle_t - \langle\sigma_k\rangle_t . \tag{6}$$

In order to make the transition to a macroscopic description of the process, we define the usual pattern overlaps. These observables (which for finite N take discrete values only) measure the similarity between the state of the system and each of the p stored patterns:

$$\boldsymbol{m}(\boldsymbol{\sigma}) = (m_1(\boldsymbol{\sigma}), \ldots, m_p(\boldsymbol{\sigma})) \qquad m_\mu(\boldsymbol{\sigma}) = \frac{1}{N}\sum_i^N \xi_i^\mu \sigma_i. \tag{7}$$

The probability density for the macroscopic variables $\boldsymbol{m}$ is given by:

$$P_t(\boldsymbol{m}) \equiv \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta[\boldsymbol{m} - \boldsymbol{m}(\boldsymbol{\sigma})]. \tag{8}$$

We next define conditional, or subshell, averages of observables $f(\boldsymbol{\sigma})$. These are averages over the statistical ensemble, with the microscopic probability distribution $p_t(\boldsymbol{\sigma})$, restricted to those microstates $\boldsymbol{\sigma}$ that obey $\boldsymbol{m}(\boldsymbol{\sigma}) = \boldsymbol{m}$ (in a distributional sense):

$$\langle f(\boldsymbol{\sigma})\rangle_{m;t} \equiv \frac{\sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta[\boldsymbol{m} - \boldsymbol{m}(\boldsymbol{\sigma})]f(\boldsymbol{\sigma})}{\sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta[\boldsymbol{m} - \boldsymbol{m}(\boldsymbol{\sigma})]}. \tag{9}$$

Note that the δ-distribution in definition (9) allows us to replace all occurrences of $\boldsymbol{m}(\boldsymbol{\sigma})$ in f simply by $\boldsymbol{m}$:

$$\langle f[\boldsymbol{\sigma}, \boldsymbol{m}(\boldsymbol{\sigma})]\rangle_{m;t} = \langle f[\boldsymbol{\sigma}, \boldsymbol{m}]\rangle_{m;t}. \tag{10}$$

2.2. The Kramers–Moyal expansion

The dynamic equation for the macroscopic variables $\boldsymbol{m}(\boldsymbol{\sigma})$ can be obtained by making the choice $f(\boldsymbol{\sigma}) = \delta[\boldsymbol{m} - \boldsymbol{m}(\sigma)]$ in equation (5)

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t(\boldsymbol{m}) = \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma}) \sum_i w_i(\boldsymbol{\sigma}) \left\{\delta\left[\boldsymbol{m} - \boldsymbol{m}(\boldsymbol{\sigma}) + \frac{2}{N}\sigma_i\boldsymbol{\xi}_i\right] - \delta[\boldsymbol{m} - \boldsymbol{m}(\boldsymbol{\sigma})]\right\}.$$

Inside this expression we make a Taylor expansion in powers of the vector $\frac{2}{N}\sigma_i\boldsymbol{\xi}_i$ and write the result in terms of subshell averages (9):

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t(\boldsymbol{m}) = \sum_{\ell\geqslant 1}\frac{1}{\ell!}\sum_{\mu_1=1}^p \cdots \sum_{\mu_\ell=1}^p \frac{\partial^\ell}{\partial m_{\mu_1}\cdots\partial m_{\mu_\ell}}\{P_t(\boldsymbol{m})F^{(\ell)}_{\mu_1,\ldots,\mu_\ell}[\boldsymbol{m};t]\} \tag{11}$$

$$F^{(\ell)}_{\mu_1,\ldots,\mu_\ell}[\boldsymbol{m};t] = \left\langle \sum_i w_i(\boldsymbol{\sigma})\left[\frac{2}{N}\sigma_i\xi_i^{\mu_1}\right]\cdots\left[\frac{2}{N}\sigma_i\xi_i^{\mu_\ell}\right]\right\rangle_{\boldsymbol{m};t} = \mathcal{O}(N^{1-\ell}). \tag{12}$$

This is the so-called Kramers–Moyal expansion, applied to the present class of models.

Since equation (11) cannot be solved exactly, we follow the standard procedure for 'large systems' (see e.g. [13]), and expand (11) in powers of $1/N$. By keeping only the two leading orders, we obtain, at least on finite timescales (i.e. on times not scaling with the system size N), the $N\to\infty$ (mean-field) equations plus the leading-order contribution due to the finite-size effects†:

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t(\boldsymbol{m}) = \sum_{\mu=1}^p \frac{\partial}{\partial m_\mu}\{P_t(\boldsymbol{m})F^{(1)}_\mu[\boldsymbol{m};t]\} + \frac{1}{2}\sum_{\mu\nu=1}^p \frac{\partial^2}{\partial m_\mu\partial m_\nu}\{P_t(\boldsymbol{m})F^{(2)}_{\mu\nu}[\boldsymbol{m};t]\} + \mathcal{O}(N^{-2}). \tag{13}$$

† Note that due to Pawula's theorem [14] we have but three options: (1) to retain only the $\mathcal{O}(N^0)$ terms of (11), (describing the infinite system), (2) to include in addition the $\mathcal{O}(N^{-1})$ terms, or (3) to keep all remaining orders in N. Retaining a finite number of terms, until and including order N^{-n} with $n > 1$ would generate solutions $P_t(\boldsymbol{m})$ which violate the obvious condition that they be positive definite (i.e. represent probability densities).

Upon insertion of the transition rates $w_i(\boldsymbol{\sigma})$ (2) and the local fields (3), written in terms of the overlaps (7), we can work out explicitly and simplify (with e.g. (10)) the various terms, giving

$$F_\mu^{(1)}[\boldsymbol{m};t] = m_\mu - \frac{1}{N}\sum_i \xi_i^\mu \tanh\beta[\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{m}+\theta_i] + \frac{\beta\Delta}{N^2}\sum_i \langle\sigma_i\rangle_{\boldsymbol{m};t}\xi_i^\mu(\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{\xi}_i)[1-\tanh^2\beta[\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{m}+\theta_i]] + \mathcal{O}(N^{-2}). \quad (14)$$

Similarly:

$$\frac{1}{2}F_{\mu\nu}^{(2)}[\boldsymbol{m};t] = \frac{1}{N^2}\sum_i \xi_i^\mu\xi_i^\nu[1-\langle\sigma_i\rangle_{\boldsymbol{m};t}\tanh\beta(\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{m}+\theta_i)] + \mathcal{O}(N^{-2}). \quad (15)$$

In the limit $N\to\infty$ equation (13) reduces to a Liouville equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t(\boldsymbol{m}) = \sum_{\mu=1}^{p}\frac{\partial}{\partial m_\mu}\{P_t(\boldsymbol{m})[m_\mu - \langle\xi_\mu\tanh\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}+\theta]\rangle_{\boldsymbol{\xi},\theta}]\}$$

with the deterministic solution

$$P_t(\boldsymbol{m}) = \delta[\boldsymbol{m}-\boldsymbol{m}^*(t)] \qquad \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{m}^*(t) = \langle\boldsymbol{\xi}\tanh\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta]\rangle_{\boldsymbol{\xi},\theta} - \boldsymbol{m}^*(t) \quad (16)$$

where we defined

$$\langle g[\boldsymbol{\xi},\theta]\rangle_{\boldsymbol{\xi},\theta} = \lim_{N\to\infty}\frac{1}{N}\sum_i g[\boldsymbol{\xi}_i,\theta_i] \qquad \boldsymbol{\xi} = (\xi_1,\ldots,\xi_p)$$

with $\boldsymbol{\xi}_i = (\xi_i^1,\ldots,\xi_i^p)$.

3. Description via rescaled variables

3.1. Derivation of Fokker–Planck equation

The stochastic vector $\boldsymbol{m}(\boldsymbol{\sigma})$ can apparently be written as the sum of a deterministic term $\boldsymbol{m}^*(t)$ and a fluctuating term, with the latter vanishing for $N\to\infty$. Since for mean-field models the overlaps can be seen as an average over N independent random variables, one would expect, from the central limit theorem, the fluctuating term to scale as $N^{-\frac{1}{2}}$. Therefore we define a new stochastic variable $\boldsymbol{q}(\boldsymbol{\sigma})$:

$$\boldsymbol{q}(\boldsymbol{\sigma}) = \sqrt{N}[\boldsymbol{m}(\boldsymbol{\sigma})-\boldsymbol{m}^*(t)] \qquad \mathcal{P}_t(\boldsymbol{q}) = \int \mathrm{d}\boldsymbol{m}\, P_t(\boldsymbol{m})\delta[\boldsymbol{q}-\sqrt{N}[\boldsymbol{m}-\boldsymbol{m}^*(t)]] \quad (17)$$

in which $\boldsymbol{m}^*(t)$ is the solution of the deterministic equation (16). Working out the temporal derivative of $\mathcal{P}_t(\boldsymbol{q})$, with the help of the macroscopic equation (13), and taking the $N\to\infty$ limit, leads to a convenient description of the leading $\mathcal{O}(N^{-\frac{1}{2}})$ finite-size effects in terms of a Fokker–Planck equation for the rescaled variables $\boldsymbol{q}$:

$$\boldsymbol{m}(\boldsymbol{\sigma}(t)) = \boldsymbol{m}^*(t) + \frac{1}{\sqrt{N}}\boldsymbol{q}(t) + \mathcal{O}\left(\frac{1}{N}\right)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_t(\boldsymbol{q}) = \sum_\mu\frac{\partial}{\partial q_\mu}\{\mathcal{P}_t(\boldsymbol{q})F_\mu[\boldsymbol{q};t]\} + \sum_{\mu\nu}\frac{\partial^2}{\partial q_\mu\partial q_\nu}\{\mathcal{P}_t(\boldsymbol{q})D_{\mu\nu}[\boldsymbol{q};t]\} \quad (18)$$

in which the flow term is given by

$$F_\mu[\boldsymbol{q};t] = q_\mu - \beta\langle\xi_\mu(\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{q})[1-\tanh^2\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta]]\rangle_{\boldsymbol{\xi},\theta}$$
$$+\lim_{N\to\infty}\sqrt{N}\Big\{\langle\xi_\mu\tanh\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta]\rangle_{\boldsymbol{\xi},\theta}$$
$$-\frac{1}{N}\sum_i\xi_i^\mu\tanh\beta[\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta_i]\Big\}$$

where we used (14), (16) and (17). The last term in $F_\mu[\boldsymbol{q};t]$ describes a 'frozen' finite-size correction to the flow field, depending on the microscopic realization of the pattern components. Similarly, we can work out the diffusion matrix:

$$D_{\mu\nu}[\boldsymbol{q};t] = \lim_{N\to\infty}\frac{1}{N}\sum_i\xi_i^\mu\xi_i^\nu[1-\langle\sigma_i\rangle_{\boldsymbol{m}^*(t);t}\tanh\beta[\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta_i]] \tag{19}$$

where we have used $\langle\sigma_i\rangle_{\boldsymbol{m},t} = \langle\sigma_i\rangle_{\boldsymbol{m}^*(t);t} + \mathcal{O}(N^{-\frac{1}{2}})$. According to (6) we may also use

$$\langle\sigma_i\rangle_{\boldsymbol{m}^*(t);t} = \sigma_i(0)\mathrm{e}^{-t} + \int_0^t \mathrm{d}s\,\mathrm{e}^{s-t}\tanh\beta[\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{m}^*(s)+\theta_i] + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{20}$$

