J. Phys. A: Math. Gen. **31** (1998) L391–L396. Printed in the UK PII: S0305-4470(98)92334-4

LETTER TO THE EDITOR

Relationship between long timescales and the static free-energy in the Hopfield model

G Biroli and R Monasson†

CNRS-Laboratoire de Physique Théorique de l'ENS, 24 rue Lhomond, 75231 Paris cedex 05, France

Received 9 March 1998

Abstract. The Glauber dynamics of the Hopfield model at low storage level is considered. We analytically derive the spectrum of relaxation times for large system sizes. The longest timescales are gathered in families, each family being in one-to-one correspondence with a stationary (not necessarily stable) point of the static mean-field free energy. Inside a family, the timescales are given by the reciprocals (of the absolute values) of the eigenvalues of the free-energy Hessian matrix.

Recently, there has been a renewal of interest for the long-time dynamics of spin-glass models, which may already exhibit nontrivial features, for example, violation of the fluctuation–dissipation relations, ageing phenomena, at the mean-field level [1]. In this context, it is natural to study the spectrum of relaxation times occuring in the Glauber dynamics of disordered Ising spin models [2]. Unfortunately, numerical investigations have so far been limited to very small sizes [4]. Moreover, to the best of our knowledge, no analytical studies of the relaxation times spectra have yet been performed owing to the technical difficulties arising in the diagonalization of the Glauber matrix.

In this letter, we focus on the Glauber dynamics of the Hopfield model at low storage [6]. This system is simple enough to be analytically solved along the lines of [5]. Our main result is that the longest timescales are gathered in families, each family being in one-to-one correspondence with a stationary (not necessarily stable) point of the static mean-field free energy. Inside a family, the timescales are given by the reciprocals of the absolute values of the eigenvalues of the free-energy Hessian matrix. As a consequence, our study strengthens the close relationship between long-time dynamics and static properties in disordered mean-field models [1].

We consider a Hopfield model including N Ising spins S_i, $i = 1, \ldots, N$ and p quenched patterns ξ_i^μ, $\mu = 1, \ldots, p$. In addition to its intrinsic interest as a paradigm for attractor neural networks [7], the Hopfield model may be seen as a spin-glass system smoothly interpolating between the Mattis model [8] (when $p = 1$) and the much more involved Sherrington–Kirpatrick model (for infinite p) [3]. Its Hamiltonian depends on the spin configuration $\boldsymbol{S} = (S_1, \ldots, S_N)$ through the set of p overlaps $m^\mu(\boldsymbol{S}) = \frac{1}{N}\sum_{i=1}^N S_i \xi_i^\mu$ and reads

$$H(\boldsymbol{S}) = -\frac{N}{2}\sum_{\mu=1}^{p} m^\mu(\boldsymbol{S})^2. \tag{1}$$

† E-mail addresses: biroli@physique.ens.fr; monasson@physique.ens.fr.

0305-4470/98/210391+06$19.50

The thermodynamics of the above Hamiltonian has been studied in [7]. At low storage level, that is when p remains finite as the size N is sent to infinity, the free-energy density is particularly simple,

$$f(\beta) = -\frac{1}{2}\sum_{\mu=1}^{p}(m^{\mu})^2 + \frac{1}{\beta}\left\langle\!\!\left\langle \ln\left[2\cosh\left(\beta\sum_{\mu=1}^{p} m^{\mu}\xi^{\mu}\right)\right]\right\rangle\!\!\right\rangle \tag{2}$$

where β is the inverse temperature and $\langle\!\langle F(\{\xi^{\mu}\})\rangle\!\rangle \equiv \frac{1}{N}\sum_{i=1}^{N} F(\{\xi_i^{\mu}\})$ denotes the site average. In the above expression (2), the overlaps are the solutions of the saddle-point equations

$$m^{\mu} = \left\langle\!\!\left\langle \xi^{\mu}\tanh\left(\beta\sum_{\nu=1}^{p} m^{\nu}\xi^{\nu}\right)\right\rangle\!\!\right\rangle \tag{3}$$

with the smallest free energy. At high temperature, the system lies in the paramagnetic phase $m^{\mu} = 0$, $\forall\mu$. Below the critical temperature $1/\beta_c = 1$, many other solutions with nonzero overlaps appear. The low-temperature phase is characterized by a condensation along some particular patterns, while spurious mixtures give rise to metastable states.

