Journal of Statistical Physics, Vol. 80, Nos. 1/2, 1995

Macroscopic Lyapunov Functions for Separable Stochastic Neural Networks with Detailed Balance

S. N. Laughton[1] and A. C. C. Coolen[1]

Received November 1, 1993; final January 18, 1995

We derive macroscopic Lyapunov functions for large, long-range, Ising-spin neural networks with separable symmetric interactions, which evolve in time according to local field alignment. We generalize existing constructions, which correspond to *deterministic* (zero-temperature) evolution and to specific choices of the interaction structure, to the case of *stochastic* evolution and arbitrary separable interaction matrices, for both parallel and sequential spin updating. We find a direct relation between the form of the Lyapunov functions (which describe dynamical processes) and the saddle-point integration that results from performing equilibrium statistical mechanical studies of the present type of model.

KEY WORDS: Neural networks; stochastic dynamics; Lyapunov functions.

1. INTRODUCTION

Long-range stochastic Ising spin models with and without detailed balance have come to play an increasingly important role in the study of information processing in neural networks. In such models spins represent neurons, interaction strengths between spins represent synaptic efficacies, and local (magnetic) alignment fields play the role of postsynaptic potentials. For a more general introduction to this field we refer to recent textbooks[1-3] or review papers.[4-7]

Since biological realism forces one to try to abandon the symmetry of the neural interactions, and since in the usual type of model interaction symmetry (in turn) is equivalent to having detailed balance (apart from some trivial exceptions), studying the dynamics directly is often the only route toward analytical results. For long-range models with separable

[1] Department of Physics-Theoretical Physics, University of Oxford, Oxford OX1 3NP, U.K.

interactions (with or without detailed balance) it turns out that in the thermodynamic limit and on finite time scales one can derive deterministic dynamical laws for a suitably chosen set of macroscopic order parameters.[8-10] The dynamics of such systems can now be studied at a macroscopic level by studying these laws, either in the form of coupled nonlinear mappings (for parallel microscopic dynamics) or in the form of coupled nonlinear differential equations (for sequential microscopic dynamics).

At present, apart from general (microscopic) results like the H-theorem,[11] the only Lyapunov functions that have been constructed for Ising-spin neural networks apply to the zero-temperature case,[12,13] to strictly positive-definite or negative-definite interaction matrices,[6,14] or to specific zero-temperature models with delays.[15] In this paper we address the problem of how to construct macroscopic Lyapunov functions corresponding to the dynamical laws which (in the thermodynamic limit) govern the evolution of order parameters for arbitrary finite stochastic noise levels and for both sequential and parallel microscopic dynamics. These macroscopic Lyapunov functions are expectation values of true state variables. We restrict ourselves to symmetric networks (i.e., with detailed balance). For both types of dynamics (sequential and parallel) we find a direct relation between the form of the macroscopic Lyapunov functions (which describe dynamical processes) and the saddle-point integration that results from performing equilibrium statistical mechanical studies of the present type of model. This emphasizes the equivalence of thermodynamic stability and dynamic stability previously demonstrated for the special case $\mathbf{A} = \mathbb{1}$ and sequential dynamics.[6] Our study clearly confirms the intuitive picture of visualizing the dynamics of symmetric neural networks as the minimization of some state variable, to be interpreted as a dynamic "free energy."

2. SEPARABLE STOCHASTIC NEURAL NETWORKS

We study systems of N Ising spins (or neurons) $s_i \in \{-1, +1\}$, $i \in \{1,..., N\}$, coupled by separable interaction matrices (or synaptic efficacies) given by the general form

$$J_{ij} = \frac{1}{N} \sum_{\mu,\nu=1}^{p} \xi_i^\mu A_{\mu\nu} \xi_j^\nu \tag{1}$$

For neural networks the vectors $\boldsymbol{\xi}^\mu = (\xi_1^\mu,..., \xi_N^\mu)$ represent stored items of information. The states of the neurons evolve according to a stochastic local field alignment with local fields given by

$$h_k(\mathbf{s}) = \sum_{i=1}^{N} J_{ki} s_i \tag{2}$$

We study two extreme cases: synchronous updating with discrete time steps, and asynchronous updating with time defined as a continuous variable.

We define the synchronous dynamics via a Markov process

$$p_{t+1}(\mathbf{s}) = \sum_{\mathbf{s}'} \omega(\mathbf{s}' \to \mathbf{s})\, p_t(\mathbf{s}')$$
$$\omega(\mathbf{s}' \to \mathbf{s}) = \frac{1}{2^N} \prod_{k=1}^{N} \{1 + \tanh[\beta s_k h_k(\mathbf{s}')]\} \tag{3}$$

where the vector $\mathbf{s} \equiv (s_1,..., s_N)$ denotes the microscopic states, and $\omega(\mathbf{s}' \to \mathbf{s})$ is the transition probability for the state change $\mathbf{s}' \to \mathbf{s}$. For asynchronous updating the transition rate is nonzero for single spin-flips only. In the latter case we define the dynamics via a master equation

$$\frac{d}{dt} p_t(\mathbf{s}) = \sum_{j=1}^{N} \{\omega_j(F_j\mathbf{s})\, p_t(F_j\mathbf{s}) - \omega_j(\mathbf{s})\, p_t(\mathbf{s})\}$$
$$\omega_k(\mathbf{s}) = \frac{1}{2} \{1 - \tanh[\beta s_k h_k(\mathbf{s})]\} \tag{4}$$

where $\omega_j(\mathbf{s})$ is the transition rate for the transition $\mathbf{s} \to F_j\mathbf{s}$ and F_j is the spin-flip operator $F_j\Phi(s_1,..., s_j,..., s_N) \equiv \Phi(s_1,..., -s_j,..., s_N)$. In both cases β is a measure of the stochasticity in the system, and plays the role of the inverse temperature.