We conclude that the diffusion matrix does not depend on $\boldsymbol{q}$:

$$D_{\mu\nu}(t) = \langle\xi_\mu\xi_\nu\rangle_{\boldsymbol{\xi},\theta} - \mathrm{e}^{-t}\lim_{N\to\infty}\frac{1}{N}\sum_i\xi_i^\mu\xi_i^\nu\sigma_i(0)\tanh\beta[\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta_i]$$
$$-\int_0^t\mathrm{d}s\,\mathrm{e}^{s-t}\langle\xi_\mu\xi_\nu\tanh\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(s)+\theta]\tanh\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta]\rangle_{\boldsymbol{\xi},\theta}. \tag{21}$$

The fact that the above limits exist and give a non-trivial flow term and diffusion matrix for (18) is the *a posteriori* justification of the ansatz (17).

We finally have to specify the pattern and threshold statistics in order to analyse the Fokker–Planck equation (18). We choose independently drawn pattern components $\xi_i^\mu \in \{-1,1\}$ (with equal probabilities) and independently drawn neural thresholds θ_i (with probability distribution $W(\theta)$):

$$\langle g[\boldsymbol{\xi},\theta]\rangle_{\boldsymbol{\xi},\theta} = 2^{-p}\sum_{\boldsymbol{\xi}\in\{-1,1\}^p}\int\mathrm{d}\theta\,W(\theta)g[\boldsymbol{\xi},\theta] \tag{22}$$

which gives convenient relations such as

$$\langle f(\boldsymbol{\xi})g(\theta)\rangle_{\boldsymbol{\xi},\theta} = 2^{-p}\sum_{\boldsymbol{\xi}\in\{-1,1\}^p}f(\boldsymbol{\xi})\int\mathrm{d}\theta\,W(\theta)g(\theta) \qquad \langle\xi_\mu\rangle_{\boldsymbol{\xi},\theta} = 0 \qquad \langle\xi_\mu\xi_\nu\rangle_{\boldsymbol{\xi},\theta} = \delta_{\mu\nu}.$$

The flow term $\boldsymbol{F}[\boldsymbol{q};t]$ in (18) can be written as the sum of two contributions; the first ($\boldsymbol{K}$) depends on the specific microscopic realization of the pattern components, whereas the second (with the matrix $\boldsymbol{L}$) depends only on the probability distribution of the pattern components:

$$F_\mu[\boldsymbol{q};t] = K_\mu(t) + \sum_\nu L_{\mu\nu}(t)q_\nu \tag{23}$$

$$K_\mu(t) = \lim_{N\to\infty}\sqrt{N}\Big\{\langle\xi_\mu\tanh\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta]\rangle_{\boldsymbol{\xi},\theta} - \frac{1}{N}\sum_i\xi_i^\mu\tanh\beta[\boldsymbol{\xi}_i\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta_i]\Big\} \tag{24}$$

$$L_{\mu\nu}(t) = \delta_{\mu\nu} - \beta\sum_\lambda\langle\xi_\mu\xi_\lambda[1-\tanh^2\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta]]\rangle_{\boldsymbol{\xi},\theta}A_{\lambda\nu}. \tag{25}$$

The diffusion matrix in (18) is symmetric, and can be simplified to

$$D_{\mu\nu}(t) = \delta_{\mu\nu} - \mathrm{e}^{-t} \lim_{N\to\infty} \frac{1}{N} \sum_i \xi_i^\mu \xi_i^\nu \sigma_i(0) \tanh\beta[\boldsymbol{\xi}_i \cdot \boldsymbol{A}\boldsymbol{m}^*(t) + \theta_i]$$
$$- \int_0^t \mathrm{d}s\, \mathrm{e}^{s-t} \langle \xi_\mu \xi_\nu \tanh\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(s)+\theta] \tanh\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^*(t)+\theta]\rangle_{\boldsymbol{\xi},\theta}. \quad (26)$$

Equation (18), with its flow term (23) which is linear in the rescaled fluctuation variables $\boldsymbol{q}$ and with its $\boldsymbol{q}$-independent diffusion matrix (26), is called the 'linear noise' Fokker–Planck equation; it describes a so-called time-dependent Ornstein–Uhlenbeck process (see e.g. [15]).

3.2. General solution

The natural solution of the Ornstein–Uhlenbeck process (18) is a Gaussian distribution:

$$\mathcal{P}_t(\boldsymbol{q}) = \frac{1}{(2\pi)^{p/2}\sqrt{\det \boldsymbol{\Xi}(t)}} \exp\{-\tfrac{1}{2}[\boldsymbol{q} - \langle \boldsymbol{q}\rangle_t] \cdot \boldsymbol{\Xi}^{-1}(t)[\boldsymbol{q} - \langle \boldsymbol{q}\rangle_t]\}. \quad (27)$$

It is fully characterized in the usual way by the time-dependent average $\langle \boldsymbol{q}\rangle_t$ and the time-dependent correlation matrix

$$\Xi_{\mu\nu}(t) = \langle q_\mu q_\nu\rangle_t - \langle q_\mu\rangle_t \langle q_\nu\rangle_t. \quad (28)$$

Here we denote averages over the distribution (27) as $\langle f(\boldsymbol{q})\rangle_t = \int \mathrm{d}\boldsymbol{q}\, \mathcal{P}_t(\boldsymbol{q}) f(\boldsymbol{q})$. Insertion of equation (27) as an ansatz into the Fokker–Planck equation (18) gives the following three necessary and sufficient conditions for (27) to be a solution:

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \boldsymbol{q}\rangle_t = -\boldsymbol{L}(t)\langle \boldsymbol{q}\rangle_t - \boldsymbol{K}(t) \quad (29)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Xi}(t) = -\boldsymbol{L}(t)\boldsymbol{\Xi}(t) - \boldsymbol{\Xi}(t)\boldsymbol{L}^\dagger(t) + 2\boldsymbol{D}(t) \quad (30)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\log\det\boldsymbol{\Xi}(t) + 2\,\mathrm{Tr}[\boldsymbol{L} - \boldsymbol{D}\boldsymbol{\Xi}^{-1}] = \boldsymbol{0} \quad (31)$$

with the (symmetric) diffusion matrix (26). Equations (29) and (30) define the evolution in time of the moments of the distribution (27). Equation (31) is then solved automatically, which can be seen by combining the Wronski identity $\frac{\mathrm{d}}{\mathrm{d}t}\log\det\boldsymbol{B} = \mathrm{Tr}[\boldsymbol{B}^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}]$ with equation (30).

Since the differential equations (29) and (30) are linear, they can be solved using standard procedures (see e.g. [13, 15]). One defines the propagator $\boldsymbol{G}(t)$ as the (matrix) solution of

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{G}(t) = -\boldsymbol{L}(t)\boldsymbol{G}(t) \qquad \boldsymbol{G}(0) = \boldsymbol{I} \qquad \text{or} \qquad \boldsymbol{G}(t) = \boldsymbol{I} - \int_0^t \mathrm{d}s\, \boldsymbol{L}(s)\boldsymbol{G}(s) \quad (32)$$

in which $\boldsymbol{I}$ denotes the unit matrix. This allows us to express the solution of (29) and (30) in the following compact way:

$$\langle \boldsymbol{q}\rangle_t = \boldsymbol{G}(t)\langle \boldsymbol{q}\rangle_0 - \boldsymbol{G}(t)\int_0^t \mathrm{d}s\, \boldsymbol{G}^{-1}(s)\boldsymbol{K}(s) \quad (33)$$

$$\boldsymbol{\Xi}(t) = \boldsymbol{G}(t)\boldsymbol{\Xi}(0)\boldsymbol{G}^\dagger(t) + 2\boldsymbol{G}(t)\int_0^t \mathrm{d}s\, \boldsymbol{G}^{-1}(s)\boldsymbol{D}(s)[\boldsymbol{G}^\dagger(s)]^{-1}\boldsymbol{G}^\dagger(t) \quad (34)$$

(as can be verified by insertion). Note, however, that calculating the propagator (32) can still be non-trivial.

