Neural Networks with a Redundant Representation: Detecting the Undetectable

Elena Agliari,[1,*] Francesco Alemanno,[2,3] Adriano Barra,[2,4] Martino Centonze,[2] and Alberto Fachechi[2,4]
[1]Dipartimento di Matematica "Guido Castelnuovo", Sapienza Università di Roma, 00185 Roma, Italy
[2]Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, 73100 Lecce, Italy
[3]C.N.R. Nanotec, 73100 Lecce, Italy
[4]Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, 73100 Lecce, Italy

(Received 25 June 2019; revised manuscript received 11 October 2019; published 13 January 2020)

We consider a three-layer Sejnowski machine and show that features learnt via contrastive divergence have a dual representation as patterns in a dense associative memory of order $P = 4$. The latter is known to be able to Hebbian store an amount of patterns scaling as N^{P-1}, where N denotes the number of constituting binary neurons interacting P wisely. We also prove that, by keeping the dense associative network far from the saturation regime (namely, allowing for a number of patterns scaling only linearly with N, while $P > 2$) such a system is able to perform pattern recognition far below the standard signal-to-noise threshold. In particular, a network with $P = 4$ is able to retrieve information whose intensity is $\mathcal{O}(1)$ even in the presence of a noise $\mathcal{O}(\sqrt{N})$ in the large N limit. This striking skill stems from a redundancy representation of patterns—which is afforded given the (relatively) low-load information storage—and it contributes to explain the impressive abilities in pattern recognition exhibited by new-generation neural networks. The whole theory is developed rigorously, at the replica symmetric level of approximation, and corroborated by signal-to-noise analysis and Monte Carlo simulations.

DOI: 10.1103/PhysRevLett.124.028301

Artificial intelligence is nearly everywhere in today's society and has rapidly changed the face of economy, communication and science. Its global success is mainly due to modern neural-network's architectures that allow deep learning [1–3] and, particularly relevant for the present Letter, pattern recognition at prohibitive noise levels [4]. Despite the pervasiveness of such technologies, a clear rationale of the underlying mechanisms is still lacking.

The statistical mechanics of disordered systems has been playing a primary role in the theoretical investigation of neural networks since the early studies by Amit, Gutfreund, and Sompolinksy on pairwise associative neural networks [5], and it still constitutes a valuable tool toward an explanation of the impressive skills of modern nets. For instance, recently, Metha and Schwab have highlighted the profound link between deep learning and renormalization group [6], Krotov and Hopfield have showed a duality between higher-order generalizations of the Hopfield model [7], referred to as *dense associative memories* (DAMs) and neural networks commonly used in (deep) learning [8,9]. Another duality was highlighted in [10–12] between the restricted Boltzmann machine and the Hopfield network: the features adaptively learned (e.g., via contrastive divergence) in the former are exactly the patterns retrieved by the latter; this idea shall be further developed in the current work.

Here we consider a basic architecture for machine learning, i.e., the *restricted Sejnowski machine* (RSM) [13], that is a third-order Boltzmann machine [1], where triples of units interact symmetrically; in the jargon of statistical mechanics, this is just a three-layer spin-glass with $(P = 3)$-wise interactions. In particular, we equip this network with one hidden layer and with two visible layers (a primary and a mirror channel; see Fig. 1 left), which possibly mimic the typical presence of two input sources in biological networks (i.e., the *eyes*). As we show, the RSM displays, as a dual representation, a bipartite DAM, i.e., a bipartite Hopfield model with $(P = 4)$-wise interactions; see Fig. 1 right. In this dual representation, the K features embedded in the RSM correspond to the K patterns stored in the DAM.