If the number of patterns p is not too large $(p \ll \sqrt{N})$,[7] then in the thermodynamic limit, deterministic equations for the evolution in time of a set of macroscopic order parameters can be derived for both parallel[9,10] and sequential[8,9] dynamics. These equations take the form of a first-order nonlinear difference equation and a first-order nonlinear differential equation, respectively,

$$\mathbf{m}_{t+1} = \langle \boldsymbol{\xi} \tanh(\beta \boldsymbol{\xi} \cdot \mathbf{A}\mathbf{m}_t) \rangle \tag{5}$$

$$\frac{d}{dt} \mathbf{m} = \langle \boldsymbol{\xi} \tanh(\beta \boldsymbol{\xi} \cdot \mathbf{A}\mathbf{m}) \rangle - \mathbf{m} \tag{6}$$

The angular brackets denote an average over the random variables ξ_i^μ: $\langle F(\boldsymbol{\xi}) \rangle \equiv (1/N) \sum_{i=1}^{N} F(\boldsymbol{\xi}_i)$, $\boldsymbol{\xi}_i \equiv (\xi_i^1,..., \xi_i^p)$. For neural networks with randomly drawn patterns $\xi_i^\mu \in \{-1, +1\}$ with equal probabilities and, since p scales as less than $\sqrt{N}$, for $N \to \infty$, averaging a function of ξ_i over the sites i is equivalent to averaging over all 2^p possible configurations of $\xi \in \{-1, +1\}^p$:

$$\langle F(\xi) \rangle \equiv 2^{-p} \sum_{\xi \in \{-1, +1\}^p} F(\xi) \tag{7}$$

The order parameters, which evolve deterministically, are defined as

$$\mathbf{m} = (m^1, \ldots, m^p), \qquad m^\mu = \frac{1}{N} \sum_{i=1}^{N} s_i \xi_i^\mu \tag{8}$$

In neural networks they give the overlaps of the current system configuration with the stored patterns ξ^μ.

In the remaining sections of this paper, we will construct Lyapunov functions $\mathscr{L}(\mathbf{m})$ of the macroscopic variables evolving according to (5), (6), for arbitrary symmetric matrices $\mathbf{A}$. From the equations of motion (5), (6), we can see that $\mathbf{m}_t \in [-1, +1]^p$ for all $t \geqslant 0$. This is immediately obvious for the discrete-time case since $-1 \leqslant \langle \xi^\mu \tanh(\beta \xi \cdot \mathbf{A}\mathbf{m}_t) \rangle \leqslant +1$. For the continuous-time case, we can see for the same reason that each component of $\mathbf{m}$ will remain within $[-1, +1]$. We mention this despite the fact that due to its physical interpretation (8), $\mathbf{m} \in [-1, +1]^p$, since it is used in our subsequent analysis, which only builds on the equations of motion (5), (6).

3. THE H-THEOREM

The H-theorem guarantees the approach of the microscopic probability distribution $p_t(\mathbf{s})$ to an equilibrium probability distribution $p_e(\mathbf{s})$, provided it exists and is positive for each microscopic state $\mathbf{s}$, for any stochastic process defined by a continuous-time master equation which obeys detailed balance. It can be quite easily proved[11] that $\mathscr{L}_t \equiv \sum_{\mathbf{s}} p_e(\mathbf{s}) f(p_t(\mathbf{s})/p_e(\mathbf{s}))$ is a monotonically decreasing function for such processes and is bounded from below, where $f(x)$ is an arbitrary convex function, i.e., $\forall x \geqslant 0$: $f(x) \geqslant 0$, $f''(x) > 0$. If the equilibrium probability distribution has the Boltzmann form, $p_e(\mathbf{s}) \sim e^{-\beta H(\mathbf{s})}$, the usual choice made is $f(x) = x \log x$. Now $\mathscr{L}_t$ becomes

$$\mathscr{L}_t = \beta \sum_{\mathbf{s}} p_t(\mathbf{s}) \left[H(\mathbf{s}) + \frac{1}{\beta} \log p_t(\mathbf{s}) \right] \tag{9}$$

This is not yet a true Lyapunov function for the evolving macroscopic state vector $\mathbf{m}$ since it is a function of the microscopic probability distribution;

however, intuitively one may hope to use (9) as a starting point for constructing a Lyapunov function for the *macroscopic* laws (5), (6).