3.3. Stationary states and detailed balance

For large times $t \to \infty$ the dependence of (26) on the microscopic initial conditions vanishes. Furthermore, for macroscopic stationary states, i.e. $\boldsymbol{m}^*(t) = \boldsymbol{m}^*$ for all t, with $\boldsymbol{m}^*$ given by the solution of the macroscopic fixed-point equation

$$\boldsymbol{m}^* = \langle \boldsymbol{\xi} \tanh \beta[\boldsymbol{\xi} \cdot \boldsymbol{A}\boldsymbol{m}^* + \theta] \rangle_{\boldsymbol{\xi},\theta} \tag{35}$$

we can in addition perform the time integration in (26). All flow and diffusion terms in the Fokker–Planck equation (18) become independent of time, the convection matrix (25) can be expressed in terms of the diffusion matrix as

$$\boldsymbol{L} = \boldsymbol{I} - \beta \boldsymbol{D}\boldsymbol{A} \tag{36}$$

and our process (18) reduces to a time-independent Ornstein–Uhlenbeck process, characterized by

$$\boldsymbol{F}[\boldsymbol{q}] = \boldsymbol{K} + [\boldsymbol{I} - \beta \boldsymbol{D}\boldsymbol{A}]\boldsymbol{q} \qquad D_{\mu\nu} = \delta_{\mu\nu} - \langle \xi_\mu \xi_\nu \tanh^2 \beta[\boldsymbol{\xi} \cdot \boldsymbol{A}\boldsymbol{m}^* + \theta] \rangle_{\boldsymbol{\xi},\theta}$$

$$\boldsymbol{K} = \lim_{N\to\infty} \sqrt{N} \left\{ \langle \boldsymbol{\xi} \tanh \beta[\boldsymbol{\xi} \cdot \boldsymbol{A}\boldsymbol{m}^* + \theta] \rangle_{\boldsymbol{\xi},\theta} - \frac{1}{N} \sum_i \boldsymbol{\xi}_i \tanh \beta[\boldsymbol{\xi}_i \cdot \boldsymbol{A}\boldsymbol{m}^* + \theta_i] \right\}.$$

Since the matrix $\boldsymbol{L}$ (36) is stationary, the propagator (32) reduces to $\boldsymbol{G}(t) = \exp[-t\boldsymbol{L}]$. Whether or not a macroscopic stationary state $\boldsymbol{m}^*(t) = \boldsymbol{m}^*$ will be reached will depend on the choice made for the matrix $\boldsymbol{A}$.

A sufficient condition for asymptotic stationarity is (microscopic) detailed balance, which states that, in addition to stationarity of the probability distribution $p_t(\boldsymbol{\sigma})$, there is no net probability current between any two configurations $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$. For the models studied in this paper this translates into symmetry of the matrix $\boldsymbol{A}$ and absence of self-interactions, i.e. $\Delta = 1$ (apart from pathological exceptions, such as systems with self-interactions only). See, e.g. [16, 17]. Note, however, that our equations (16), (18) show that presence or absence of self-interactions does not yet play a role in the first two leading orders in the system size. We will now inspect the conditions for the Fokker–Planck equation (18) to have a stationary solution, and show that for this solution to obey detailed balance (i.e. for there to be no net probability current in $\boldsymbol{q}$-space) we must again require symmetry of the matrix $\boldsymbol{A}$. Equation (18) can be written as a continuity equation for the probability density $P_t(\boldsymbol{q})$:

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t(\boldsymbol{q}) + \boldsymbol{\nabla} \cdot \boldsymbol{J}_t(\boldsymbol{q}) = 0$$

with $\boldsymbol{\nabla} = (\frac{\partial}{\partial q_1}, \ldots, \frac{\partial}{\partial q_p})$ and with, in the case of macroscopic stationarity,

$$\boldsymbol{J}_t(\boldsymbol{q}) = P_t(\boldsymbol{q})\{\boldsymbol{D}\boldsymbol{\Xi}^{-1}(\boldsymbol{q} - \langle \boldsymbol{q} \rangle_t) - \boldsymbol{K} - \boldsymbol{L}\boldsymbol{q}\}. \tag{37}$$

From (29) and (30) we deduce that for (33) to be a stationary solution, i.e. $\frac{\mathrm{d}}{\mathrm{d}t}\langle \boldsymbol{q} \rangle = \boldsymbol{0}$ and $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Xi} = \boldsymbol{0}$, we must require

$$\langle \boldsymbol{q} \rangle = -\boldsymbol{L}^{-1}\boldsymbol{K} \qquad \tfrac{1}{2}\{\boldsymbol{L}\boldsymbol{\Xi} + (\boldsymbol{L}\boldsymbol{\Xi})^\dagger\} = \boldsymbol{D}. \tag{38}$$

For such stationary states the probability current (37) reduces to

$$\boldsymbol{J}(\boldsymbol{q}) = P(\boldsymbol{q})[\boldsymbol{D}\boldsymbol{\Xi}^{-1} - \boldsymbol{L}](\boldsymbol{q} + \boldsymbol{L}^{-1}\boldsymbol{K}). \tag{39}$$

We conclude that detailed balance, i.e. a vanishing current, requires in addition to (38) that

$$\boldsymbol{L}\boldsymbol{\Xi} = \boldsymbol{D}. \tag{40}$$

Combination of (38) and (40) leads to the condition $\boldsymbol{D}\boldsymbol{L}^{\dagger} = \boldsymbol{L}\boldsymbol{D}$, which, with identity (36), translates into $\boldsymbol{D}\boldsymbol{A}^{\dagger}\boldsymbol{D} = \boldsymbol{D}\boldsymbol{A}\boldsymbol{D}$. We now use the symmetry and non-negativity of the stationary diffusion matrix $\boldsymbol{D}$, i.e. $\boldsymbol{x}\cdot\boldsymbol{D}\boldsymbol{x} = \langle(\boldsymbol{x}\cdot\boldsymbol{\xi})^2[1-\tanh^2\beta[\boldsymbol{\xi}\cdot\boldsymbol{A}\boldsymbol{m}^* + \theta]]\rangle_{\boldsymbol{\xi},\theta} \geqslant 0$ (with $\boldsymbol{x}\cdot\boldsymbol{D}\boldsymbol{x} = 0$ only for $\boldsymbol{x} = \boldsymbol{0}$). We denote with $\{|n\rangle\}$ the orthogonal basis of normalized eigenvectors of $\boldsymbol{D}$, and with $\{d_n\}$ the corresponding (positive) eigenvalues. This allows us to derive from $\boldsymbol{D}\boldsymbol{A}^{\dagger}\boldsymbol{D} = \boldsymbol{D}\boldsymbol{A}\boldsymbol{D}$ that $\forall n, m:\ d_n d_m\langle n|[\boldsymbol{A}^{\dagger} - \boldsymbol{A}]|m\rangle = 0$. This implies that $\boldsymbol{A} = \boldsymbol{A}^{\dagger}$, which thus is found to be not only a sufficient condition, but also a necessary condition for a stationary solution of equation (18) to obey detailed balance.

4. Application to associative memories

Our first application is an associative memory model, which generalizes the standard model of [1] by allowing for patterns to be stored with different embedding strengths: $J_{ij} = \frac{1}{N}[1-\Delta\delta_{ij}]\sum_{\mu} w_{\mu}\xi_i^{\mu}\xi_j^{\mu}$ (with $0 < w_{\mu} \leqslant 1$ for all μ). This model, due to [18], corresponds to the choice $A_{\mu\nu} = w_{\mu}\delta_{\mu\nu}$ in the language of (4), and thus obeys detailed balance. For simplicity we choose zero thresholds, i.e. $W(\theta) = \delta(\theta)$. We will only study finite size corrections to the so-called 'pure states', where $m^*(t) = m(t)\delta_{\lambda\mu}$, which are the most important macroscopic solutions from both a thermodynamic and an information processing point of view. Without loss of generality we can choose $m(0) > 0$ and $\lambda = 1$ (as long as we refrain from ordering the embedding strengths with respect to magnitude), so

$$m_{\mu}^*(t) = m^*(t)\delta_{\mu,1} \qquad \frac{\mathrm{d}}{\mathrm{d}t}m^*(t) = \tanh\beta[w_1 m^*(t)] - m^*(t). \tag{41}$$

The above mean-field equation (41) will always evolve towards a fixed point, given by the solution of $m^* = \tanh\beta[w_1 m^*]$. Above the critical temperature $T_c = w_1$ the macroscopic fixed point is paramagnetic, i.e. $m^* = 0$, below T_c one finds an ordered state, i.e. $m^* > 0$, which represents retrieval of pattern one. Both fixed points, however, need not be stable against perturbations in the direction of non-nominated patterns [18]. We have to define initial conditions that will generate a pure macroscopic state, for which we choose

$$p_0(\boldsymbol{\sigma}) = \prod_i \{\tfrac{1}{2}[1+m(0)]\delta_{\sigma_i,\xi_i^1} + \tfrac{1}{2}[1-m(0)]\delta_{\sigma_i,-\xi_i^1}\}. \tag{42}$$

This indeed gives $m_{\mu}^*(0) = m(0)\delta_{\mu,1}$, as it should.

4.1. Statistics of finite-size effects

The restriction to 'pure' macroscopic states simplifies our finite size analysis considerably. The relevant objects in the Fokker–Planck equation (18) become

$$\boldsymbol{K}(t) = K(t)\boldsymbol{R} \qquad L_{\mu\nu}(t) = \ell_{\mu}(t)\delta_{\mu\nu} \qquad \boldsymbol{D}(t) = D(t)\boldsymbol{I}$$

with the scalar functions

$$K(t) = -\tanh\beta[w_1 m^*(t)] \qquad \ell_{\mu}(t) = 1 - \beta w_{\mu}[1-\tanh^2\beta[w_1 m^*(t)]] \tag{43}$$

$$D(t) = 1 - \tanh\beta[w_1 m^*(t)]\left\{\mathrm{e}^{-t}m(0) + \int_0^t \mathrm{d}s\ \mathrm{e}^{s-t}\tanh\beta[w_1 m^*(s)]\right\} \tag{44}$$

and with the (stationary) vector $\boldsymbol{R}$, defined as

$$R_1 = 0 \qquad R_{\mu>1} = \lim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_i \xi_i^{\mu}\xi_i^1. \tag{45}$$

The propagator (32) now becomes trivial:

$$G_{\mu\nu}(t) = \delta_{\mu\nu} \mathrm{e}^{-\int_0^t \mathrm{d}s\, \ell_\mu(s)}. \tag{46}$$

