A Scaling Limit of the Glauber Dynamics for a Class of Neural Network Models

Hans Koch *
Department of Mathematics, University of Texas at Austin
Austin, TX 78712

Abstract. We consider the time evolution of d mean field variables for networks of N neurons whose connection matrices J_N have d distinct rows. Certain assumptions are made about the large N behavior of J_N, which guarantee the convergence of a free energy density function. These assumptions are known to be satisfied e.g. in the Hopfield model with p stored patterns, for $d = 2^p$. It is proved that in a scaling limit, where N tends to infinity and d stays fixed, the time evolution approaches that of a diffusion process in $\mathbb{R}^d$. This process describes in detail, and for times up to $\mathcal{O}(N^{3/2})$ iterations, the dynamics of the mean field fluctuations near a local minimum of the free energy density.

* Supported in Part by the National Science Foundation under Grant No. DMS–8802590.

1. Introduction and Results

We consider the Glauber dynamics for a systems of N Ising spins $S_i \in \{1,-1\}$, in a limit where $N \to \infty$. For any given N, the dynamics is governed by a Hamiltonian of the form

$$H_N(S) \equiv -\frac{1}{2}\sum_{i,j=1}^{N} J_{Nij} S_i S_j \,, \tag{1.1}$$

and certain assumptions will be made about the sequences $N \mapsto J_{Nij}$ of coupling constants.

The Glauber dynamics at temperature $1/\beta$ is a random walk on the set $\mathcal{S}_N \equiv \{1,-1\}^N$ of spin configurations, whose transition probabilities $W_N(S',S)$ are defined as follows. If S and S' are two spin configurations such that S' can be obtained from S by flipping a single spin, then

$$W_N(S',S) \equiv \frac{1}{N}\cdot\frac{\omega(S')}{\omega(S)+\omega(S')}\,, \tag{1.2}$$

where ω denotes the normalized Gibbs distribution

$$\omega(S) \equiv \Bigl(\sum_{S\in\mathcal{S}_N} e^{-\beta H_N(S)}\Bigr)^{-1}\cdot e^{-\beta H_N(S)}\,. \tag{1.3}$$

All other off–diagonal entries of the transition matrix W_N are zero, and the diagonal entries are determined by requiring that the probabilities $W_N(S',S)$, for fixed S, add up to one.

Notice that in the limit $\beta \to \infty$, we get $W_N(S',S) = 0$ whenever $H_N(S') > H_N(S)$. Thus, at zero temperature, every local minimum of H_N is an attractor for the dynamics W_N. For many of the Hamiltonians used in neural network models, a similar behavior is observed at low but positive temperatures, on the time scale of a typical numerical experiment (retrieval of a pattern). Our goal is to give a detailed and quantitative description of this behavior, at the level of mean field fluctuations. We note that on a very large time scale the situation is different: At positive temperature, if ρ is any probability distribution on $\mathcal{S}_N$ then $(W_N)^n\rho$ converges to ω as $n \to \infty$, where

$$(W_N\rho)(S') \equiv \sum_{S\in\mathcal{S}_N} W_N(S',S)\,\rho(S)\,. \tag{1.4}$$

Assumption (A1). *For any N, the matrix J_N is symmetric and has at most d distinct rows, where d is some fixed positive integer.*

This assumption holds e.g. in neural network models with a fixed number of stored patterns $\xi^1,\ldots,\xi^p \in \mathcal{S}_N$, if the learning rule leads to coupling constants of the form $J_{Nij} = F_N[(\xi_i^1,..,\xi_i^p);(\xi_j^1,..,\xi_j^p)]$, for some symmetric function F_N on $\mathcal{S}_p \times \mathcal{S}_p$ [1]. This includes e.g. the Hopfield model [2], and models based on the learning rule of Personnaz, *et. al.* [3–5]. It is easy to see that the matrix J_N has at most $|\mathcal{S}_p| = 2^p$ distinct rows, if the number of patterns is fixed to p.

If (A1) is satisfied, we can find a partition of the set $\{1, \ldots, N\}$ into non–empty blocks $\mathcal{B}_{N,1}, \ldots, \mathcal{B}_{N,d}$, such that two rows i_1 and i_2 of J_N agree whenever i_1 and i_2 lie in the same block. Following an idea of [1], we associate a mean field m_k to each block $\mathcal{B}_{N,k}$. Define

$$m_k(S) \equiv \frac{1}{\lambda_k} \sum_{i \in \mathcal{B}_{N,k}} S_i - \eta_k \,, \qquad 1 \leq k \leq d \,, \tag{1.5}$$

where η_k and $\lambda_k > 0$ are scaling parameters which will be specified later. Clearly, the energy $H_N(S)$ of a spin configuration S only depends on the values of the mean fields $m_k(S)$. Furthermore, as was observed in [6], the Glauber dynamics induces a random walk on the set $\mathcal{Z}_N$ of all possible mean field configurations $m(S)$: If for every $z \in \mathcal{Z}_N$ we denote by $\mathcal{S}_N(z)$ the set of spin configurations S for which $m(S) = z$, then the random walk on $\mathcal{Z}_N$ is given by the transition probabilities

$$\overline{W}_N(z', z) \equiv \sum_{S' \in \mathcal{S}_N(z')} W_N(S', S) \,, \qquad S \in \mathcal{S}_N(z) \,. \tag{1.6}$$

These probabilities do not depend on the choice of $S \in \mathcal{S}_N(z)$. Consequently, the matrix $\overline{W}_N$ describes the time evolution of a distribution $U_N \rho$ on $\mathcal{Z}_N$,

$$(U_N \rho)(z) \equiv \left[\prod_{k=1}^{d} 2\lambda_k \right] \sum_{S \in \mathcal{S}_N(z)} \rho(S) \,, \tag{1.7}$$

in the sense that $U_N W_N = \overline{W}_N U_N$.

