Journal Pre-proof

Generalized Guerra's interpolation schemes for dense associative neural networks

Elena Agliari, Francesco Alemanno, Adriano Barra, Alberto Fachechi

Please cite this article as: E. Agliari, F. Alemanno, A. Barra et al., Generalized Guerra's interpolation schemes for dense associative neural networks. *Neural Networks* (2020), doi: https://doi.org/10.1016/j.neunet.2020.05.009.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Generalized Guerra's interpolation schemes for dense associative neural networks

Elena Agliari[a] **Francesco Alemanno**[b,c] **Adriano Barra**[b,d] **Alberto Fachechi**[b,d]

[a] *Dipartimento di Matematica* Guido Castelnuovo, *Italy*

[b] *Dipartimento di Matematica e Fisica* Ennio De Giorgi, *Università del Salento, Italy*

[c] *C.N.R. Nanotec Lecce, Italy*

[d] *I.N.F.N., Sezione di Lecce, Italy*

ABSTRACT: In this work we develop analytical techniques to investigate a broad class of associative neural networks set in the high-storage regime. These techniques translate the original statistical-mechanical problem into an analytical-mechanical one which implies solving a set of partial differential equations, rather than tackling the canonical probabilistic route. We test the method on the classical Hopfield model − where the cost function includes only two-body interactions (i.e., quadratic terms) − and on the "relativistic" Hopfield model − where the (expansion of the) cost function includes p-body (i.e., of degree p) contributions. Under the replica symmetric assumption, we paint the phase diagrams of these models by obtaining the explicit expression of their free energy as a function of the model parameters (i.e., noise level and memory storage). Further, since for non-pairwise models ergodicity breaking is non necessarily a critical phenomenon, we develop a fluctuation analysis and find that criticality is preserved in the relativistic model.

Contents

1   Introduction

In the last few years, the introduction of a new-generation of computational devices [1, 2], the development of better performing learning schemes [3, 4] and the implementation of massive data repositories allowing efficient training [5, 6] naturally raised the quest for appropriate mathematical tools and frameworks able to describe, address and possibly explain the related information processing capabilities [4].

Since the 80s', the statistical mechanics of spin glasses [7] has been playing a primary role in the investigation of neural networks, as for both their learning skills [8, 9] and their retrieval properties [10, 11]. Along the past decades, beyond the bulk of results achieved via heuristic approaches like the replica trick [7, 10], a considerable amount of rigorous results exploiting alternative routes was also developed (see e.g. [12–26]).

This paper goes in the last direction and aims bridging between a statistical-mechanics approach and a mechanical approach, the latter possibly more familiar to a wider community. In particular, we shall focus on the Hopfield model (as a reference framework) and its "relativistic" generalization [27]. The latter, exhibiting a cost function that is an (infinite and convergent) series of monomials in the microscopic variables (i.e., the neural activities) offers not only a perfect "playground" where testing our methods, but also an interesting example of dense architectures [28, 29]. The translation of the statistical-mechanics problem into a mechanical framework is based on the analogy between the variational principles in statistical mechanics (e.g., maximum entropy and minimal energy) and the least-action principle in analytical mechanics: this route was already paved for ferromagnetic models [30–32], for spin glasses [33–35] and for (simpler) neural networks [12, 27]. A main advantage is that it allows painting the phase diagrams of the model under study by relying upon tools originally developed in the analytical counterpart (i.e. mainly differential equation techniques) almost without any knowledge of the statistical mechanics of complex systems.

We stress that in this work we deal with models in the high-storage regime, namely, we set the number P of stored patterns to be retrieved scaling linearly with the number N of neurons making

up the network, i.e., the load $\alpha := \lim_{N \to \infty} P/N$ is non vanishing. This regime is notoriously difficult to handle rigorously, given the emergence of glassy phenomena impairing the retrieval capabilities of the system (see e.g., [10, 11]). In fact, the "relativistic" Hopfield model has already been addressed in the low-storage regime, namely when $\alpha = 0$, while when $\alpha > 0$ the control of the model was still unsatisfactory [27]. Here we overcome that limit by generalizing Guerra's interpolation technique in order to be able to work with a broad class of Hamiltonians, that are expressable as a smooth function $\mathcal{F}$ of the squared Mattis magnetization $\boldsymbol{m}$, that is, $H = \mathcal{F}(\boldsymbol{m}^2) = \mathcal{F}(\sum_\mu m_\mu^2)$, where m_μ measures the retrieval of the μ-th stored pattern ($\mu = 1, ..., P$). Remarkably, this general model also includes a subset of "dense" networks, where "dense" refers to the presence of p-wise ($p > 2$) interactions among neurons.

This paper is structured as follows: as a preamble, in Sec. 2, we solve the classical Hopfield model in the high-storage regime via the mechanical analogy and, by solving a set of differential equations, we recover the standard Amit-Gutfreund-Sompolinsky (AGS) quenched free-energy and related self-consistencies for the order parameters [10]; while in the original statistical mechanical setting this result emerges through a rather tricky and lengthy procedure, here it is just an almost trivial exercise. In Sec. 3 the "relativistic" Hopfield model in the high-storage regime is introduced and embedded in its statistical mechanical framework. In Sec. 4, by relying upon a generalization of the mechanical analogy, we calculate explicitly its quenched free energy (at the replica symmetric level, as standard) and we provide a first picture of the phase diagram of the model. Next, in Sec. 5, we study the fluctuations of the order parameters to inspect ergodicity breaking and its possible critical nature. By combining results from Secs. 4 and 5 a consistent picture of the phase diagram of the relativistic model is finally obtained. Then, Sec. 6 is left for conclusions and speculations on the broad applicability of these new mathematical techniques. Finally, in Appendix A we present the solution of the "relativistic" Hopfield model via replica trick.

2   The classical Hopfield network via the mechanical analogy

The Hopfield model is the prototype model for neural networks performing pattern recognition; soon after the seminal paper by J.J. Hopfield [36], the statistical-mechanics analysis by Amit, Gutfreund and Sompolinsky [37, 38] highlighted a very rich phenomenology which prompted an upsurge of interest among physicists and mathematicians. Admittedly, the statistical-mechanics approach is strongly grounded on frustrated systems with quenched disorder (i.e., *spin-glasses*), typically studied just in some fields of Theoretical Physics and Applied Mathematics. Alternative investigation routes, possibly based on more widespread frameworks, may allow for a deeper comprehension and may facilitate interdisciplinarity with a new wave of ideas and inspirations. Hereafter we solve the Hopfield model by exploiting differential equation theory, standard in many transport problems.
In the next definitions we introduce the fundamental quantities we will deal with.

Definition 1. *Set $\alpha \in \mathbb{R}^+$ and let $\boldsymbol{\sigma} \in \{-1, +1\}^N$ be a configuration of N binary neurons. Given $P = \alpha N$ random patterns $\{\boldsymbol{\xi}^\mu\}_{\mu=1,...,P}$, each made of N digital entries identically and independently drawn with probability $P(\xi_i^\mu = +1) = P(\xi_i^\mu = -1) = 1/2$, for $i = 1, ..., N$, the classical Hopfield cost-function (or "Hamiltonian" to preserve a physical jargon) is*

$$H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) := -\frac{1}{2N} \sum_{\mu=1}^P \sum_{i,j=1}^{N,N} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j. \tag{2.1}$$

Definition 2. *The partition function related to the cost-function (1) is given by*

$$Z_N(\beta, \boldsymbol{\xi}) := \sum_{\{\boldsymbol{\sigma}\}} \exp\left[-\beta H_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\right] = \sum_{\{\boldsymbol{\sigma}\}} \exp\left(\frac{\beta}{2N} \sum_{\mu=1}^{P} \sum_{i,j=1}^{N,N} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j\right),$$ (2.2)

where $\beta \geq 0$ is a real number accounting for the fast noise (or "inverse temperature") in the network and such that for $\beta \to 0$ the probability distribution for the neural configuration is uniformly spread while for $\beta \to \infty$ it is sharply peaked at the minima of the cost function (1).

Definition 3. *The Boltzmann average induced by the partition function (2) is denoted with $\omega_{\boldsymbol{\xi}}$ and, for an arbitrary observable $O(\boldsymbol{\sigma})$, is defined as*

$$\omega_{\boldsymbol{\xi}}(O) := \frac{\sum_{\boldsymbol{\sigma}} O(\boldsymbol{\sigma}) e^{-\beta H_N(\boldsymbol{\sigma}|\boldsymbol{\xi})}}{Z_N(\beta, \boldsymbol{\xi})}.$$ (2.3)

This can be further averaged over the realization of the $\boldsymbol{\xi}$'s, also referred to as quenched average and denoted as $\mathbb{E}$, to get

$$\langle O \rangle := \mathbb{E}\omega_{\boldsymbol{\xi}}(O).$$ (2.4)

Further, we introduce the product state $\Omega_{s,\boldsymbol{\xi}} = \omega_{\boldsymbol{\xi}}^{(1)} \times \omega_{\boldsymbol{\xi}}^{(2)} \times ... \times \omega_{\boldsymbol{\xi}}^{(s)}$ over s replicas of the system, characterized by the same realization $\boldsymbol{\xi}$ of disorder. In the following, we shall use the product state over two replicas only, hence we shall neglect the index s without ambiguity; also, to lighten the notation, we shall also omit the subscript $\boldsymbol{\xi}$ in $\omega_{\boldsymbol{\xi}}$ and in $\Omega_{\boldsymbol{\xi}}$. Thus, for an arbitrary observable $O(\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)})$

$$\langle O \rangle := \mathbb{E}\Omega(O) = \mathbb{E}\frac{\sum_{\boldsymbol{\sigma}} O(\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}) e^{-\beta[H_N(\boldsymbol{\sigma}^{(1)}|\boldsymbol{\xi}) + H_N(\boldsymbol{\sigma}^{(2)}|\boldsymbol{\xi})]}}{Z_N^2(\beta, \boldsymbol{\xi})},$$ (2.5)

where $\boldsymbol{\sigma}^{(1,2)}$ is the configuration pertaining to the replica labelled as $1, 2$.

Definition 4. *The intensive quenched pressure of the classical Hopfield model (1) is defined as*

$$A_N(\alpha, \beta) := \frac{1}{N}\mathbb{E} \log Z_N(\beta, \boldsymbol{\xi}),$$ (2.6)

and its thermodynamic limit is referred to as

$$A(\alpha, \beta) := \lim_{N \to \infty} A_N(\alpha, \beta).$$ (2.7)

Remark 1. *The quenched pressure $A_N(\alpha, \beta)$ is related to the quenched free-energy $F_N(\alpha, \beta)$ as $A_N(\alpha, \beta) = -\beta F_N(\alpha, \beta)$ and here it is chosen for mathematical convenience.*

Before proceeding it is worth recalling the *universality property* of the quenched noise in spin-glasses: when dealing with mean-field spin-glasses [39] and mean-filed bipartite spin-glasses [40], the coupling distribution is proved not to affect the resulting pressure, provided that it is centered, symmetrical and with finite variance. This property was extended to the Hopfield model in [12], where it was shown that the quenched noise contribution appearing in the expression for the pressure (which stems from the $P-1$ non-retrieved patterns and which tends to inhibit retrieval), exhibits the very same shape, regardless of the nature of the pattern entries (that is, digital − e.g., Boolean − or analog − e.g., Gaussian).

Focusing on pure retrieval states, we will assume without loss of generality that the candidate pattern to be retrieved $\boldsymbol{\xi}^1$ is a Boolean vector of N entries, while $\boldsymbol{\xi}^\mu$, $\mu = 2, ..., P$ are real vectors whose N entries are i.i.d. standard Gaussian. Accordingly, the average $\mathbb{E}$ acts as a Boolean average over $\boldsymbol{\xi}^1$ and as a Gaussian average over $\boldsymbol{\xi}^2 \cdots \boldsymbol{\xi}^P$.

Definition 5. *The order parameters used to describe the macroscopic behavior of the model are the standard ones [10, 11], namely, the Mattis magnetization*

$$m(\boldsymbol{\sigma}) := m(\boldsymbol{\sigma}|\boldsymbol{\xi}) := \frac{1}{N} \sum_{i=1}^{N} \xi_i^1 \sigma_i \qquad (2.8)$$

to quantify the retrieval capabilities of the network, and the two-replica overlap in the $\boldsymbol{\sigma}$'s variables

$$q_{12}(\boldsymbol{\sigma}) := \frac{1}{N} \sum_{i=1}^{N} \sigma_i^{(1)} \sigma_i^{(2)} \qquad (2.9)$$

to quantify the level of slow noise the network must cope with when performing pattern recognition. Further, as an additional set of variables $\{\tau_\mu\}_{\mu=1,...,P}$ shall be introduced (vide infra), we accordingly define the related two-replica overlap

$$p_{11}(\boldsymbol{\tau}) := \frac{1}{P} \sum_{\mu=1}^{P} \tau_\mu^{(1)} \tau_\mu^{(1)}, \quad p_{12}(\boldsymbol{\tau}) := \frac{1}{P} \sum_{\mu=1}^{P} \tau_\mu^{(1)} \tau_\mu^{(2)} \qquad (2.10)$$

which as well captures the level of the noise due to pattern inference.

Definition 6. *Given five real interpolating parameters (t, x, y, z, w) to be set a posteriori and $N + P$ auxiliary quenched i.i.d. random variables $J_i \sim \mathcal{N}[0,1], i \in (1, ..., N)$ and $\tilde{J}_\mu \sim \mathcal{N}[0,1], \mu \in (1, ..., P)$, the interpolating pressure – also known as Guerra's action – for the classical Hopfield model (1) in the high-storage regime is defined as*

$$\mathcal{A}_N(t, x, y, z, w) := \frac{1}{N} \mathbb{E} \log \sum_{\{\boldsymbol{\sigma}\}} \int \mathcal{D}\tau \exp\Big[\frac{\sqrt{t}}{\sqrt{N}} \sum_{i,\mu>1}^{N,P} \xi_i^\mu \sigma_i \tau_\mu + \frac{tN}{2} m^2(\boldsymbol{\sigma}) + \sqrt{x} \sum_{i=1}^{N} J_i \sigma_i +$$

$$+ \sqrt{y} \sum_{\mu=1}^{P} \tilde{J}_\mu \tau_\mu + z \sum_{\mu=1}^{P} \frac{\tau_\mu^2}{2} + wNm(\boldsymbol{\sigma}) \Big], \qquad (2.11)$$

where $\mathcal{D}\tau := \prod_{\mu=1}^{P} \frac{e^{-\tau_\mu^2/2}}{\sqrt{2\pi}}$ is the standard Gaussian measure over the P real auxiliary variables τ_μ, $\mu \in (1, ..., P)$, and the expectation $\mathbb{E}$ is now meant over $\boldsymbol{\xi}$, $\boldsymbol{J}$, and $\tilde{\boldsymbol{J}}$.