This will enable us to calculate the moments of the distribution $\mathcal{P}_t(\boldsymbol{q})$ at any time explicitly. Since (42) describes statistically independent initial components σ_i, the initial distribution $\mathcal{P}_0(\boldsymbol{q})$ is Gaussian. From (42) it follows that $\langle \boldsymbol{q} \rangle_0 = m(0)\boldsymbol{R}$ and $\boldsymbol{\Xi}(0) = [1-m^2(0)]\boldsymbol{I}$. The moments at any time $t \geqslant 0$ then follow from (33), (34):

$$\langle q_\mu \rangle_t = R_\mu \left\{ m(0) \mathrm{e}^{-\int_0^t \mathrm{d}s\, \ell_\mu(s)} + \int_0^t \mathrm{d}s\, \mathrm{e}^{-\int_s^t \mathrm{d}u\, \ell_\mu(u)} \tanh \beta [w_1 m^*(s)] \right\} \tag{47}$$

$$\Xi_{\mu\nu}(t) = \delta_{\mu\nu} \Xi_\mu(t) \qquad \Xi_\mu(t) = [1-m^2(0)] \mathrm{e}^{-2\int_0^t \mathrm{d}s\, \ell_\mu(s)} + 2\int_0^t \mathrm{d}s\, \mathrm{e}^{-2\int_s^t \mathrm{d}s\, \ell_\mu(s)} D(s). \tag{48}$$

This gives the full distribution

$$\mathcal{P}_t(\boldsymbol{q}) = \left[\prod_\mu [2\pi\, \Xi_\mu(t)] \right]^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_\mu [q_\mu - \langle q_\mu \rangle_t]^2 / \Xi_\mu(t)} \tag{49}$$

which describes uncoupled fluctuations, together with 'frozen' finite size corrections to the overlaps corresponding to uncondensed patterns. The above results obviously break down when the propagator develops runaway solutions, which is likely to happen at phase transitions.

4.2. Near the ordering transition

We will inspect the behaviour of the finite-size effects in stationary states close to the phase transition separating the paramagnetic from the ordered state. For stationary states, where $m^*(t) = m^*$ for all $t \geqslant 0$ (with m^* given by the solution of $m^* = \tanh \beta [w_1 m^*]$), we have in the asymptotic region (i.e. for $t \to \infty$) $K(t) = K$, $\ell_\mu(t) = \ell_\mu$ and $D(t) = D$, with

$$K = -m^* \qquad \ell_\mu = 1 - \beta w_\mu [1 - (m^*)^2] \qquad D = 1 - (m^*)^2.$$

This gives

$$\langle q_\mu \rangle_\infty = \frac{m^* R_\mu}{1 - \beta w_\mu [1-(m^*)^2]} \qquad \Xi_\mu(\infty) = \frac{1-(m^*)^2}{1-\beta w_\mu [1-(m^*)^2]}.$$

In the paramagnetic state, i.e. $m^* = 0$, we thus find

$$\langle q_\mu \rangle_\infty = 0 \qquad \Xi_\mu(\infty) = \frac{T}{T - w_\mu}.$$

The fluctuations diverge for $T \downarrow T_c = \max_\nu w_\nu$, which indeed marks the temperature where the paramagnetic state destabilizes, in favour of a non-trivial pure state. In a non-trivial pure state, i.e. for $m^* > 0$, we find that both the asymptotic average $\langle q_\mu \rangle_\infty$ and the asymptotic variance $\Xi_\mu(\infty)$ for a fluctuation direction μ diverge when $\ell_\mu \to 0$. This again makes sense: the condition $\ell_\mu > 0$ for all $\mu > 1$ is the condition for the pure state corresponding to pattern one to be macroscopically stable, with $\ell_\mu = 0$ signalling destabilization of this pure state in favour of an alternative pure state $\mu > 1$ (see [18]). Expansion of the macroscopic fixed-point equation close to T_c, for that particular pure state which is the first

to order as the temperature is lowered (i.e. we now assume $\max_\nu w_\nu = w_1$ and $T_c = w_1$), gives

$$0 < \kappa = \frac{T_c - T}{T_c} \ll 1 \qquad \beta w_1 = 1 + \kappa + \mathcal{O}(\kappa^2) \qquad m^* = \sqrt{3\kappa} + \mathcal{O}(\kappa).$$

This allows us to expand the finite-size terms in powers of the rescaled distance κ from the critical temperature:

$$\langle q_1 \rangle_\infty = 0 \qquad \langle q_{\mu>1} \rangle_\infty = \frac{\sqrt{3\kappa} R_\mu w_1}{w_1 - w_\mu} + \mathcal{O}(\kappa)$$

$$\Xi_1(\infty) = \frac{1}{2\kappa} + \mathcal{O}(\kappa^0) \qquad \Xi_{\mu>1}(\infty) = \frac{w_1}{w_1 - w_\mu}.$$

The approach of the transition $T \uparrow T_c$ is signalled by diverging fluctuations, as it should.

Finally we test some of the above predictions against numerical simulations. The simplest model of our (symmetric) class is the one where all embedding strengths are equal: $w_\mu = 1$ for all μ. Here we know that at $T > 1$ the system will be paramagnetic, whereas at $T = 1$ a second-order thermodynamic transition occurs to a pure low-temperature state. If we denote the non-negative solution of the macroscopic fixed-point equation $m^* = \tanh \beta[m^*]$ by $m(T)$, with $m(T > 1) = 0$, we arrive at the following predictions:

μ condensed: $\qquad \langle q_\mu \rangle_\infty = 0 \qquad \langle q_\mu^2 \rangle_\infty - \langle q_\mu \rangle_\infty^2 = \dfrac{T[1 - m^2(T)]}{T - 1 + m^2(T)}$

μ uncondensed: $\qquad \langle q_\mu \rangle_\infty R_\mu^{-1} = \dfrac{T m(T)}{T - 1 + m^2(T)} \qquad \langle q_\mu^2 \rangle_\infty - \langle q_\mu \rangle_\infty^2 = \dfrac{T[1 - m^2(T)]}{T - 1 + m^2(T)}.$

In figure 1 we show these predicted equilibrium moments as functions of temperature, together with results from numerical simulations carried out for $N = 10\,000$ for $T \leqslant 0.6$ and $N = 50\,000$ for $T > 0.6$, in systems with three stored patterns ($p = 3$). Each point was obtained upon performing $n = 1000$ simulations (following different initializations). The accuracy in an observation which is averaged over n trials, $\langle q_\mu \rangle_n$, is given by the relation $\langle q_\mu \rangle_{n\to\infty} = \langle q_\mu \rangle_n [1 \pm \sqrt{\Xi_{\mu\mu}/n}]$. If we calculate the error bars for the data in figure 1, we find that these are negligible for $T \leqslant 0.8$, that for $T = 0.9$ they are of the order of 5% and that for $T = 0.95$ they increase to about 9%. As we can see, the agreement between theory and computer simulations is quite satisfactory.

4.3. Zero temperature

For $T \to 0$ (or $\beta \to \infty$) the mean-field equation for the amplitude of pure states reduces to $\frac{\mathrm{d}}{\mathrm{d}t} m(t) = 1 - m(t)$, given our convention $m(0) > 0$, with solution

$$m(t) = m(0)\mathrm{e}^{-t} + 1 - \mathrm{e}^{-t}. \tag{50}$$

Since $m(t) > 0$ for any time (i.e. we will always be away from the $T = 0$ discontinuities at $m = 0$), we may deal with the non-trivial terms in our problem by using

$$\lim_{\beta\to\infty} \beta[1 - \tanh^2 \beta[w_1 m(t)]] = w_1^{-1} \lim_{m\to m(t)} \lim_{\beta\to\infty} \frac{\mathrm{d}}{\mathrm{d}m} \tanh[\beta w_1 m] = 2w_1^{-1} \lim_{m\to m(t)} \delta(m) = 0.$$

With this identity we obtain for $T = 0$:

$$K(t) = -1 \qquad \ell_\mu(t) = 1 \qquad D(t) = \mathrm{e}^{-t}[1 - m(0)]$$

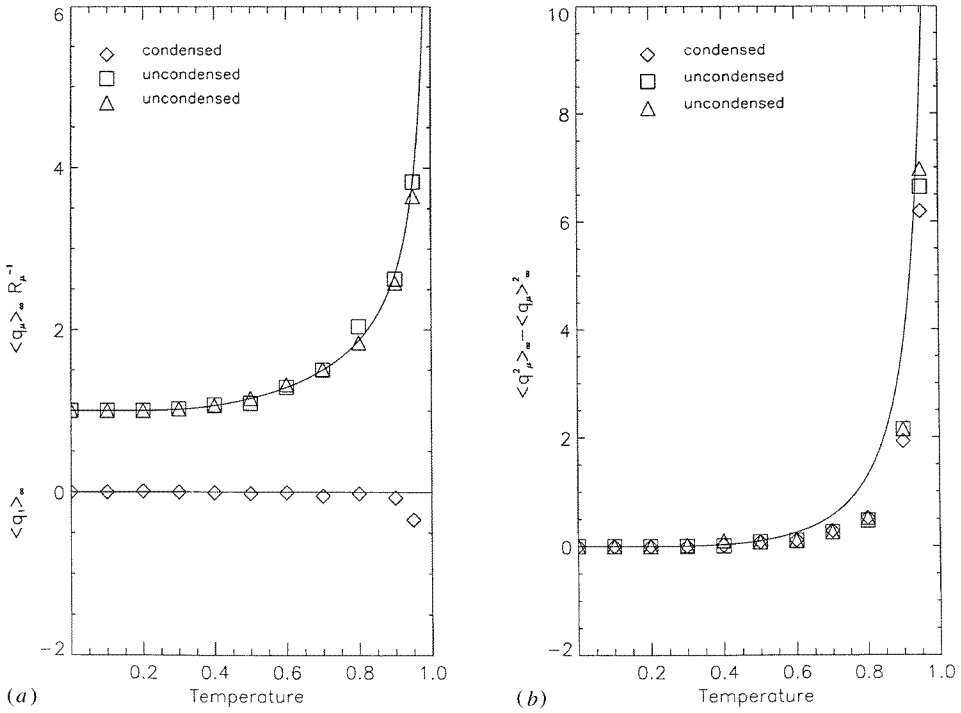

Figure 1. Comparison between the theoretical predictions for the moments of the finite-size contributions to the pattern overlaps and numerical simulations, in stationary states. (*a*) Normalized averages $\langle q_1\rangle$ and $\langle q_\mu\rangle_\infty R_\mu^{-1}$ ($\mu > 1$). (*b*) Variances $\langle q_\mu^2\rangle_\infty - \langle q_\mu\rangle_\infty^2$. Full curves: theoretical predictions. Markers: simulation results, for $N = 10\,000$ for $T \leqslant .6$ and $N = 50\,000$ for $T > 0.6$.