Denote by $\ell^1(\mathcal{Z}_N)$ the vector space of real–valued functions on $\mathcal{Z}_N$, equipped with the following norm.

$$\|f\|_N \equiv \left[\prod_{k=1}^{d} 2\lambda_k \right]^{-1} \sum_{z \in \mathcal{Z}_N} |f(z)| \,. \tag{1.8}$$

Then it is easy to see that $\overline{\omega} \equiv U_N \omega$ is the unique normalized equilibrium distribution for $\overline{W}_N$, and that $\overline{\omega}$ is the Gibbs distribution for the free energy function

$$\overline{H}_N(z) \equiv H_N(S) - \frac{1}{\beta} \ln |\mathcal{S}_N(z)| \,, \qquad S \in \mathcal{S}_N(z) \,. \tag{1.9}$$

In our analysis of the mean field dynamics $\overline{W}_N$, we will focus on particular regions by choosing the scaling parameters η_k and λ_k appropriately. In order to specify these regions, let us first look at the global picture by setting temporarily $\eta_k = 0$ and $\lambda_k = |\mathcal{B}_{N,k}|$. Define

$$P_{km}^{(N)} \equiv \frac{1}{N} \sum_{i \in \mathcal{B}_{N,k}} \sum_{j \in \mathcal{B}_{N,m}} J_{Nij} \,, \qquad 1 \leq k \leq d \,. \tag{1.10}$$

Assumption (A2). *The relative block sizes $\ell_k^{(N)} \equiv |\mathcal{B}_{N,k}|/N$ and the mean field coupling constants $P_{km}^{(N)}$ have limits $\ell_k^{(\infty)} > 0$ and $P_{km}^{(\infty)}$, respectively, as $N \to \infty$, for $1 \leq k, m \leq d$.*

Under this assumption, the configuration space $\mathcal{Z}_N$ is approximately $[-1,1]^d$ for large N, and the free energy function can be written as $\overline{H}_N = Nh^{(N)} + \mathcal{O}(\ln N)$, where

$$h^{(N)}(y) \equiv -\frac{1}{2}\sum_{k,m=1}^{d} P_{km}^{(N)} y_k y_m + \frac{1}{\beta}\sum_{k=1}^{d}\ell_k^{(N)}\int_0^{y_k} dt\,\tanh^{-1}(t) - \frac{\ln 2}{\beta}\,. \tag{1.11}$$

Since the equilibrium distribution $\overline{\omega}$ is approximately proportional to $\exp(-Nh^{(N)})$, it is clear that on a sufficiently large time scale, a random walk for $\overline{W}_N$ spends most of its time near local minima of the free energy density $h^{(N)}$. We now restrict our attention to a small neighborhood of such a local minimum. Define a function $h^{(\infty)}$ on $[-1,1]^d$ by setting $N=\infty$ in the equation (1.11).

Assumption (A3). *There is a point $y^{(\infty)}$ in $(-1,1)^d$ where the free energy density $h^{(\infty)}$ has a local minimum with a positive second derivative $D^2h^{(\infty)}(y^{(\infty)})$.*

We note that the last two assumptions are known to hold in the Hopfield model with an "unbiased" set of p stored patterns, for all $\beta \neq 1$. In this model, we have for example $\ell_k^{(\infty)} = 1/d$ for $1 \le k \le d = 2^p$, and if $\beta > 1$, then the function $h^{(\infty)}$ has $2p$ global minima, which are in 1–1 correspondence with the stored patterns and their negatives; other local minima exist as well, but only at sufficiently low temperatures [6–8]. For results on the Hopfield model at zero temperature we refer to [9–11].

Assuming (A3), we can find a sequence of points $y^{(N)}$ in $(-1,1)^d$, converging to $y^{(\infty)}$, such that $h^{(N)}$ has a local minimum at $y^{(N)}$ with a positive second derivative $D^2h^{(N)}(y^{(N)})$, if N is sufficiently large. The corresponding Hessian matrix is $(Q^{(N)})^{-1}B^{(N)}$, where

$$Q_{km}^{(N)} \equiv \frac{1-\left(y_k^{(N)}\right)^2}{\ell_k^{(N)}}\delta_{km}\,, \qquad B^{(N)} \equiv \mathrm{I} - \beta Q^{(N)}P^{(N)}\,. \tag{1.12}$$

Here, and in what follows, the case $N=\infty$ is meant to be included whenever we define a quantity $X^{(N)}$. For every $t>0$ define

$$C(t)^{(N)} \equiv \left(\mathrm{I} - e^{-2tB^{(N)}}\right)\left(B^{(N)}\right)^{-1}Q^{(N)}\,, \tag{1.13}$$

and let $x \mapsto \phi^{(N)}(t,x)$ be the normalized Gaussian distribution on $\mathbb{R}^d$ with mean zero and covariance $C(t)^{(N)}$, i.e.

$$\phi^{(N)}(t,x) \equiv \left[2\pi\cdot\det\left(C(t)^{(N)}\right)\right]^{-1/2}\exp\left(-\frac{1}{2}\left\langle x,\left(C(t)^{(N)}\right)^{-1}x\right\rangle\right), \tag{1.14}$$

In order to describe the dynamics near $y^{(N)}$, for large N, we will now fix the values of the scaling parameters η_k and λ_k as follows.

$$\eta_k = \sqrt{N}y_k^{(N)}\,, \qquad \lambda_k = \sqrt{N}\ell_k^{(N)}\,, \qquad 1 \le k \le d\,. \tag{1.15}$$

If f is an integrable function on $\mathbb{R}^d$, denote by $J_N f$ the function on $\mathcal{Z}_N$ whose value at z coincides with the average of $f(x)$ in the region $|x_k - z_k| \leq 1/\lambda_k$, $1 \leq k \leq d$, for every point $z \in \mathcal{Z}_N$.

Theorem 1.1. *Assume that (A1)–(A3) hold, and let $N \mapsto n(N)$ be a sequence of positive integers such that $N^{-3/2} n(N) \to 0$ as $N \to \infty$. Then for every integrable function f on $\mathbb{R}^d$, we have*

$$\big\| \big(\overline{W}_N\big)^{n(N)} J_N f - J_N T^{(\infty)}_{n(N)/N} f \big\|_N \longrightarrow 0 \qquad \text{as} \quad N \to \infty\,, \tag{1.16}$$

where

$$\big(T_t^{(N)} f\big)(x) \equiv \int d^d u \, \phi^{(N)}\big(t, x - e^{-tB^{(N)}} u\big) f(u)\,, \qquad t > 0\,. \tag{1.17}$$

The maps $T_t^{(\infty)}$, for $t > 0$, define a contraction semigroup on $L^1(\mathbb{R}^d, d^d x)$ which by equation (1.16) can be considered the scaling limit for the mean field dynamics near $y^{(\infty)}$. The generator for $T_t^{(N)}$, $N \leq \infty$, is the differential operator

$$H^{(N)}_{osc} \equiv -\mathrm{tr}\big(B^{(N)}\big) - \sum_{k=1}^{d} \big(B^{(N)} x\big)_k \frac{\partial}{\partial x_k} - \sum_{k=1}^{d} Q^{(N)}_{kk} \frac{\partial^2}{\partial x_k^2}\,. \tag{1.18}$$

To be more precise, denote by $\Omega_0^{(N)}(x)$ the limit as $t \to \infty$ of $\phi^{(N)}(t, x)$. Then the following holds.