These equilibrium states may be seen as the attractors of a retrieval dynamics [2] starting from a given initial configuration (which is often but not necessarily chosen as a corrugated version of the quenched patterns). One possible evolution scheme is provided by the usual Glauber dynamics: given a configuration $\boldsymbol{S}$, the spin at site i is flipped with probability rate $w(\boldsymbol{S}, i) = \frac{1}{2}(1 - S_i\tanh[\beta h_i(\boldsymbol{S})])$ and left unchanged with rate $1 - w$. The instantaneous magnetic field h_i reads from (1),

$$h_i(\boldsymbol{S}) = \sum_{j(\neq i)}\left(\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^{\mu}\xi_j^{\mu}\right)S_j = \sum_{\mu=1}^{p}\xi_i^{\mu}m^{\mu}(\boldsymbol{S}) - \frac{p}{N}S_i. \tag{4}$$

The probability distribution $P(\boldsymbol{S}, t)$ satisfies the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t}P(\boldsymbol{S}, t) = \sum_{i=1}^{N}(w(\boldsymbol{S}^i, i)P(\boldsymbol{S}^i, t) - w(\boldsymbol{S}, i)P(\boldsymbol{S}, t)) \tag{5}$$

where the spin configuration $\boldsymbol{S}^i$ is obtained from $\boldsymbol{S}$ by flipping spin i. Due to its linear structure, equation (5) may be recast in a more suitable operator formalism. To each configuration $\boldsymbol{S}$ is associated a vector $|\boldsymbol{S}\rangle$, which form the basis of a 2^N-dimensional linear space $\mathcal{V}$. The probability distribution then becomes a vector $|P(t)\rangle = \sum_{\boldsymbol{S}} P(\boldsymbol{S}, t)|\boldsymbol{S}\rangle$. We now define the spin–flip and magnetization operators, respectively denoted by σ_i^x and σ_i^z through $\sigma_i^x|\boldsymbol{S}\rangle = |\boldsymbol{S}^i\rangle$ and $\sigma_i^z|\boldsymbol{S}\rangle = S_i|\boldsymbol{S}\rangle$. It is easy to check that the vectorial operator $\frac{1}{2}\boldsymbol{\sigma} = \frac{1}{2}(\sigma^x, \sigma^y \equiv i\sigma^x\sigma^z, \sigma^z)$ satisfies the usual commutation relations for a quantum spin $\frac{1}{2}$. The master equation (5) now acquires the following compact form

$$\frac{\mathrm{d}}{\mathrm{d}t}|P(t)\rangle = W|P(t)\rangle \qquad W = \frac{1}{2}\sum_{i=1}^{N}(\sigma_i^x - 1)(1 - \sigma_i^z\tanh(\beta h_i)) \tag{6}$$

where the local field operator h_i is obtained by replacing in (4) the spins S_i with their corresponding magnetization operators σ_i^z. Due to the detailed balance conditions, the operator $W^s = \exp(\beta H/2)W\exp(-\beta H/2)$ is Hermitian and reads

$$W^s = -\frac{N}{2} + \frac{1}{2}\sum_{i=1}^{N}[\sigma_i^x(\cosh(\beta h_i))^{-1} + \sigma_i^z\tanh(\beta h_i)]. \tag{7}$$

We shall now analyse the spectrum of W^s, which is the same as W's one. From a rigorous point of view, all eigenvalues are strictly negative except the zero mode corresponding to the

equilibrium Gibbs measure. First, as far as long-time dynamics is concerned, all large (in absolute value) eigenvalues, that are associated with fast decaying modes, may be discarded. Second, for very large system sizes ($N \gg 1$), indeed the limit of physical interest, the zero eigenvalue of W^s is expected to become almost degenerate when $\beta > \beta_c$. Dynamical processes taking place over exponentially large times, or more generally times diverging with N will not be considered. We therefore focus only on long but finite timescales.