which, in turn, gives the simple propagator $\boldsymbol{G}(t) = \boldsymbol{I}\mathrm{e}^{-t}$ and the following moments for the finite-size corrections:

$$\mu \text{ condensed: } \langle q_\mu\rangle_t = 0 \qquad \langle q_\mu^2\rangle_t - \langle q_\mu\rangle_t^2 = [1-m(0)]\mathrm{e}^{-t}\{2-[1-m(0)]\mathrm{e}^{-t}\}$$

$$\mu \text{ uncondensed: } \langle q_\mu\rangle_t R_\mu^{-1} = m(t) \qquad \langle q_\mu^2\rangle_t - \langle q_\mu\rangle_t^2 = [1-m(0)]\mathrm{e}^{-t}\{2-[1-m(0)]\mathrm{e}^{-t}\}$$

with $m(t)$ given in (50). The 'frozen' correction to the mean-field laws, i.e. the term $\langle q_\mu\rangle_t$, increases in absolute strength as time progresses. The fluctuations, which at $T = 0$ have their origin purely in the randomness of the order of the single-spin updates, decrease to zero exponentially. In figure 2 we show these predicted zero-temperature moments as functions of time, together with results from numerical simulations carried out for $N = 5000$ in a system with three stored patterns. As it can be seen in figure 2(*b*), the largest standard deviation was found at $t = 0$, where the error bars of all $\langle q_\mu\rangle$ (figure 2(*a*)) are of the order of ± 0.02, which amounts to approximately 3% of the overlap of the uncondensed patterns. The error bars decrease at larger times. So, again the agreement between theory and computer simulations is quite satisfactory.

5. Application to non-equilibrium models

In this section we will apply our theory to non-symmetric systems, i.e. $A_{\mu\nu} \neq A_{\nu\mu}$, for which detailed balance does not hold. We will restrict ourselves to the case $p = 2$ and

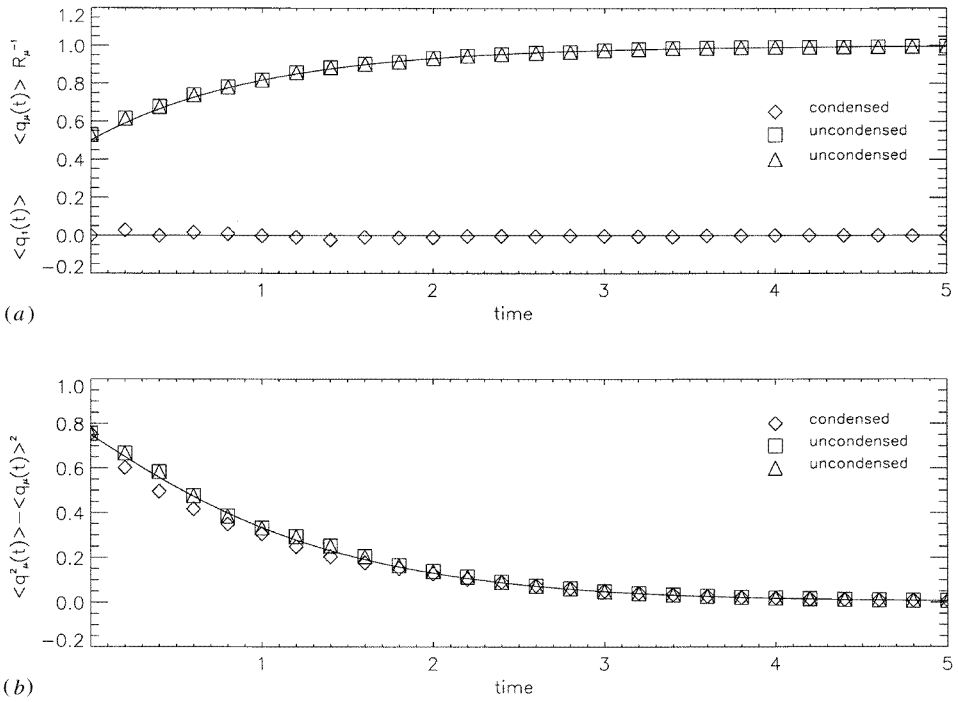

Figure 2. Comparison between the theoretical predictions for the moments of the finite-size contributions to the pattern overlaps and numerical simulations, at zero temperature, as functions of time. (*a*) Normalized averages $\langle q_1\rangle$ and $\langle q_\mu\rangle R_\mu^{-1}$ ($\mu > 1$). (*b*) Variances $\langle q_\mu^2\rangle - \langle q_\mu\rangle^2$. Full curves: theoretical predictions. Markers: simulation results, for $N = 5000$.

$W(\theta) = \delta(\theta)$ for simplicity; extensions to larger values of p and/or non-zero external fields are straightforward and are not likely to generate new physics. We define an initial microscopic distribution with statistically independent spins, in order to guarantee a Gaussian shape for $\mathcal{P}_0(\boldsymbol{q})$, given by

$$p_0(\boldsymbol{\sigma}) = \prod_i \{\tfrac{1}{2}[1 - m_1(0) - m_2(0)] + m_1(0)\delta_{\sigma_i,\xi_i^1} + m_2(0)\delta_{\sigma_i,\xi_i^2}\}. \tag{51}$$

Obviously, we have to restrict ourselves to the physical region, defined by the two conditions $|m_1(0) + m_2(0)| \leqslant 1$ and $|m_1(0) - m_2(0)| \leqslant 1$. Our definition generates the required initial macroscopic observables, $\lim_{N\to\infty}\langle m_\mu(\boldsymbol{\sigma})\rangle_0 = m_\mu(0)$ ($\mu = 1, 2$), and gives the following initial moments for the finite size variables:

$$\begin{pmatrix}\langle q_1\rangle_0\\ \langle q_2\rangle_0\end{pmatrix} = R\begin{pmatrix}m_2(0)\\ m_1(0)\end{pmatrix} \qquad \boldsymbol{\Xi}(0) = \begin{pmatrix}1-\boldsymbol{m}^2(0) & -2m_1(0)m_2(0)\\ -2m_1(0)m_2(0) & 1-\boldsymbol{m}^2(0)\end{pmatrix} \tag{52}$$

with $\boldsymbol{m}(0) = (m_1(0), m_2(0))$ and with $R = \lim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_i \xi_i^1\xi_i^2$.

5.1. Non-equilibrium stationary states

In this section we will study the class of networks where the matrix $\boldsymbol{A}$ has the form

$$\boldsymbol{A} = \begin{pmatrix}1 & \epsilon\\ 0 & 1\end{pmatrix}$$

with $\epsilon \geqslant 0$. These systems obey detailed balance only for $\epsilon = 0$. The mean-field equations, describing the overlap evolution in the $N \to \infty$ limit, are given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{m}^* = -\boldsymbol{m}^* + \frac{1}{2}\tanh\beta[m_1^* + (\epsilon+1)m_2^*]\begin{pmatrix}1\\1\end{pmatrix} + \frac{1}{2}\tanh\beta[m_1^* + (\epsilon-1)m_2^*]\begin{pmatrix}1\\-1\end{pmatrix}. \tag{53}$$

In the low-temperature regime these equations have two types of fixed points. First, for $T < 1$ one finds the two non-trivial fixed points $\boldsymbol{m}^* = \pm(m^*, 0)$, related only to pattern one, where m^* is the positive solution of $m^* = \tanh[\beta m^*]$. The second set of fixed points is related to pattern two. A more detailed analysis of the space of states of the system as a function of ϵ, and the behaviour of the system within the various basins of attraction will be published elsewhere [19]. We will study the finite-size effects for the pure fixed point $\boldsymbol{m}^* = (m^*, 0)$, and their dependence on the parameter ϵ which can be interpreted as measuring the degree of violation of detailed balance. To assess the macroscopic stability of this pure state we study the effect of perturbations: $\boldsymbol{m}^*(t) = (m^*, 0) + (\delta_1(t), \delta_2(t))$, with $|\delta_1(t)| \ll 1$ and $|\delta_2(t)| \ll 1$. Linearization of the mean-field laws gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\delta_1\\\delta_2\end{pmatrix} = \left\{[\beta[1-(m^*)^2]-1]\boldsymbol{I} + \beta\epsilon[1-(m^*)^2]\begin{pmatrix}0&1\\0&0\end{pmatrix}\right\}\begin{pmatrix}\delta_1\\\delta_2\end{pmatrix} + \cdots$$

with the solution

$$\begin{pmatrix}\delta_1(t)\\\delta_2(t)\end{pmatrix} = \mathrm{e}^{[\beta[1-(m^*)^2]-1]t}\begin{pmatrix}\delta_1(0) + \beta\epsilon t[1-(m^*)^2]\delta_2(0)\\\delta_2(0)\end{pmatrix}.$$

Therefore the pure states $\boldsymbol{m}^* = (m^*, 0)$ are (globally) stable if and only if $\beta[1-(m^*)^2]-1 < 0$. This condition is met by the solution of $m^* = \tanh[\beta m^*]$ as soon as it is non-zero, i.e. for all $T < 1$.