If the Hamiltonian $H(\mathbf{s})$ depends on the microscopic state $(\mathbf{s})$ only through the values of the order parameters $\mathbf{m}$ (as is the case for the class of models considered in this paper), we obtain, by putting $H(\mathbf{s}) = N\varepsilon(\mathbf{m}(\mathbf{s}))$,

$$\frac{\mathscr{L}_t}{\beta N} = \int d\mathbf{m}\, \mathscr{P}_t(\mathbf{m}) \left[\varepsilon(\mathbf{m}) + \frac{\sum_{\mathbf{s}} \delta[\mathbf{m} - \mathbf{m}(\mathbf{s})]\, p_t(\mathbf{s})(1/\beta N) \log p_t(\mathbf{s})}{\sum_{\mathbf{s}} \delta[\mathbf{m} - \mathbf{m}(\mathbf{s})]\, p_t(\mathbf{s})} \right] \tag{10}$$

where $\mathscr{P}_t(\mathbf{m}) \equiv \sum_{\mathbf{s}} p_t(\mathbf{s})\, \delta[\mathbf{m} - \mathbf{m}(\mathbf{s})]$ is the *macroscopic* probability distribution. For the case where the order parameters $\mathbf{m}$ evolve in time deterministically in the thermodynamic limit giving rise to Eqs. (5), (6), $\mathscr{P}_t(\mathbf{m}) \to \delta[\mathbf{m} - \mathbf{m}_t]$, where $\mathbf{m}_t$ evolves according to (5), (6). The quantity (10), however, cannot be reduced to a function of $\mathbf{m}_t$ without using additional information about the underlying microscopic states, or by making additional assumptions or approximations, because of the appearance of the entropic term $\log p_t(\mathbf{s})$. Such microscopic information would require knowledge of the solution of Eqs. (3), (4) which is exactly what one tries to avoid in deriving the macroscopic equations (5), (6). An approximation which could be made would be to (incorrectly) assume equipartitioning of probability in the $\mathbf{m}$-subshells of the ensemble, which would imply making in (10) the replacement

$$p_t^{-1}(\mathbf{s}) \sim \sum_{\mathbf{s}} \delta[\mathbf{m} - \mathbf{m}(\mathbf{s})] = e^{-Nc^*(\mathbf{s})} \qquad (N \to \infty)$$

resulting in the appealing expression

$$\frac{\mathscr{L}_t}{\beta N} \to \frac{\mathscr{L}_t^{\text{equip}}}{\beta N} = \varepsilon(\mathbf{m}) + \frac{1}{\beta} c^*(\mathbf{m}) \qquad \text{for} \quad N \to \infty \tag{11}$$

where $c^*(\mathbf{m})$ is the Legendre transform of the cumulant generating function, derived from large-deviations theory[6]:

$$c^*(\mathbf{m}) = \sup_{\mathbf{x}} (\mathbf{m} \cdot \mathbf{x} - \langle \log \cosh(\mathbf{x} \cdot \boldsymbol{\xi}) \rangle) \tag{12}$$

Unfortunately, away from equilibrium the assumption of subshell-equipartitioning of probability is unsustainable. $p_t(\mathbf{s})$ is the full complicated solution of the microscopic laws (3), (4) and will exhibit equipartitioning in the macroscopic subshells *only in equilibrium*. To see this, assume that at $t = 0$ we prepare an ensemble distribution $p_0(\mathbf{s})$ obeying equipartitioning within the $\mathbf{m}$ subshells: $p_0(\mathbf{s}) = f[\mathbf{m}(\mathbf{s})]$ for some function f. According to the master equation (4), we now find

$$\frac{d}{dt}p_t(\mathbf{s})\bigg|_{t=0}=\frac{1}{2}\sum_i\left\{f\left[\mathbf{m}(\mathbf{s})-\frac{2}{N}\xi_i s_i\right]-f[\mathbf{m}(\mathbf{s})]\right\}$$
$$+\frac{1}{2}\sum_i s_i\tanh[\beta h_i(\mathbf{s})]\left\{f\left[\mathbf{m}(\mathbf{s})-\frac{2}{N}\xi_i s_i\right]+f[\mathbf{m}(\mathbf{s})]\right\} \qquad (13)$$

Expanding in powers of N and using $h_i(\mathbf{s})=\xi_i\cdot\mathbf{Am}(\mathbf{s})-(1/N)\,\xi_i\cdot\mathbf{A}\xi_i s_i$, we get

$$\frac{d}{dt}p_t(\mathbf{s})\bigg|_{t=0}=-\nabla_{\mathbf{m}}f[\mathbf{m}(\mathbf{s})]\cdot\{\mathbf{m}(\mathbf{s})+\langle\xi\tanh[\beta\xi\cdot\mathbf{Am}(\mathbf{s})]\rangle_\xi\}$$
$$-f[\mathbf{m}(\mathbf{s})]\frac{\beta}{N}\sum_i[\xi_i\cdot\mathbf{A}\xi_i]\{1-\tanh^2[\beta\xi_i\cdot\mathbf{Am}(\mathbf{s})]\}$$
$$+f[\mathbf{m}(\mathbf{s})]\sum_i s_i\tanh[\beta\xi_i\cdot\mathbf{Am}(\mathbf{s})]+\mathcal{O}(N^{-1}) \qquad (14)$$

Equipartitioning at $t=0$ is sustained for $t>0$ only if the above expression depends on the microvariables $\mathbf{s}$ only through $\mathbf{m}(\mathbf{s})$. For the present class of models the last of the above terms violates this requirement (except for the trivial case $p=1$, where we recover the ∞-range ferromagnet). Therefore equipartitioning at $t=0$ does not imply equipartitioning at $t>0$. This is in contrast to a formulation in terms of sublattice magnetizations $m_\xi=(1/I_\xi)\sum_{i\in I_\xi}s_i$, where the sublattices I_ξ consist of all those sites for which $\xi_i=\xi$. Deterministic evolution of the latter order parameters (which are akin to the magnetization in the ∞-range ferromagnet), however, requires $p\ll\log N$, rather than the much greater number afforded by our description at the level of overlaps, requiring only $p^2\ll N$.