Remark 2. *We interpret $\boldsymbol{r} = (x, y, z, w)$ as the spatial coordinate in a four-dimensional Euclidean space and t as the temporal variable. According to the mechanical analogy, we will consider the spatial coordinates as functions of t (we will specify their form in the next theorem) and we require that $\boldsymbol{r}(t = \beta) \overset{!}{=} 0$, in such a way that, when the temporal coordinate t is equal to β, we recover the original pressure (2.6), namely $\mathcal{A}_N(t = \beta, \boldsymbol{r} = 0) = A_N(\alpha, \beta)$.*

Definition 7. *Since the variables x, y, z, w are functions of t, we will refer to the Boltzmann average stemming from the interpolating system as $\omega_t(\cdot)$, that is,*

$$\omega_t(\cdot) = \frac{\sum_{\boldsymbol{\sigma}} \int \mathcal{D}\tau \ (\cdot) \ B_{t,\boldsymbol{r}}(\boldsymbol{\sigma}, \boldsymbol{\tau})}{Z_{t,\boldsymbol{r}}}, \qquad (2.12)$$

where $B_{t,\boldsymbol{r}}$ is the generalized Boltzmann factor

$$B_{t,\boldsymbol{r}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) := \exp\Big[\frac{\sqrt{t}}{\sqrt{N}} \sum_{i,\mu>1}^{N,P} \xi_i^\mu \sigma_i \tau_\mu + \frac{tN}{2} m^2(\boldsymbol{\sigma}) + \sqrt{x} \sum_i J_i \sigma_i + \sqrt{y} \sum_\mu \tilde{J}_\mu \tau_\mu + z \sum_\mu \frac{\tau_\mu^2}{2} + wNm(\boldsymbol{\sigma}) \Big],$$

and

$$Z_{t,\boldsymbol{r}} = \sum_{\boldsymbol{\sigma}} \int \mathcal{D}\tau B_{t,\boldsymbol{r}}(\boldsymbol{\sigma}, \boldsymbol{\tau}). \qquad (2.13)$$

In analogy with the non-deformed average, we also pose

$$\langle \cdot \rangle_t := \mathbb{E}\omega_t(\cdot), \tag{2.14}$$

where now $\mathbb{E}$ means also the average over the $\boldsymbol{J}$ and $\tilde{\boldsymbol{J}}$ variables.

Lemma 1. *The partial derivatives of the Guerra Action (2.11) w.r.t. t, x, y, z, w can be expressed in terms of the generalized expectations of the order parameters as*

$$\frac{\partial \mathcal{A}_N}{\partial t} = \frac{\alpha}{2}\big[\langle p_{11}\rangle_t - \langle p_{12}q_{12}\rangle_t\big] + \frac{1}{2}\langle m^2\rangle_t \tag{2.15}$$

$$\frac{\partial \mathcal{A}_N}{\partial x} = \frac{1}{2}\big[1 - \langle q_{12}\rangle_t\big], \tag{2.16}$$

$$\frac{\partial \mathcal{A}_N}{\partial y} = \frac{\alpha}{2}\big[\langle p_{11}\rangle_t - \langle p_{12}\rangle_t\big], \tag{2.17}$$

$$\frac{\partial \mathcal{A}_N}{\partial z} = \frac{\alpha}{2}\langle p_{11}\rangle_t, \tag{2.18}$$

$$\frac{\partial \mathcal{A}_N}{\partial w} = \langle m\rangle_t, \tag{2.19}$$

Proof. The previous equations can be obtained by straightforward computations. We provide details for (2.15) only, the other derivations work analogously. Deriving $\mathcal{A}_N(t, \boldsymbol{r})$ with respect to t we get

$$\frac{\partial \mathcal{A}_N}{\partial t} = \frac{1}{N}\mathbb{E}\frac{\sum_{\boldsymbol{\sigma}}\int \mathcal{D}\tau \left[\frac{1}{2\sqrt{tN}}\sum_{i,\mu>1}^{N,P}\xi_i^\mu\sigma_i\tau_\mu + \frac{N}{2}m^2(\boldsymbol{\sigma})\right]B_{t,\boldsymbol{r}}(\boldsymbol{\sigma},\boldsymbol{\tau})}{Z_{t,\boldsymbol{r}}} =$$

$$= \frac{1}{2N\sqrt{tN}}\mathbb{E}\frac{\sum_{\boldsymbol{\sigma}}\int \mathcal{D}\tau \sum_{i,\mu>1}^{N,P}\xi_i^\mu\sigma_i\tau_\mu B_{t,\boldsymbol{r}}(\boldsymbol{\sigma},\boldsymbol{\tau})}{Z_{t,\boldsymbol{r}}} + \frac{1}{2}\langle m^2\rangle_t. \tag{2.20}$$

Now, we still need to handle the first term in the right-hand-side of eq. (2.20); by applying Wick's theorem $\mathbb{E}[\xi_i^\mu f(\boldsymbol{\xi})] = \mathbb{E}\partial_{\xi_i^\mu}f(\boldsymbol{\xi})$ to that term, we get

$$\frac{\partial \mathcal{A}_N}{\partial t} = \frac{1}{2N^2}\sum_{i,\mu>1}^{N,P}\big(\langle\tau_\mu^2\rangle_t - \langle\sigma_i\tau_\mu\rangle_t^2\big) + \frac{1}{2}\langle m^2\rangle_t =$$

$$= \frac{\alpha}{2}\big(\langle p_{11}\rangle_t - \langle p_{12}q_{12}\rangle_t\big) + \frac{1}{2}\langle m^2\rangle_t \tag{2.21}$$

where we used the definitions (2.9) and (2.10) and replaced $P - 1$ with αN (which is valid in the thermodynamic limit). $\qquad\square$

Definition 8. *Under the replica-symmetry (RS) assumption, in the thermodynamic limit, the distribution of the generic order parameter X is centered at its expectation value $\bar{X}(t)$ with vanishing fluctuations for all t, that is, it converges to a Dirac delta: $\lim_{N\to\infty} P_t(X) = \delta(X - \bar{X}(t))$. Otherwise stated, being*

$$\Delta X := X - \bar{X}(t), \tag{2.22}$$

for all $t \in \mathbb{R}^+$ we have $\langle\Delta X(t)\rangle_t \xrightarrow{N\to\infty} 0$ and $\langle\Delta X\Delta Y\rangle_t = 0$ for any generic pair of order parameters X, Y. We therefore define the following expectation values

$$\bar{m}(t) = \lim_{N\to\infty}\langle m\rangle_t, \tag{2.23}$$

$$\bar{q}(t) = \lim_{N\to\infty}\langle q_{12}\rangle_t, \tag{2.24}$$

$$\bar{p}(t) = \lim_{N\to\infty}\langle p_{12}\rangle_t. \tag{2.25}$$

Proposition 1. *The Guerra Action* (2.11) *obeys the following differential equation:*

$$\frac{\partial \mathcal{A}_N}{\partial t} - \alpha \bar{p}(t) \frac{\partial \mathcal{A}_N}{\partial x} - \bar{q}(t) \frac{\partial \mathcal{A}_N}{\partial y} - \left(1 - \bar{q}(t)\right) \frac{\partial \mathcal{A}_N}{\partial z} - \bar{m}(t) \frac{\partial \mathcal{A}_N}{\partial w} = -\frac{\alpha}{2} \bar{p}(t) \left(1 - \bar{q}(t)\right) - \frac{1}{2} \bar{m}^2(t) + V_N(t), \quad (2.26)$$

where

$$V_N(t) = \frac{1}{2} \langle (\Delta m)^2 \rangle_t - \frac{1}{2} \langle \Delta p_{12} \Delta q_{12} \rangle_t, \tag{2.27}$$

and, in the thermodynamic limit, under the RS assumption, one has

$$\frac{\partial \mathcal{A}}{\partial t} - \alpha \bar{p}(t) \frac{\partial \mathcal{A}}{\partial x} - \bar{q}(t) \frac{\partial \mathcal{A}}{\partial y} - \left(1 - \bar{q}(t)\right) \frac{\partial \mathcal{A}}{\partial z} - \bar{m}(t) \frac{\partial \mathcal{A}}{\partial w} = -\frac{\alpha}{2} \bar{p}(t) \left(1 - \bar{q}(t)\right) - \frac{1}{2} \bar{m}^2(t). \tag{2.28}$$

Proof. First, we write the generalized averages $\langle \cdot \rangle_t$ appearing in (2.15) in terms of the Δ operator (2.22) as

$$\langle m^2 \rangle_t = -\bar{m}^2(t) + 2\bar{m} \langle m \rangle_t + \langle (\Delta m)^2 \rangle_t, \tag{2.29}$$

$$\langle p_{12} q_{12} \rangle_t = -\bar{p}(t) \bar{q}(t) + \bar{p}(t) \langle q_{12} \rangle_t + \bar{q}(t) \langle p_{12} \rangle_t + \langle \Delta p_{12} \Delta q_{12} \rangle_t, \tag{2.30}$$

and, by plugging these espressions into (2.15), we get

$$\frac{\partial A_N}{\partial t} = \frac{\alpha}{2} \langle p_{11} \rangle_t - \frac{\alpha}{2} \left(-\bar{p}(t) \bar{q}(t) + \bar{p}(t) \langle q_{12} \rangle_t + \bar{q}(t) \langle p_{12} \rangle_t \right) + \frac{1}{2} \left(\bar{m}^2(t) + 2\bar{m} \langle m \rangle_t \right) + V_N(t). \tag{2.31}$$

Then, the generalized averages appearing in the previous expression can be recast in terms of the spatial derivatives of the Guerra Action $\mathcal{A}_N$ using the relations (2.16)-(2.19); after some trivial computations we get (2.26).

Now, under the RS assumption, the averages $\langle \Delta X \rangle_t$ and $\langle \Delta X \Delta Y \rangle_t$ for any generic pair of order parameters X, Y and all $t \in \mathbb{R}^+$ vanish in the thermodynamic limit. As a consequence, we have $\lim_{N \to \infty} V_N(t) = 0$, and (2.26) reduces to (2.28). $\square$

Proposition 2. *The unique solution to the differential equation* (2.28) *is given by the functional*

$$\begin{aligned}
\mathcal{A}_{[\bar{m}, \bar{p}, \bar{q}]}^{(t, \boldsymbol{r})} &= \log 2 + \mathbb{E}_J \log \cosh \left[w + \int_0^t \bar{m}(\tilde{t}) \, d\tilde{t} + J \sqrt{x + \alpha \int_0^t \bar{p}(\tilde{t}) \, d\tilde{t}} \right] + \\
&\quad + \frac{\alpha}{2} \frac{y + \int_0^t \bar{q}(\tilde{t}) \, d\tilde{t}}{1 - z - \int_0^t [1 - \bar{q}(\tilde{t})] \, d\tilde{t}} - \frac{\alpha}{2} \log \left(1 - z - \int_0^t [1 - \bar{q}(\tilde{t})] \, d\tilde{t} \right) + \\
&\quad - \frac{1}{2} \int_0^t d\tilde{t} \, [\alpha \bar{p}(\tilde{t}) \left(1 - \bar{q}(\tilde{t})\right) + \bar{m}^2(\tilde{t})],
\end{aligned} \tag{2.32}$$

which is a function of the parameters t, x, y, z, w and a functional of the expectactions of the order parameters $(\bar{m}, \bar{p}, \bar{q})(t)$.

Proof. The differential equation in (2.28) can be solved with different methods. The simplest is perhaps the method of characteristics, whose key step is to find the "characteristic curves" of the PDE along which the PDE turns into an ODE. In this case, the characteristic curves are expressed

in terms of the expectaction values of the order parameters as

$$x(t) = x_0 - \alpha \int_{t_0}^{t} \bar{p}(\tilde{t}) d\tilde{t}, \tag{2.33}$$

$$y(t) = y_0 - \int_{t_0}^{t} \bar{q}(\tilde{t}) d\tilde{t}, \tag{2.34}$$

$$z(t) = z_0 - \int_{t_0}^{t} [1 - \bar{q}(\tilde{t})] d\tilde{t}, \tag{2.35}$$

$$w(t) = w_0 - \int_{t_0}^{t} \bar{m}(\tilde{t}) d\tilde{t}, \tag{2.36}$$

where $(x_0, y_0, z_0, w_0) = \boldsymbol{r}(t = t_0) = \boldsymbol{r_0}$. We can easily verify that, along these curves, $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{A}(t, \boldsymbol{r}(t))$ gives us exactly the l.h.s. of (2.28). Therefore, we can write

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{A}(t, \boldsymbol{r}(t)) = -\frac{\alpha}{2}\bar{p}(t)\Big(1 - \bar{q}(t)\Big) - \frac{1}{2}\bar{m}^2(t). \tag{2.37}$$

By integrating between t_0 and t the previous equation we get

$$\mathcal{A}(t, \boldsymbol{r}(t)) = \mathcal{A}(t_0, \boldsymbol{r}_0) - \frac{1}{2} \int_{t_0}^{t} d\tilde{t}\Big[\alpha\bar{p}(\tilde{t})\Big(1 - \bar{q}(\tilde{t})\Big) + \bar{m}^2(\tilde{t})\Big]. \tag{2.38}$$

Here, the parameter t_0 is free and can be chosen in order to simplify the computations. By inspecting (2.11), we see that the most convenient choice is $t_0 = 0$, leaving us with the evaluation of a 1-body problem which can be directly solved:

$$\mathcal{A}(t_0 = 0, \boldsymbol{r}_0) = \log 2 + \mathbb{E}_J \log \cosh \Big[w_0 + J\sqrt{x_0}\Big] + \frac{\alpha}{2}\frac{y_0}{1 - z_0} - \frac{\alpha}{2}\log\Big(1 - z_0\Big). \tag{2.39}$$