Proposition 1.2. *If f is an integrable function on $\mathbb{R}^d$, then for $N = \infty$ or $N < \infty$ sufficiently large, the function $(t, x) \mapsto \big(T_t^{(N)} f\big)(x)$ is C^∞ on $(0, \infty) \times \mathbb{R}^d$, and*

(a) $\partial/\partial t \; T_t^{(N)} f = -H^{(N)}_{osc} T_t^{(N)} f\,, \quad$ *for* $t > 0$.

(b) $\big\| T_t^{(N)} f - f \big\|_{L^1} \to 0 \quad$ *as* $t \downarrow 0$.

(c) $\big\| T_t^{(N)} f - \int d^d x f(x) \cdot \Omega_0^{(N)} \big\|_{L^1} \to 0 \quad$ *as* $t \to \infty$.

Part (c) of this proposition, combined with (1.16), shows that the local minima $y^{(N)}$ of $h^{(N)}$ behave like attractors for the mean field dynamics, on a time scale of $n = o(N^{3/2})$ iterations. With respect to the rescaled time $t = n/N$, the rates of attraction are given approximately by the eigenvalues of $H^{(\infty)}_{osc}$. These eigenvalues can be computed explicitly, since $H^{(\infty)}_{osc} + \frac{1}{2}\mathrm{tr}(B^{(\infty)})$ is the Hamiltonian of d uncoupled harmonic oscillators whose frequencies are the eigenvalues of the matrix $B^{(\infty)}$ (which by (A3) are all positive). For more details, including the proofs of Theorem 1.1 and Proposition 1.2, we refer to Section 2.

Remarks.

1) What we call "mean field dynamics" depends on the choice of the scaling parameters λ_k and η_k in the definition (1.5) of the mean field. Our choice (1.15) leads to a description of the time evolution of probability distributions for the Markov chain $Z^{(N)}(n)$, defined by the equation

$$y_k^{(N)} + \tfrac{1}{\sqrt{N}} Z_k^{(N)}(n) = M_k^{(N)}(n) \equiv |\mathcal{B}_{N,k}|^{-1} \sum_{i \in \mathcal{B}_{N,k}} S_i(n)\,. \tag{1.19}$$

$Z^{(N)}$ represents the fluctuation (or finite size correction) of the mean field $M^{(N)}$, at a point $y^{(N)}$ in $(-1,1)^d$ which minimizes locally the free energy density $h^{(N)}$. Theorem 1.1 compares $Z^{(N)}(tN)$ with a Markov process $Z = Z(t)$ that satisfies the stochastic differential equation

$$\dot{Z} = -B^{(N)}Z + \big(Q^{(N)}\big)^{1/2}\dot{\mathcal{W}}\,, \tag{1.20}$$

where $\mathcal{W}$ is a d–dimensional Wiener Process.

2) If the block average $M^{(N)}$ is initially distributed near some arbitrary point in $(-1,1)^d$, then the constant vector $y^{(N)}$ in (1.19) should be replaced by a sequence of vectors $Y^{(N)}(n) \approx \mathbb{E}[M^{(N)}(n)]$. We expect that $Y^{(N)}(tN) \approx Y(t)$ and $Z^{(N)}(tN) \approx Z(t)$, with $Y = Y(t)$ and $Z = Z(t)$ satisfying the equation [12]

$$\dot{Y}_k = \tanh\big(\beta(\ell_k^{(N)})^{-1}(P^{(N)}Y)_k\big) - Y_k \tag{1.21}$$

and (1.20), respectively. In this case, the (now time dependent) coefficients in equation (1.20) are obtained by substituting $Y + \dot{Y}$ for $y^{(N)}$ in the definition (1.12).

3) If the free energy density $h^{(N)}$ has local minima at more than one point, then almost every trajectory $\big(M^{(N)}(0), M^{(N)}(1), \ldots\big)$, which starts in the "free energy valley" of one of these points, will eventually escape that valley. However, this happens on a time scale which is much larger than the one we have considered here: The expected value of the first exit time T_N should be roughly the exponential of the free energy difference between the bottom and the lowest passage out of the valley. Since it takes a large number of (essentially independent) trials before one of the escape attempts succeeds, the distribution of $T_N/\mathbb{E}[T_N]$ should be exponential with mean one, in the limit $N \to \infty$. Such a result has been proved for the Curie–Weiss model (the Hopfield model with one stored pattern) at low temperature, by Cassandro *et. al.* [13].

2. Proofs

We start with some immediate consequences of (A1)–(A3). For any given pair (ℓ, P) in $\mathbb{R}^d \times \mathrm{L}(\mathbb{R}^d, \mathbb{R}^d)$ consider the equation $Dh_{\ell,P}(y) = 0$, where $h_{\ell,P}$ is the function on $(-1,1)^d$ defined by

$$h_{\ell,P}(y) \equiv -\frac{1}{2}\sum_{k,m=1}^{d} P_{km}y_k y_m + \frac{1}{\beta}\sum_{k=1}^{d}\ell_k \int_0^{y_k} dt\,\tanh^{-1}(t) - \frac{\ln 2}{\beta}\,. \tag{2.1}$$

By Assumption (A3), this equation has a solution $y = y^{(\infty)}$ for $(\ell, P) = (\ell^{(\infty)}, P^{(\infty)})$, and $D^2 h_{\ell^{(\infty)},P^{(\infty)}}(y^{(\infty)})$ is positive. Thus, by the implicit function theorem, there exists an open neighborhood $\mathcal{O}$ of $(\ell^{(\infty)}, P^{(\infty)})$, and a C^1 map $(\ell, P) \mapsto y(\ell, P)$ on $\mathcal{O}$, such that $y(\ell^{(\infty)}, P^{(\infty)}) = y^{(\infty)}$ and $Dh_{\ell,P}(y(\ell,P)) = 0$ for all $(\ell, P) \in \mathcal{O}$. Assuming also (A2), it follows that if N is sufficiently large then $y^{(N)} \equiv y(\ell^{(N)}, P^{(N)})$ lies in $(-1,1)^d$ and