To do so, we shall perform an expansion of W_s in powers of $1/\sqrt{N}$ and only keep the nonvanishing terms in the thermodynamical limit. This expansion is made possible by the mean-field nature of the Hopfield Hamiltonian (1). For a given configuration $\boldsymbol{S}$, the dominant contributions $H_i \equiv \sum_\mu \xi_i^\mu m^\mu(\boldsymbol{S})$ to the local fields h_i (4) take only 2^p different values when scanning all sites i. Let us partition the N sites in 2^p blocks $G_{\boldsymbol{\tau}}$ where $\boldsymbol{\tau}$ is a p binary component vector; site i belongs to the only group $G_{\boldsymbol{\tau}}$ such that H_i equals $H_{\boldsymbol{\tau}} = \sum_\mu \tau^\mu m^\mu(\boldsymbol{S})$ [9]. We restrict ourselves to quenched patterns that differ from each other on a macroscopic number of sites, so each block $G_{\boldsymbol{\tau}}$ contains a number $Nc_{\boldsymbol{\tau}}$ of sites of the order of N; note that if the patterns are randomly drawn from an unbiased distribution $c_{\boldsymbol{\tau}} = 2^{-p}$ in the limit of large sizes.

After some simple algebra, the evolution operator may be rewritten as $W^s = \sum_{\boldsymbol{\tau}} W^s_{\boldsymbol{\tau}}$ with

$$W^s_{\boldsymbol{\tau}} = -\frac{N}{2}c_{\boldsymbol{\tau}} + J^x_{\boldsymbol{\tau}}(\cosh(\beta H_{\boldsymbol{\tau}}))^{-1} + J^z_{\boldsymbol{\tau}}\tanh(\beta H_{\boldsymbol{\tau}}) - \frac{\beta p}{2}c_{\boldsymbol{\tau}}(1-\tanh(\beta H_{\boldsymbol{\tau}})^2) + \mathrm{O}\left(\frac{J^y_{\boldsymbol{\tau}}}{N}\right) \tag{8}$$

where the angular momenta and the field operators respectively read

$$\boldsymbol{J}_{\boldsymbol{\tau}} = \frac{1}{2}\sum_{i\in G_{\boldsymbol{\tau}}}\boldsymbol{\sigma}_i \qquad \text{and} \qquad H_{\boldsymbol{\tau}} = \frac{2}{N}\sum_{\boldsymbol{\eta}}\sum_{\mu=1}^{p}\tau^\mu\eta^\mu J^z_{\boldsymbol{\eta}}. \tag{9}$$

The last two terms on the r.h.s of (8) stem from the contributions pS_i/N to the local fields (4). The angular momenta $\boldsymbol{J}_{\boldsymbol{\tau}}$ and $\boldsymbol{J}_{\boldsymbol{\tau}'}$ commute for any two different groups $\boldsymbol{\tau} \neq \boldsymbol{\tau}'$ as does any $\boldsymbol{J}_{\boldsymbol{\tau}}$ with W^s. Thus, the angular momenta may be used as quantum numbers labelling the different invariant subspaces.

It can be checked that the subspace $\mathcal{V}^s$ corresponding to the maximum value of each angular momentum $J_{\boldsymbol{\tau}} = \frac{1}{2}Nc_{\boldsymbol{\tau}}$ contains the totally symmetric vectors $|\{M_{\boldsymbol{\tau}}\}\rangle$ having magnetizations $M_{\boldsymbol{\tau}} = -Nc_{\boldsymbol{\tau}}, -Nc_{\boldsymbol{\tau}}+2, \ldots, Nc_{\boldsymbol{\tau}}$ in blocks $G_{\boldsymbol{\tau}}$. Indeed, the vector $|\{M_\tau = Nc_{\boldsymbol{\tau}}\}\rangle = |+,+,\cdots,+\rangle$ belongs to $\mathcal{V}^s$ since its angular momentum component along the z-axis equals $N/2$ and is maximal. Moreover, the lowering operators $J^-_{\boldsymbol{\tau}}$ leave $\mathcal{V}^s$ invariant and $|\{Nc_{\boldsymbol{\tau}} - 2d_{\boldsymbol{\tau}}\}\rangle = \otimes(J^-_{\boldsymbol{\tau}})^{d_{\boldsymbol{\tau}}}|\{Nc_{\boldsymbol{\tau}}\}\rangle$. In addition, the equilibrium Gibbs vector belongs to $\mathcal{V}^s$. In the following we shall see that not only the Gibbs mode but also all of the eigenvectors corresponding to the smallest eigenvalues lie in $\mathcal{V}^s$.

