A Statistical Physics approach to a multi-channel Wigner spiked model

Published Version:
Alberici, D., Camilli, F., Contucci, P., Mingione, E. (2022). A Statistical Physics approach to a multi-channel Wigner spiked model. EUROPHYSICS LETTERS, 136(4), p1-p6 [10.1209/0295-5075/ac4794].

Availability:
This version is available at: https://hdl.handle.net/11585/906994 since: 2022-11-24

Published:
DOI: http://doi.org/10.1209/0295-5075/ac4794

A Statistical Physics approach to a multi-channel Wigner spiked model

Diego Alberici[1], Francesco Camilli[2], Pierluigi Contucci[2] and Emanuele Mingione[2]

[1] *Communication Theory Laboratory, École Polytechnique Fédérale de Lausanne - Lausanne, Switzerland*
[2] *Dipartimento di Matematica, Università di Bologna - Bologna, Italy*

Abstract –In this letter we present a finite temperature approach to a high-dimensional inference problem, the Wigner spiked model, with group dependent signal-to-noise ratios. For two classes of convex and non-convex network architectures the error in the reconstruction is described in terms of the solution of a mean-field spin-glass on the Nishimori line. In the cases studied the order parameters do not fluctuate and are the solution of finite dimensional variational problems. The deep architecture is optimized in order to confine the high temperature phase where reconstruction fails.

Consider the inference task of recovering a signal, made of N bits belonging to K groups, sent through channels with a group dependent noise. In the present paper we consider $N \gg 1$, $K = O(1)$. In order to ease its reconstruction the signal is encoded with some redundancy, *i.e.*, not only bit by bit but also in products of bit pairs. In particular we focus on the special case of Gaussian noise and signals made up of ± 1 components drawn independently with probability 1/2. The original signal (*ground truth* from now on) is therefore an N-components binary vector denoted with $\sigma^* = \{\sigma_i^*\}_{i\in\Lambda}$, where $\Lambda = \{1, 2, \dots, N\}$ is a set of indices. More specifically, the statistician that has to infer the ground truth receives the following $N + N^2$ observations

$$\begin{aligned} \tilde{y}_i(\sigma^*) &= \sqrt{h_r}\,\sigma_i^* + \tilde{z}_i\,, \quad i \in \Lambda_s \\ y_{ij}(\sigma^*) &= \sqrt{\frac{\mu_{rs}}{2N}}\,\sigma_i^*\sigma_j^* + z_{ij}\,, \quad (i,j) \in \Lambda_r \times \Lambda_s\,, \end{aligned} \tag{1}$$

where $\{\Lambda_r\}_{r=1}^K$ is a partition of Λ with $N_r = |\Lambda_r|$. The random variables $z_{ij}\,, \tilde{z}_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ represent additive Gaussian noise causing the corruption of the signal, while the parameters $\mu_{rs} = \mu_{sr} \geq 0$, $h_r \geq 0$ (called *signal-to-noise ratios*) tune the amplitude of the signal. In the following we will use the notations $\mathbf{h} = (h_r)_{r=1}^K$, $\mu = (\mu_{rs})_{r,s=1}^K$.

The problem of reconstructing a ground truth signal from noisy observations has been investigated in different models of high-dimensional Statistical Inference. Equations (1) generalize the so called *Wigner spiked model* [1–3] by grouping signal bits in K sets and allowing for group-dependent signal-to-noise ratios: the Wigner spiked model corresponds to the case $K = 1$ and describes the idealised condition of homogeneous noise. Recently, in the context of neural networks with associative memory features [4, 5] the classical *Hebbian rule* to store and retrive P patterns of N bits each (ground truth) has been modified to take into account a so-called "synaptic noise" that corrupts the information stored into the couplings between neurons. The effects on the retrieval capability of the network have been analyzed in presence of different realisations of the synaptic noise. In particular, one of those realisations for $P = 1$ would give couplings between neurons of the form of the y_{ij}'s in (1) and for a special choice of the temperature the resulting Statistical Mechanics model is equivalent to standard Wigner spiked model. We also refer the interested reader to the seminal paper [6].

The simplest way to account for inhomogeneity of the noise is to weaken the permutation symmetry among the observations components, preserving it only inside each set of a given partition $\{\Lambda_r\}_{r=1}^K$. We call the resulting model *multi-channel* Wigner spiked model (see also [7] for a similar model studied in different limits). From equations (1) one can see that the choice of the sizes N_r and the parameters μ, $\mathbf{h}$ completely determines the distribution of the observed channels. The purpose of this work is a systematic study of the problem for different choices of these parameters.