Now, from (2.33)-(2.36) we get $x_0 = x(t) + \alpha \int_0^t \bar{p}(\tilde{t})d\tilde{t}$, and similarly for y_0, w_0, z_0. These expressions can then be plugged in (2.39) and the resulting formula can be used in (2.38), finally obtaining (2.32). $\qquad\square$

Theorem 1. *The replica-symmetry quenched pressure of the Hopfield model* (2.6) *in the infinite volume limit is given by*

$$A_{\mathrm{RS}}(\alpha, \beta) = \log 2 + \mathbb{E}_J \log \cosh \Big[\int_0^{\beta} \bar{m}(t)\,dt + J\sqrt{\alpha \int_0^{\beta} \bar{p}(t)\,dt}\Big] + \frac{\alpha}{2}\frac{\int_0^{\beta} \bar{q}(t)\,dt}{1 - \int_0^{\beta}[1 - \bar{q}(t)]\,dt} \tag{2.40}$$
$$-\frac{\alpha}{2}\log\Big(1 - \int_0^{\beta}[1 - \bar{q}(t)]\,dt\Big) - \frac{1}{2}\int_0^{\beta} dt\,[\alpha\bar{p}(t)\big(1 - \bar{q}(t)\big) + \bar{m}^2(t)].$$

The self-consistency equations obtained from the quenched pressure (2.40) *for* $t' \in [0, \beta]$ *are:*

$$\bar{m}(t') = \mathbb{E}_J \tanh \Big[\int_0^{\beta} \bar{m}(t)\,dt + J\sqrt{\alpha \int_0^{\beta} \bar{p}(t)\,dt}\Big],$$

$$\bar{q}(t') = \mathbb{E}_J \tanh^2 \Big[\int_0^{\beta} \bar{m}(t)\,dt + J\sqrt{\alpha \int_0^{\beta} \bar{p}(t)\,dt}\Big], \tag{2.41}$$

$$\bar{p}(t') = \frac{\int_0^{\beta} \bar{q}(t)\,dt}{\Big(1 - \int_0^{\beta}[1 - \bar{q}(t)]\,dt\Big)^2}.$$

Proof. As anticipated, by setting the interpolating parameters as $t = \beta$ and $x, y, z, w = 0$, we recover the Hopfield picture, thus we get (2.40) by evaluating (2.32) in that point. The derivation of the self-consistency equations is based on the extremization of the functional (2.40) with respect to functional variations of the trajectories $m(t), q(t), p(t)$. Therefore, we must set to zero the functional derivatives of (2.40) w.r.t. all the order parameters appearing as functions of the temporal variable t. The self-consistency equation for $\bar{m}(t)$ can be found in the following way:

$$
\begin{aligned}
\frac{\delta A_{\mathrm{RS}}}{\delta \bar{m}(t')} &= \mathbb{E}_J \tanh \Big[\int_0^\beta \bar{m}(t)\,\mathrm{d}t + J\sqrt{\alpha \int_0^\beta \bar{p}(t)\,\mathrm{d}t} \Big] \int_0^\beta \frac{\delta \bar{m}(t)}{\delta \bar{m}(t')}\,\mathrm{d}t - \int_0^\beta \bar{m}(t)\frac{\delta \bar{m}(t)}{\delta \bar{m}(t')}\,\mathrm{d}t = \\
&= \mathbb{E}_J \tanh \Big[\int_0^\beta \bar{m}(t)\,\mathrm{d}t + J\sqrt{\alpha \int_0^\beta \bar{p}(t)\,\mathrm{d}t} \Big] \int_0^\beta \delta(t-t')\,\mathrm{d}t - \int_0^\beta \bar{m}(t)\delta(t-t')\,\mathrm{d}t = \\
&= \mathbb{E}_J \tanh \Big[\int_0^\beta \bar{m}(t)\,\mathrm{d}t + J\sqrt{\alpha \int_0^\beta \bar{p}(t)\,\mathrm{d}t} \Big] - \bar{m}(t') = 0,
\end{aligned}
\tag{2.42}
$$

thus we have

$$
\bar{m}(t') = \mathbb{E}_J \tanh \Big[\int_0^\beta \bar{m}(t)\,\mathrm{d}t + J\sqrt{\alpha \int_0^\beta \bar{p}(t)\,\mathrm{d}t} \Big].
\tag{2.43}
$$

The remaining self-consistency equations can be obtained by following the same lines and are uniquely determined. On the other hand, as well known, the self-consistent equations can display multiple solutions corresponding to neural configurations which are not necessarily stable as they may possibly be minima for the pressure. $\qquad\square$

Corollary 1. *The replica-symmetry quenched pressure* (2.40) *returns the standard AGS theory:*

$$
\begin{aligned}
A_{\mathrm{RS}}(\bar{m}, \bar{p}, \bar{q}) &= \log 2 + \mathbb{E}_J \log \cosh \Big[\beta\bar{m} + J\sqrt{\alpha\beta\bar{p}} \Big] - \frac{\beta}{2}[\alpha\bar{p}(1-\bar{q}) + \bar{m}^2] + \\
&\quad + \frac{\alpha}{2}\frac{\beta\bar{q}}{1-\beta[1-\bar{q}]} - \frac{\alpha}{2}\log\Big(1 - \beta[1-\bar{q}]\Big).
\end{aligned}
\tag{2.44}
$$

As a consequence, the self-consistency equations (2.41) *recovers the standard AGS result:*

$$
\begin{aligned}
\bar{p} &= \frac{\beta\bar{q}}{\Big(1-\beta(1-\bar{q})\Big)^2}, \\
\bar{m} &= \mathbb{E}_J \tanh\Big[\beta\bar{m} + J\sqrt{\alpha\beta\bar{p}}\Big], \\
\bar{q} &= \mathbb{E}_J \tanh^2\Big[\beta\bar{m} + J\sqrt{\alpha\beta\bar{p}}\Big].
\end{aligned}
\tag{2.45}
$$

Proof. By inspecting (2.41), we can see that the r.h.s.'s do not depend on the temporal coordinate t'. Thus, (2.41) actually predict constant functional order parameters, i.e.

$$
\bar{m}(t) = \bar{m}, \quad \bar{q}(t) = \bar{q}, \quad \bar{p}(t) = \bar{p} \quad \forall t \in \mathbb{R}^+.
\tag{2.46}
$$

Thus, by replacing these values in (2.40) and in (2.41) we recover the standard AGS picture. $\qquad\square$

3 The relativistic Hopfield network

The "relativistic" Hopfield model has been introduced in [27], where its investigation was restricted to the low storage (see also [41, 42]). The appellation "relativistic" can be understood via the

mechanical analogy: the Hamiltonian of the classical Hopfield model (1) in terms of the Mattis magnetizations reads as $H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) = -N \sum_{\mu=1}^{P} m_\mu^2/2$ and its statistical mechanical picture can be mapped into the dynamics of a fictitious particle of unitary mass living in a $(P+1)$-dimension Minkowsky space [27]; in this analogy the Mattis magnetization plays the role of the momentum, and the Hopfield Hamiltonian (1) reads as the (classical) kinetic energy for this particle. As the underlying metric of the space-time is Minkowskian, it is quite natural to generalize the classical expression of the kinetic energy into its relativistic counterpart.

Here, our interest in the "relativistic" Hopfield model is two-fold: on the one hand it provides an extension of the classical model which includes also p-body interactions and, in this sense, it constitutes an example of dense networks, on the other hand, it allows us to test our generalized mathematical techniques.

Definition 9. *Set $\alpha, \lambda \in \mathbb{R}^+$ and let $\boldsymbol{\sigma} \in \{-1, +1\}^N$ be a configuration of N binary neurons. Given $P = \alpha N$ random patterns $\{\boldsymbol{\xi}^\mu\}_{\mu=1,...,P}$, each made of N digital entries identically and independently drawn with probability $P(\xi_i^\mu = +1) = P(\xi_i^\mu = -1) = 1/2$, for $i = 1, ..., N$, the "relativistic" Hopfield cost-function (or "Hamiltonian" to preserve a physical jargon) is*

$$H_N^{rel}(\boldsymbol{\sigma}|\boldsymbol{\xi}, \lambda) := H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}, \lambda) = -\frac{N}{\lambda} \sqrt{1 + \lambda \sum_{\mu=1}^{P} \Big(\sum_{i=1}^{N} \frac{\xi_i^\mu \sigma_i}{N} \Big)^2}. \tag{3.1}$$

The additional parameter λ is such that for $\lambda = 1$ we recover the pure relativistic scenario, while for $\lambda \neq 1$ we get a neural network with enriched computational skills with respect to the classical one (1); this point has been extensively addressed in [27], while here we focus on the mathematical backbone of the theory.

Remark 3. *By expanding the relativistic model (3.1) we obtain*

$$\frac{H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}, \lambda)}{N} \sim -\frac{1}{\lambda} - \frac{1}{2N^2} \sum_{i,j}^{N,N} (\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu) \sigma_i \sigma_j + \frac{\lambda}{8N^4} \sum_{i,j,k,l}^{N,N,N,N} (\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu)(\sum_{\nu=1}^{P} \xi_k^\nu \xi_l^\nu) \sigma_i \sigma_j \sigma_k \sigma_l + ... \tag{3.2}$$

namely a series of (denser and denser) associative networks. Note also the alternate signs of the terms making up the series: the attractive ones contribute to memory storage, while the repulsive ones contribute to memory erasure [18, 27, 43].

Remark 4. *In order to get more intuition on the model, we can look at the internal field h_i acting on neuron i and which, in a dynamical picture, can be used to ascertain whether the neuronal state σ_i is likely to be updated or not (see e.g., [11, 44]). By definition, $H_N = -\sum_i h_i \sigma_i$, in such a way that the internal field reads as*

$$h_i = \left(\sum_{\nu=1}^{P} \xi_i^\nu m_\nu \right) \sum_{k=0}^{\infty} \mathcal{B}\left(\frac{1}{2}, k \right) \lambda^{k-1} \left(\sum_{\mu=1}^{P} m_\mu^2 \right)^{k-1} = -\frac{H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}, \lambda)}{N} \left(\sum_{\nu=1}^{P} \xi_i^\nu m_\nu \right) \left(\sum_{\mu=1}^{P} m_\mu^2 \right)^{-1},$$

where $m_\mu = \frac{1}{N} \sum_i \xi_i^\mu \sigma_i$ is the Mattis magnetization associated to the pattern ξ^μ, while $\mathcal{B}(n, k)$ is the general binomial coefficient, namely $\mathcal{B}(n, k) = \Gamma(n+1)/[\Gamma(k+1)\Gamma(1-k+n)]$, $n, k \in \mathbb{R}$. One can see that configurations corresponding to simultaneously large magnitudes of the magnetization entries (i.e., $\sum_{\mu=1}^{P} m_\mu^2$ large) yield to weaker fields; also, by increasing λ the internal field decreases and the system gets more susceptible to the effects of external noise sources.

Definition 10. *The infinite volume limit of the quenched pressure of the "relativistic" Hopfield network, $A^{rel}(\alpha, \beta, \lambda)$, can be written in terms of the partition function of the model $Z_N^{rel}(\boldsymbol{\xi}, \beta, \lambda)$ as*

$$A^{rel}(\alpha, \beta, \lambda) := A(\alpha, \beta, \lambda) = \lim_{N \to \infty} A_N(\alpha, \beta, \lambda) = \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \ln Z_N(\boldsymbol{\xi}, \beta, \lambda), \tag{3.3}$$

$$Z_N^{rel}(\boldsymbol{\xi}, \beta, \lambda) := Z_N(\boldsymbol{\xi}, \beta, \lambda) := \sum_{\{\boldsymbol{\sigma}\}} \exp\left(-\beta H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}, \lambda)\right). \tag{3.4}$$

As in the previous section, we use the subscript N to stress that we are working at finite size, while when we omit it we mean that we are evaluating quantities in the thermodynamic limit.

Definition 11. *Given $O(\boldsymbol{\sigma})$ as a generic function of the neuron configuration, we define the Boltzmann average $\omega_{\boldsymbol{\xi}}(O)$, its replicated product state over s replicas Ω and its quenched expectation $\langle O \rangle$ respectively as*

$$\omega_{\boldsymbol{\xi}}(O) := \frac{\sum_{\{\boldsymbol{\sigma}\}} O(\boldsymbol{\sigma}) e^{-\beta H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}, \lambda)}}{Z_N(\boldsymbol{\xi}, \beta, \lambda)}, \tag{3.5}$$

$$\Omega(O) := \omega_{\boldsymbol{\xi}}^{(1)}(O) \times \omega_{\boldsymbol{\xi}}^{(2)}(O) \times ... \times \omega_{\boldsymbol{\xi}}^{(s)}(O), \tag{3.6}$$

$$\langle O \rangle := \mathbb{E}\Omega(O). \tag{3.7}$$

In the next section, we will exploit a novel interpolation strategy to obtain an explicit expression for the quenched pressure of this model, then, in Sec. 5, preserving the same interpolation (but focusing on the variances of the order parameters instead of the pressure), we will provide a detailed picture of its (possible) critical behavior.