At infinite temperature, the system is purely diffusive and the equilibrium state $|0\rangle$ equals the normalized sum of all $|\boldsymbol{S}\rangle$s. The average values of the angular momentum read $\langle 0|J^y|0\rangle = \langle 0|J^z|0\rangle = 0$, $\langle 0|J^x|0\rangle = N/2$. For low-lying excitations $|\omega\rangle$, that is plane waves on the N-dimensional hypercube, the mean value of the angular momentum $\langle\omega|\boldsymbol{J}|\omega\rangle$ equals that in the equilibrium state up to nonextensive terms. The same picture holds at finite temperature, but not at phase transition boundaries. More precisely, equilibrium or quasistationary modes $|0\rangle$ are expected to correspond to a well-defined direction of the angular momenta $\boldsymbol{J}_{\boldsymbol{\tau}}$ in the three-dimensional space. This direction varies smoothly when exploring the excited modes slightly above $|0\rangle$†. To capture the spectrum of low-

† Activated processes between wells cannot be obtained this way, but they correspond to infinite timescales which we do not investigate here.

lying states, we therefore proceed in two steps: first identify the preferred directions of the angular momentum corresponding to quasiequilibrium modes and then compute the fluctuations around these positions.

This approach, reminiscent of semiclassical approximations and spin-wave theory [11], may be made more precise in the following way. Consider the Holstein–Primakoff (HP) representation of the angular momentum vectorial operator [10] in terms of Bose operators

$$\begin{aligned} L^x_\tau &= \tfrac{1}{2}Nc_\tau - a^\dagger_\tau a_\tau \\ L^y_\tau &= \frac{1}{2i}\left[a^\dagger_\tau\sqrt{Nc_\tau - a^\dagger_\tau a_\tau} - \sqrt{Nc_\tau - a^\dagger_\tau a_\tau}\,a_\tau\right] \\ L^z_\tau &= \tfrac{1}{2}\left[a^\dagger_\tau\sqrt{Nc_\tau - a^\dagger_\tau a_\tau} + \sqrt{Nc_\tau - a^\dagger_\tau a_\tau}\,a_\tau\right] \end{aligned} \tag{10}$$

which is valid for fixed angular momentum $\frac{1}{2}Nc_\tau$ and as long as the number of quanta does not exceed Nc_τ. The Bose operators fulfil the usual commutation relation $[a_\tau, a^\dagger_{\tau'}] = \delta_{\tau,\tau'}$. The large N expansion will be consistent provided that we enforce the number of quanta to keep finite. To do so, we align the HP representations (10) along the quasiequilibrium directions of the angular momenta

$$J^x_\tau = \cos\theta_\tau L^x_\tau - \sin\theta_\tau L^z_\tau \qquad J^y_\tau = L^y_\tau \qquad J^z_\tau = \sin\theta_\tau L^x_\tau + \cos\theta_\tau L^z_\tau \tag{11}$$

in every block. We may now expand the evolution operator $W^s = N\cdot W^s_0 + \sqrt{N}\cdot W^s_1 + W^s_2 + \mathrm{O}(1/\sqrt{N})$ and find

$$\begin{aligned} W^s_0 &= \frac{1}{2}\sum_\tau c_\tau\left(\frac{\cos\theta_\tau}{\cosh K_\tau} + \sin\theta_\tau\tanh K_\tau - 1\right) \\ W^s_1 &= \frac{\sqrt{c_\tau}}{2\cosh K_\tau}(\cos\theta_\tau\sinh K_\tau - \sin\theta_\tau) \\ &\quad\times\left(a^\dagger_\tau + a_\tau - \sum_\eta\sum_{\mu=1}^{p}\tau^\mu\eta^\mu\sqrt{c_\eta c_\tau}\,\frac{\cos\theta_\eta}{\cosh K_\tau}(a^\dagger_\eta + a_\eta)\right) \end{aligned} \tag{12}$$

where

$$K_\tau = \beta\sum_{\mu=1}^{p}\tau^\mu\sum_\eta c_\eta\sin\theta_\eta\eta^\mu. \tag{13}$$