$D^2h^{(N)}(y^{(N)}) = D^2h_{\ell^{(N)},P^{(N)}}(y^{(N)})$ is positive. By using the fact that $(Q^{(N)})^{-1}B^{(N)}$ is the Hessian matrix for $D^2h^{(N)}(y^{(N)})$, together with the trivial identity

$$\left(Q^{(N)}\right)^{-1/2}B^{(N)}\left(Q^{(N)}\right)^{1/2} = \left(Q^{(N)}\right)^{1/2}\left[\left(Q^{(N)}\right)^{-1}B^{(N)}\right]\left(Q^{(N)}\right)^{1/2}, \qquad (2.2)$$

we see that the matrix $B^{(N)}$ is diagonalizable, and that all eigenvalues of $B^{(N)}$ are positive. The identity (2.2) also shows that the matrix $C(t)^{(N)}$ is positive for all $t > 0$; hence the function $\phi^{(N)}$ is well defined on $(0,\infty) \times \mathbb{R}^d$. Here, we have assumed that N is either finite and sufficiently large, or $N = \infty$. Under the same assumption, and for any $(u,s) \in \mathbb{R}^d \times (0,\infty)$, we can define

$$\psi_{u,s}^{(N)}(x) \equiv \phi^{(N)}\left(s, x - e^{-sB^{(N)}}u\right), \qquad x \in \mathbb{R}^d. \qquad (2.3)$$

Lemma 2.1. *Given any compact subset $\mathcal{U}$ of $\mathbb{R}^d$, and constants $\delta, s_o > 0$, there is a positive integer N_o such that*

$$\sum_{k,m=1}^{d}\left|C(s)_{km}^{(N)} - C(s)_{km}^{(\infty)}\right|, \ \left\|\psi_{u,s}^{(N)} - \psi_{u,s}^{(\infty)}\right\|_{L^1} \ \leq \ \delta, \qquad (2.4)$$

whenever $u \in \mathcal{U}$, $s \geq s_o$, and $N \geq N_o$.

Proof. The bound on $C(s)^{(N)} - C(s)^{(\infty)}$ follows since the map $(\ell, P) \mapsto y(\ell, P)$ is C^1 near $(\ell^{(\infty)}, P^{(\infty)})$. To prove the second bound, we define functions $\psi_{u,s}^{[r]}$ as in (2.3) and (1.14), but with the matrices $C(t)^{(N)}$ and $B^{(N)}$ replaced by the interpolated matrices $C(t)^{[r]} \equiv (1-r)C(t)^{(\infty)} + rC(t)^{(N)}$ and $B^{[r]} \equiv (1-r)B^{(\infty)} + rB^{(N)}$, respectively, for $0 \leq r \leq 1$. From the definition (1.13) it is clear that the matrices $C(s)^{[r]}$, $\left(C(s)^{[r]}\right)^{-1}$, and their s–derivatives, are uniformly bounded in the region $0 \leq r \leq 1$, $s \geq s_o$. As a consequence, we find e.g. that, given any integer $n \geq 0$, there are positive constants K and τ, such that

$$\left|\frac{d^n}{ds^n}\psi_{u,s}^{[r]}(z)\right|, \ \left|\frac{\partial^n}{\partial z_1^{n_1}\cdots\partial z_d^{n_d}}\psi_{u,s}^{[r]}(z)\right| \ \leq \ Ke^{-\tau\|z\|^2}, \qquad n_1 + \ldots + n_d = n, \qquad (2.5)$$

for all $r \in [0,1]$, $s \in [s_o, \infty)$, $u \in \mathcal{U}$, $z \in \mathbb{R}^d$, and for N sufficiently large. If $u = 0$, the difference $\psi_{u,s}^{(N)} - \psi_{u,s}^{(\infty)}$ can now be bounded as claimed in (2.4), by using the identity

$$\frac{d}{dr}\phi_{0,s}^{[r]}(z) = \sum_{k,m=1}^{d}\left[\frac{1}{2}\frac{d}{dr}\left(C(s)^{[r]}\right)_{km}\right]\frac{\partial^2}{\partial z_k \partial z_m}\phi_{0,s}^{[r]}(z). \qquad (2.6)$$

The extension to $u \in \mathcal{U}$ is straightforward. $\blacksquare$

We continue by estimating the elements of the matrix $N(\overline{W}_N - \mathrm{I})$, which will turn out to be a discrete approximation for $H_{osc}^{(N)}$. A straightforward computation, using the definition of $\overline{W}_N$, shows that the off–diagonal elements of $\overline{W}_N$ are given by

$$\overline{W}_N(z,z') = \begin{cases} w_{N,k}^{\pm}\big(\frac{1}{\sqrt{N}}, z\big), & \text{if } z' = z \pm \frac{2}{\sqrt{N}\ell_k^{(N)}}\delta_k\,, \\ 0, & \text{otherwise,} \end{cases} \tag{2.7}$$

where δ_k denotes the unit vector in the direction of the k-th coordinate axis, and where

$$w_{N,k}^{\pm}(\varepsilon,x) \equiv \frac{\frac{1}{2}\ell_k^{(N)}\Big[1 \pm y_k^{(N)} \pm \varepsilon x_k + 2\varepsilon^2\big(\ell_k^{(N)}\big)^{-1}\Big]}{1+\exp\Big[\pm 2\beta\big(\ell_k^{(N)}\big)^{-1}\big[P^{(N)}\big(y^{(N)}+\varepsilon x\big)\big]_k + 2\varepsilon^2\beta\big(\ell_k^{(N)}\big)^{-2}P_{kk}^{(N)}\Big]}\,, \tag{2.8}$$

for all $x \in \mathbb{R}^d$ and $\varepsilon \in \mathbb{R}$. The diagonal entries of $\overline{W}_N$ are determined by the fact that $\overline{W}_N$ is a matrix of transition probabilities:

$$\overline{W}_N(z,z) = 1 - \sum_{k=1}^{d} w_{N,k}^{+}\big(\frac{1}{\sqrt{N}}, z - \frac{2}{\sqrt{N}\ell_k^{(N)}}\delta_k\big) - \sum_{k=1}^{d} w_{N,k}^{-}\big(\frac{1}{\sqrt{N}}, z + \frac{2}{\sqrt{N}\ell_k^{(N)}}\delta_k\big)\,. \tag{2.9}$$