4 The mechanical generalization of Guerra's interpolation scheme

In this Section we achieve the RS expression for the quenched pressure of the relativistic Hopfield model again exploiting a differential-equation-based approach. The main problem consists in the fact that, unlike the classical counterpart, here the cost function (3.2) is not monomial in its degrees of freedom $\{\sigma_i\}_{i=1,...,N}$, but it is a (convergent) infinite series. In order to deal with this, we take advantage of the integral representation of the Dirac delta directly on the cost function (3.1), rather than on the order parameters as usual [11]. This is represented by the following

Proposition 3. *The partition function of the "relativistic" Hopfield model (3.1) displays the following integral representation*

$$Z_N(\boldsymbol{\xi}, \beta, \lambda) \propto \sum_{\{\boldsymbol{\sigma}\}} \int dX \, dK \, \mathcal{D}\tau \, \exp\left[\frac{\beta N}{\lambda}\sqrt{1 + \lambda X} - \frac{KX\beta N}{2} + \right.$$
$$\left. + \frac{\beta N}{2}\Big(\frac{\sqrt{K}}{N}\sum_{i=1}^{N}\xi_i^1 \sigma_i\Big)^2 + \sqrt{\frac{\beta K}{N}}\sum_{\mu=2}^{P}\sum_{i=1}^{N}\xi_i^\mu \sigma_i \tau_\mu\right]. \tag{4.1}$$

Proof. The proof works by direct construction. In fact, by definition, the partition function for the Hamiltonian (3.1) is

$$Z_N(\boldsymbol{\xi}, \beta, \lambda) = \sum_{\{\boldsymbol{\sigma}\}} \exp\left[\frac{\beta N}{\lambda}\sqrt{1 + \frac{\lambda}{N^2}\sum_{\mu=1}^{P}\Big(\sum_{i=1}^{N}\xi_i^\mu \sigma_i\Big)^2}\right], \tag{4.2}$$

and, expanding the square and exploiting the integral representation of the Dirac delta, we get

$$
Z_N(\boldsymbol{\xi}, \beta, \lambda) = \sum_{\{\boldsymbol{\sigma}\}} \int dX \delta\Big(X - \frac{1}{N^2} \sum_{\mu,i,j=1}^{P,N,N} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j\Big) \exp\Big(\frac{\beta N}{\lambda}\sqrt{1+\lambda X}\Big)
$$

$$
\propto \sum_{\{\boldsymbol{\sigma}\}} \int dX \, dK \, \exp\Big(\frac{\beta N}{\lambda}\sqrt{1+\lambda X} + iKX - i\frac{K}{N^2}\sum_{\mu,i,j=1}^{P,N,N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j\Big). \tag{4.3}
$$

Notice that the auxiliary variable X allows us to move the sum out of the square root and the auxiliary variable K allows us to recast the δ function in a mathematically more convenient expression. Now, we rescale $K \to i\frac{\beta N}{2}K$ as standard [11] and we apply a Hubbard-Stratonovich transformation, in such a way that eq. (4.3) can be recast as[1]

$$
Z_N(\boldsymbol{\xi}, \beta, \lambda) \propto \sum_{\{\boldsymbol{\sigma}\}} \int dX \, dK \, \exp\Big(\frac{\beta N}{\lambda}\sqrt{1+\lambda X} - \frac{KX\beta N}{2} + \frac{K\beta}{2N}\sum_{\mu,i,j=1}^{P,N,N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j\Big)
$$

$$
= \sum_{\{\boldsymbol{\sigma}\}} \int dX \, dK \, \mathcal{D}\tau \, \exp\Big[\frac{\beta N}{\lambda}\sqrt{1+\lambda X} - \frac{KX\beta N}{2} +
$$

$$
+ \frac{\beta N}{2}\Big(\frac{\sqrt{K}}{N}\sum_{i=1}^{N}\xi_i^1\sigma_i\Big)^2 + \sqrt{\frac{\beta K}{N}}\sum_{\mu=2}^{P}\sum_{i=1}^{N}\xi_i^\mu\sigma_i\tau_\mu\Big], \tag{4.4}
$$

where, we recall, $\mathcal{D}\tau := \prod_{\mu=1}^{P}\frac{e^{-\tau_\mu^2/2}}{\sqrt{2\pi}}$. Here, under the assumption that a single pattern (say $\boldsymbol{\xi}^1$) is candidate for retrieval, we split the signal term from the (quenched) noise stemming from all the remaining patterns ($\mu > 1$). We also disregarded pre-factors (due to the integral form of Dirac's delta and to rescaling) linear in N and whose contribution in the intensive pressure is vanishing in the thermodynamic limit. $\qquad\square$

Remark 5. *Here, X and K play the role of auxiliary order parameters. In particular, by construction (notice the delta function), the parameter X is fixed to the value*

$$
X \equiv \frac{1}{N^2}\sum_{\mu,i,j=1}^{P,N,N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j = \Big(\frac{1}{N}\sum_{i=1}^{N}\xi_i^1\sigma_i\Big)^2 + \frac{1}{N^2}\sum_{\mu\geq2}^{P}\sum_{i,j=1}^{N,N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j, \tag{4.5}
$$

where the signal and the noise terms are highlighted. The fact that these display the same form as in the Hopfield model suggests that even in the "relativistic" model the number of storable patterns is $P \sim \alpha N$.

Definition 12. *Besides the auxiliary order parameters X and K, the order parameters needed to handle the model (3.1) are the natural relativistic extension of the standard ones, namely the Mattis magnetization and the two-replica overlaps introduced in Definition 5, defined as follows:*

$$
\tilde{m}_1(\boldsymbol{\sigma}) := m_1(\boldsymbol{\sigma}|\boldsymbol{\xi}) := \frac{1}{N}\sum_{i=1}^{N}\xi_i^1\sigma_i, \tag{4.6}
$$

$$
\tilde{q}_{12}(\boldsymbol{\sigma}) := \frac{1}{N}\sum_{i=1}^{N}\sigma_i^{(1)}\sigma_i^{(2)}, \tag{4.7}
$$

$$
p_{12}(\boldsymbol{\tau}) := \frac{1}{P-1}\sum_{\mu>1}^{P}\tau_\mu^{(1)}\tau_\mu^{(2)}, \quad p_{11}(\boldsymbol{\tau}) := \frac{1}{P-1}\sum_{\mu>1}^{P}(\tau_\mu^{(1)})^2. \tag{4.8}
$$

---

[1] Notice that the complex shift for the variable K formally converts fluctuating functions in exponential functions. However, this does not lead to divergences of the whole partition function, since also integration bounds change: the K variable is therefore integrated along the imaginary axis, so functions in K are still fluctuating.

Remark 6. *The reason for the tilde notation is that, in the following, we will introduce a transformation (by a multiplicative factor) on parameters $\tilde{m}$ and $\tilde{q}$ and the transformed variables shall be denoted as m and q, respectively. Also notice that, unlike the previous approach for the classical Hopfield model, here we need also the diagonal term p_{11}, hence we will not drop the indices.*

Definition 13. *Assuming that the order parameters display well defined, replica-symmetric expectations in the thermodynamic limit, we define*

$$\bar{\tilde{m}} := \lim_{N \to \infty} \langle m_1 \rangle, \tag{4.9}$$

$$\bar{\tilde{q}}_{12} := \lim_{N \to \infty} \langle q_{12} \rangle, \tag{4.10}$$

$$\bar{p}_{12} := \lim_{N \to \infty} \langle p_{12} \rangle, \tag{4.11}$$

$$\bar{p}_{11} := \lim_{N \to \infty} \langle p_{11} \rangle. \tag{4.12}$$

Analogously, for the auxiliary order paramaters,

$$\bar{X} := \lim_{N \to \infty} \langle X \rangle, \tag{4.13}$$

$$\bar{K} := \lim_{N \to \infty} \langle K \rangle. \tag{4.14}$$

We can now discuss our strategy to solve the "relativistic" Hopfield model. We anticipate that the main idea is to introduce an interpolating pressure $\mathcal{A}_N(t)$ that recovers the original model for $t = 1$, while for $t = 0$ it corresponds to the pressure of a simpler model analytically addressable; then, the expression for $\mathcal{A}_N(t)$ is obtained by exploiting the fundamental theorem of calculus

$$\mathcal{A}_N(t = 1) = \mathcal{A}_N(t = 0) + \int_0^1 \dot{\mathcal{A}}_N(\tilde{t}) \mathrm{d}\tilde{t}.$$

The expression for $\mathcal{A}_N(t)$ can be figured out recalling that it can display a number of additional variables (effective only for $t \neq 1$), which can be set a posteriori; with a suitable choice of these variables, the derivative of $\mathcal{A}_N(t)$ could be written in terms of the correlation functions of the order parameters of the original model in such a way that, at least under some assumptions (e.g., replica symmetry), the integral of the derivative of the interpolating pressure can be solved.

The preliminary steps in this path are therefore the introduction of an interpolating pressure (Definition 14), the evaluation of its streaming (Lemma 2) and of the Cauchy condition (Lemma 3).

Definition 14. *Given a scalar interpolating parameter $t \in [0,1]$, $N + P$ auxiliary quenched i.i.d. random variables $J_i \sim \mathcal{N}[0,1], i \in (1,...,N)$ and $\tilde{J}_\mu \sim \mathcal{N}[0,1], \mu \in (1,...,P)$, and five real constants $C_m,\ C_\sigma,\ C_z, V_\sigma, V_\tau$ to be set a posteriori, we define the Guerra Generalized Action $\mathcal{A}_N(t)$ as*

$$
\mathcal{A}_N(t) := \frac{1}{N} \mathbb{E} \log \sum_{\{\boldsymbol{\sigma}\}} \int dX \, dK \, \mathcal{D}\tau \exp\Big[\frac{\beta N}{\lambda}\sqrt{1 + \lambda X} - \frac{KX\beta N}{2} + t\frac{\beta N}{2}m_1^2 +
$$
$$
+ (1-t)C_m N m_1 + \sqrt{t}\sqrt{\frac{\beta}{N}}\sum_{\mu,i=1}^{P,N} \xi_i^\mu \sqrt{K}\sigma_i \tau_\mu + \sqrt{1-t}\Big(C_\sigma \sum_{i=1}^N J_i \sqrt{K}\sigma_i + \tag{4.15}
$$
$$
+ C_\tau \sum_{\mu=1}^P \tilde{J}_\mu \tau_\mu\Big) + (1-t)V_\tau \sum_{\mu=1}^P \frac{\tau_\mu^2}{2} + \frac{(1-t)}{2}V_\sigma N K \Big],
$$

where now $\mathbb{E}$ averages over all the quenched random variables involved in the above expression.

Remark 7. *In the following, instead of working with t-dependent expectations for the order parameters and to check a posteriori that they do not depend on the interpolation parameter t, for the generic observable X we will directly assume that $\lim_{N \to \infty} P_t(X) = \delta(X - \bar{X})$ for all $t \in [0,1]$. However, in the Appendix, we check that the results of our interpolation procedure are consistent with the replica trick computations, which do not make use of any interpolation scheme.*

To simplify calculations it is convenient to introduce a new set of order parameters, as stated in the following

Definition 15. *The Mattis magnetization and the two-replica overlap defined, respectively, in (4.6) and (4.7), are transformed as*

$$m(\boldsymbol{\sigma}) := \sqrt{K}\tilde{m}(\boldsymbol{\sigma}), \tag{4.16}$$

$$q_{12}(\boldsymbol{\sigma}) := \sqrt{K^{(1)}K^{(2)}}\tilde{q}_{12}(\boldsymbol{\sigma}), \tag{4.17}$$

where $K^{(a)}$ means that the variable K is evaluated in the a-th replica of the system. The corresponding RS expectations are denoted with the upper bar, i.e., $\bar{m}$ and $\bar{q}_{12}$. Conversely, the variables X, K, p_{11} and p_{12} are left unchanged.

As in the previous section dealing with the classical Hopfield model, we exploit the *universality property* of the quenched noise in the thermodynamic limit and retain the signal as Boolean while the remaining the contribution from the remaining $P-1$ patterns is treated as a Gaussian variable [12, 39, 40].

Lemma 2. *In the infinite volume limit, the t-derivative of the Guerra Generalized Action, under the replica symmetric ansatz, reads as*

$$\frac{d\mathcal{A}_{\mathrm{RS}}(t)}{dt} = -\frac{\beta}{2}\bar{m}^2 - \frac{\beta\alpha}{2}(\bar{p}_{11}\bar{K} - \bar{p}_{12}\bar{q}_{12}). \tag{4.18}$$

Proof. We evaluate directly the t-derivative of $\mathcal{A}_N(t)$ and get

$$\frac{d\mathcal{A}_N(t)}{dt} = \frac{1}{N}\left[\frac{\beta N}{2}\langle m_1^2\rangle_t + \frac{\beta}{2N}\sum_{\mu,i}^{P,N}\left(\langle K\tau_\mu^2\rangle_t - \langle\sqrt{K}\sigma_i\tau_\mu\rangle_t^2\right) - C_m N\langle m_1\rangle_t - \frac{V_\tau}{2}\sum_\mu^P\langle\tau_\mu^2\rangle_t + \right.$$
$$\left. - \frac{C_\sigma^2}{2}\sum_i\left(\langle K\rangle_t - \langle\sqrt{K}\sigma_i\rangle_t^2\right) - \frac{V_\sigma N}{2}\langle K\rangle_t - \frac{C_\tau^2}{2}\sum_u^P\left(\langle\tau_\mu^2\rangle_t - \langle\tau_\mu\rangle_t^2\right)\right], \tag{4.19}$$

where $\langle\cdot\rangle_t$ generalizes (3.7) to the interpolating framework (4.15) and $\langle\cdot\rangle_{t=1} = \langle\cdot\rangle$. Using the definitions (4.6)-(4.8) and (4.16)-(4.17) for the order parameters, the streaming becomes

$$\frac{d\mathcal{A}_N(t)}{dt} = \left[\frac{\beta}{2}\langle m_1^2\rangle_t + \frac{\beta\alpha}{2}\langle Kp_{11}\rangle_t - \frac{\beta\alpha}{2}\langle q_{12}p_{12}\rangle_t - C_m\langle m_1\rangle_t - \frac{V_\tau\alpha}{2}\langle p_{11}\rangle_t + \right.$$
$$\left. - \frac{C_\sigma^2}{2}\langle K\rangle_t + \frac{C_\sigma^2}{2}\langle q_{12}\rangle_t - \frac{V_\sigma}{2}\langle K\rangle_t - \frac{\alpha C_\tau^2}{2}\langle p_{11}\rangle_t + \frac{\alpha C_\tau^2}{2}\langle p_{12}\rangle_t\right]. \tag{4.20}$$