Both W^s_0 and W^s_1 vanish when the angles are such as $\sin\theta_\tau = \tanh K_\tau$ for all blocks $\boldsymbol{\tau}$. Expression (13) implies that the 2^p angles may be computed through the knowledge of the p parameters $m^\mu = \sum_\eta c_\eta\sin\theta_\eta\eta^\mu$. It turns out that the resulting self-consistent equations for the m^μs are identical to the equilibrium saddle-point condition (3), as could be expected from the above discussion.

As a consequence, the smallest eigenvalues gather together in different families $\mathcal{F}$, in one-to-one correspondence with the stationary points of the free energy (2). In each family $\mathcal{F}$, we are left with the spin-wave part of the evolution operator

$$W^s_2 = -\tfrac{1}{4}\sum_{\tau,\tau'}[(a_\tau + a^\dagger_\tau)(O^2 - 1)_{\tau,\tau'}(a_{\tau'} + a^\dagger_{\tau'})] - \sum_\tau[a^\dagger_\tau a_\tau + \tfrac{1}{2}(1 - O_{\tau,\tau})] \tag{14}$$

where the $2^p\times 2^p$ matrix O reads

$$O_{\tau,\tau'} = 1_{\tau,\tau'} - \beta\sum_{\mu=1}^{p}\left(\frac{\sqrt{c_\tau}}{\cosh K_\tau}\boldsymbol{\tau}^\mu\right)\left(\frac{\sqrt{c_{\tau'}}}{\cosh K_{\tau'}}\boldsymbol{\tau}'^\mu\right). \tag{15}$$

At least $2^p - p$ eigenvalues of O equal one. A straigthforward calculation shows that the difference between the traces of the nth powers of O and of the free-energy Hessian matrix, see (2) and (3),

$$H_{\mu,\nu} = \left.\frac{\partial^2 f}{\partial m^\mu \partial m^\nu}\right|_{\mathcal{F}} \tag{16}$$

equals $2^p - p$ for all integers n. In the above expression, the Hessian depends on the particular family $\mathcal{F}$, i.e. saddle point under consideration. Therefore, the *a priori* nonunit eigenvalues of O, hereafter called λ_js, $j = 1, \ldots, p$ are precisely the eigenvalues of the second derivatives matrix of the static free energy.

To achieve the calculation of the spectrum of relaxation times, we now perform an orthogonal transformation over the Bose operators to diagonalize O and obtain

$$W_2^s = -\sum_{j=1}^{p}[\tfrac{1}{4}(\lambda_j^2 - 1)(b_j + b_j^\dagger)^2 + b_j^\dagger b_j + \tfrac{1}{2}(1 - \lambda_j)] - \sum_{j=p+1}^{2^p} b_j^\dagger b_j. \tag{17}$$

The first term on the r.h.s. of (17) includes p interacting bosons which can be decoupled by means of a Bogoliubov transformation

$$\begin{pmatrix} c_j \\ c_j^\dagger \end{pmatrix} = \begin{pmatrix} \cosh\phi_j & \sinh\phi_j \\ \sinh\phi_j & \cosh\phi_j \end{pmatrix} \begin{pmatrix} b_j \\ b_j^\dagger \end{pmatrix} \tag{18}$$

where $\phi_j = \frac{1}{2}\ln|\lambda_j|$. We can finally rewrite W_2^s in a completely diagonal form,

$$W_2^s = -\sum_{i=1}^{p}|\lambda_j|(c_j^\dagger c_j + \Delta_j) - \sum_{j=p+1}^{2^p} b_j^\dagger b_j \tag{19}$$

where $\Delta_j = 1$ if $\lambda_j < 0$ and $\Delta_j = 0$ if $\lambda_j > 0$. In this form, the slow eigenmodes of the Glauber matrix are seen to correspond to the eigenvectors of a quantum harmonic oscillator and can be easily calculated [5]. Note that transformation (18) requires that no eigenvalue vanishes. Indeed, for a marginal saddle-point $\mathcal{F}$, the large N expansion has to take into account that the number of quanta $a^\dagger a$ becomes of the order of $\sqrt{N}$.