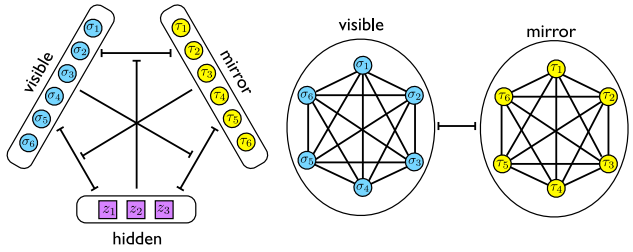

FIG. 1. Schematic representations of the restricted Seinowskj machine (left) and its dual representation in terms of a bipartite dense associative network (right). In the former, neurons i, μ, ρ interact 3-wisely through the coupling $\xi^{\rho}_{i\mu}$ [see also Eq. (1)], while, in the latter, neurons i, μ, j, ν interact 4-wisely through the coupling $J^{j\nu}_{i\mu}$ [see also Eq. (5)].

It is worth recalling that, for a P-spin associative memory built of N binary neurons, the largest number of storable patterns scales as N^{P-1} (a result found by Baldi and Venkatesh [14] and made rigorous by Bovier and Niederhauser [15]); clearly for $P=2$ we recover the standard Hopfield scenario [5,16]. In the last decades, the quest for enhanced storage capacities has strongly biased the statistical mechanical investigations, possibly limiting alternative inspections of the computational capabilities of these networks, which is the main focus of this work, as will be summarized hereafter.

In the standard Hopfield model it is possible to retrieve a number K of patterns that is extensive in N (i.e., $K=\alpha N$ with $\alpha \leq 0.14$) by pushing the signal-to-noise ratio to its limit, namely by letting the magnitude $\mathcal{S}$ of the signal—stemming from the pattern to be retrieved—and the magnitude $\mathcal{N}$ of the (quenched) noise—stemming from the remaining patterns providing an intrinsic glassiness—share the same order. Should the information encoded by patterns be affected by some source of noise, the condition $\mathcal{S}/\mathcal{N} \sim \mathcal{O}(1)$ would be deranged in favor of the noise and retrieval capabilities would be lost. On the other hand, as we show, if we let dense ($P=4$) networks operate with a load $K=\alpha N$ (with $\alpha>0$), these turn out to be able to retrieve the information [$\sim\mathcal{O}(1)$] encoded by patterns perturbed by extensive noise [$\sim\mathcal{O}(\sqrt{N})$]. This is ultimately due to the possibility of redundant representation of patterns [17,18], which implies a storage cost of $\mathcal{O}(N^2)$ bits per pattern. In the following we give more technical details to prove the previous statements.

The RSM [13] considered here is built on three layers, two of which—referred to as *visible* and *mirror*, respectively (see Fig. 1, left panel)—are digital and made up of N Ising neurons per layer, $\boldsymbol{\sigma} \in \{-1,+1\}^N$ and $\boldsymbol{\tau} \in \{-1,+1\}^N$, while the third layer—referred to as *hidden*—is analog and made of K neurons $\boldsymbol{z}$, whose states are i.i.d. Gaussians $\mathcal{N}(0,\beta^{-1})$ ($\beta>0$ tuning the level of the fast noise in the net [5]). The model presents third-order interactions among neurons of different layers but no intralayer interactions (whence the *restriction*). Its cost function H_{RSM} is given by

$$H_{\mathrm{RSM}}(\boldsymbol{\sigma},\boldsymbol{\tau},z|\boldsymbol{\xi}) = -\frac{1}{N^{3/2}}\sum_{i,\mu,\rho=1}^{N,N,K}\xi_{i\mu}^{\rho}\sigma_i\tau_\mu z_\rho, \tag{1}$$

with $i,\mu=1,\ldots,N$ and $\rho=1,\ldots,K$. In the thermodynamic limit each layer size diverges such that $\lim_{N\to\infty} K/N=\alpha>0$ and the factor $N^{-3/2}$ keeps the mean value of the cost function (under the quenched Gibbs measure [19]) linearly extensive in N. The interaction between each triplet of neurons is encoded in the $K\times N\times N$ tensor $\boldsymbol{\xi}$ whose ρth element will be written as

$$\xi_{i\mu}^{\rho} = \xi_i^{\rho}\xi_\mu^{\rho}, \quad i,\mu=1,\ldots,N, \tag{2}$$

where $\xi_i^{\rho}\in\{-1,+1\}$ is meant as the ith entry of the ρth pattern to be retrieved in the dual bipartite DAM. Notice that the factorization (2) ensures the symmetry of $\xi_{i\mu}^{\rho}$ for any ρ and it lies at the core of the pattern redundancy scheme pursued here. In fact, the information contained into a set of K binary patterns of length N is inflated into a symmetric tensor of size KN^2.