Recalling that $m_1 = \bar{m} + \Delta m$ (see 2.22), and similarly for the other order parameters, we can write

$$\langle m_1^2\rangle_t = \langle(\bar{m} + \Delta m)^2\rangle_t = -\bar{m}^2 + 2\bar{m}\langle m_1\rangle_t + \langle(\Delta m)^2\rangle_t,$$
$$\langle q_{12}p_{12}\rangle_t = \langle(\bar{q}_{12} + \Delta q_{12})(\bar{p}_{12} + \Delta p_{12})\rangle_t = -\bar{q}_{12}\bar{p}_{12} + \bar{p}_{12}\langle q_{12}\rangle_t + \bar{q}_{12}\langle p_{12}\rangle_t + \langle\Delta q_{12}\Delta p_{12}\rangle_t,$$
$$\langle Kp_{11}\rangle_t = \langle(\bar{K} + \Delta K)(\bar{p}_{11} + \Delta p_{11})\rangle_t = -\bar{K}\bar{p}_{11} + \bar{p}_{11}\langle K\rangle_t + \bar{K}\langle p_{11}\rangle_t + \langle\Delta K\Delta p_{11}\rangle_t.$$

Now, under the assumption $\langle\Delta X\rangle_t \to 0$ and $\langle\Delta X\Delta Y\rangle_t \to 0$ for all the observables involved, the t-streaming of the Guerra Generalized Action reads as

$$\frac{d\mathcal{A}_{\mathrm{RS}}(t)}{dt} = -\frac{\beta}{2}\bar{m}^2 - \frac{\beta\alpha}{2}(\bar{p}_{11}\bar{K} - \bar{p}_{12}\bar{q}_{12}) + \left[\beta\bar{m}\langle m_1\rangle_t + \frac{\beta\alpha}{2}\bar{K}\langle p_{11}\rangle_t + \frac{\beta\alpha}{2}\bar{p}_{11}\langle K\rangle_t\right.$$
$$- \frac{\beta\alpha}{2}\bar{q}_{12}\langle p_{12}\rangle_t + -\frac{\beta\alpha}{2}\bar{p}_{12}\langle q_{12}\rangle_t - C_m\langle m_1\rangle_t - \frac{V_\tau\alpha}{2}\langle p_{11}\rangle_t - \frac{C_\sigma^2}{2}\langle K\rangle_t + \frac{C_\sigma^2}{2}\langle q_{12}\rangle_t \quad (4.21)$$
$$\left. - \frac{V_\sigma}{2}\langle K\rangle_t + -\frac{\alpha C_\tau^2}{2}\langle p_{11}\rangle_t + \frac{\alpha C_\tau^2}{2}\langle p_{12}\rangle_t\right],$$

and, by choosing

$$C_m = \beta\bar{m}, \quad C_\sigma = \sqrt{\alpha\beta\bar{p}_{12}}, \quad C_\tau = \sqrt{\beta\bar{q}_{12}}, \quad V_\tau = \beta(\bar{K} - \bar{q}_{12}), \quad V_\sigma = \beta\alpha(\bar{p}_{11} - \bar{p}_{12}), \quad (4.22)$$

we can simplify further the streaming by putting to zero the terms in square brackets, therefore obtaining the statement of the lemma. □

Lemma 3. *In the infinite volume limit, the Cauchy condition $\mathcal{A}_{\mathrm{RS}}(t=0)$ of the Guerra Generalized Action, under the replica symmetric ansatz reads as*

$$\mathcal{A}_{\mathrm{RS}}(t=0) = \log 2 - \frac{\beta\bar{K}\bar{X}}{2} + \frac{\beta\alpha}{2}\bar{K}(\bar{p}_{11} - \bar{p}_{12}) + \frac{\beta}{\lambda}\sqrt{1 + \lambda\bar{X}} + \mathbb{E}_J \log\cosh\left[\beta\bar{m}\sqrt{\bar{K}} + J\sqrt{\beta\alpha\bar{p}_{12}\bar{K}}\right]$$
$$- \frac{\alpha}{2}\log\left[1 - \beta(\bar{K} - \bar{q}_{12})\right] + \frac{\alpha}{2}\frac{\beta\bar{q}_{12}}{1 - \beta(\bar{K} - \bar{q}_{12})}. \quad (4.23)$$

Proof. The 1-body term we need to evaluate is

$$\mathcal{A}_{RS}(t=0) := \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log\sum_{\{\boldsymbol{\sigma}\}} \int dX\, dK\, \mathcal{D}\tau \exp\left[\frac{\beta N}{\lambda}\sqrt{1 + \lambda X} - \frac{KX\beta N}{2} + \frac{1}{2}V_\sigma NK\right.$$
$$\left. + C_m N m_1 + \left(C_\sigma \sum_{i=1}^{N} J_i\sqrt{K}\sigma_i + C_\tau \sum_{\mu=1}^{P}\tilde{J}_\mu\tau_\mu\right) + V_\tau \sum_{\mu=1}^{P}\frac{\tau_\mu^2}{2}\right], \quad (4.24)$$

where we recall that the values of the constants C_m, C_σ, C_z, V_σ, V_τ are fixed by the condition (4.22). By inspecting this quantity, we can see that the σ-dependent terms are linear in the spins, while the integral over the variables $\boldsymbol{\tau}$ is of Gaussian type. Therefore, the $(\boldsymbol{\sigma}, \boldsymbol{\tau})$-dependent part can be easily evaluated, with the result being a non-trivial function of X, K and the other order parameters. However, the remaining integral (over the variables K and X) is of Laplace form, so the leading behaviour in the thermodynamic limit can be computed by applying the saddle point method. By following this route, with straightforward computations, we reach the thesis (4.23). □

Theorem 2. *The infinite volume limit of the replica symmetric expression of the quenched pressure of the "relativistic" Hopfield model in the high-storage regime reads as*

$$A_{\mathrm{RS}}(\alpha, \beta, \lambda) = \log 2 - \frac{\beta K}{2}m^2 - \frac{\beta\alpha K}{2}p(1-q) - \frac{\beta K X}{2} + \frac{\beta}{\lambda}\sqrt{1 + \lambda X} + \frac{\alpha}{2}\frac{\beta K q}{1 - \beta K(1-q)} +$$
$$+ \mathbb{E}_J \log\cosh\left[\beta m K + J\sqrt{\beta\alpha p K}\right] - \frac{\alpha}{2}\log\left[1 - \beta K(1-q)\right], \quad (4.25)$$

whose extremization returns the following self-consistencies for the order parameters

$$K = \frac{1}{\sqrt{1 + \lambda X}}, \quad (4.26)$$

$$X = m^2 + \alpha p(1-q) + \frac{\alpha}{1 - K\beta(1-q)}, \quad (4.27)$$

$$q = \mathbb{E}_J \tanh^2\left(K\beta m + J\sqrt{K\alpha\beta p}\right), \quad (4.28)$$

$$p = \frac{K\beta q}{[1 - K\beta(1-q)]^2}, \quad (4.29)$$

$$m = \mathbb{E}_J \tanh\left(K\beta m + J\sqrt{K\alpha\beta p}\right). \quad (4.30)$$

Proof. The explicit expression of the infinite volume limit of the quenched pressure of the "relativistic" Hopfield network, in terms of the natural order parameters of the theory, can be obtained via the fundamental theorem of Calculus as

$$A(\alpha, \beta, \lambda) = \lim_{N\to\infty}\mathcal{A}_N(t=1) = \lim_{N\to\infty}\left(\mathcal{A}_N(t=0) + \int_0^1 \frac{d\mathcal{A}_N}{dt}\bigg|_{t=t'} dt'\right). \quad (4.31)$$

By using the sum rule (4.31) and taking into account Proposition 2 and Proposition 3, in the thermodynamic limit we can write

$$
\begin{aligned}
A(\alpha, \beta, \lambda) = & \log 2 - \frac{\beta}{2}\bar{m}^2 - \frac{\beta\alpha}{2}\bar{p}_{12}(K - \bar{q}_{12}) - \frac{\beta\bar{K}\bar{X}}{2} + \frac{\beta}{\lambda}\sqrt{1 + \lambda\bar{X}} - \frac{\alpha}{2}\log\left[1 - \beta(\bar{K} - \bar{q}_{12})\right] \\
& + \mathbb{E}_J \log\cosh\left[\beta\bar{m}\sqrt{\bar{K}} + J\sqrt{\beta\alpha\bar{p}_{12}\bar{K}}\right] + \frac{\alpha}{2}\frac{\beta\bar{q}_{12}}{1 - \beta(\bar{K} - \bar{q}_{12})}.
\end{aligned}
$$
(4.32)

Finally, to restore the original order parameters (4.6), we perform a rescaling $\bar{q}_{12} = \bar{K}\bar{\bar{q}}$, $\bar{m} = \sqrt{\bar{K}}\bar{\tilde{m}}$, therefore obtaining Eq. 4.25 (notice that we are omitting the upper bar and tilde to lighten the notation). The self-consistency equations follow by extremization of the pressure over the order parameters as standard. □

Remark 8. *By forcing $\alpha = 0$ in the eq. (4.32) we recover the low-storage scenario of the relativistic Hopfield network [27, 41], as expected.*

Remark 9. *As it happens in Mechanics when performing the classical limit of a relativistic theory, here the classical limit of the quenched free energy of the relativistic Hopfield model collapses as well to the classical Hopfield model addressed in Sec. 2.*

5 Fluctuation theory, criticality and ergodicity breaking

A main contribution of the statistical-mechanics approach to neural networks is the synthesis of their rich phenomenology, as parameters are varied, in terms of a phase diagram. In particular, the Hopfield model is known to exhibit an *ergodic* (E) phase where fast noise prevails and the neuron state is random, a *spin-glass* (SG) phase where the slow noise due to pattern interference prevails and the neuron state gets stuck in spurious states, and a *retrieval* phase where the system can work as an associative memory and perform pattern recognition; the retrieval phase can be further split into a sub-region (R) where the retrieval state corresponds to a global minimum of the free energy and another sub-region (MR) where the retrieval state corresponds to a local minimum of the free energy (see Fig. 1). These regions can be distinguished by looking at the expectation value of the order parameters: in the ergodic region the expectation of all the order parameters is zero, in the spin-glass region the expectation of the Mattis magnetization is zero while the expectation of the overlap is non-null, and in the retrieval region the expectation of the Mattis magnetization and of the overlap is non null. The transition between the retrieval and the spin-glass phase is first-order, as evidenced by the abrupt drop of the Mattis magnetization, while the transition between the spin-glass phase and the ergodic phase is second-order (or critical), as evidenced by the continuous drop of the overlap and by the divergence of their fluctuations. In fact, in pairwise models the onset of ergodicity is typically signalled by a critical behavior whose analytical investigation is feasible by a fluctuation analysis. On the other hand, in pure p-spin models (e.g., a $p = 4$ spin glass) the onset of ergodicity is first-order and the study of this kind of transition is much more challenging.

The relativistic model we are addressing in this paper is neither a pure pairwise model nor a pure p-spin model, whence the quest for a deep analysis of ergodicity breaking, which will be addressed hereafter. In order to accomplish this task, we focus on the two-replica overlaps and, once centered them around their expectations, we study the evolution of their variances (suitably amplified by their volumes, see Def. 16) in the space of the tunable parameters $\{\alpha, \beta, \lambda\}$ to inspect if and where these diverge, eventually marking the onset of criticality. To this goal we retain the interpolation defined by the Guerra Generalized Action (4.15) and we use it to evaluate the expectations of the

order parameters fluctuations and correlations at $t = 1$. As before, to achieve this result, we start the evaluation at $t = 0$ (where a fictitious, but mathematically treatable, environment is experienced by the neurons) and then propagate this solution up to $t = 1$ (where the real surrounding is perceived by the neurons).

Definition 16. *Using (l, m) to label replicas, the centered and rescaled overlap fluctuations θ_{lm} and ρ_{lm} are introduced as*

$$
\begin{aligned}
\theta_{lm} &:= \sqrt{N}\big[q_{lm} - \delta_{lm}Q - (1 - \delta_{lm})q\big], \\
\rho_{lm} &:= \sqrt{P}\big[p_{lm} - \delta_{lm}P' - (1 - \delta_{lm})p\big],
\end{aligned}
\tag{5.1}
$$

where, Q and q are the expectation values of, respectively, q_{11} and q_{12}, and, analogously, P' and p are the expectation values of, respectively, p_{11} and p_{12}.

Proposition 4. *Given O as a smooth function of overlaps $\{q_{lm}, p_{lm}\}_{l,m=1,\dots,s}$ involving s distinct replicas, the following streaming equation holds*

$$
d_T\langle O\rangle = \frac{1}{2}\sum_{a,b}^{s}\langle O \cdot g_{a,b}\rangle - s\sum_{a=1}^{s}\langle O \cdot g_{a,s+1}\rangle + \frac{s(s+1)}{2}\langle O \cdot g_{s+1,s+2}\rangle - \frac{s}{2}\langle O \cdot g_{s+1,s+1}\rangle, \tag{5.2}
$$

where we posed

$$
g_{a,b} = \theta_{ab}\rho_{ab}
$$

and

$$
d_T := \frac{1}{\beta\sqrt{\alpha}}\frac{d}{dt}
$$

to simplify notation.