When the distribution used to generate the ground truth σ^* and the parameters $\mathbf{h}$, μ are known, the inference problem is said to be in the Bayesian Optimal Setting (BOS), meaning that one can write the posterior distribution given the observations. In the BOS the posterior is

$$\begin{aligned} &P(\sigma^*=\sigma|y,\tilde{y}) \propto \\ &\propto \exp\left[-\frac{1}{2}\sum_{r,s=1}^{K}\sum_{(i,j)\in\Lambda_r\times\Lambda_s}\left(y_{ij}-\sqrt{\frac{\mu_{rs}}{2N}}\sigma_i\sigma_j\right)^2\right]\times \\ &\quad\times\exp\left[-\frac{1}{2}\sum_{r=1}^{K}\sum_{i\in\Lambda_r}\left(\tilde{y}_i-\sqrt{h_r}\sigma_i\right)^2\right] \end{aligned} \tag{2}$$

where we have used the Bayes rule together with the fact that conditionally on $\sigma^*=\sigma$ the observations have independent Gaussian distributions: $y_{ij}\sim\mathcal{N}\left(\sqrt{\frac{\mu_{rs}}{2N}}\,\sigma_i\sigma_j,1\right)$, $\tilde{y}_i\sim\mathcal{N}\left(\sqrt{h_r}\,\sigma_i,1\right)$ for $(i,j)\in\Lambda_r\times\Lambda_s$.

By replacing the observations $y,\tilde{y}$ with their definitions (1) and absorbing the terms that depend on $z,\tilde{z},\sigma^*$ only into the normalization we can rewrite the posterior as a random Boltzmann-Gibbs distribution $P(\sigma^*=\sigma|\,y,\tilde{y})=\exp\left(-\tilde{H}_N(\sigma)\right)/Z_N$ where:

$$\begin{aligned} \tilde{H}_N(\sigma) = &-\sum_{r,s=1}^{K}\sum_{ij\in\Lambda_r\times\Lambda_s}\left[\sqrt{\frac{\mu_{rs}}{2N}}z_{ij}\sigma_i\sigma_j+\frac{\mu_{rs}}{2N}\sigma_i\sigma_j\sigma_i^*\sigma_j^*\right] \\ &-\sum_{r=1}^{K}\sum_{i\in\Lambda_r}\left[\sqrt{h_r}\tilde{z}_i\sigma_i+h_r\sigma_i\sigma_i^*\right]. \end{aligned} \tag{3}$$

In this work we are interested in the high dimensional regime of the inference problem, with $N\to\infty$ and form factors $N_r/N\to\alpha_r$. The main goal is the computation of the limiting rescaled mutual information which coincides, up to an additive constant, with the Statistical Mechanics quenched free energy (at $\beta=1$) of the model with Hamiltonian (3):

$$f_N=-\frac{1}{N}\,\mathbb{E}\log\sum_{\sigma\in\{-1,1\}^N}\exp\left(-\tilde{H}_N(\sigma)\right), \tag{4}$$

where $\mathbb{E}$ denotes the expectation w.r.t. the independent quenched variables $\sigma^*,z,\tilde{z}$. We denote the quenched averaged expectation with $\mathbb{E}\langle\cdot\rangle_{\sigma^*}$. In our approach, also called *finite temperature* in Statistical Mechanics [8,9], the estimator for the ground truth will be chosen to be the average w.r.t. the posterior measure. The quality of the signal reconstruction is quantified by the minimum mean square error (MMSE) defined by

$$\mathrm{MMSE}=\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\langle(\sigma_i^*-\sigma_i)^2\rangle_{\sigma^*}=2\left(\sum_{r=1}^{K}\frac{\partial f_N}{\partial h_r}+1\right). \tag{5}$$

Thanks to the $\mathbb{Z}_2$-gauge symmetry

$$\tilde{z}_i\mapsto\tilde{z}_i\sigma_i^*\,,\quad z_{ij}\mapsto z_{ij}\sigma_i^*\sigma_j^*\,,\quad \sigma_i\mapsto\sigma_i\sigma_i^* \tag{6}$$

the Hamiltonian $\tilde{H}_N$ becomes independent of the ground truth signal (it can be evaluated at $\sigma_i^*=1$ for any $i\in\Lambda$), more precisely

$$\begin{aligned} \tilde{H}_N(\sigma)\mapsto &-\sum_{r,s=1}^{K}\sum_{(i,j)\in\Lambda_r\times\Lambda_s}\left[\sqrt{\frac{\mu_{rs}}{2N}}z_{ij}+\frac{\mu_{rs}}{2N}\right]\sigma_i\sigma_j+ \\ &-\sum_{r=1}^{K}\sum_{i\in\Lambda_r}\left[\sqrt{h_r}\tilde{z}_i+h_r\right]\sigma_i \end{aligned} \tag{7}$$