Therefore the H-theorem cannot be used to prove *a priori* that (11) is a true macroscopic Lyapunov function in the thermodynamic limit for the macroscopic laws (5), (6). Nevertheless we will show in Section 5 that for separable systems this is indeed the case, and can even be generalized to discrete-time parallel dynamics.

4. STRICTLY POSITIVE OR STRICTLY NEGATIVE MATRICES *A*

It is reasonably easy to find a macroscopic Lyapunov function $\mathscr{L}(\mathbf{m})$ for the differential equation (6) when the symmetric matrix $\mathbf{A}$ is either positive definite, $\mathscr{L}^+(\mathbf{m})$, or negative definite, $\mathscr{L}^-(\mathbf{m})$,[6,14,16]

$$\mathscr{L}^\pm(\mathbf{m})\equiv\pm\frac{1}{2}\mathbf{m}\cdot\mathbf{Am}\mp\frac{1}{\beta}\langle\log\cosh(\beta\xi\cdot\mathbf{Am})\rangle \qquad (15)$$

since

$$\frac{d}{dt}\mathscr{L}^{\pm} = \pm(\mathbf{m} - \langle \boldsymbol{\xi} \tanh(\beta\boldsymbol{\xi}\cdot\mathbf{Am})\rangle)\cdot\mathbf{A}\frac{d\mathbf{m}}{dt} = \mp\frac{d\mathbf{m}}{dt}\cdot\mathbf{A}\frac{d\mathbf{m}}{dt} \leqslant 0 \quad (16)$$

$(d/dt)\,\mathscr{L}^{\pm}$ can only be zero when $(d/dt)\,\mathbf{m} = 0$, since $\mathbf{A}$ is strictly positive or strictly negative definite. It is also easy to convince oneself that $\mathscr{L}^{\pm}$ is bounded from below, since $\mathbf{m}$ is only defined on the interval $[-1, +1]^p$, and the expression (15) contains no singularities. Hence $\mathscr{L}^{\pm}$ is a Lyapunov function for the macroscopic equations (6) derived for asynchronous updating and strictly positive or strictly negative matrices $\mathbf{A}$, respectively.

For synchronous updating we know from a microscopic analysis that at $T=0$ the network will again settle into an equilibrium configuration for $\mathbf{A}$ positive definite and a period-two cycle for $\mathbf{A}$ negative definite.[1] Here we show that the Lyapunov function $[\mathscr{L}^{+}$, (15)] of the sequential case is also a Lyapunov function of the macroscopic equations (5) derived for the parallel case for $\mathbf{A}$ positive definite. First we note that the macroscopic parallel dynamics (5) can be written

$$\mathbf{m}_{t+1} = \mathbf{m}_t - \mathbf{A}^{-1}\cdot\nabla\mathscr{L}^{+}(\mathbf{m}_t) \quad (17)$$

We then use the identities

$$\begin{aligned}\mathscr{L}^{+}(\mathbf{x}+\mathbf{y}) &= \mathscr{L}^{+}(\mathbf{x}) + \sum_{\mu} y^{\mu}\partial_{\mu}\mathscr{L}^{+}(\mathbf{x}) \\ &\quad + \int_0^1 \lambda\,d\lambda \int_0^1 d\rho \sum_{\mu,\nu} y^{\mu}y^{\nu}\partial^2_{\mu\nu}\mathscr{L}^{+}(\mathbf{x}+\lambda\rho\mathbf{y}) \end{aligned} \quad (18)$$

$$\begin{aligned}\partial^2_{\rho\alpha}\mathscr{L}^{+}(\mathbf{x}) &= A_{\rho\alpha} - \beta\sum_{\mu,\eta}\Gamma_{\mu\eta}(\mathbf{x})\,A_{\eta\rho}A_{\mu\alpha}, \\ \Gamma_{\mu\eta}(\mathbf{x}) &= \langle \xi^{\mu}\xi^{\eta}[1-\tanh^2(\beta\boldsymbol{\xi}\cdot\mathbf{Ax})]\rangle \end{aligned} \quad (19)$$

[where the symmetric matrix $\boldsymbol{\Gamma}(\mathbf{x})$ has only nonnegative eigenvalues], to obtain for symmetric and invertible $\mathbf{A}$

$$\begin{aligned}\Delta\mathscr{L}^{+} &= \mathscr{L}^{+}(\mathbf{m}_{t+1}) - \mathscr{L}^{+}(\mathbf{m}_t) = \mathscr{L}^{+}(\mathbf{m}_t - \nabla\mathbf{A}^{-1}\mathscr{L}^{+}_{\mathbf{m}_t}) - \mathscr{L}^{+}(\mathbf{m}_t) \\ &= -\nabla\mathscr{L}^{+}\cdot\left\{\int_0^1 \lambda\,d\lambda\int_0^1 d\rho\,\mathbf{A}^{-1} + \beta\boldsymbol{\Gamma}(\mathbf{m}_t - \lambda\rho\mathbf{A}^{-1}\nabla\mathscr{L}^{+})\right\}\nabla\mathscr{L}^{+} \\ &\leqslant 0 \end{aligned} \quad (20)$$

Equality only holds when $\partial_\mu \mathscr{L}^+ = 0$, which reduces to exactly the condition for a fixed point $\mathbf{m} = \langle \boldsymbol{\xi} \tanh(\beta \boldsymbol{\xi} \cdot \mathbf{A}\mathbf{m}) \rangle$. Hence $\mathscr{L}^+$ decreases monotonically until a fixed point is reached.