We will not report the proof of this proposition as passages are quite lengthy but rather standard (see [19, 47]), rather, to understand the physics underlying this streaming, it is enough to show how the core of these rules can be obtained in general. The simplest (already s-replicated) interpolating structure reads as

$$
Z^s(t) = \sum_{\{\sigma^{(a)}\}}\int d\pi(K^{(a)})\, e^{-\beta\sum_{a=1}^{s}H(\sqrt{K^{(a)}},\sigma^{(a)})}e^{\sqrt{\frac{t}{N}}\sum_{a,i=1}^{s,N}J_i\sqrt{K^{(a)}}\sigma_i^{(a)}} \tag{5.3}
$$

where $d\pi(K^{(a)})$ denotes a generic continuous distribution for $K^{(a)}$, for instance, in (4.1) the measure is the complex exponential. Then, the streaming of the generalized expectaction for a generic smooth function O of the overlaps among s replicas is

$$
\begin{aligned}
\partial_t\langle O\rangle &= \partial_t\mathbb{E}\frac{\sum_{\sigma^{(a)}}\int d\pi(K^{(a)})Oe^{-\beta\sum_{a=1}^{s}H(\sqrt{K^{(a)}},\sigma^{(a)})}e^{\sqrt{\frac{t}{N}}\sum_{a,i=1}^{s,N}J_i\sqrt{K^{(a)}}\sigma_i^{(a)}}}{\sum_{\sigma^{(a)}}\int d\pi(K^{(a)})e^{-\beta\sum_a^{s}H(\sqrt{K^{(a)}},\sigma^{(a)})}e^{\sqrt{\frac{t}{N}}\sum_{a,i}^{s,N}J_i\sqrt{K^{(a)}}\sigma_i^{(a)}}} \\
&= \frac{1}{2\sqrt{t}N}\mathbb{E}\sum_{i=1}^{N}J_i\sum_{a=1}^{s}\Big[\Omega_t(O\sqrt{K^{(a)}}\sigma_i^{(a)}) - \Omega_t(O)\Omega_t(\sqrt{K^{(a)}}\sigma_i^{(b)})\Big] \\
&= \frac{1}{2\sqrt{t}N}\mathbb{E}\sum_{i=1}^{N}\partial_{J_i}\sum_{a=1}^{s}\Big[\Omega_t(O\sqrt{K^{(a)}}\sigma_i^{(a)}) - \Omega_t(O)\Omega_t(\sqrt{K^{(a)}}\sigma_i^{(a)})\Big] \\
&= \frac{1}{2N}\mathbb{E}\sum_{i=1}^{N}\sum_{a,b=1}^{s,s}\Big[\Omega_t(O\sqrt{K^{(a)}K^{(b)}}\sigma_i^{(a)}\sigma_i^{(b)}) - 2\Omega_t(O\sqrt{K^{(a)}}\sigma_i^{(a)})\Omega_t(\sqrt{K^{(b)}}\sigma_i^{(b)})+ \\
&\qquad + 2\Omega_t(O)\Omega_t(\sqrt{K^{(a)}}\sigma_i^{(a)})\Omega_t(\sqrt{K^{(b)}}\sigma_i^{(b)}) - \Omega_t(O)\Omega_t(\sqrt{K^{(a)}K^{(b)}}\sigma_i^{(a)}\sigma_i^{(b)})\Big].
\end{aligned}
\tag{5.4}
$$

The factorized Boltzmann averages can now be represented as averages over distinct replicas, e.g. $\Omega_t(O)\Omega_t(\sqrt{K^{(a)}K^{(b)}}\sigma_i^{(a)}\sigma_i^{(b)}) = \Omega_t(\sqrt{K^{(s+1)}K^{(s+2)}}O\sigma_i^{(s+1)}\sigma_i^{(s+2)})$ for $a \neq b$ and the average $N^{-1}\sum_i$ then produces overlaps like $g_{s+1,s+2}$.

We are interested in finding, in the (α, β, λ) space, the critical surface for ergodicity breaking *from the high noise limit* (where no correlations persist) we can treat θ_{ab}, ρ_{ab} as Gaussian variables with zero mean (this allows us to apply Wick theorem inside the averages) and we also treat both $\sqrt{K}\sigma_i$ and z_μ as zero mean random variables (thus all averages involving uncoupled fields are vanishing): this considerably simplifies the evaluation of the critical surface in the (α, β, λ) space.

We can thus reduce the analysis of the rescaled overlap fluctuations to

$$
\begin{aligned}
\langle\theta_{12}^2\rangle_t = C(t), \quad \langle\theta_{12}\rho_{12}\rangle_t = D(t), \qquad \langle\rho_{12}^2\rangle_t = G(t), \\
\langle\theta_{11}^2\rangle_t = H(t), \quad \langle\theta_{11}\rho_{11}\rangle_t = I(t), \qquad \langle\rho_{11}^2\rangle_t = L(t), \\
\langle\theta_{11}\rho_{22}\rangle_t = Q(t), \quad \langle\theta_{11}\theta_{22}\rangle_t = R(t), \quad \langle\rho_{11}\rho_{22}\rangle_t = S(t).
\end{aligned}
\tag{5.5}
$$

As stated, the strategy is to evaluate $\langle\theta_{12}^2\rangle_{t=1}$ as well as $\langle\theta_{12}\rho_{12}\rangle_{t=1}$ and $\langle\rho_{12}^2\rangle_{t=1}$ by calculating their values at $t = 0$ (the Cauchy condition) and then propagating the solution up to $t = 1$ again via the Fundamental Theorem of Calculus: the surface, in the (α, β, λ) space, where these quantities diverge marks the onset of criticality and it is depicted by the next

Theorem 3. *The ergodic region of the relativistic Hopfield model in the high-storage is delimited by the following critical surface in the (α, β, λ) space of the tunable parameters*

$$
\beta_c = \frac{\sqrt{1 + \sqrt{\alpha}\lambda + \alpha\lambda}}{1 + \sqrt{\alpha}}.
\tag{5.6}
$$

Proof. According to Proposition (4), by inspecting the $t-$evolution of $C(t) = \langle\theta_{12}^2\rangle_t$, $D(t) = \langle\theta_{12}\rho_{12}\rangle_t$ and $G(t) = \langle\rho_{12}^2\rangle_t$, we can write down the following system of coupled ODE

$$
\begin{aligned}
d_T C &= 2CD, \\
d_T D &= D^2 + CG, \\
d_T G &= 2GD.
\end{aligned}
\tag{5.7}
$$

Suitably combining C and G in (5.7) we can write

$$
d_T \ln\frac{C}{G} = 0 \implies C(T) = r^2 G(T), \quad r^2 = \frac{C(0)}{G(0)}.
\tag{5.8}
$$

Now we are left with

$$
\begin{aligned}
d_T D &= D^2 + r^2 G^2, \\
d_T G &= 2GD.
\end{aligned}
\tag{5.9}
$$

The trick here is to complete the square by summing $d_T D + r d_T G$, thus obtaining

$$
\begin{aligned}
d_T Y &= Y^2, \\
Y &= D + rG, \\
d_T G &= 2G(Y - rG).
\end{aligned}
\tag{5.10}
$$

The solution of the above system of coupled ODE is trivial and it is given by

$$
Y(T) = \frac{Y_0}{1 - TY_0}, \quad Y_0 = D(0) + \sqrt{C(0)G(0)}.
\tag{5.11}
$$

Hence we are left with the evaluation of the correlations at $t = 0$: namely the Cauchy condition related to the solution coded by eq. (5.11). To this task we introduce, inside the $\mathcal{A}(t = 0)$ term, a one-body generating function of the momenta of $\tau, \sqrt{K}\sigma$ as follows

$$\mathcal{A}(t = 0) = \log \sum_{\{\boldsymbol{\sigma}\}} \int \mathcal{D}z \exp\left[\beta K \sum_\mu \tau_\mu^2 / 2 + \sum_i j_i \sqrt{K}\sigma_i + \sum_\mu J_\mu \tau_\mu\right]. \tag{5.12}$$

More precisely, the source fields j_i and J_μ provide a mathematical tool to generate the average values $\langle(\sqrt{K}\sigma)^P\rangle$ and $\langle z_m u^P\rangle$ by suitable derivatives of $\mathcal{A}(t = 0)$; for instance, we can generate $\langle\tau_\mu\rangle$ as $\partial_{J_\mu}\mathcal{A}(t = 0)$ and finally setting $J, j = 0$. We stress that this trick is particularly useful because it allows calculating the various mean values necessary to get eq (5.12) by means of simple derivatives of $\mathcal{A}(t = 0)$. For the sake of clearness, we highlight that the relevant terms in j, J are

$$\mathcal{A}(t = 0) \propto \sum_i \log\cosh(j_i\sqrt{K}) + \frac{1}{2(1 - \beta K)} \sum_\mu J_\mu^2. \tag{5.13}$$

All the averages needed at $t = 0$ can now be calculated simply as its derivatives and *after* performing the derivatives by setting $(j = 0, J = 0)$. This operation leads to

$$\begin{aligned}
D(t = 0) &= \sqrt{NP}\left(\partial_j\mathcal{A}\right)^2\left(\partial_J\mathcal{A}\right)^2\Big|_{j,J=0} = 0, \\
C(t = 0) &= \left(\partial_j^2\mathcal{A}\right)^2\Big|_{j,J=0} = K^2, \\
G(t = 0) &= \left(\partial_J^2\mathcal{A}\right)^2\Big|_{j,J=0} = (1 - \beta K)^{-2}.
\end{aligned} \tag{5.14}$$

Inserting this result in (5.11), we get

$$Y(T) = \frac{\frac{K}{1 - \beta K}}{1 - T\frac{K}{1 - \beta K}}. \tag{5.15}$$

Upon evaluating $Y(T)$ for $T = \beta\sqrt{\alpha}t$ at $t = 1$ we obtain

$$Y(t = 1) = \frac{\frac{K}{1 - \beta K}}{1 - \beta\sqrt{\alpha}\frac{K}{1 - \beta K}} = \frac{K}{1 - \beta K(1 + \sqrt{\alpha})}. \tag{5.16}$$

Since we are interested in obtaining the critical surface where ergodicity breaks down, namely where fluctuations (i.e. $Y(t = 1)$) grow arbitrarily large, we can check where the denominator at the r.h.s. of eq. (5.16) vanishes. This leads to

$$\beta_c = \frac{1}{K(1 + \sqrt{\alpha})}, \tag{5.17}$$

$$K = \frac{1}{\sqrt{1 + \lambda X}}, \tag{5.18}$$

$$X = \frac{\alpha}{1 - K\beta_c}. \tag{5.19}$$

The system above can be rearranged explicitly in order to get the critical surface $\beta_c(\alpha, \lambda)$, as stated in the theorem. $\qquad\square$

Remark 10. *The above expression, for $\lambda = 1$ (i.e. in the true relativistic framework), generalizes the standard AGS critical line [10] and collapses to the latter in the classical limit as it should.*

Remark 11. *In the standard Hopfield model the critical noise diverges as $T_c := 1/\beta_c \sim \sqrt{\alpha}$, as the load α is made larger and larger. Conversely, in the relativistic setup, the critical temperature settles on the constant value $T_c := 1/\lambda$ for large values of the load α. Also notice that, for $\lambda < 2$, the critical temperature for ergodicity breaking is a non-monotonous function of α: T_c increases as long as $\alpha < (\lambda - 2)^2/2$, then, at $\alpha = (\lambda - 2)^2/2$ it exhibits a maximum, after which the critical temperature slowly decreases towards its asymptotic limit. For $\lambda = 2$, the maximum is placed at $\alpha = 0$.*

Using the fluctuation theory, we were able to highlight a critical behavior and to provide an explicit equation for the critical surface (see eq. 5.6). Now, we try to get the same result by Taylor-expanding the self-consistency (4.25) near $q = 0$ and $m = 0$, in fact, this is a standard procedure in models displaying criticality and here it is used as a consistency check. By keeping only leading order terms for every order parameter we write

$$K = \frac{1}{\sqrt{1 + \lambda X}}, \tag{5.20}$$

$$X = \alpha \bar{p}(1 - \bar{q}) + \frac{\alpha}{1 - K\beta(1 - \bar{q})} = \frac{\alpha}{1 - \beta K} + \mathcal{O}(\bar{q}^2), \tag{5.21}$$

$$\bar{q} = \mathbb{E}_y \tanh^2\left(y\sqrt{K\alpha\beta\bar{p}}\right) = K\alpha\beta\bar{p} + \mathcal{O}(\bar{p}^2), \tag{5.22}$$

$$\bar{p} = \frac{K\beta\bar{q}}{[1 - K\beta(1 - \bar{q})]^2} = \frac{\beta K\bar{q}}{(1 - \beta K)^2} + \mathcal{O}(\bar{q}^2). \tag{5.23}$$

After some rearrangements we get

$$K^2 = \frac{1 - \beta K}{1 - \beta K + \alpha\lambda} \tag{5.24}$$

$$\beta K = \frac{1}{1 + \sqrt{\alpha}} \tag{5.25}$$

whose solution recovers the previous expression for the critical surface, see eq. (5.6). In Fig. 1, we reported the phase diagrams of the "relativistic" Hopfield model for various values of $\lambda(=1, 2$ and $5)$. As a first comment, we see that increasing the parameter λ, the critical line for the transition to the ergodic region changes convexity, and the ergodic region gets wider towards smaller values of the noise (this features is in common with [47]). While the critical storage capacity remains $\alpha_c(\lambda, \beta \to \infty) \sim 0.14$, we see that the corresponding critical line (i.e. the transition to the spin-glass phase) is stretched toward the α axis, meaning that the relativistic model is less stable w.r.t. thermal noise. This can be understood with the fact that the relativistic cost-function presents shallower potential wells w.r.t to the standard Hopfield model, and this feature is amplified for higher and higher values of λ. This means that less thermal noise is needed for the system to escape from these wells, thus reducing the associative power of the relativistic network.