We can follow the same procedure to diagonalize W^s in the subspace labelled by angular momenta $\{J_\tau = \frac{1}{2}Nc_\tau - j_\tau\}$ where all j_τs are finite; we obtain for W_τ^s expression (19) plus the constant term $-j_\tau$. Therefore, the smallest eigenvalues do not belong to such subspaces. If the j_τs are not finite as N goes to infinity, the diverging evolution operators W_0^s and W_1^s can never be simultaneously cancelled through adequate rotations.

Starting from a given initial condition, the evolution in the parameter space $\{m^\mu\}$ will be deterministic [12] and follow the macroscopic law

$$\frac{\mathrm{d}}{\mathrm{d}t}m^\mu(t) = -\frac{\partial f}{\partial m^\mu}(\{m^\nu(t)\}) \tag{20}$$

giving rise to a gradient descent of the free-energy landscape down to some local minimum. When the number p of patterns increases, this free-energy landscape becomes very complex, for example, the number of stationary points scales exponentially with p [7]. Each fixed point has a basin of attraction; a point for a local maximum, a line for points with only one stable direction, ..., a manifold of dimension $N - I$ for a stationary point of index I. However, because of pure entropic effects, some stationary points have much wider basins of attractions than other ones, that is the number of microscopic configurations associated with them may be extremely large. For instance, a random and unbiased initial condition would fall into the basin of the paramagnetic saddle-point $m^\mu = 0$, $\forall\mu$.

Generally speaking, when the starting point lies in a basin of attraction of a stationary point $\{m^\mu\}_\mathcal{F}$ of the free energy, the system relaxes toward this stationary point. Later on, the order parameters remain stuck to their saddle-point values for all but finite times, even if $\mathcal{F}$ is unstable from a thermodynamical standpoint. This description is reminiscent of the *scenario* proposed in [13]. The timescale which we have calculated here govern the $1/\sqrt{N}$ fluctuations around the stationary point. Fluctuations in the stable directions, that are related to positive λ_js, decay with relaxation times equal to $1/\lambda_j$ and eventually equilibrate ($\Delta_j = 0$). As for fluctuations along the unstable directions associated with negative λ_js, they exponentially grow ($\Delta_j = 1$) with some characteristic 'escape' times $1/|\lambda_j|$. In both cases, the number of affected spins is of the order of $\sqrt{N}$ and the intensive quantities remain unchanged.

References

[1] For a recent review, see Bouchaud J P, Cugliandolo L, Mézard M and Kurchan J 1997 *Out of Equilibrium Dynamics in Spin-glasses and Other Glassy Systems* ed A P Young (Singapore: World Scientific) and references therein

[2] Glauber R J 1963 *J. Math. Phys.* **4** 294–307

[3] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792

[4] Mélin R 1996 *J. Physique* **6** 469
Yoshino H, Hukushima K and Takayama H 1997 Relaxational modes and ageing in the Glauber dynamics of the Sherrington–Kirkpatrick model *Prog. Theor. Phys.* supplement 126

[5] Ruijgrok Th W, Tjon J A 1973 *Physica* **65** 539

[6] Hopfield J J 1982 *Proc. Natl. Acad. Sci. USA* **79** 2554

[7] Amit D J 1989 *Modeling Brain Function* (Cambridge: Cambridge University Press)

[8] Mattis D C 1976 *Phys. Lett.* **56A** 421

[9] van Hemmen J L, Grensing D, Huber A and Kühn R 1986 *Z. Phys.* B **65** 53

[10] Holstein T and Primakoff H 1940 *Phys. Rev.* **58** 1908

[11] Mattis D C 1981 *The Theory of Magnetism (Springer Series in Solid State Sciences)* vol 17 (Berlin: Springer)

[12] Van Kampen N G 1961 *Can. J. Phys.* **39** 551

[13] Kurchan J and Laloux L 1996 *J. Phys. A: Math. Gen.* **29** 1929