5. CONSTRUCTION FOR ARBITRARY SEPARABLE MODELS

In equilibrium we know that the properties of the network are determined by using as a generating function the free energy for the sequential case, and Peretto's pseudo free energy[13] for the parallel case. Consider the free energy per neuron, f, with H as either the Ising Hamiltonian or Peretto's pseudo-Hamiltonian (as appropriate)

$$f = -\frac{1}{\beta N} \log Z = -\frac{1}{\beta N} \log \left(\sum_{\{\mathbf{s}\}} e^{-\beta H(\mathbf{s})} \right) \tag{21}$$

If the (extensive) Hamiltonian depends on the system state (s) only through the macroscopic variables (m^μ), then we can write $H(\{\mathbf{s}\}) = N\varepsilon(\mathbf{m}(\mathbf{s}))$. Along with the density of microscopic states

$$\mathscr{D}(\mathbf{m}) = \sum_{\{\mathbf{s}\}} \delta(\mathbf{m} - \mathbf{m}(\mathbf{s})) = e^{-Nc^*(\mathbf{m})} \qquad (N \to \infty)$$

where $c^*(\mathbf{m})$ is the Legendre transform of the cumulant generating function, derived from large-deviations theory[6]

$$c^*(\mathbf{m}) = \sup_{\mathbf{x}} (\mathbf{m} \cdot \mathbf{x} - \langle \log \cosh(\mathbf{x} \cdot \boldsymbol{\xi}) \rangle) \tag{22}$$

we can now write the free energy for large N as

$$\begin{aligned} f &= -\frac{1}{\beta N} \log \left(\int d\mathbf{m}\, e^{-N(c^*(\mathbf{m}) + \beta\varepsilon(\mathbf{m}))} \right) \\ &= \frac{1}{\beta} \min_{\mathbf{m} \in \mathscr{R}^p} \mathscr{L}(\mathbf{m}) = \frac{1}{\beta} c^*(\tilde{\mathbf{m}}) + \varepsilon(\tilde{\mathbf{m}}) \end{aligned} \tag{23}$$

where $\tilde{\mathbf{m}}$ is the value of $\mathbf{m}$ which minimizes $c^*(\mathbf{m}) + \beta\varepsilon(\mathbf{m})$, since the integral will be extremally dominated. We will prove that $(1/\beta)\, c^*(\mathbf{m}) + \varepsilon(\mathbf{m})$, with $\varepsilon(\mathbf{s})$ as appropriate to the dynamics, is a Lyapunov function for both types of macroscopic laws (5), (6) with symmetric $\mathbf{A}$.

We will first prove a property of $c^*(\mathbf{m})$, namely that for physically realizable $\mathbf{m}$ [i.e., $m^\mu = (1/N) \sum_i \xi_i^\mu s_i$] the supremum is satisfied by a finite critical point rather than at $\mathbf{x} = \infty$. This result is required in the proofs to follow. If we write $\mathbf{x} = \hat{\mathbf{x}} x$, where $\hat{\mathbf{x}}$ is a unit vector in the direction of $\mathbf{x}$ and

x is its modulus, and use the identity $\log\cosh(z) = -\log 2 + |z| + \log(1+e^{-2|z|})$, then the quantity to be maximized becomes

$$x(\mathbf{m}\cdot\hat{\mathbf{x}} - \langle|\boldsymbol{\xi}\cdot\hat{\mathbf{x}}|\rangle) + \log 2 - \langle\log(1+e^{-2x|\boldsymbol{\xi}\cdot\hat{\mathbf{x}}|})\rangle \tag{24}$$

We can write $\mathbf{m}$ in terms of the sublattice magnetizations $m_\xi \in [-1, +1]$, $\mathbf{m} = \langle\boldsymbol{\xi} m_\xi\rangle$, where $m_\xi(\mathbf{s}) = (1/I_\xi)\sum_{i\in I_\xi} s_i$ and the sublattices I_ξ consist of those sites i for which $\boldsymbol{\xi}_i = \boldsymbol{\xi}$. As a result, for large x the expression in (24) behaves as

$$x\langle\boldsymbol{\xi}\cdot\hat{\mathbf{x}} m_\xi - |\boldsymbol{\xi}\cdot\hat{\mathbf{x}}|\rangle + \log 2 = x\langle|\boldsymbol{\xi}\cdot\hat{\mathbf{x}}|[m_\xi \operatorname{sgn}(\boldsymbol{\xi}\cdot\hat{\mathbf{x}}) - 1]\rangle + \log 2 \tag{25}$$

Clearly this is not maximized in the limit $x \to \infty$ for physical m_ξ ($\in [-1, +1]$). Hence the supremum is bounded, and so must be realized by a finite critical point.

5.1. Parallel Dynamics

The microscopic properties of the models with synchronous updating (5) are obtained by using the Peretto pseudo-Hamiltonian[13]

$$\begin{aligned} H(\{\mathbf{s}\}) &= -\frac{1}{\beta}\sum_i \log\left(2\cosh\beta\sum_{j\neq i} J_{ij}s_j\right) \\ &= -\frac{N}{\beta}\langle\log 2\cosh[\beta\boldsymbol{\xi}\cdot\mathbf{A}\mathbf{m}(\mathbf{s})]\rangle + \mathcal{O}(1) \end{aligned} \tag{26}$$

so that

$$\varepsilon(\mathbf{m}) = -\frac{1}{\beta}\langle\log 2\cosh(\beta\boldsymbol{\xi}\cdot\mathbf{A}\mathbf{m})\rangle \tag{27}$$

and

$$\mathscr{L}_{\text{par}}(\mathbf{m}) = c^*(\mathbf{m}) - \langle\log 2\cosh(\beta\boldsymbol{\xi}\cdot\mathbf{A}\mathbf{m})\rangle \tag{28}$$

Theorem 1. $\mathscr{L}_{\text{par}}(\mathbf{m})$ is a monotonically decreasing function of the process (5) if $\mathbf{A}^\dagger = \mathbf{A}$, only stationary when in a period-two cycle.