6   Conclusions

In this paper we developed mathematical techniques to deal with dense neural networks, possibly without extensively relying upon the statistical mechanics of spin glasses, but rather by exploiting differential-equation theory.
We first addressed the classical Hopfield network, whose solution was recovered by relying on a *mechanical analogy*. The knowledge of statistical mechanics of complex system is rather marginal in this setting and this may trigger further research by applied mathematicians that are more familiar with differential-equation theory but possibly less acquainted with spin glasses.
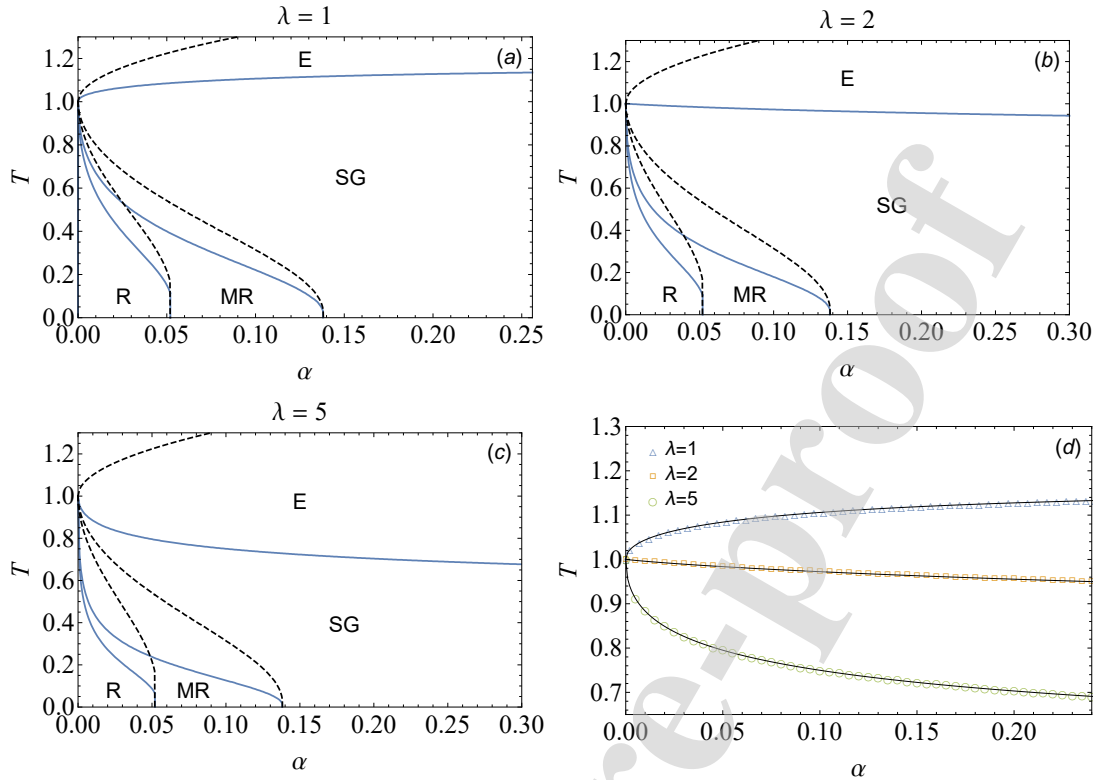
Figure 1. Phase diagram for the relativistic model for different values of λ, namely $\lambda = 1$ (panel a), $\lambda = 2$ (panel b) and $\lambda = 5$ (panel c). The phase diagrams are obtained by numerically solving the self-consistency Eqs. (4.26)-(4.26): the retrieval region (R+MR) is characterized by both non-vanishing m and q. The spin-glass phase (SG) is instead given by $m = 0$ but non vanishing Edward-Anderson parameter q. The distinction between the pure retrieval (R) and mixed retrieval (MR) stands for the fact that, respectively, retrieval and spin-glass states correspond to maxima in the pressure. Finally, in the ergodic region (E) $m = q = 0$. Notice that the transition line corresponding to the boundary between the retrieval region (R+MR) and the spin-glass phase (SG) as well as the transition line corresponding to the boundary between the pure retrieval region (R) and the mixed retrieval phase (MR) are close to those found for the classical Hopfield model (black dashed lines) and, as λ increases, the retrieval regions shrinks. On the other hand, the critical line signalling the emergence of ergodicity is qualitatively different: the ergodic region engulfs part of the SG region. Notice the, in the $\lambda = 1$ phase diagram, we only see the increasing branch of the ergodicity-breaking critical temperature (we decided not to include this behavior in the plot since the critical temperature start to decrease only at $\alpha = 1$ for $\lambda = 1$, and the retrieval regions would not be transparent by inspecting also the region). Finally, panel d shows a comparison between the transition lines to ergodicity found with numerical solutions of self-consistency Eqs. (4.26)-(4.26) (symbols) and the analytical prediction Eq. 5.17 for different values of λ as explained by the legend: the agreement is excellent.

In the rest of the paper, we considered the "relativistic" Hopfield network as an example of dense network allowing for p-body interactions; still by preserving a *mechanical setting* we got an explicit expression of its quenched free energy in the high-load regime. We stress that the cost-function of this model is no longer a simple monomial expression of the neurons and the patterns, and handling this cost-function required a non-trivial generalization of the above differential-equation approach.

At the end of this journey we can finally consider very broad classes of cost functions

$$H = \mathcal{F}\Big[\sum_{\mu=1}^{P} m_\mu^2\Big], \tag{6.1}$$

where $\mathcal{F}$ is a generic smooth function, and obtain their related pressure (at least under the replica symmetry assumption) and thus all the properties of the model under consideration (e.g., the phase diagram) by following the scheme pursued here. In fact, for the generic cost function (6.1), the quenched pressure in the infinite volume limit reads as

$$A(\alpha, \beta, \lambda) = \log 2 - \frac{\beta K}{2}\bar{m}^2 - \frac{\beta\alpha K}{2}p(1-q) - \frac{\beta K X}{2} + \frac{\beta}{2}\mathcal{F}[X] + \frac{\alpha}{2}\frac{\beta K q}{1 - \beta K(1-q)} +$$
$$+ \mathbb{E}_y \log\cosh\Big[\beta m K + y\sqrt{\beta\alpha pK}\Big] - \frac{\alpha}{2}\log\big[1 - \beta K(1-q)\big], \tag{6.2}$$

whose extremization returns the following self-consistencies for the order parameters

$$K = \partial_X \mathcal{F}[X], \tag{6.3}$$

$$X = \bar{m}^2 + \alpha\bar{p}(1-\bar{q}) + \frac{\alpha}{1 - K\beta(1-\bar{q})}, \tag{6.4}$$

$$\bar{q} = \mathbb{E}_y \tanh^2\left(K\beta\bar{m} + y\sqrt{K\alpha\beta\bar{p}}\right), \tag{6.5}$$

$$\bar{p} = \frac{K\beta\bar{q}}{(1 - K\beta(1-\bar{q}))^2}, \tag{6.6}$$

$$\bar{m} = \mathbb{E}_y \tanh\left(K\beta\bar{m} + y\sqrt{K\alpha\beta\bar{p}}\right). \tag{6.7}$$

Furthermore, if the leading contribution in (6.1) is provided by pairwise interactions, an ergodic surface is expected and its expression is given by

$$\beta = \frac{1}{(\sqrt{\alpha} + 1)\,\partial_X \mathcal{F}(\alpha + \sqrt{\alpha})}. \tag{6.8}$$

If $\mathcal{F}(x) = -Nx/2$ we recover the classical Hopfield scenario, while, if $\mathcal{F}[x] = -N\sqrt{1+x}$ we recover the relativistic Hopfield scenario, but the method itself can be applied in full generality to various cost function and can be particularly useful to address dense networks.

Acknowledgments

A   Relativistic Hopfield solution via replica trick route

In this Appendix, we re-derive the quenched pressure in the thermodynamic limit for the "relativistic" Hopfield model by using standard replica trick computations. To do this, we start with representation of the partition function (4.4), which we recall here:

$$Z_N(\boldsymbol{\xi}, \beta, \lambda) = \sum_{\{\boldsymbol{\sigma}\}} \int dX dK \exp\Big(\frac{\beta N}{\lambda}\sqrt{1 + \lambda X} - \frac{KX\beta N}{2} + \frac{K\beta}{2N}\sum_{\mu=1}^{P}\sum_{i,j=1}^{N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j\Big). \tag{A.1}$$

In the hypotesis of retrieval of a single pattern (say $\boldsymbol{\xi^1}$), we can separate the signal contribution from the noise generated by all the others patterns. Then, the partition function can be rewritten in the form

$$Z_N(\boldsymbol{\xi}, \beta, \lambda) = \sum_{\{\boldsymbol{\sigma}\}} \int dX dK \exp\Big(\frac{\beta N}{\lambda}\sqrt{1+\lambda X} - \frac{KX\beta N}{2} + \frac{K\beta}{2N}\sum_{i,j=1}^{N}\xi_i^1\xi_j^1\sigma_i\sigma_j +$$
$$+ \frac{K\beta}{2N}\sum_{\mu>1}^{P}\sum_{i,j=1}^{N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j\Big). \tag{A.2}$$

We can now linearize the quadratic terms in the $\boldsymbol{\sigma}$ variables by introducing a set of auxiliary (Gaussian) variables z_μ, so that

$$Z_N(\boldsymbol{\xi}, \beta, \lambda) = \sum_{\{\boldsymbol{\sigma}\}} \int dX dK \mathcal{D}\tau_1 \exp\Big(\frac{\beta N}{\lambda}\sqrt{1+\lambda X} - \frac{KX\beta N}{2} + \sqrt{\frac{K\beta}{N}}\sum_{i=1}^{N}\xi_i^1\tau_1\sigma_i\Big) \times$$
$$\times \int\Big(\prod_{\mu>1}\mathcal{D}\tau_\mu\Big)\exp\Big(\sqrt{\frac{K\beta}{2N}}\sum_{\mu>1}^{P}\sum_{i=1}^{N}\xi_i^\mu\tau_\mu\sigma_i\Big). \tag{A.3}$$

The replica trick approach implies the computation of the quantity $\mathbb{E}'Z_N^n$, where n is the number of replicas and $\mathbb{E}'$ is the average over non-recalled patterns, that is

$$\mathbb{E}'Z_N^n = \mathbb{E}'\sum_{\boldsymbol{\sigma}^{(1)}}\cdots\sum_{\boldsymbol{\sigma}^{(n)}}\int\Big(\prod_{a=1}^{n}dX^{(a)}dK^{(a)}\mathcal{D}\tau_1^{(a)}\Big)\exp\Big(\frac{\beta N}{\lambda}\sum_a\sqrt{1+\lambda X^{(a)}} - \frac{\beta N}{2}\sum_a K^{(a)}X^{(a)}$$
$$+ \sqrt{\frac{\beta}{N}}\sum_a\sqrt{K^{(a)}}\sum_{i=1}^{N}\xi_i^1\tau_1^{(a)}\sigma_i^{(a)}\Big) \times \int\Big(\prod_{\mu>1}\prod_{a=1}^{n}\mathcal{D}\tau_\mu^{(a)}\Big)\exp\Big(\sqrt{\frac{\beta}{N}}\sum_a\sqrt{K^{(a)}}\sum_{\mu>1}^{P}\sum_{i=1}^{N}\xi_i^\mu\tau_\mu^{(a)}\sigma_i^{(a)}\Big),$$

where $\mathcal{D}\tau$ is the usual Gaussian measure. The only term depending on the non-retrieved patterns is the last integral, so we can directly compute the average $\mathbb{E}'$. Then,

$$\int\Big(\prod_{\mu>1}\prod_{a=1}^{n}\mathcal{D}\tau_\mu^{(a)}\Big)\mathbb{E}'\exp\Big(\sqrt{\frac{\beta}{N}}\sum_a\sqrt{K^{(a)}}\sum_{\mu>1}^{P}\sum_{i=1}^{N}\xi_i^\mu\tau_\mu^{(a)}\sigma_i^{(a)}\Big)$$
$$= \int\Big(\prod_{\mu>1}\prod_{a=1}^{n}\mathcal{D}\tau_\mu^{(a)}\Big)\exp\Big(\sum_{\mu>1}\sum_{i=1}^{N}\log\cosh(\sqrt{\frac{\beta}{N}}\sum_a\sqrt{K^{(a)}}\xi_i^\mu\tau_\mu^{(a)}\sigma_i^{(a)})\Big) \tag{A.4}$$
$$\simeq \int\Big(\prod_{\mu>1}\prod_{a=1}^{n}\mathcal{D}\tau_\mu^{(a)}\Big)\exp\Big(\frac{\beta}{2N}\sum_{\mu>1}\sum_{i=1}^{N}\sum_{a,b=1}^{n}\sqrt{K^{(a)}K^{(b)}}\tau_\mu^{(a)}\tau_\mu^b\sigma_i^{(a)}\sigma_i^{(b)}\Big),$$

where in the last line we used the fact that $\log\cosh x = \frac{x^2}{2} + \mathcal{O}(x^3)$, which is a reasonable approximation since we want to evaluate the thermodynamic limit of the partition function. At this point,

we can introduce the overlap order parameters by inserting a Dirac delta in the integral:

$$
\int \Big(\prod_{\mu>1} \prod_{a=1}^{n} \mathcal{D}\tau_\mu^{(a)} \Big) \exp \Big(\frac{\beta}{2N} \sum_{\mu>1} \sum_{i=1}^{N} \sum_{a,b=1}^{n} \sqrt{K^{(a)}K^{(b)}} \tau_\mu^{(a)} \tau_\mu^{(b)} \sigma_i^{(a)} \sigma_i^{(b)} \Big)
$$

$$
= \int \Big(\prod_{\mu>1} \prod_{a=1}^{n} \mathcal{D}\tau_\mu^{(a)} \Big) \Big(\prod_{a,b=1}^{n} dq_{ab} \delta(q_{ab} - \tfrac{1}{N}\sqrt{K^{(a)}K^{(b)}} \sum_i \sigma_i^{(a)} \sigma_i^{(b)}) \Big) \exp \Big(\frac{\beta}{2N} \sum_{\mu>1} \sum_{i=1}^{N} \sum_{a,b=1}^{n} \tau_\mu^{(a)} \tau_\mu^{(b)} q_{ab} \Big)
$$

$$
= \int \Big(\prod_{\mu>1} \prod_{a=1}^{n} \mathcal{D}\tau_\mu^{(a)} \Big) \Big(\prod_{a,b=1}^{n} dq_{ab} \frac{N dp_{ab}}{2\pi} \Big) \exp \Big[iN \sum_{a,b} p_{ab} q_{ab}
$$

$$
- i \sum_{a,b,i} p_{ab} \sqrt{K^{(a)}K^{(b)}} \sigma_i^{(a)} \sigma_i^{(b)} + \frac{\beta}{2N} \sum_{\mu>1} \sum_{i=1}^{N} \sum_{a,b=1}^{n} \tau_\mu^{(a)} \tau_\mu^{(b)} q_{ab} \Big],
$$

where in the last line we used the Fourier representation of the Dirac delta by introducing the q-conjugated order parameters p_{ab}. At this point, we can directly compute the Gaussian integral over the z variables, whose result is

$$
\int \Big(\prod_{\mu>1} \prod_{a=1}^{n} \mathcal{D}\tau_\mu^{(a)} \Big) \exp \Big(\frac{\beta}{2N} \sum_{\mu>1} \sum_{i=1}^{N} \sum_{a,b=1}^{n} \tau_\mu^{(a)} \tau_\mu^{(b)} q_{ab} \Big) = \exp \Big(-\frac{\alpha N}{2} \log \det(\mathbf{1} - \beta \mathbf{q}) \Big), \quad (A.5)
$$

where $\mathbf{1}$ is the $n \times n$ identity matrix and $\mathbf{q}$ is the overlap matrix and we replaced $P-1$ with αN, which is possible since we are analyzing the thermodynamic limit. Then, after rescaling $p_{ab} \to \frac{\alpha\beta}{2} p_{ab}$ and $\tau_1^{(a)} \to \sqrt{\beta N} \tau_1^{(a)}$, we have

$$
\mathbb{E}' Z_N^n = \sum_{\boldsymbol{\sigma}^{(1)}} \cdots \sum_{\boldsymbol{\sigma}^{(n)}} \int \Big(\prod_{a=1}^{n} dX^{(a)} dK^{(a)} \sqrt{\frac{\beta N}{2\pi}} d\tau_1^{(a)} \Big) \Big(\prod_{a,b=1}^{n} dq_{ab} \frac{i\alpha\beta N dp_{ab}}{4\pi} \Big) \exp \Big(-\frac{\beta N}{2} \sum_a (\tau_1^{(a)})^2
$$

$$
+ \frac{\beta N}{\lambda} \sum_a \sqrt{1 + \lambda X^{(a)}} - \frac{\beta N}{2} \sum_a K^{(a)} X^{(a)} + \beta \sum_a \sqrt{K^{(a)}} \sum_{i=1}^{N} \xi_i^1 \tau_1^{(a)} \sigma_i^{(a)} - \frac{\alpha\beta N}{2} \sum_{a,b} p_{ab} q_{ab}
$$

$$
+ \frac{\alpha\beta}{2} \sum_{a,b,i} p_{ab} \sqrt{K^{(a)}K^{(b)}} \sigma_i^{(a)} \sigma_i^{(b)} - \frac{\alpha N}{2} \log \det(\mathbf{1} - \beta \mathbf{q}) \Big).
$$