Here, as well as in (2.7), the arguments of $\overline{W}_N$ are assumed to be points in $\mathcal{Z}_N$, i.e. vectors in $\mathbb{R}^d$ whose components z_k range between $\sqrt{N}\big(-y_k^{(N)} \pm 1\big)$, in increments of $2\big(\sqrt{N}\ell_k^{(N)}\big)^{-1}$. If f is a function on $\mathcal{Z}_N$, define

$$\begin{aligned} (D_{N,k}f)(z) &\equiv \frac{1}{4}\sqrt{N}\ell_k^{(N)}\Big[f\big(z + \frac{2}{\sqrt{N}\ell_k^{(N)}}\delta_k\big) - f\big(z - \frac{2}{\sqrt{N}\ell_k^{(N)}}\delta_k\big)\Big], \\ (\Delta_{N,k}f)(z) &\equiv \frac{1}{4}N\big(\ell_k^{(N)}\big)^2\Big[f\big(z + \frac{2}{\sqrt{N}\ell_k^{(N)}}\delta_k\big) + f\big(z - \frac{2}{\sqrt{N}\ell_k^{(N)}}\delta_k\big) - 2f(z)\Big], \end{aligned} \tag{2.10}$$

with the convention that $f(x) = 0$ whenever $x \notin \mathcal{Z}_N$. Then, by using the identities (2.7) and (2.9), we find the following representation for $\overline{W}_N f$.

$$\begin{aligned} \big(\overline{W}_N f\big)(z) = f(z) + \frac{1}{N}\sum_{k=1}^{d}\Big[& a_k\big(1/\sqrt{N}, z\big)f(z) + b_k\big(1/\sqrt{N}, z\big)\big(D_{N,k}f\big)(z) \\ & + q_k\big(1/\sqrt{N}, z\big)\big(\Delta_{N,k}f\big)(z)\Big]\,, \end{aligned} \tag{2.11}$$

where the coefficients a_k, b_k, q_k are given by the equations

$$\begin{aligned} a_k(\varepsilon,x) &\equiv \varepsilon^{-2}\Big[w_{N,k}^{+}(\varepsilon,x) - w_{N,k}^{-}\big(\varepsilon, x + \frac{2\varepsilon}{\ell_k^{(N)}}\delta_k\big) + w_{N,k}^{-}(\varepsilon,x) - w_{N,k}^{+}\big(\varepsilon, x - \frac{2\varepsilon}{\ell_k^{(N)}}\delta_k\big)\Big], \\ b_k(\varepsilon,x) &\equiv 2\big(\varepsilon\ell_k^{(N)}\big)^{-1}\big[w_{N,k}^{+}(\varepsilon,x) - w_{N,k}^{-}(\varepsilon,x)\big], \\ q_k(\varepsilon,x) &\equiv 2\big(\ell_k^{(N)}\big)^{-2}\big[w_{N,k}^{+}(\varepsilon,x) + w_{N,k}^{-}(\varepsilon,x)\big]. \end{aligned} \tag{2.12}$$

Lemma 2.2. *There is a polynomial p such that for $0 < |\varepsilon| < 1$, $1 \le k \le d$, $x \in \mathbb{R}^d$, and for $N < \infty$ sufficiently large,*

$$\left|a_k(\varepsilon, x) - B_{kk}^{(N)}\right| \,,\; \left|b_k(\varepsilon, x) - \left(B^{(N)}x\right)_k\right| \,,\; \left|q_k(\varepsilon, x) - Q_{kk}^{(N)}\right| \le \varepsilon p(x). \tag{2.13}$$

Proof. The estimate for a_k is obtained by using a Taylor expansion about $\varepsilon = 0$ for the functions $w_{N,k}^{\pm}$, with a remainder of the order ε^3. After substituting this expansion into the definition (2.12) of $a_k(\varepsilon, x)$, one finds that the negative powers of ε cancel, due to the identity

$$\exp\left[2\beta\left(\ell_k^{(N)}\right)^{-1}\left(P^{(N)}y^{(N)}\right)_k\right] = \frac{1 + y_k^{(N)}}{1 - y_k^{(N)}}, \qquad 1 \le k \le d\,, \tag{2.14}$$

which expresses the fact that $y^{(N)}$ is a critical point of $h^{(N)}$. The coefficient of ε^0 turns out to be $B_{kk}^{(N)}$, and the modulus of the remainder satisfies (2.13) since all derivatives of $w_{N,k}^{\pm}$ are polynomially bounded. The estimates for b_k and q_k are obtained similarly.

■

Next, we compare the two maps $\overline{W}_N$ and $T_{1/N}^{(N)}$, when applied to the functions $\psi_{u,s}^{(N)}$. For any function f on $\mathbb{R}^d$, denote by $R_N f$ the restriction of f to the set $\mathcal{Z}_N$.

Lemma 2.3. *Given a compact subset $\mathcal{U}$ of $\mathbb{R}^d$, and $s_o > 0$, there are constants $K, \kappa > 0$ such that if N is sufficiently large,*

$$\left|\left(\overline{W}_N R_N \psi_{u,s}^{(N)}\right)(z) - \psi_{u,s+1/N}^{(N)}(z)\right| \le K N^{-3/2} e^{-\kappa\|z\|^2}\,, \tag{2.15}$$

for all $u \in \mathcal{U}$, $s \ge s_o$, and $z \in \mathcal{Z}_N$.