For finite N the mean-field picture will be modified by finite-size effects. In the pure fixed point $\boldsymbol{m}^* = (m^*, 0)$, obtained following a pure initialization $\boldsymbol{m}^*(0) = (m^*(0), 0)$, the stationary Ornstein–Uhlenbeck process is characterized by

$$\boldsymbol{L} = \boldsymbol{I} - \beta[1-(m^*)^2]\boldsymbol{A} \qquad \boldsymbol{D} = [1-(m^*)^2]\boldsymbol{I} \qquad \boldsymbol{K} = -m^* R\begin{pmatrix}0\\1\end{pmatrix}$$

with R given by equation (45) with $\mu = 2$. We work out the propagator $\boldsymbol{G}(t)$ (32) by splitting the matrix $\boldsymbol{L}$ into two commuting parts, such that $\boldsymbol{G}(t) = \exp[-t\boldsymbol{L}]$ factorizes into two separate matrix exponentiations:

$$\begin{aligned}\boldsymbol{G}(t) &= \exp\left\{-t[1-\beta[1-(m^*)^2]]\boldsymbol{I} + \epsilon\beta t[1-(m^*)^2]\begin{pmatrix}0&1\\0&0\end{pmatrix}\right\}\\ &= \mathrm{e}^{-t[1-\beta[1-(m^*)^2]]}\left[\boldsymbol{I} + \epsilon\beta t[1-(m^*)^2]\begin{pmatrix}0&1\\0&0\end{pmatrix}\right].\end{aligned}$$

The condition for the propagator to be well behaved is identical to the condition for the fixed point under consideration to be macroscopically stable: $\beta[1-(m^*)^2] < 1$. By working out (33) and (34) we arrive at the moments of the distribution $\mathcal{P}_t(\boldsymbol{q})$:

$$\langle \boldsymbol{q}\rangle_t = m^*(0)R\mathrm{e}^{-t\boldsymbol{L}}\begin{pmatrix}0\\1\end{pmatrix} + m^* R\boldsymbol{L}^{-1}[\boldsymbol{I} - \mathrm{e}^{-t\boldsymbol{L}}]\begin{pmatrix}0\\1\end{pmatrix}$$

$$\begin{aligned}\boldsymbol{\Xi}(t) = \mathrm{e}^{-t\boldsymbol{L}}&\begin{pmatrix}1-\boldsymbol{m}^2(0) & -2m_1(0)m_2(0)\\ -2m_1(0)m_2(0) & 1-\boldsymbol{m}^2(0)\end{pmatrix}\mathrm{e}^{-t\boldsymbol{L}^\dagger}\\ &+2[1-(m^*)^2]\int_0^t \mathrm{d}s\,\mathrm{e}^{(s-t)\boldsymbol{L}}\mathrm{e}^{(s-t)\boldsymbol{L}^\dagger}.\end{aligned}$$

The limits $t \to \infty$ are given by:

$$\langle \boldsymbol{q} \rangle_\infty = m^* R \left\{ [1-\beta[1-(m^*)^2]]\boldsymbol{I} - \epsilon\beta[1-(m^*)^2] \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$= \frac{m^* R}{[1-\beta[1-(m^*)^2]]^2} \begin{pmatrix} \epsilon\beta[1-(m^*)^2] \\ 1-\beta[1-(m^*)^2] \end{pmatrix} \tag{54}$$

$$\boldsymbol{\Xi}(\infty) = 2[1-(m^*)^2] \lim_{t\to\infty} \int_0^t \mathrm{d}s\, \mathrm{e}^{2(s-t)[1-\beta[1-(m^*)^2]]} \left\{ \boldsymbol{I} - \epsilon\beta(s-t)[1-(m^*)^2] \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right.$$
$$\left. + \epsilon^2\beta^2(s-t)^2[1-(m^*)^2]^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$
$$= H(T) \begin{pmatrix} 1+\frac{1}{2}\epsilon^2\beta^2 H^2(T) & \frac{1}{2}\epsilon\beta H(T) \\ \frac{1}{2}\epsilon\beta H(T) & 1 \end{pmatrix} \tag{55}$$

in which

$$H(T) = \frac{1-(m^*)^2}{1-\beta[1-(m^*)^2]} \qquad H(0)=0,\ \ H(1)=\infty.$$

Note that $H(T>1) = T/(T-1)$. Apart from inducing a non-zero stationary correction $\langle q_1 \rangle_\infty$ to the overlap with pattern one, violation of detailed balance (i.e. having $\epsilon > 0$ rather than $\epsilon = 0$) leads to an increase in the fluctuations of the non-trivial overlap, and a coupling of the fluctuations in the q_1 and q_2 directions, which in the case of detailed balance would have been statistically independent.

We can appreciate most clearly the effects of the correlations in the fluctuations by examining the curl of the probability current $\boldsymbol{J}(\boldsymbol{q})$ in $\boldsymbol{q}$-space. In doing so we can use the stationarity condition (38) for the correlation matrix, i.e. $\frac{1}{2}[\boldsymbol{L}\boldsymbol{\Xi} + (\boldsymbol{L}\boldsymbol{\Xi})^\dagger] = \boldsymbol{D}$ (which allows us to put $\boldsymbol{x}\cdot[\boldsymbol{D}-\boldsymbol{L}\boldsymbol{\Xi}]\boldsymbol{x} = 0$ for each $\boldsymbol{x} \in \mathbb{R}^2$), as well as the symmetry of both the correlation matrix $\boldsymbol{\Xi}$ and its inverse. For stationary states the probability current (37), which must be divergence-free, reduces to

$$\boldsymbol{J}(\boldsymbol{q}) = \mathcal{P}(\boldsymbol{q})[\boldsymbol{D}\boldsymbol{\Xi}^{-1} - \boldsymbol{L}](\boldsymbol{q} - \langle \boldsymbol{q} \rangle_\infty).$$

Its curl is found to be

$$\boldsymbol{\nabla} \times \boldsymbol{J}(\boldsymbol{q}) = \mathcal{P}(\boldsymbol{q})\{\boldsymbol{\nabla} \times [\boldsymbol{D}\boldsymbol{\Xi}^{-1} - \boldsymbol{L}](\boldsymbol{q} - \langle \boldsymbol{q} \rangle_\infty) - (\boldsymbol{q} - \langle \boldsymbol{q} \rangle_\infty) \cdot (\boldsymbol{\Xi}^{-1})^\dagger$$
$$\times [\boldsymbol{D} - \boldsymbol{L}\boldsymbol{\Xi}]\boldsymbol{\Xi}^{-1}(\boldsymbol{q} - \langle \boldsymbol{q} \rangle_\infty)\}$$
$$= \mathcal{P}(\boldsymbol{q})\boldsymbol{\nabla} \times \{[\boldsymbol{D}\boldsymbol{\Xi}^{-1} - \boldsymbol{L}](\boldsymbol{q} - \langle \boldsymbol{q} \rangle_\infty)\}$$
$$= -\epsilon\beta[1-(m^*)^2]\, \mathcal{P}(\boldsymbol{q}).$$

Violation of detailed balance, due to the asymmetry of the matrix $\boldsymbol{A}$ for $\epsilon \neq 0$, produces a stationary rotational current in the space of the finite-size variables $\boldsymbol{q}$. The magnitude of this current is proportional to the magnitude of the parameter ϵ. For $T < 1$ the prefactor $\beta[1-(m^*)^2]$ is a monotonically increasing function of temperature, starting at zero for $T = 0$ and approaching one for $T \to 1$. The rotational current persists above the critical temperature, where we find $\boldsymbol{\nabla} \times \boldsymbol{J}(\boldsymbol{q}) = -\epsilon\beta\mathcal{P}(\boldsymbol{q})$.

Finally we test the predictions (54), (55) for the moments of $\mathcal{P}_\infty(\boldsymbol{q})$ against numerical simulations. In figure 3 we show these predicted equilibrium moments as functions of temperature, together with results from numerical simulations carried out for $N = 50\,000$ and $\epsilon \in \{0, 0.2, 0.5\}$. We performed $n = 10\,000$ simulation experiments in all cases, therefore the observation errors in the observed fluctuation average are given by one

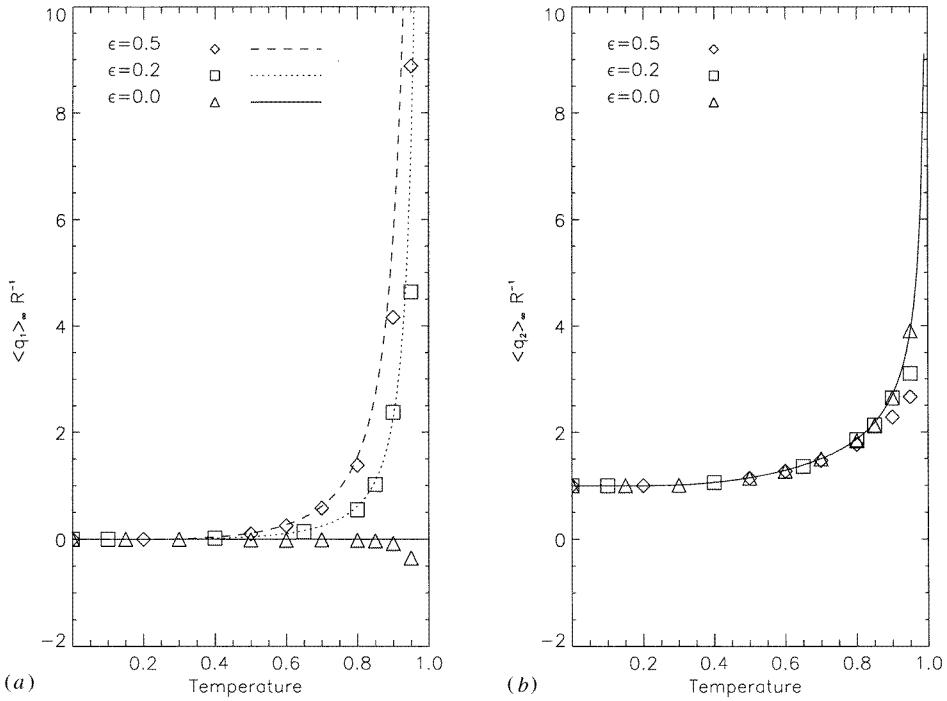

Figure 3. Comparison between the theoretical predictions for the first-order moments of the finite-size contributions to the pattern overlaps and numerical simulations, in stationary states, as functions of temperature and for $\epsilon \in \{0, 0.2, 0.5\}$. (*a*) Normalized average $\langle q_1 \rangle_\infty R^{-1}$. (*b*) Normalized average $\langle q_2 \rangle_\infty R^{-1}$. Full curves: theoretical predictions. Markers: simulation results, for $N = 50\,000$ and $n = 10\,000$ samples.

hundredth of the observed standard deviation times the fluctuation average. The present non-equilibrium model is found to require larger system sizes for our fluctuation theory to hold (i.e. neglected higher orders in $N^{-\frac{1}{2}}$ are more prominent) than the equilibrium models studied in earlier sections. In addition the time required for transient effects to have died out is longer. This explains why the agreement between theory and experiment, as observed in figure 3, although still reasonable, is less than that observed in our previous simulations.