From now on, we will denote the integration measure over the order parameters simply with $d\mu$ (we stress that the contribution of the prefactors in the integration measures is negligible in the thermodynamic limit). With straightforward computation, it is easy to prove that

$$
\sum_{\boldsymbol{\sigma}^{(1)}} \cdots \sum_{\boldsymbol{\sigma}^{(n)}} \exp \Big(\beta \sum_a \sqrt{K^{(a)}} \sum_{i=1}^{N} \xi_i^1 \tau_1^{(a)} \sigma_i^{(a)} + \frac{\alpha\beta}{2} \sum_{a,b,i} p_{ab} \sqrt{K^{(a)}K^{(b)}} \sigma_i^{(a)} \sigma_i^{(b)} \Big) =
$$
$$
\exp \Big(N \mathbb{E} \log \sum_{\boldsymbol{\sigma}} \exp \big(\beta \sum_a \sqrt{K^{(a)}} \xi^1 \tau_1^{(a)} \sigma^{(a)} + \frac{\alpha\beta}{2} \sum_{a,b} p_{ab} \sqrt{K^{(a)}K^{(b)}} \sigma^{(a)} \sigma^{(b)} \big) \Big),
$$
(A.6)

where we used the fact that, in the thermodynamic limit, $\lim_{N\to\infty} \frac{1}{N} f(\boldsymbol{\xi}_i) = \mathbb{E} f(\boldsymbol{\xi})$. Then, we have the final form

$$
\mathbb{E}' Z_N^n = \int d\mu \exp \Big(-\frac{\beta N}{2} \sum_a (\tau_1^a)^2 + \frac{\beta N}{\lambda} \sum_a \sqrt{1 + \lambda X^{(a)}} - \frac{\beta N}{2} \sum_a K^{(a)} X^{(a)} - \frac{\alpha\beta N}{2} \sum_{ab} p_{ab} q_{ab}
$$

$$
- \frac{\alpha N}{2} \log \det(\mathbf{1} - \beta \mathbf{q}) + N \mathbb{E} \log \sum_{\boldsymbol{\sigma}} \exp \big(\beta \sum_a \sqrt{K^{(a)}} \xi^1 \tau_1^a \sigma^{(a)} + \frac{\alpha\beta}{2} \sum_{a,b} p_{ab} \sqrt{K^{(a)}K^{(b)}} \sigma^{(b)} \sigma^{(b)} \big) \Big).
$$

To go forward, we must assume the replica symmetry, which in our case turns out to be

$$K^{(b)} = K, \tag{A.7}$$
$$X^{(a)} = X, \tag{A.8}$$
$$\tau_1^{(a)} = \sqrt{K}m, \tag{A.9}$$
$$q_{ab} = K\delta_{ab} + Kq(1 - \delta_{ab}), \tag{A.10}$$
$$p_{ab} = P\delta_{ab} + p(1 - \delta_{ab}). \tag{A.11}$$

With this ansatz, we get

$$\mathbb{E}'Z_N^n = \int d\mu \exp\Big[-\frac{\beta N n}{2} K m^2 + \frac{\beta N n}{\lambda}\sqrt{1 + \lambda X} - \frac{\beta N n}{2}KX - \frac{\alpha \beta N n}{2}Kp(1 - q)$$
$$-\frac{\alpha N n}{2}\Big(\log[1 - \beta K(1 - q)] - \frac{\beta Kq}{1 - \beta K(1 - q)}\Big) + nN \int Dy \log 2 \cosh(\beta Km + y\sqrt{\alpha\beta Kp})\Big].$$

Then, we can apply the basic replica trick formula $A = \lim_{N\to\infty, n\to 0}(\mathbb{E}'Z_N^n - 1)/nN$. To do this, we first compute the leading contribution of the partition function in the $N \to \infty$ limit by means of Laplace method, then we expand the exponential up to the order $\mathcal{O}(n)$. With simple algebra, we arrive to the result

$$A_{\mathrm{RS}}(\alpha, \beta, \lambda) = -\frac{\beta}{2}Km^2 + \frac{\beta}{\lambda}\sqrt{1 + \lambda X} - \frac{\beta KX}{2} - \frac{\alpha\beta}{2}Kp(1 - q) - \frac{\alpha}{2}\log[1 - \beta K(1 - q)]$$
$$+ \frac{\alpha}{2}\frac{\beta Kq}{1 - \beta K(1 - q)} + \mathbb{E}_y \log 2 \cosh(\beta Km + y\sqrt{\alpha\beta Kp}), \tag{A.12}$$

which is precisely the quenched pressure reported in (4.25).

References

[1] J. Sanders, E. Kandrot, *CUDA by example: an introduction to general-purpose GPU programming*, Addison-Wesley Professional, (2010).

[2] R. Collobert, J. Weston, *A unified architecture for natural language processing: Deep neural networks with multitask learning*, Proc. 25th Int. Conf. on Mach. Learn. ACM, (2008).

[3] R. Salakhutdinov, G. Hinton, *Deep Boltzmann machines*, Artificial Intelligence and Statistics (2009).

[4] Y. Le Cun, Y. Bengio, G. Hinton, *Deep learning*, Nature **521**:436-444, (2015).

[5] I.A.T. Hashem, et al., *The rise of "big data" on cloud computing: Review and open research issues*, Infor. Sys. **47**:98, (2015).

[6] L. Zhou, et al., *Machine learning on big data: Opportunities and challenges*, Neurocomputing **237**:350, (2017).

[7] M. Mezard, G. Parisi, M.A. Virasoro, *Spin glass theory and beyond: an introduction to the replica method and its applications*, World Scientific, Singapore (1987)

[8] A. Engel, C. Van den Broeck, *Statistical mechanics of learning*, Cambridge University Press (2001).

[9] H.S. Seung, H. Sompolinsky, N. Tishby, *Statistical mechanics of learning from examples*, Phys. Rev. A **45**(8):6056, (1992).

[10] D.J. Amit, *Modeling brain functions*, Cambridge Univ. Press (1989).

[11] A.C.C. Coolen, R. Kuhn, P. Sollich, *Theory of neural information processing systems*, Oxford Press (2005).

[12] E. Agliari, A. Barra, C. Longo, D. Tantari, *Neural Networks retrieving binary patterns in a sea of real ones*, J. Stat. Phys. **168**, 1085, (2017).

[13] E. Agliari, A. Barra, B. Tirozzi, *Free energies of Boltzmann Machines: self-averaging, annealed and replica symmetric approximations in the thermodynamic limit*, J. Stat. Mech. 033301 (2019).

[14] A. Bovier, *Statistical Mechanics of Disordered Systems*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge (2006).

[15] A. Bovier, V. Gayrard, *Hopfield models as generalized random mean field models*, Mathematical aspects of spin glasses and neural networks, 3-89, Birkhauser, Boston (1998).

[16] A. Bovier, V. Gayrard, P. Picco, *Gibbs states of the Hopfield model in the regime of perfect memory*, Prob. Theor. & Rel. Fields **100**(3):329, (1994).

[17] A. Bovier, V. Gayrard, P. Picco, *Gibbs states of the Hopfield model with extensively many patterns*, J. Stat. Phys. **79**(1-2):395, (1995).

[18] A. Fachechi, E. Agliari, A. Barra, *Dreaming neural networks: forgetting spurious memories and reinforcing pure ones*, Neural Net. **112**, 24, (2019).

[19] A. Barra, G. Genovese, F. Guerra, *The replica symmetric approximation of the analogical neural network*, J. Stat. Phys. **140**(4):784, (2010).

[20] A. Barra, G. Genovese, F. Guerra, *Equilibrium statistical mechanics of bipartite spin systems*, J. Phys. A **44**, 245002, (2011).

[21] V. Dotsenko, N.D. Yarunin, E.A. Dorotheyev, *Statistical mechanics of Hopfield-like neural networks with modified interactions*, J. Phys. A **24**, 2419, (1991).

[22] V. Dotsenko, B. Tirozzi, *Replica symmetry breaking in neural networks with modified pseudo-inverse interactions*, J. Phys. A **24**:5163-5180, (1991).

[23] M. Talagrand, *Rigorous results for the Hopfield model with many patterns*, Prob. Theor. & Rel. Fiel. **110**(2):177, (1998).

[24] M. Talagrand, *Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model*, Ann. Prob. 1393-1469, (2000).

[25] L. Pastur, M. Shcherbina, B. Tirozzi, *On the replica symmetric equations for the Hopfield model*, J. Math. Phys. **40**(8): 3930, (1999).

[26] L. Pastur, M. Shcherbina, B. Tirozzi, *The replica-symmetric solution without replica trick for the Hopfield model*, J. Stat. Phys. **74**(5-6):1161, (1994).

[27] A. Barra, M. Beccaria, A. Fachechi, *A new mechanical approach to handle generalized Hopfield neural networks*, Neural Networks (2018).

[28] D. Krotov, J.J. Hopfield, *Dense associative memory for pattern recognition*, Adv. Neur. Inform. Proc. Sys. 1172–1180, (2016).

[29] D. Krotov, J.J. Hopfield, *Dense associative memory is robust to adversarial inputs*, Neur. Comp. 30.(12):3151, (2018).

[30] N. Bogolyubov, et al., *Some classes of exactly soluble models of problems in quantum statistical mechanics: the method of the approximating Hamiltonian*, Russ. Math. Surv. **39**, 1, (1984).

[31] J.G. Brankov, A.S. Shumovsky, V.A. Zagrebnov, *On model spin Hamiltonians including long-range ferromagnetic interaction*, Physica **78**, 183, (1974).

[32] A. Barra, A. Di Lorenzo, F. Guerra, A. Moro, *On quantum and relativistic mechanical analogues in mean-field spin models*, Proc. Royal Soc. A **470**(2172), 20140589, (2014).

[33] A. Barra, *The mean field Ising model trough interpolating techniques*, J. Stat. Phys. **132**.5:787, (2008).

[34] A. Barra, A. Di Biasio, F. Guerra, *Replica symmetry breaking in mean-field spin glasses through the Hamilton–Jacobi technique*, JSTAT 09:P09006, (2010).

[35] F. Guerra, *Sum rules for the free energy in the mean field spin glass model*, Fiel. Inst. Comm. **30**, 11, (2001).

[36] J.J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proceedings of the national academy of sciences 79.8 (1982): 2554-2558.

[37] D. Amit, H. Gutfreund, H. Sompolinsky, *Spin-glass models of neural networks*, Phys. Rev. A **32**.2:1007, (1985).

[38] D. Amit, H. Gutfreund, H. Sompolinsky, *Storing infinite numbers of patterns in a spin-glass model of neural networks*, Phys. Rev. Lett. **55**.14:1530, (1985).

[39] P. Carmona, Y. Hu, *Universality in Sherrington–Kirkpatrick's spin glass model*, Ann. Henri Poincarè **42**, 2, (2006).

[40] G. Genovese, *Universality in bipartite mean field spin glasses*, J. Math. Phys. **53**(12):123304, (2012).

[41] E. Agliari, A. Barra, M. Notarnicola, *The relativistic Hopfield network: Rigorous results*, J. Math. Phys. **60**.3:033302, (2019).

[42] E. Agliari, A. Fachechi, C. Marullo, *The "relativistic" Hopfield network with correlated patterns*, submitted (2020).

[43] K. Nokura, *Paramagnetic unlearning in neural network models*, Phys. Rev. E **54**(5):5571, (1996).

[44] J. Hertz, A. Krogh, R.G. Palmer, *Introduction to the theory of neural computation*, CRC Press Taylor & Francis Group (1991).

[45] E. Agliari, D. Migliozzi, and D. Tantari, *Non-convex multi-species Hopfield models*, J. Stat. Phys. **172**(5):1247, (2018).

[46] F. Guerra, *Broken replica symmetry bounds in the mean field spin glass model*, Comm. Math. Phys. **233**(1), 1, (2003).

[47] E. Agliari, F. Alemanno, A. Barra, A. Fachechi, *Dreaming neural networks: rigorous results*, J. Stat. Mech. 083503 (2019).

Declaration of interests

χ Regarding the paper "Generalized Guerra's interpolation schemes for dense associative networks", submitted to Neural Networks (Elsevier) in November 2019, the authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Prof. Adriano Barra (on the behalf of all the Authors)
Dipartimento di Matematica & Fisica, Università del Salento