Proof. By using Taylor's formula, together with the bound (2.5), we obtain

$$\left|\left(D_{N,k}R_N\psi_{u,s}^{(N)}\right)(z) - \frac{\partial}{\partial x_k}\psi_{u,s}^{(N)}(z)\right|, \left|\left(\Delta_{N,k}R_N\psi_{u,s}^{(N)}\right)(z) - \frac{\partial^2}{\partial x_k^2}\psi_{u,s}^{(N)}(z)\right| \le \frac{K'}{\sqrt{N}} e^{-2\kappa\|z\|^2}, \tag{2.16}$$

uniformly on $\{(u,s) : u \in \mathcal{U}, s \ge s_o\}$, for some constants $K', \kappa > 0$. Here, we have assumed that z is an interior point of $\mathcal{Z}_N$, i.e. a point in $\mathcal{Z}_N$ for which $z \pm 2\left(\sqrt{N}\ell_k^{(N)}\right)^{-1}\delta_k$ lies in $\mathcal{Z}_N$ for all k. If z is a non–interior point of $\mathcal{Z}_N$, then we have $\|z\|^2 > cN$ for some fixed constant $c > 0$, so that the right hand side of (2.5) may be replaced by

$$const \cdot e^{-\kappa c N} e^{-2\kappa\|z\|^2}. \tag{2.17}$$

The resulting inequality shows that (2.16) also holds for non–interior points of $\mathcal{Z}_N$. By using the identity (2.11), the left hand side of (2.15) can be written as follows

$$
\begin{aligned}
&\big|(\overline{W}_N R_N \psi_{u,s}^{(N)})(z) - \psi_{u,s+1/N}^{(N)}(z)\big| \\
&= \Big|\frac{1}{N}\Big\{\sum_{k=1}^{d} a_k\big(1/\sqrt{N},z\big)\psi_{u,s}^{(N)}(z) + \sum_{k=1}^{d} b_k\big(1/\sqrt{N},z\big)\big(D_{N,k}\psi_{u,s}^{(N)}\big)(z) \\
&+ \sum_{k=1}^{d} q_k\big(1/\sqrt{N},z\big)\big(\Delta_{N,k}\psi_{u,s}^{(N)}\big)(z) + \big(H_{osc}^{(N)}\psi_{u,s}^{(N)}\big)(z)\Big\} \\
&- \Big[\psi_{u,s+1/N}^{(N)}(z) - \psi_{u,s}^{(N)}(z) - \frac{1}{N}\frac{d}{ds}\psi_{u,s}^{(N)}(z)\Big] - \frac{1}{N}\Big(H_{osc}^{(N)}\psi_{u,s}^{(N)} + \frac{d}{ds}\psi_{u,s}^{(N)}\Big)(z)\Big|.
\end{aligned}
\tag{2.18}
$$

The term in curly brackets can be bounded by using (2.16), (2.5), and Lemma 2.2 :

$$
\Big|\frac{1}{N}\Big\{\ldots\Big\}\Big| \le \frac{const}{N\sqrt{N}} p(z) e^{-2\kappa\|z\|^2} \le const N^{-3/2} e^{-\kappa\|z\|^2}. \tag{2.19}
$$

To bound the term $[\ldots]$ in (2.18), we apply Taylor's formula, together with (2.5), and obtain

$$
\big|[\ldots]\big| \le const N^{-2} e^{-\kappa\|z\|^2}. \tag{2.20}
$$

The assertion now follows, since $\big(H_{osc}^{(N)} + d/ds\big)\psi_{u,s}^{(N)} = 0$, as can be verified by direct computation; see also Lemma 2.6.

■

Due to the factor $N^{-3/2}$ in (2.15), we can repeat this bound n times to estimate $\big(\overline{W}_N\big)^n$ for $n = o(N^{3/2})$.

Lemma 2.4. *Let $N \mapsto n(N)$ be a sequence of positive integers such that $N^{-3/2}n(N) \to 0$ as $N \to \infty$. Then for every compact subset $\mathcal{U}$ of $\mathbb{R}^d$, and for every $\delta, s_o > 0$, there exists a constant $N_o > 0$ such that*

$$
\big\|\big(\overline{W}_N\big)^{n(N)} J_N \psi_{u,s}^{(\infty)} - J_N T_{n(N)/N}^{(\infty)} \psi_{u,s}^{(\infty)}\big\|_N \le \delta\,, \tag{2.21}
$$

whenever $u \in \mathcal{U}$, $s \ge s_o$ and $N \ge N_o$.

Proof. First, we note that $T_t^{(N)}\psi_{u,s}^{(N)} = \psi_{u,s+t}^{(N)}$ for $s, t > 0$ and for N large (including ∞). By using the triangle inequality, we estimate the left hand side of (2.21) as follows.

$$
\begin{aligned}
\big\|\big(\overline{W}_N\big)^n J_N \psi_{u,s}^{(\infty)} - J_N \psi_{s+n/N}^{(\infty)}\big\|_N &\le \big\|\big(\overline{W}_N\big)^n J_N\big(\psi_{u,s}^{(\infty)} - \psi_{u,s}^{(N)}\big)\big\|_N \\
&+ \big\|J_N\big(\psi_{u,s+n/N}^{(N)} - \psi_{u,s+n/N}^{(\infty)}\big)\big\|_N + \big\|\big(\overline{W}_N\big)^n\big(J_N - R_N\big)\psi_{u,s}^{(N)}\big\|_N \\
&+ \big\|\big(R_N - J_N\big)\psi_{u,s+n/N}^{(N)}\big\|_N + \big\|\big(\overline{W}_N\big)^n R_N \psi_{u,s}^{(N)} - R_N \psi_{u,s+n/N}^{(N)}\big\|_N\,,
\end{aligned}
\tag{2.22}
$$

where $n = n(N)$. The first two terms on the right hand side of (2.22) can be bounded as desired by using Lemma 2.1, together with the fact that the maps $J_N :\ L^1(\mathbb{R}^d) \to \ell^1(\mathcal{Z}_N)$

and $\overline{W}_N : \ell^1(\mathcal{Z}_N) \to \ell^1(\mathcal{Z}_N)$ are of norm ≤ 1. To get the necessary estimates for the third and fourth term, we can apply the mean value theorem and then bound the derivatives using (2.5). The last term in (2.22) is controlled by using Lemma 2.3 :

$$\begin{aligned} &\big\|\big(\overline{W}_N\big)^n R_N \psi^{(N)}_{u,s} - R_N \psi^{(N)}_{u,s+n/N}\big\|_N \\ &\qquad = \Big\| \sum_{m=1}^{n} \big(\overline{W}_N\big)^{n-m} \big[\overline{W}_N R_N \psi^{(N)}_{s+(m-1)/N} - R_N \psi^{(N)}_{s+m/N}\big] \Big\|_N \\ &\qquad \leq \sum_{m=1}^{n} \big\|\overline{W}_N R_N \psi^{(N)}_{s+(m-1)/N} - R_N \psi^{(N)}_{s+m/N}\big\|_N \leq n \cdot const\, N^{-3/2}\,. \end{aligned} \tag{2.23}$$