Given a small learning rate $\epsilon>0$, we obtain for this network the following contrastive-divergence [20] learning rule [21]

$$\Delta\xi_{i\mu}^{\rho} = \epsilon\beta(\langle\sigma_i\tau_\mu z_\rho\rangle_+ - \langle\sigma_i\tau_\mu z_\rho\rangle_-), \tag{3}$$

where the subscript "+" means that both visible and mirror layers are set at the data input (i.e., they are *clamped*), while the subscript "−" means that all neurons in the network are left free to evolve; importantly, while clamped, visible and mirror layers are always exposed to the same information (i.e., $\boldsymbol{\sigma}=\boldsymbol{\tau}=\boldsymbol{\xi}^{\rho}$).

Using the symbol Dz_ρ to denote the Gaussian measure with variance β^{-1} [i.e., $Dz_\rho \equiv dz_\rho \exp(-\beta z_\rho^2/2)\sqrt{\beta/2\pi}$], the partition function Z related to the cost function (1) reads

$$Z = \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}}\int\prod_{\rho=1}^{K}Dz_\rho \exp\left(\frac{\beta}{N^{3/2}}\sum_{i,\mu,\rho=1}^{N,N,K}\xi_{i\mu}^{\rho}\sigma_i\tau_\mu z_\rho\right). \tag{4}$$

By construction, the couplings are symmetric ($\xi_{i\mu}^{\rho}=\xi_{\mu i}^{\rho}$) and a detailed balance ensures that the long term relaxation of any (not-pathological) neural dynamics is described by the related Gibbs measure [5,10]. Marginalizing over the hidden layer,

$$P(\boldsymbol{\sigma},\boldsymbol{\tau}|\boldsymbol{\xi}) = \frac{\int Dz e^{-\beta H_{\mathrm{RSM}}(\boldsymbol{\sigma},\boldsymbol{\tau},z|\boldsymbol{\xi})}}{Z} \equiv \frac{e^{-\beta H_{\mathrm{DAM}}(\boldsymbol{\sigma},\boldsymbol{\tau}|\boldsymbol{\xi})}}{Z},$$

where the last equation tacitly defines the cost function of the DAM, namely

$$\begin{aligned} H_{\mathrm{DAM}}(\boldsymbol{\sigma},\boldsymbol{\tau}|\boldsymbol{\xi}) &= -\frac{1}{2N^3}\sum_{\rho=1}^{K}\left(\sum_{i,\mu=1}^{N,N}\xi_{i\mu}^{\rho}\sigma_i\tau_\mu\right)^2, \\ &= -\frac{1}{2N^3}\sum_{i,j=1}^{N,N}\sum_{\mu,\nu=1}^{N,N}J_{i\mu}^{j\nu}\sigma_i\sigma_j\tau_\mu\tau_\nu, \end{aligned} \tag{5}$$

where $J_{i\mu}^{j\nu}=(\sum_\rho \xi_{i\mu}^{\rho}\xi_{j\nu}^{\rho})$. This decomposition shows that the $\boldsymbol{\xi}$s play as eigenvectors for the tensor $\boldsymbol{J}$, whose symmetry with respect to an exchange of indices (i,μ) and (j,ν) mirrors the symmetry between the $\boldsymbol{\sigma}$ and the $\boldsymbol{\tau}$ variables underlying the learning rule (3). Notice that H_{DAM} corresponds to a ($P=4$)-wise bipartite Hopfield model (see Fig. 1, right panel), namely a minimal generalization of the Hebbian kernel in the classic Hopfield reference (quite similar to autoencoders in engineering

jargon [28]). Also, this equivalence extends the duality between restricted Boltzmann machines and (pairwise) Hopfield neural networks [10–12].