Proof. Consider $\mathscr{L}_{\text{par}}(\mathbf{m}_{t+1})$. Taking the gradient of the argument of the supremum and subsituting the dynamics (5), we find that it is realized among those $\mathbf{x}$ satisfying $\langle\boldsymbol{\xi}\tanh(\beta\boldsymbol{\xi}\cdot\mathbf{A}\mathbf{m}_t)\rangle = \langle\boldsymbol{\xi}\tanh\boldsymbol{\xi}\cdot\mathbf{x}\rangle$. Since the supremum is realized at a finite critical point we can write for those $\mathbf{x}$ satisfying the extremization criterion

$$0 = \langle \xi[\tanh(\beta\xi\cdot\mathbf{Am}) - \tanh(\xi\cdot\mathbf{x})] \rangle$$
$$= \left\langle \xi \int_{\xi\cdot\mathbf{x}}^{\beta\xi\cdot\mathbf{Am}} dz\,(1-\tanh^2 z) \right\rangle \tag{29}$$

Then using the substitution $z = \xi\cdot[\mathbf{x} + \lambda(\beta\mathbf{Am} - \mathbf{x})]$ and taking the inner product with $\beta\mathbf{Am} - \mathbf{x}$ gives

$$0 = \left\langle [\xi\cdot(\beta\mathbf{Am} - \mathbf{x})]^2 \int_0^1 d\lambda\,(1 - \tanh^2\{\xi\cdot[\mathbf{x} - \lambda(\beta\mathbf{Am} - \mathbf{x})]\}) \right\rangle \tag{30}$$

This implies $\mathbf{x} = \beta\mathbf{Am}_t$, since the integral can never be zero. If we use this to calculate $\Delta\mathscr{L}_{\text{par}} = \mathscr{L}_{\text{par}}(\mathbf{m}_{t+1}) - \mathscr{L}_{\text{par}}(\mathbf{m}_t)$, then

$$\begin{aligned}
\Delta\mathscr{L}_{\text{par}} &= \beta\mathbf{m}_{t+1}\cdot\mathbf{Am}_t - \langle \log\cosh(\beta\xi\cdot\mathbf{Am}_t)\rangle - \langle\log\cosh(\beta\xi\cdot\mathbf{Am}_{t+1})\rangle \\
&\quad - \sup_{\mathbf{y}}(\mathbf{m}_t\cdot\mathbf{y} - \langle\log\cosh(\xi\cdot\mathbf{y})\rangle) + \langle\log\cosh(\beta\xi\cdot\mathbf{Am}_t)\rangle \\
&= \mathbf{m}_t\cdot(\beta\mathbf{Am}_{t+1}) - \langle\log\cosh(\beta\xi\cdot\mathbf{Am}_{t+1})\rangle \\
&\quad - \sup_{\mathbf{y}}(\mathbf{m}_t\cdot\mathbf{y} - \langle\log\cosh(\xi\cdot\mathbf{y})\rangle) \\
&\leqslant 0
\end{aligned} \tag{31}$$

$\Delta\mathscr{L}_{\text{par}} = 0$ requires $\mathbf{y} = \beta\mathbf{Am}_{t+1}$, the supremum condition for $\mathbf{y}$ is $\mathbf{m}_t = \langle\xi\tanh(\xi\cdot\mathbf{y})\rangle$, and hence a stationary $\mathscr{L}_{\text{par}}$ implies

$$\mathbf{m}_t = \langle\xi\tanh(\beta\xi\cdot\mathbf{Am}_{t+1})\rangle = \mathbf{m}_{t+2} \tag{32}$$

i.e., $\mathscr{L}_{\text{par}}$ is a monotonically decreasing function of the macroscopic dynamics (5) for $\mathbf{A} = \mathbf{A}^\dagger$ and is only stationary when the system is in a period-two cycle. Since $\mathbf{m}$ is only defined on $[-1, +1]^p$ and $\mathscr{L}_{\text{par}}$ has no poles, it must be bounded from below, and hence satisfies the necessary conditions to be a Lyapunov function of the macroscopic synchronous dynamics (5).

5.2. Sequential Dynamics

For the asynchronous Glauber dynamics (6) we must use the usual Ising Hamiltonian

$$\begin{aligned}
H(\{\mathbf{s}\}) &= -\frac{1}{2N}\sum_{i\neq j} J_{ij}s_is_j = -\frac{1}{2N}\sum_{i,j}\sum_{\mu,\nu}\xi_i^\mu A_{\mu\nu}\xi_j^\nu s_is_j + \text{const} \\
&= -\frac{N}{2}\mathbf{m}\cdot\mathbf{Am} + \text{const}
\end{aligned} \tag{33}$$

so that

$$\varepsilon(\mathbf{m}) = -\tfrac{1}{2}\mathbf{m}\cdot\mathbf{Am} \tag{34}$$

and

$$\mathscr{L}_{\text{seq}}(\mathbf{m}) = c^*(\mathbf{m}) - \frac{\beta}{2}\mathbf{m}\cdot\mathbf{Am} \tag{35}$$

Theorem 2. $\mathscr{L}_{\text{seq}}(\mathbf{m})$ is a monotonically decreasing function of the process (6) if $\mathbf{A}^\dagger = \mathbf{A}$, only stationary when in a fixed point.