■

Proof of Theorem 1.1. Assume that the hypothesis of Theorem 1.1. are satisfied, and let $\delta > 0$. Since by Proposition 1.2, $\big\|T^{(\infty)}_s f - f\big\|_{L^1} \to 0$ as $s \downarrow 0$, we can find $s > 0$, and functions $c_1 \psi^{(\infty)}_{u_1,s}, \ldots, c_m \psi^{(\infty)}_{u_m,s}$, such that

$$\Big\| f - \sum_{j=1}^{m} c_j \psi^{(\infty)}_{u_j,s} \Big\|_{L^1} < \frac{\delta}{4}. \tag{2.24}$$

By Lemma 2.4, there exists $N_o > 0$ such that

$$|c_j| \cdot \big\|\big(\overline{W}_N\big)^{n(N)} J_N \psi^{(\infty)}_{u_j,s} - J_N T^{(\infty)}_{n(N)/N} \psi^{(\infty)}_{u_j,s}\big\|_N \leq \frac{\delta}{2m} \tag{2.25}$$

for $1 \leq j \leq m$, and $N \geq N_o$. As a consequence, we have

$$\begin{aligned} &\big\|\big(\overline{W}_N\big)^{n(N)} J_N f - J_N T^{(\infty)}_{n(N)/N} f\big\|_N \\ &\qquad \leq \Big\| \Big[\big(\overline{W}_N\big)^{n(N)} J_N - J_N T^{(\infty)}_{n(N)/N}\Big] \Big(f - \sum_{j=1}^{m} c_j \psi^{(\infty)}_{u_j,s}\Big) \Big\|_N \\ &\qquad + \sum_{j=1}^{m} |c_j| \cdot \big\|\big(\overline{W}_N\big)^{n(N)} J_N \psi^{(\infty)}_{u_j,s} - J_N T^{(\infty)}_{n(N)/N} \psi^{(\infty)}_{u_j,s}\big\|_N \\ &\qquad \leq 2 \cdot \frac{\delta}{4} + m \cdot \frac{\delta}{2m} = \delta\,. \end{aligned} \tag{2.26}$$

for all $N \geq N_o$. Since $\delta > 0$ was arbitrary, the assertion follows.

■

In what follows, N is assumed to fixed and sufficiently large (or ∞), such that the matrices $Q^{(N)}$ and $\big(Q^{(N)}\big)^{-1} B^{(N)}$ are well defined and positive. To simplify notation, the index ${}^{(N)}$ will be omitted from now on. Since Q and $Q^{-1}B$ are positive, we can find a positive diagonal matrix Λ and an orthogonal matrix V such that

$$V^*\Lambda V = \big(Q^{-1}B\big)^{1/2} Q \big(Q^{-1}B\big)^{1/2}. \tag{2.27}$$

In order to discuss the properties of the time evolution $f \mapsto T_t f$, it is convenient to introduce the Hilbert space $\mathcal{H} \equiv L^2\big(\mathbb{R}^d, \Omega_0(x)^{-1} d^d x\big)$ with the inner product

$$\langle f, g\rangle \equiv \int d^d x\, \Omega_0(x)^{-1} f(x)\overline{g(x)}\,, \qquad f, g \in \mathcal{H}, \tag{2.28}$$

where Ω_0 is the function given by

$$\Omega_0(x) = \big[2\pi \cdot \det\big(B^{-1}Q\big)\big]^{-1/2} \exp\Big(-\frac{1}{2}\big\langle x, Q^{-1}Bx\big\rangle\Big)\,, \qquad x \in \mathbb{R}^d. \tag{2.29}$$

Denote by $\mathcal{D}$ the (dense) subspace of $\mathcal{H}$ spanned by the function Ω_0 and all its partial derivatives, and define operators $A_1, \ldots, A_d$ on $\mathcal{D}$ as follows.

$$\big(A_m f\big)(x) \equiv \sum_{k=1}^{d} M_{mk} \Big[\frac{\partial}{\partial x_k} + \big(Q^{-1}Bx\big)_k\Big] f(x)\,, \qquad f \in \mathcal{D}, \quad x \in \mathbb{R}^d, \tag{2.30}$$

where $M \equiv V\big(Q^{-1}B\big)^{-1/2}$. A short calculation shows that the adjoint of A_m satisfies

$$\big(A_m^* f\big)(x) = -\sum_{k=1}^{d} M_{mk} \frac{\partial}{\partial x_k} f(x)\,, \qquad f \in \mathcal{D}, \quad x \in \mathbb{R}^d. \tag{2.31}$$

Lemma 2.5. *For every $f \in \mathcal{D}$, and for $1 \le m, k \le d$, the following holds.*
(a) $\big[A_m, A_k\big] f = \big[A_m^*, A_k^*\big] f = 0$.
(b) $\big[A_m, A_k^*\big] f = \delta_{mk} f$.
(c) $H_{osc} f = \sum_{k=1}^{d} \Lambda_{kk} A_k^* A_k f$.

Proof. Part (a) follows immediately from (2.31), and the other two identities are easy to prove by using that $M^*\Lambda M = Q$ and $MQ^{-1}BM^* = \mathrm{I}$.

■

This shows that H_{osc} is (up to a constant) the Hamiltonian of d harmonic oscillators with frequencies Λ_{kk}. The eigenvalues of H_{osc} are linear combinations $E_\alpha \equiv \sum_{k=1}^{d} \alpha_k \Lambda_{kk}$ of these frequencies, with nonnegative integer coefficients α_k, and the corresponding eigenvectors are the Hermite functions Ω_α,

$$\Omega_\alpha \equiv \prod_{k=1}^{d} \big(A_k^*\big)^{\alpha_k} \Omega_0\,, \tag{2.32}$$

which form an orthogonal basis for $\mathcal{H}$.