To start dealing with network's capabilities, it is convenient to introduce generalized Mattis order parameters M_ρ defined as

$$M_\rho \equiv \frac{1}{N^2}\sum_{i,\mu=1}^{N,N} \xi_{i\mu}^\rho \sigma_i \tau_\mu. \tag{6}$$

The signal-to-noise analysis for this system can be obtained by requiring the dynamic stability of the neural state recalling, without loss of generality, the pattern $\rho = 1$, that is, $\sigma_i \tau_\mu = \xi_{i\mu}^1$. Therefore, denoting with $h_{i\mu}$ the internal field acting on σ_i and τ_μ we get

$$\begin{aligned}\sigma_i \tau_\mu h_{i\mu} &= \mathcal{S} + \mathcal{N} = \frac{1}{2N}\sum_{\rho=1}^{K} M_\rho \xi_{i\mu}^\rho \xi_{i\mu}^1 \\ &= \frac{1}{2N}\left[M_1 + \sum_{\rho>1}^{K} M_\rho \xi_{i\mu}^\rho \xi_{i\mu}^1\right].\end{aligned} \tag{7}$$

As the signal term inside the brackets in (7) is $M_1 \sim \mathcal{O}(1)$, while the noise term corresponds to a sum of $(K-1)$ stochastic and uncorrelated contributions, each of order $\mathcal{O}(N^{-1})$, exploiting the central limit theorem it is immediate to check that the quenched noise due to nonretrieved patterns can be amplified by a factor $\sqrt{N}$ still preserving the stability condition $\mathcal{S}/\mathcal{N} \sim \mathcal{O}(1)$. We can therefore introduce noisy patterns yielding to the noisy tensor $\boldsymbol{\eta}$ with entries

$$\eta_{i\mu}^\rho \equiv \xi_{i\mu}^\rho + \sqrt{K}\tilde{\xi}_{i\mu}^\rho, \tag{8}$$

where the information is carried by the Boolean entries of $\xi_{i\mu}^\rho$, while the noise is coded in the real $\tilde{\xi}_{i\mu}^\rho$ that are i.i.d. standard Gaussian variables for $i, \mu = 1, \ldots, N$ and $\rho = 1, \ldots, K$. Notice that the information encoded by the patterns is perturbed by adding a stochastic term $\tilde{\boldsymbol{\xi}}$ on $\boldsymbol{\xi}$ [Eq. (8)] rather than directly on $\boldsymbol{J}$; the latter choice would have a lower impact on network capacity and is therefore less challenging. In analogy with (6) we also define

$$\tilde{M}_\rho \equiv \frac{1}{N^2}\sum_{i,\mu=1}^{N,N} \tilde{\xi}_{i\mu}^\rho \sigma_i \tau_\mu. \tag{9}$$

Replacing the Boolean tensor (2) in Eq. (5) with the noisy tensor (8) and exploiting the definitions (6) and (9), we get $H_{\mathrm{DAM}} = -(N/2)\sum_\rho (M_\rho + \sqrt{K}\tilde{M}_\rho)^2$. Then, in the limit of large N, splitting the signal and the noise contributions, the Boltzmann factor in Eq. (4) reads as [21]

$$\exp\left(-\beta H_{\mathrm{RSM}}\right) \underset{N\to\infty}{\sim} \exp\left(\beta\frac{N}{2}M_1^2 + \beta\frac{\alpha N^2}{2}\sum_{\rho\geq 2}^{K}\tilde{M}_\rho^2\right).$$