Proof. The supremum is realized among those $\mathbf{x}$ satisfying $\mathbf{m} = \langle \xi \tanh \xi\cdot\mathbf{x}\rangle$, since the argument of $\mathscr{L}_{\text{seq}}$ remains bounded. If we substitute this relation into $\mathscr{L}_{\text{seq}}(\mathbf{m})$, we get

$$\mathscr{L}_{\text{seq}}(\mathbf{m}) = \langle \xi\cdot\mathbf{x}\tanh(\xi\cdot\mathbf{x}) - \log\cosh(\xi\cdot\mathbf{x})\rangle - \tfrac{1}{2}\beta\mathbf{m}\cdot\mathbf{Am} \tag{36}$$

$$= \left\langle \int_0^{\xi\cdot\mathbf{x}} dy\, y(1-\tanh^2 y)\right\rangle - \tfrac{1}{2}\beta\mathbf{m}\cdot\mathbf{Am} \tag{37}$$

Differentiating with respect to one component of $\mathbf{m}$, we find, using $\mathbf{A}^\dagger = \mathbf{A}$,

$$\frac{\partial\mathscr{L}_{\text{seq}}(\mathbf{m})}{\partial\mathbf{m}^\alpha} = \sum_\gamma \langle \xi\cdot\mathbf{x}\xi^\gamma[1-\tanh^2(\xi\cdot\mathbf{x})]\rangle \frac{\partial x^\gamma}{\partial\mathbf{m}^\alpha} - \beta(\mathbf{Am})^\alpha \tag{38}$$

The supremum, however, is realized among those $\mathbf{x}$ satisfying $\mathbf{m} = \langle \xi\tanh\xi\cdot\mathbf{x}\rangle$; differentiating this equation gives

$$\delta_{\mu\alpha} = \frac{\partial\mathbf{m}^\mu}{\partial\mathbf{m}^\alpha} = \sum_\nu \langle \xi^\mu\xi^\nu[1-\tanh^2(\xi\cdot\mathbf{x})]\rangle \frac{\partial x^\nu}{\partial\mathbf{m}^\alpha} \tag{39}$$

Combining these two equations, we see that the $\mathbf{x}$ realizing the supremum satisfies $\nabla_{\mathbf{m}}\mathscr{L}_{\text{seq}}(\mathbf{m}) = \mathbf{x} - \beta\mathbf{Am}$. Substituting this into our expression for $\mathscr{L}_{\text{seq}}(\mathbf{m})$ gives

$$\begin{aligned}\mathscr{L}_{\text{seq}}(\mathbf{m}) &= \tfrac{1}{2}\beta\mathbf{m}\cdot\mathbf{Am} + \mathbf{m}\cdot\nabla_{\mathbf{m}}\mathscr{L}_{\text{seq}}(\mathbf{m}) \\ &\quad - \langle \log\cosh[\beta\xi\cdot\mathbf{Am} + \xi\cdot\nabla_{\mathbf{m}}\mathscr{L}_{\text{seq}}(\mathbf{m})]\rangle \end{aligned}\tag{40}$$

Now $d\mathscr{L}_{\text{seq}}(\mathbf{m})/dt = \nabla_{\mathbf{m}}\mathscr{L}_{\text{seq}}(\mathbf{m})\cdot d\mathbf{m}/dt$; using the equation of motion for $\mathbf{m}$, (6), we obtain from this

$$\begin{aligned}\frac{d\mathscr{L}_{\text{seq}}(\mathbf{m})}{dt} &= -\mathbf{m}\cdot\nabla_{\mathbf{m}}\mathscr{L}_{\text{seq}}(\mathbf{m}) + \langle \nabla_{\mathbf{m}}\mathscr{L}_{\text{seq}}(\mathbf{m})\cdot\xi\tanh(\beta\xi\cdot\mathbf{Am})\rangle \\ &= -\mathscr{L}_{\text{seq}}(\mathbf{m}) - \langle \log\cosh[\beta\xi\cdot\mathbf{Am} + \xi\cdot\nabla_{\mathbf{m}}\mathscr{L}_{\text{seq}}(\mathbf{m})]\rangle \\ &\quad + \frac{1}{2}\beta\mathbf{m}\cdot\mathbf{Am} + \langle \nabla_{\mathbf{m}}\mathscr{L}_{\text{seq}}(\mathbf{m})\cdot\xi\tanh(\beta\xi\cdot\mathbf{Am})\rangle \end{aligned}\tag{41}$$