5.2. Escape times controlled by system size

As a final application of our finite-size theory we turn to a model which at $T = 0$ is exactly critical, in the sense that the mean-field flow is such that the asymptotic value $\boldsymbol{m}^*(\infty)$ of the overlap vector is exactly *on* a regional boundary in $\boldsymbol{m}$-space which separates qualitatively different macroscopic flow domains:

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

As a result one finds in this particular system that the relevant escape times, which dictate whether and when the state vector can leave a given domain, are controlled entirely by the size N of the system.

At non-zero noise levels $0 < T < 1$ the solutions of the mean-field equations for $\boldsymbol{m}^*$, describing the overlap evolution in the $N \to \infty$ limit, show evolution into a stable

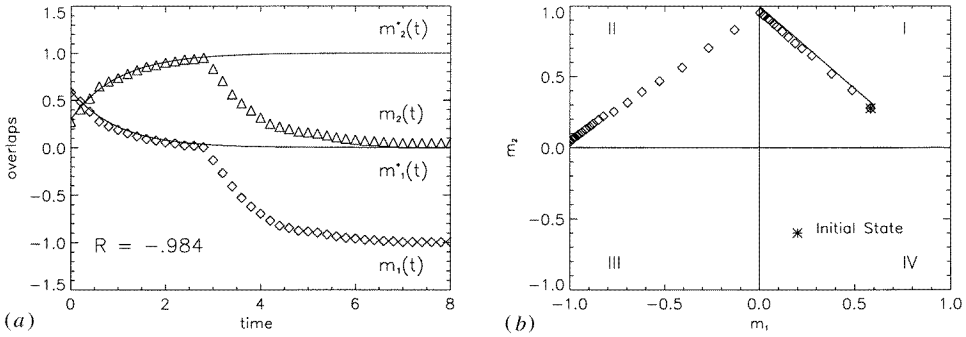

Figure 4. (*a*) Time evolution of the order parameters $m_1(t)$ (squares) and $m_2(t)$ (triangles). For $R < 0$ the system will eventually cross the $m_1 = 0$ line and jump into the neighbouring basin of attraction (region II). (*b*) In overlap space the system initially follows a straight line from any initial state in region I towards the fixed point $(0, 1)$, slowing down exponentially. If $R < 0$, the finite system (squares) eventually jumps to region II and the flow is redirected towards the point $(-1, 0)$.

limit-cycle [10]. For $T = 0$ the mean-field equations reduce to

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{m}^* = \frac{1}{2}\,\mathrm{sgn}[m_1^*]\begin{pmatrix}1\\1\end{pmatrix} + \frac{1}{2}\,\mathrm{sgn}[m_2^*]\begin{pmatrix}-1\\1\end{pmatrix} - \boldsymbol{m}^*$$

giving

$$\begin{array}{lll}
\text{region I} & m_1^*(0) > 0,\ m_2^*(0) > 0 & \boldsymbol{m}^*(t) = \boldsymbol{m}^*(0)\mathrm{e}^{-t} + (0,1)[1-\mathrm{e}^{-t}] \\
\text{region II} & m_1^*(0) < 0,\ m_2^*(0) > 0 & \boldsymbol{m}^*(t) = \boldsymbol{m}^*(0)\mathrm{e}^{-t} + (-1,0)[1-\mathrm{e}^{-t}] \\
\text{region III} & m_1^*(0) < 0,\ m_2^*(0) < 0 & \boldsymbol{m}^*(t) = \boldsymbol{m}^*(0)\mathrm{e}^{-t} + (0,-1)[1-\mathrm{e}^{-t}] \\
\text{region IV} & m_1^*(0) > 0,\ m_2^*(0) < 0 & \boldsymbol{m}^*(t) = \boldsymbol{m}^*(0)\mathrm{e}^{-t} + (1,0)[1-\mathrm{e}^{-t}].
\end{array} \tag{56}$$

There are four qualitatively different macroscopic flow regions, separated by the two lines $m_1^* = 0$ and $m_2^* = 0$. In all four cases the macroscopic flow is directed towards a state which is exactly *at* the regional boundary, such that the asymptotics of the system (i.e. whether or not the state vector will escape to another region) will be determined purely by the finite-size effects. Figure 4 illustrates the time evolution of the escape process from region I to region II. Data obtained by simulation of a single finite system (markers) are compared with the $N \to \infty$ theory (continuous lines); for $N \to \infty$ or $R > 0$ the system gets trapped in region I.

Due to the overall symmetry of our models with respect to the transformation $\forall \boldsymbol{\sigma}$: $p_t(\boldsymbol{\sigma}) \to p_t(-\boldsymbol{\sigma})$ (at least: in the absence of external fields), the properties of regions I and III and of regions II and IV are pairwise identical. Furthermore, the properties of region II follow from those of region I via the transformation $\boldsymbol{\xi}^1 \to -\boldsymbol{\xi}^1$, and the properties of region IV follow from those of region III via the transformation $\boldsymbol{\xi}^1 \to -\boldsymbol{\xi}^1$. This implies that without loss of generality we can restrict our quantitative analysis to region I. We choose $T = 0$ and the initial state $\boldsymbol{m}^*(0) = (m^*(0), 0)$, with $0 < m^*(0) < 1$ (i.e. in region I). The relevant quantities in the Fokker–Planck equation (18) are then given by

$$\boldsymbol{K}(t) = -\begin{pmatrix}R\\0\end{pmatrix} \qquad \boldsymbol{L}(t) = \boldsymbol{I} \qquad \boldsymbol{D}(t) = \mathrm{e}^{-t}\begin{pmatrix}1 & -m^*(0)\\ -m^*(0) & 1\end{pmatrix}$$

with R given by equation (45) with $\mu = 2$. We find the simple propagator $\boldsymbol{G}(t) = \mathrm{e}^{-t}\boldsymbol{I}$. Note that working out the relevant moments (33), (34) of $\mathcal{P}_t(\boldsymbol{q})$ for the present model,

following the initial conditions (52), gives

$$\langle \boldsymbol{q}\rangle_t = \mathrm{e}^{-t} R m^*(0) \begin{pmatrix} 0 \\ 1 \end{pmatrix} + [1-\mathrm{e}^{-t}] R \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\boldsymbol{\Xi}(t) = \mathrm{e}^{-2t} \begin{pmatrix} 1-(m^*(0))^2 & 0 \\ 0 & 1-(m^*(0))^2 \end{pmatrix} + 2\mathrm{e}^{-t}[1-\mathrm{e}^{-t}] \begin{pmatrix} 1 & -m^*(0) \\ -m^*(0) & 1 \end{pmatrix}.$$

We want to calculate the probability that at time t the system will have escaped from region I to region II. To this aim we first define

$$\Pi_t(M) = \mathrm{Prob}[m_1(\boldsymbol{\sigma}(t)) < M] = \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\theta[M - m_1(\boldsymbol{\sigma})]. \tag{57}$$

Note that in both regions I and II we have $m_2(\boldsymbol{\sigma}) > 0$ (in fact the escape process I→II happens close to $m_2(\boldsymbol{\sigma}) = 1$). The time derivative of $\Pi_t(M)$ follows from (5), which for $T = 0$ reduces to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Pi_t(M) &= \left\langle \left\{ M - \frac{1}{N}\sum_i \mathrm{sgn}[M - m_2(\boldsymbol{\sigma}) + \xi_i^1\xi_i^2[M + m_2(\boldsymbol{\sigma})]] \right\} \delta[M - m_1(\boldsymbol{\sigma})] \right\rangle_t \\ &\quad + \mathcal{O}(N^{-1}) \\ &= \left\langle \left\{ M + \theta[-M] - \frac{R}{\sqrt{N}}\theta[M] \right\} \delta[M - m_1(\boldsymbol{\sigma})] \right\rangle_t + \mathcal{O}(N^{-1})\end{aligned}$$

with the usual definition $R = \frac{1}{\sqrt{N}}\sum_i \xi_i^1\xi_i^2$. It follows that $\frac{\mathrm{d}}{\mathrm{d}t}\Pi_t(M)$ is discontinuous at $M = 0$:

$$\begin{aligned}\lim_{M\to 0_+} \frac{\mathrm{d}}{\mathrm{d}t}\Pi_t(M) &= -\frac{R}{\sqrt{N}}\langle \delta[m_1(\sigma)]\rangle_t + \mathcal{O}(N^{-1}) \\ \lim_{M\to 0_-} \frac{\mathrm{d}}{\mathrm{d}t}\Pi_t(M) &= \langle \delta[m_1(\sigma)]\rangle_t + \mathcal{O}(N^{-1}).\end{aligned}$$

The escape process requires times sufficiently large to allow finite-size effects to come into play, i.e. $\mathrm{e}^{-t} = \mathcal{O}(N^{-\frac{1}{2}})$. We are thus led to the introduction of the new time variable†: $\gamma = \mathrm{e}^t/\sqrt{N}$. For such times the average $\langle \delta[m_1(\sigma)]\rangle_t$ can be written as

$$\begin{aligned}\langle \delta[m_1(\sigma)]\rangle_t &= \int \mathrm{d}\boldsymbol{q}\, \mathcal{P}_t(\boldsymbol{q})\delta[m^*(0)\mathrm{e}^{-t} + \frac{1}{\sqrt{N}}q_1 + \mathcal{O}(N^{-1})] \\ &= \frac{\sqrt{N}}{\sqrt{2\pi\,\Xi_{11}}}\mathrm{e}^{-\frac{1}{2}[\langle q_1\rangle_t + \frac{m^*(0)}{\gamma} + \mathcal{O}(N^{-\frac{1}{2}})]^2/\Xi_{11}(t)}.\end{aligned}$$