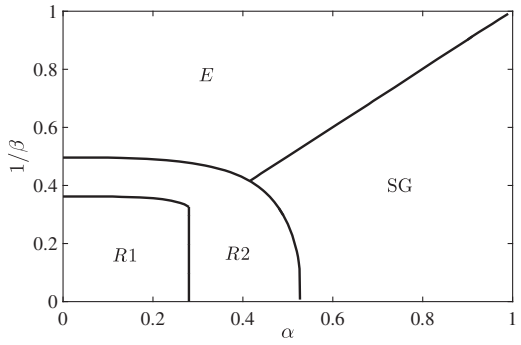

FIG. 2. Phase diagram for the DAM with $(P = 4)$-wise interactions among the N neurons and a load $K = \alpha N$, as a function of the capacity α and of the noise level $1/\beta$. This diagram was obtained by solving the self-consistent Eqs. (11)–(15) and by identifying the retrieval phase as the region where each neural configuration corresponding to a stored pattern is a maximum of the pressure—either global ($R1$), or local ($R2$)—the spin-glass (SG) phase as the region where retrieval capabilities are lost due to prevailing "slow noise" α, and the ergodic (E) phase as the region where retrieval capabilities are lost due to prevailing "fast noise" $1/\beta$ [21]. This diagram was found for a noise regime $\mathcal{O}(\sqrt{N})$ [see Eq. (8)], where the standard Hopfield would fail.

Let us now handle the two terms appearing as argument of the exponential in the rhs: exploiting the redundancy $\xi_{i\mu}^1 = \xi_i^1 \xi_\mu^1$ and calling m_σ and m_τ the Mattis magnetization related to the visible layer $\boldsymbol{\sigma}$ and to the mirror layer $\boldsymbol{\tau}$ respectively, we get $M_1 = ([1/N]\sum_i \xi_i^1 \sigma_i)([1/N]\sum_\mu \xi_\mu^1 \tau_\mu) \equiv m_\sigma m_\tau$, in such a way that $\beta N M_1^2/2 = \beta N m_\sigma^2 m_\tau^2/2$; by performing a Hubbard-Stratonovich transformation, the quenched noise given by the nonretrieved $K-1$ patterns is linearized as $\sqrt{\alpha\beta}\sum_{i,\mu,\rho\geq 2}\tilde{\xi}_{i\mu}^\rho \sigma_i \tau_\mu z_\rho/N$. After these passages one can address the evaluation of the intensive quenched pressure of the model, defined as,

$$A(\alpha,\beta) \equiv \lim_{N\to\infty}\frac{1}{N}\mathbb{E}_{\boldsymbol{\eta}} \ln \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}} \int \prod_{\rho=1}^{K} Dz_\rho \exp(-\beta H_{\mathrm{RSM}}),$$

exploiting Guerra's interpolation techniques [10,29]. Under the replica symmetric (RS) ansatz, the quenched pressure reads as [21]

$$\begin{aligned}A^{\mathrm{RS}} = {} & 2\ln 2 + \frac{\alpha^2\beta^2}{2}p(2qr - r - q) - \frac{3}{2}\beta\bar{m}_\sigma^2\bar{m}_\tau^2 \\ & + \int Dx \ln\cosh\left(\alpha\beta x\sqrt{rp} + \beta\bar{m}_\sigma\bar{m}_\tau^2\right) \\ & + \int Dx \ln\cosh\left(\alpha\beta x\sqrt{qp} + \beta\bar{m}_\sigma^2\bar{m}_\tau\right) \\ & - \frac{\alpha}{2}\ln[1-\alpha\beta(1-qr)] + \frac{\alpha^2\beta}{2}\frac{qr}{1-\alpha\beta(1-qr)},\end{aligned} \tag{10}$$

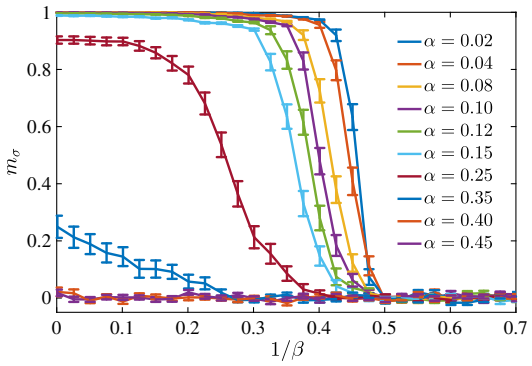

FIG. 3. Expected Mattis magnetization obtained from Monte Carlo simulations run for $N = 150$ and for different values of α, as a function of $1/\beta$. Notice that, as α is tuned from 0.25 to 0.35, the magnetization abruptly drops even at small values of $1/\beta$, consistently with the transition from the region $R1$ to the region $R2$ found theoretically (see Fig. 2 and [21]).