which is distributionally equivalent to

$$H_N(\sigma)=-\sum_{r,s=1}^{K}\sum_{(i,j)\in\Lambda_r\times\Lambda_s}J_{ij}^{rs}\,\sigma_i\sigma_j-\sum_{r=1}^{K}\sum_{i\in\Lambda_r}h_i^r\sigma_i \tag{8}$$

with $J_{ij}^{rs}\overset{\text{iid}}{\sim}\mathcal{N}(\mu_{rs}/2N,\mu_{rs}/2N)$ and $h_i^r\overset{\text{iid}}{\sim}\mathcal{N}(h_r,h_r)$. The previous Hamiltonian represents a multi-species Sherrington-Kirkpatrick model [10, 11] on the Nishimori line [12,13], where interactions and bias disorders are tied to have means identical to their variances. We denote the corresponding Boltzmann-Gibbs measure with $\langle\cdot\rangle$. When $K=1$ the model reduces to the standard SK model on the Nishimori line. For generic K the model enjoys a block permutation symmetry and the interactions among blocks are tuned by the parameters μ. When the matrix μ is positive definite we say that the system has a convex architecture. The meaning of such architecture is that the interaction between spins belonging to the same block dominates those between different blocks. Such condition is violated when the internal interactions of each block are absent. A special non-convex case is the so called deep architecture (a feature of deep Boltzmann machines) introduced in the seminal paper [14] in the context of Machine Learning. In this case the system is divided in K layers and only interactions between two consecutive layers are allowed. In this work we consider and solve the Statistical Mechanics problem for both convex and deep architectures.

The distributionally equivalent models (3) and (8) fulfill two families of identities, called the Nishimori identities [9, 15], that are mapped into each other by the gauge transformation (6). For the model (8) one finds for instance $\mathbb{E}\langle\sigma_i\rangle^2=\mathbb{E}\langle\sigma_i\rangle$ which in particular implies

$$\mathrm{MMSE}=\sum_{r=1}^{K}\frac{N_r}{N}\,\mathbb{E}[1-\langle m_r\rangle]\,,\quad m_r:=\frac{1}{N_r}\sum_{i\in\Lambda_r}\sigma_i\,. \tag{9}$$

The MMSE represents the information-theoretical lower bound for the error that any algorithm can achieve in the task of signal reconstruction. Therefore (9) provides a way to characterize the best possible performance of an algorithm in terms of the average magnetization of a spin glass system. For this reason the study of the phase diagram of the system, and in particular the location of the phase transitions, is a crucial a priori information for algorithms design. In the region of the phase space where the magnetization is zero the MMSE turns out to be maximum and equal to one. This implies there is no way to find an algorithm that performs better than a random guess drawn from the ground truth distribution.

The high dimensional limit of the MMSE, and in particular of the vector $(\mathbb{E}\langle m_r\rangle)_{r=1}^K$, can be derived from the computation of the limiting free energy, which is done by steps starting from the non interacting case where the signal to noise ratio matrix $\mu = 0$. This is easily solved by a convex combination of free energies:

$$\psi(\mathbf{h}) := \lim_{N\to\infty} f_N^{(0)} = -\sum_{r=1}^K \alpha_r \, \mathbb{E} \log 2\cosh\left(z\sqrt{h_r} + h_r\right) \tag{10}$$

where z is a standard Gaussian.

When instead $\mu \neq 0$ we solve the problem for two wide classes of architectures, namely the topological structure identified by μ, by showing that the solution is given by an ordinary variational principle in K dimensions. Define the matrix $\alpha := \mathrm{diag}(\alpha_1, \dots, \alpha_K)$ and consider the variational free energy

$$\begin{aligned} f(\mu,\mathbf{h};\mathbf{x}) =& \frac{\mathbf{x}\cdot(\alpha\mu\alpha)\mathbf{x}}{2} - \frac{(1-\mathbf{x})\cdot(\alpha\mu\alpha)(1-\mathbf{x})}{4} + \\ &+ \psi(\mu\alpha\mathbf{x}+\mathbf{h}) \ , \end{aligned} \tag{11}$$

where $\mathbf{x} \in \mathbb{R}^K_{\geq 0}$. Notice that, when μ is invertible, the stationary points of f fulfill

$$\mathbf{x} = \mathbb{E}_z \tanh\left(z\sqrt{\mu\alpha\mathbf{x}+\mathbf{h}} + \mu\alpha\mathbf{x} + \mathbf{h}\right) \ , \quad z \sim \mathcal{N}(0,1) \ . \tag{12}$$

The proof of the following Propositions are outlined in the Supplementary Material.