Lemma 2.6. *On $\mathcal{H}$ we have $T_t = e^{-tH_{osc}}$ for all $t > 0$.*

Proof. Let $t > 0$. As the discussion above shows, H_{osc} is essentially self–adjoint on $\mathcal{D}$ and positive. Thus $e^{-tH_{osc}}$ is well defined and bounded. Furthermor, by using Hölder's inequality, we find that

$$\big|\langle g, T_t f\rangle\big|^2 \leq \|g\|^2 \|f\|^2 \int d^d x \int d^d u \, \phi\big(t, x - e^{-tB} u\big)^2 \Omega_0(u)/\Omega_0(x)\,, \tag{2.33}$$

for every $g, f \in \mathcal{H}$. This proves that T_t defines a bounded linear operator on $\mathcal{H}$; in fact $\|T_t\| \leq 1$ since the double integral in (2.33) equals 1. Thus, it suffices to verify that $T_t \Omega_\alpha = e^{-tH_{osc}} \Omega_\alpha$ for all α. The case $\alpha = 0$ is easy to check. To get the other cases we proceed by induction. By using the identity $MB^* M^{-1} = \Lambda$, we obtain

$$\begin{aligned}
\big(T_t A_m^* \Omega_\alpha\big)(x) &= -\sum_{k=1}^{d} M_{mk} \int d^d u \, \phi\big(t, x - e^{-tB} u\big) \frac{\partial}{\partial u_k} \Omega_\alpha(u) \\
&= -\sum_{k,l=1}^{d} M_{mk} \big(e^{-tB}\big)_{lk} \frac{\partial}{\partial x_l} \int d^d u \, \phi\big(t, x - e^{-tB} u\big) \Omega_\alpha(u) \\
&= \sum_{m,n=1}^{d} \big(M e^{-tB^*} M^{-1}\big)_{mn} \big(A_n^* T_t \Omega_\alpha\big)(x) \\
&= e^{-t\Lambda_{mm}} e^{-tE_\alpha} \big(A_m^* \Omega_\alpha\big)(x) = \big(e^{-tH_{osc}} A_m^* \Omega_\alpha\big)(x)\,,
\end{aligned} \tag{2.34}$$

for $x \in \mathbb{R}^d$, $1 \leq m \leq d$, and for every multi–index α. This proves the assertion.

■

Proof of Proposition 1.2. The smoothness of the function $(t, x) \mapsto \big(T_t f\big)(x)$ follows from the definition (1.17) of T_t, given the fact that at every point in $(0, \infty) \times \mathbb{R}^d$, the partial derivatives of the functions $\phi_u : (t, x) \mapsto \phi\big(t, x - e^{-tB} u\big)$ are bounded uniformly in u. Furthermore, by Lemma 2.6, we have $\big(H_{osc} + \partial/\partial t\big)\phi_u = 0$ for every $u \in \mathbb{R}^d$. This proves assertion (a). To prove (b), fix $\varepsilon > 0$. Since $\mathcal{H}$ can be identified with a dense linear subspace of $L^1\big(\mathbb{R}^d, d^d x\big)$, we can find a function $g \in \mathcal{H}$ such that $\|f - g\|_{L^1} < \epsilon/3$. Thus,

$$\begin{aligned}
\big\|(T_t - \mathrm{I}) f\big\|_{L^1} &\leq \big\|(T_t - \mathrm{I}) g\big\|_{L^1} + \big\|(T_t - \mathrm{I})(f - g)\big\|_{L^1} \\
&\leq \big\|\big(e^{-tH_{osc}} - \mathrm{I}\big) g\big\|_{\mathcal{H}} + 2 \cdot \varepsilon/3 \leq \varepsilon\,,
\end{aligned} \tag{2.35}$$

if $t > 0$ is sufficiently small. Here, we have used that T_t is a bounded liner operator on $L^1\big(\mathbb{R}^d, d^d x\big)$ of norm one, and that $\|g\|_{L^1} = \langle \Omega_0, |g|\rangle \leq \|g\|_{\mathcal{H}}$, for every function $g \in \mathcal{H}$. Since $\varepsilon > 0$ was arbitrary, (b) follows. The proof of (c) is similar.

■

References

[1] J.L. van Hemmen, R. Kühn, *Nonlinear Neural Networks*, Phys. Rev. Lett. **57**, 913 (1986).

[2] J.J. Hopfield, *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[3] L. Personnaz, I. Guyon, G. Dreyfus, *Information Storage and Retrieval in Spin–Glass like Neural Networks*, J. Physique Lett. **46**, L-359 (1985).

[4] S. Diederich, M. Opper, *Learning of Correlated Patterns in Spin–Glass Networks by Local Learning Rules*, Phys. Rev. Lett. **58**, 949 (1987).

[5] K.W. Berryman, M.E. Inchiosa, A.M. Jaffe, S.A. Janowsky, *Convergence of an Iterative Neural Network Learning Algorithm for Linearly Dependent Patterns*, Harvard University Preprint HUTMP 89/B237 (1989).

[6] H. Koch, J. Piasko, *Some Rigorous Results on the Hopfield Neural Network Model*, J. Stat. Phys. **55**, 903 (1989).

[7] D.J. Amit, *The Properties of Models of Simple Neural Networks*, in "Heidelberg Colloquium on Glassy Dynamics (1986)", J.L. van Hemmen and I. Morgenstern (eds.), Lecture Notes in Physics **275**, 430 (1987).

[8] J.L. van Hemmen, D. Grensing, A. Huber, R. Kühn, *Nonlinear Neural Networks:*
I. General Theory, J. Stat. Phys. **50**, 231 (1987).
II. Information Processing, J. Stat. Phys. **50**, 259 (1987).

[9] R.J. McEliece, E.C. Posner, E.R. Rodemich, S.S. Venkatesh, *The Capacity of the Hopfield Associative Memory*, IEEE Transactions on Information Theory **33**, 461 (1986).

[10] C.M. Newman, *Memory Capacity in Neural Network Models: Rigorous Lower Bounds*, Neural Networks **1**, 223 (1988).

[11] J. Komlós, R. Paturi, *Convergence Results in the Hopfield Model*, Preprint, UC San Diego (1987).

[12] U. Riedel, R. Kühn, J.L. van Hemmen, *Temporal Sequences and Chaos in Neural Nets*, Phys. Rev. A **38**, 1105 (1988).

[13] M. Cassandro, A. Galves, E. Olivieri, M.E. Vares, *Metastable Behavior of Stochastic Dynamics: A Pathwise Approach*, J. Stat. Phys. **35**, 603 (1984).