In terms of the new time variable we write $\Pi_t(M) = P_\gamma(M)$; using $\frac{\mathrm{d}}{\mathrm{d}t} = \gamma\frac{\mathrm{d}}{\mathrm{d}\gamma}$ we then arrive at

$$\begin{aligned}\lim_{M\to 0_+} \frac{\mathrm{d}}{\mathrm{d}\gamma}P_\gamma(M) &= -\frac{R}{\gamma\sqrt{2\pi\,\Xi_{11}}}\mathrm{e}^{-\frac{1}{2}[\langle q_1\rangle_t + \frac{m^*(0)}{\gamma} + \mathcal{O}(N^{-\frac{1}{2}})]^2/\Xi_{11}(t)} + \mathcal{O}(N^{-1}) \\ \lim_{M\to 0_-} \frac{\mathrm{d}}{\mathrm{d}\gamma}P_\gamma(M) &= \frac{\sqrt{N}}{\gamma\sqrt{2\pi\,\Xi_{11}}}\mathrm{e}^{-\frac{1}{2}[\langle q_1\rangle_t + \frac{m^*(0)}{\gamma} + \mathcal{O}(N^{-\frac{1}{2}})]^2/\Xi_{11}(t)} + \mathcal{O}(N^0).\end{aligned}$$

† Here there could be a potential conflict with the assumptions of the theory. However, inspection shows that our derivation of the Fokker–Planck equation as the correct description of the leading-order finite-size effects requires $\lim_{N\to\infty} t/\sqrt{N} = 0$, which means that the theory still applies if $t = \mathcal{O}(\log(N))$.

Note that on the γ timescales $\langle q_1\rangle_t = R + \mathcal{O}(N^{-\frac{1}{2}})$ and $\Xi_{11}(t) = \mathcal{O}(N^{-\frac{1}{2}})$. Consequently

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}P_\gamma(0_+) = -\frac{R}{\gamma}\delta\left[R + \frac{m^*(0)}{\gamma}\right] + \mathcal{O}(N^{-\frac{1}{2}})$$
$$\frac{\mathrm{d}}{\mathrm{d}\gamma}P_\gamma(0_-) = \frac{\sqrt{N}}{\gamma}\delta\left[R + \frac{m^*(0)}{\gamma}\right] + \mathcal{O}(N^0).$$

Since $P_0(0_\pm) = 0$ (for we start with $m_1^*(0) > 0$) we are led to the following predictions. If $R > 0$ the state vector never escapes region I; if $R < 0$, on the other hand, the state vector will ultimately escape. Once past the regional boundary, the state vector cannot return due to the boost described by $P_\gamma(0_-)$. Integration over γ gives the explicit form

$$P_\gamma(0^+) = -R\int_{m^*(0)/\gamma}^{\infty}\frac{\mathrm{d}z}{z}\delta[R+z] + \mathcal{O}(N^{-\frac{1}{2}})$$
$$= \theta[-R]\,\theta\left[\gamma - \frac{m^*(0)}{|R|}\right] + \mathcal{O}(N^{-\frac{1}{2}}).$$

So, provided $R < 0$, the state vector leaves region I precisely when $\gamma = m^*(0)/|R|$. Translation back into the original time variable gives the following escape time:

$$t_{\mathrm{esc}} = \frac{1}{2}\log N + \log\left[\frac{m^*(0)}{|R|}\right] + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{58}$$

If we use the simple transformations that relate the properties of the four regions, we arrive at the following picture. If $R < 0$ the system will be able to make the transitions I $\to$ II and III $\to$ IV, but never the transitions II $\to$ III or IV $\to$ I. For $R > 0$, on the other hand, the transitions II $\to$ III and IV $\to$ I will be observed, but never the transitions I $\to$ II or III $\to$ IV. When $R = 0$, the case where the two patterns are orthogonal in the first two leading orders in N, the escape properties will be controlled by the $\mathcal{O}(N^{-1})$ finite size effects.

Finally we compare the prediction (58) for the escape time with the results of numerical simulations. Figure 5 shows the average escape time as a function of $\log[m^*(0)/|R|]$, where the broken curve corresponds to the order of magnitude of neglected orders ($\pm 1/\sqrt{N}$). The agreement between theory and computer simulations is quite satisfactory.

6. Discussion

We have performed a systematic study of finite-size effects in separable recurrent neural network models away from saturation. Since our approach is based on analysis of the dynamics, our results apply to models with detailed balance (i.e. with symmetric synaptic interactions) and to models without detailed balance (with non-symmetric synaptic interactions). In leading order in the system size ($N^{-\frac{1}{2}}$) the finite-size effects turn out to be governed by a time-dependent Ornstein–Uhlenbeck process, and their time-dependent probability density can be calculated in explicit form. The leading-order finite-size effects are found to come in two distinct forms: they show up as 'frozen' corrections to the mean field laws (dependent on the details of the correlations between the randomly drawn stored patterns) and as fluctuations, which have their origin in thermal noise in the local field alignment as well as the randomness in the selection of the neuron to be updated.

We use our theory to work out several specific but characteristic examples, including symmetric attractor neural network models, used as associative memories, and non-equilibrium models (with non-symmetric interactions). For detailed balance models

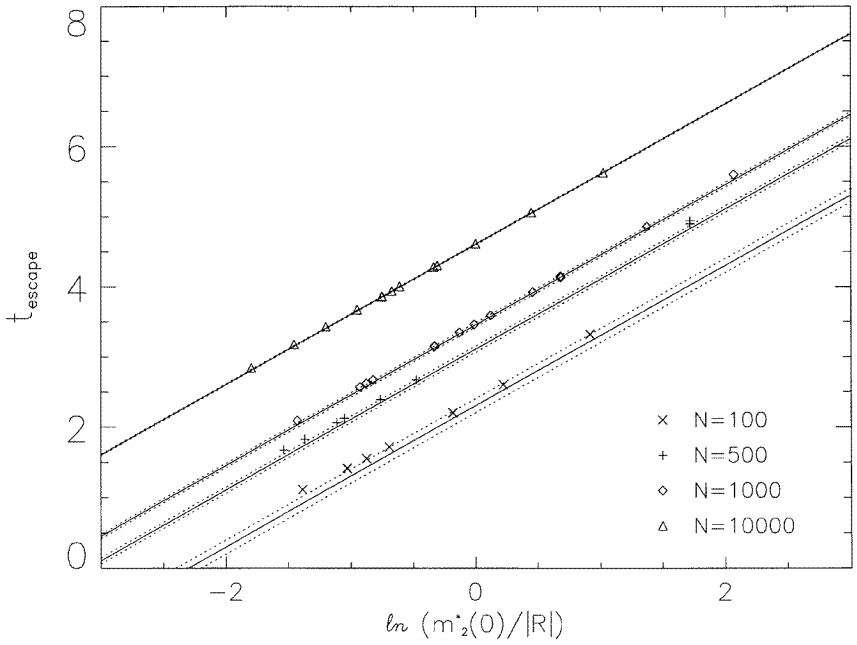

Figure 5. Comparison between the theoretical prediction (58) for the escape time from the initial region and numerical simulations for $n = 10\,000$ systems in each case. Full curves: leading two orders according to the theory. Broken curves: indication of the potential magnitude of subsequent (neglected) orders ($\pm N^{-\frac{1}{2}}$). Markers: simulation results.

we quantify within our fluctuation theory the familiar features of equilibrium statistical mechanics, e.g. diverging fluctuations near phase transitions and absent probability currents in the stationary state. For non-equilibrium models, in contrast, we find persistent rotational currents in the stationary state. One of our non-equilibrium examples involves the calculation of escape times which are purely controlled by finite-size effects, which is a nice example of a problem where the finite-size effects are significantly more than simply a correction to the corresponding result for an infinite system. More extensive applications will be published in [20]. Comparison with extensive numerical simulations confirms the theoretical predictions in all cases.

Acknowledgments

ACCC would like to thank the UNAM Institute of Physics in Ensenada (now Centro de Ciencias de la Materia Condensada) for their kind hospitality. AC is grateful to the Consejo Nacional de Ciencia y Tecnología México (CONACYT) for a fellowship no 85630. This work was partially supported by grant DGAPA IN100895 (UNAM).

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
[2] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev.* A **32** 1007
[3] Domany E, van Hemmen J L and Schulten K (ed) 1991 *Models of Neural Networks* vol I (Berlin: Springer)

[4] Coolen A C C and Sherrington D 1993 *Mathematical Approaches to Neural Networks* ed J G Taylor (Amsterdam: North-Holland) p 293
[5] Kohring G A 1990 *J. Phys. A: Math. Gen.* **23** 2237
[6] Ohira T and Cowan J D 1996 *Mathematics of Neural Networks* ed S W Ellacott, J C Mason and I J Anderson (London: Kluwer) p 290
[7] Horner H, Bormann D, Frick M, Kinzelbach H and Schmidt A 1989 *Z. Phys.* B **76** 381
[8] Rieger H, Schreckenberg M and Zittartz J 1988 *Z. Phys.* B **72** 523
[9] Coolen A C C, Laughton S N and Sherrington D 1996 *Phys. Rev.* B **53** 8184
[10] Coolen A C C and Ruijgrok Th W 1988 *Phys. Rev.* A **38** 4253
[11] Cugliandolo L F and Tsodyks M V 1994 *J. Phys. A: Math. Gen.* **27** 741
[12] Laughton S N and Coolen A C C 1995 *Phys. Rev.* E **51** 2581
[13] Van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* (Amsterdam: North-Holland)
[14] Pawula R F 1967 *Phys. Rev.* **162** 186
[15] Gardiner C W 1990 *Handbook of Stochastic Methods* (Berlin: Springer)
[16] Laughton S N and Coolen A C C 1995 *J. Stat. Phys.* **80** 375
[17] Coolen A C C 1997 *Statistical Mechanics of Neural Networks* (London: King's College Lecture Notes)
[18] Viana L 1988 *J. Physique* **49** 167
[19] Castellanos A and Viana L in preparation
[20] Castellanos A *PhD Thesis* CICESE, México, in preparation