where $\bar{m}_\sigma$ and $\bar{m}_\tau$ are the RS values of the Mattis magnetizations, while q, p, and r are the RS values for the two-replica overlaps for each layer (visible, hidden, and mirror, respectively). Its extremization returns the following self-consistency equations for the order parameters

$$q = \int Dx \tanh^2(\alpha\beta\sqrt{rp}x + \beta\bar{m}_\sigma\bar{m}_\tau^2), \tag{11}$$

$$r = \int Dx \tanh^2(\alpha\beta\sqrt{qp}x + \beta\bar{m}_\sigma^2\bar{m}_\tau), \tag{12}$$

$$p = \frac{\alpha qr}{[1-\alpha\beta(1-qr)]^2}, \tag{13}$$

$$\bar{m}_\sigma = \int Dx \tanh(\alpha\beta\sqrt{rp}x + \beta\bar{m}_\sigma\bar{m}_\tau^2), \tag{14}$$

$$\bar{m}_\tau = \int Dx \tanh(\alpha\beta\sqrt{qp}x + \beta\bar{m}_\sigma^2\bar{m}_\tau), \tag{15}$$

whose solution paints the phase diagram in Fig. 2. There emerge four regions corresponding to qualitatively different solutions: an ergodic (E) one, where the order parameters are all vanishing, a spin-glass (SG) one, where the magnetizations are zero but the overlaps are positive, and two retrieval (R_1, R_2) regions, where the magnetization of the retrieved pattern is non vanishing (in R_2 this solution is only metastable) [21].

The theory is also corroborated via Monte Carlo simulations; a sample of this analysis is shown in Fig. 3, while more extensive discussions can be found in [21].

To summarize, we considered a Sejnowski machine equipped with two visible layers and we showed that it can perform pattern-redundant representation via a suitable generalization of the standard contrastive divergence. Further, we proved that this machine has a dual representation in terms of a bipartite DAM in such a way that the features learned by the former correspond to the patterns stored in the latter and, whatever the learning mode (adaptive vs Hebbian), in the operational mode these networks achieve pattern recognition always in a Hebbian fashion. We studied these nets via statistical mechanical tools obtaining (under the RS ansatz) a phase diagram, where their remarkable capabilities shine. In particular, there exists a region in the parameter space where they can retrieve patterns although these are (apparently) overpowered by the noise. This may contribute to explain the high-rate ability of deep or dense networks in pattern recognition, as empirically evidenced in a variety of tasks. Indeed, at finite volumes (as standard dealing with real datasets), it is not obvious which regime of operation the network is actually set at: to see this one can notice that at finite N and K one has only access to the ratio $\alpha(K,N) = K/N$, which can possibly be compatible with different scalings (e.g., $K = \alpha_1 N^{P-1}$ or $K = \alpha_2 N$). Hence, we speculate that such impressive detection skills emerge when these nets are away from the memory storage saturation. Further, we have shown by a pure statistical mechanical perspective, how pattern recognition power and memory storage are strongly related.

For the sake of completeness, we report that also in the purely engineering counterpart, pattern redundancy is exploited to cope with high noise rate (e.g., in white Gaussian additive channels [30,31]). In particular, our approach is close to the so called *channel access method* in telecommunications, namely a setup where more than two terminals connected to the same transmission medium are allowed to share its capacity.

The authors are grateful to Unisalento and Sapienza University of Rome (RG11715C7CC31E3D) for partial funding.