If we let $\xi \cdot \nabla_{\mathbf{m}} \mathscr{L}_{\text{seq}} = \lambda_{\xi}$, then the maximum of $d\mathscr{L}_{\text{seq}}/dt + \mathscr{L}_{\text{seq}}$ with respect to variation of λ_{ξ} is satisfied for $\lambda_{\xi} = 0 \ \forall \xi$. Hence we can write

$$\begin{aligned}
\frac{d\mathscr{L}_{\text{seq}}(\mathbf{m})}{dt} &\leqslant -\mathscr{L}_{\text{seq}}(\mathbf{m}) + \frac{1}{2}\beta \mathbf{m}\cdot\mathbf{A}\mathbf{m} - \langle \log\cosh(\beta\xi\cdot\mathbf{A}\mathbf{m})\rangle \\
&\leqslant -\mathscr{L}_{\text{seq}}(\mathbf{m}) + \bigg\{ [\beta\mathbf{m}\cdot\mathbf{A}\mathbf{m} - \langle \log\cosh(\beta\xi\cdot\mathbf{A}\mathbf{m})\rangle] \\
&\quad - \frac{1}{2}\beta\mathbf{m}\cdot\mathbf{A}\mathbf{m}\bigg\} \\
&\leqslant \bigg\{ [\beta\mathbf{m}\cdot\mathbf{A}\mathbf{m} - \langle \log\cosh(\beta\xi\cdot\mathbf{A}\mathbf{m})\rangle] - \frac{1}{2}\beta\mathbf{m}\cdot\mathbf{A}\mathbf{m}\bigg\} \\
&\quad - \bigg\{ \sup_{\mathbf{x}}[\mathbf{m}\cdot\mathbf{x} - \langle \log\cosh(\mathbf{x}\cdot\xi)\rangle] - \frac{\beta}{2}\mathbf{m}\cdot\mathbf{A}\mathbf{m}\bigg\} \\
&\leqslant 0
\end{aligned} \tag{42}$$

$\mathscr{L}_{\text{seq}}(\mathbf{m})$ is only stationary when the above-mentioned maximum of $d\mathscr{L}_{\text{seq}}/dt + \mathscr{L}_{\text{seq}}$ is realized, i.e., $\xi \cdot \nabla_{\mathbf{m}} \mathscr{L}_{\text{seq}} = 0$ for all ξ. From this we conclude that $\nabla_{\mathbf{m}} \mathscr{L}_{\text{seq}} = 0$ and hence $\mathbf{x} = \beta\mathbf{A}\mathbf{m}$. Combination with the supremum criterion $\mathbf{m} = \langle \xi \tanh(\xi\cdot\mathbf{x})\rangle$ subsequently gives

$$\mathbf{m} = \langle \xi \tanh(\beta\xi\cdot\mathbf{A}\mathbf{m})\rangle \tag{43}$$

which is a fixed point of the dynamics (6). The only remaining constraint on $\mathscr{L}_{\text{seq}}(\mathbf{m})$ for it to be a Lyapunov function is that it is bounded from below, which is obviously the case, since $\mathbf{m}$ only exists in the range $[-1, +1]^p$, and $\mathscr{L}_{\text{seq}}(\mathbf{m})$ has no poles. Hence $\mathscr{L}_{\text{seq}}(\mathbf{m})$ is a Lyapunov function of the macroscopic asynchronous dynamics (6).

6. CONCLUSION

We have shown that for Ising spin models of neural networks with long-range separable symmetric interactions, macroscopic Lyapunov functions exist. We have generalized existing Lyapunov functions of the "free energy" type to finite temperatures and arbitrary separable symmetric embedding matrices. We have shown that the proposed Lyapunov functions correspond exactly to the scalar surfaces that are encountered in the saddle-point integration resulting from equilibrium statistical mechanical studies, emphasizing the equivalence of thermodynamic and dynamic stability. We can therefore interpret the dynamics of the present type of symmetric network as (not necessarily gradient) descent on a "free energy" surface, for both parallel and sequential updating of the individual spins. The macroscopic dynamical equations which form the basis of our calculations are only strictly valid for $p \ll \sqrt{N}$, $N \to \infty$, but we may suspect that similar results can be obtained for the equations governing the behavior for $p = \alpha N$ as suggested by recent dynamical studies.[17]

ACKNOWLEDGMENT

We acknowledge financial support from the EPSRC under grant numbers 92310910 and GRH73028.

REFERENCES

1. D. J. Amit, *Modelling Brain Function* (Cambridge University Press, Cambridge, 1989).
2. B. Müller and J. Reinhardt, *Neural Networks—An Introduction* (Springer-Verlag, Berlin, 1991).
3. J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, California, 1991).
4. L. G. Abbott, *Network* **1**:105 (1990).
5. D. Sherrington, in *Mathematical Approaches to Neural Networks*, J. G. Taylor, ed. (North-Holland, Amsterdam, 1993), Chapter 8, p. 261.
6. J. L. van Hemmen and R. Kühn, in *Models of Neural Networks*, E. Domany, J. L. van Hemmen, and K. Schulten, eds. (Springer-Verlag, Berlin, 1991), Chapter 1.
7. A. C. C. Coolen and D. Sherrington, in *Mathematical Approaches to Neural Networks*, J. G. Taylor, ed. (North-Holland, Amsterdam, 1993), Chapter 9, p. 293.
8. A. C. C. Coolen and T. W. Ruijgrok, *Phys. Rev. A* **38**:4253 (1988).
9. U. Riedel, R. Kühn, and J. L. van Hemmen, *Phys. Rev. A* **38**:1105 (1988).
10. O. Bernier, *Europhys. Lett.* **16**:531 (1991).
11. N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
12. J. J. Hopfield, *Proc. Nat. Acad. Sci. USA* **79**:2554 (1982).
13. P. Peretto, *Biol. Cybernet.* **50**:51 (1984).
14. D. Bollé, P. Dupont, and B. Vinck, *J. Phys. A* **25**:2859 (1992).
15. A. V. M. Herz, Z. Li, and J. L. van Hemmen, *Phys. Rev. Lett.* **66**:1370 (1991).
16. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **32**:1007 (1985).
17. A. C. C. Coolen and D. Sherrington, *Phys. Rev. Lett.* **71**:3886 (1993).