*agliari@mat.uniroma1.it

[1] R. Salakhutdinov and G. Hinton, Deep Boltzmann machines, in *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics,Hilton Clearwater Beach Resort, Florida, USA* (PMLR, 2009), p. 448–455.
[2] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature (London) **521**, 436 (2015).
[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (M.I.T. Press, Cambridge, 2017).
[4] R. Tandra and A. Sahai, SNR walls for signal detection, IEEE Sel. Top. Sign. Proc. **2**, 4 (2008).
[5] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, 1989).
[6] P. Mehta and D. J. Schwab, An exact mapping between the variational renormalization group and deep learning, arXiv: 1410.3831.
[7] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

[8] D. Krotov and J. J. Hopfield, Dense associative memory for pattern recognition, in *Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain* (Curran Associates Inc., Red Hook, 2016), p. 1180–1188.
[9] D. Krotov and J. J. Hopfield, Dense associative memory is robust to adversarial inputs, Neural Comput. **30**, 3151 (2018).
[10] A. Barra, A. Bernacchia, E. Santucci, and P. Contucci, On the equivalence among Hopfield neural networks and restricted Boltzmann machines, Neural Netw. **34**, 1 (2012).
[11] E. Agliari, A. Barra, A. De Antoni, and A. Galluzzi, Parallel retrieval of correlated patterns: From Hopfield networks to Boltzmann machines, Neural Netw. **38**, 52 (2013).
[12] A. Barra, G. Genovese, P. Sollich, and D. Tantari, Phase diagrams of restricted Boltzmann machines and generalized Hopfield models, Phys. Rev. E **97**, 022310 (2018).
[13] T. J. Sejnowski, Higher-order Boltzmann machines, in *AIP Conference Proceedings 151 on Neural Networks for Computing, Snowbird, Utah, USA* (American Institute of Physics Inc., USA, 1986), p. 398–403.
[14] P. Baldi and S. S. Venkatesh, Number of Stable Points for Spin-Glasses and Neural Networks of Higher Orders, Phys. Rev. Lett. **58**, 913 (1987).
[15] A. Bovier and B. Niederhauser, The spin-glass phase-transition in the Hopfield model with p-spin interactions, Adv. Theor. Math. Phys. **5**, 1001 (2001).
[16] E. Gardner, The space of interactions in neural network models, J. Phys. A **21**, 257 (1988).
[17] M. Elad, Sparse and redundant representation modeling: What next?, IEEE Signal Process. Lett. **19**, 922 (2012).
[18] H. Barlow, Redundancy reduction revisited, Netw. Comput. Neural Syst. **12**, 241 (2001).
[19] A. Barra, G. Genovese, and F. Guerra, The replica symmetric approximation of the analogical neural network, J. Stat. Phys. **140**, 784 (2010).
[20] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, A learning algorithm for Boltzmann machines, Cogn. Sci. **9**, 147 (1985).
[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.028301 for details on its derivation and performances (Appendixes A, B, C, D, and E), which includes Refs. [22–27].
[22] J. Barbier, Overlap matrix concentration in optimal Bayesian inference, https://arxiv.org/abs/1904.02808.
[23] F. Guerra, Broken replica symmetry bounds in the mean field spin glass model, Commun. Math. Phys. **233**, 1 (2003).
[24] A. Barra, Irreducible free energy expansion and overlaps locking in mean field spin glasses, J. Stat. Phys. **123**, 601 (2006).
[25] A. Barra, The mean-field Ising model trought interpolating techniques, J. Stat. Phys. **132**, 787 (2008).
[26] H. Jacquin and A. Rançon, Resummed mean-field inference for strongly coupled data, Phys. Rev. E **94**, 042118 (2016).
[27] V. Sessak and R. Monasson, Small-correlation expansions for the inverse Ising problem, J. Phys. A **42**, 055001 (2009).
[28] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science **313**, 504 (2006).
[29] E. Agliari, A. Barra, C. Longo, and D. Tantari, Neural networks retrieving binary patterns in a sea of real ones, J. Stat. Phys. **168**, 1085 (2017).
[30] B. Li, R. Di Fazio, and A. Zeira, A low bias algorithm to estimate negative SNRs in an AWGN channel, IEEE Commun. Lett. **6**, 469 (2002).
[31] J. Lehnert and M. B. Pursley, Error probabilities for binary direct-sequence spread-spectrum communications with random signature sequences, IEEE Trans. Commun. **35**, 87 (1987).