Proposition 1 (Convex architecture [12])**.** *Let μ be positive definite. Then :*

$$\lim_{N\to\infty} f_N = \inf_{\mathbf{x}\in\mathbb{R}^K_{\geq 0}} f(\mu,\mathbf{h};\mathbf{x}) \, . \tag{13}$$

If the spectral radius of $\mu\alpha$ is smaller than 1, then $f(\mu,\mathbf{h};\mathbf{x})$ is convex in $\mathbf{x}$ and the minimizer is unique.

Remark 1. When $K = 1$ one can easily see that the solution $\bar{x}(\mu, h)$ of (12) is continuous at $(\mu, h) = (1, 0)$ with $\bar{x}(1, 0) = 0$. Moreover, the critical behaviours turn out to be

$$\bar{x}(\mu,0) = \frac{\mu-1}{\mu^2}(1+o(1)) \, , \quad \mu \to 1_+ \tag{14}$$

i.e. $\bar{x}$ vanishes linearly in $\mu - 1$,

$$\bar{x}(1,h) = \sqrt{h}(1+o(1)) \, , \quad h \to 0_+ \tag{15}$$

and for a $\lambda > 0$

$$\bar{x}(\mu,\lambda(\mu-1)) = \sqrt{\frac{\lambda(\mu-1)}{\mu^2}}(1+o(1)) \, , \quad \mu \to 1_+ \, . \tag{16}$$

Proposition 2 (Deep architecture [13])**.** *Let μ be a tridiagonal matrix with zero diagonal:*

$$\mu = \begin{pmatrix} 0 & \mu_{12} & 0 & \cdots & 0 \\ \mu_{21} & 0 & \mu_{23} & \cdots & 0 \\ 0 & \mu_{32} & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \mu_{K-1,K} \\ 0 & 0 & 0 & \mu_{K,K-1} & 0 \end{pmatrix} \tag{17}$$

and let K be even. Then

$$\lim_{N\to\infty} f_N = \inf_{\mathbf{x}_o} \sup_{\mathbf{x}_e} f(\mu,\mathbf{h};\mathbf{x}) \tag{18}$$

where $\mathbf{x}_o$ and $\mathbf{x}_e$ denote the odd and even components of the vector $\mathbf{x} \in \mathbb{R}^K_{\geq 0}$ respectively. If the spectral radius of the $K/2 \times K/2$ matrix obtained from $(\mu\alpha)^2$ by erasing the even rows and columns is smaller than 1, then the solution of the variational problem (18) *is unique. If h_s is positive for every s, then the solution of* (18) *is unique.*

The conditions on the spectral radii in Propositions 1, 2 identify the phase transitions in the corresponding architectures when $\mathbf{h} = 0$. Indeed, in this case $\mathbf{x} = 0$ is always a solution to the consistency equations (12) both in the convex and deep architectures. However, $\mathbf{x} = 0$ is the stable solution of the optimization problem only when the mentioned spectral radii are smaller than 1. Otherwise, in both cases an arbitrarily small $h_r > 0$ would bring the system towards the stable solution which is a vector of positive entries $\bar{\mathbf{x}} \in \mathbb{R}^K_{>0}$ that can be identified with the vector of the limiting averaged magnetizations $(\mathbb{E}\langle m_r\rangle)_{r=1}^K$.

Hence, by (9) the limiting MMSE is

$$\lim_{N\to\infty} \mathrm{MMSE} = \sum_{r=1}^K \alpha_r (1-\bar{x}_r) \, . \tag{19}$$

It is possible moreover to show that the magnetization is a self-averaging quantity [12] (see also [16]) as suggested by the finite dimensional nature of the variational principles (13) and (18).

The free energies in Propositions 1 and 2 refer to the $N \to \infty$ limit. One may want to investigate the order in N of the finite size corrections to the variational principles. For the simpler case $K = 1$ this was done in [17] by means of the cavity method and the interpolation technique. The extension of this approach to the present multi-channel setting is left for future work. The main idea, as for the single channel case, would be to obtain a finer control on the variance of the magnetization in the quenched measure, namely to identify the correct N-scaling at which its fluctuations are significant. Once this is done by means of the cavity method, one can use the interpolation technique to evaluate the gap between the targeted variational principle and the finite N pressure, which is usually expressed as an integral of the variance of the magnetization.

In the next Proposition, we deal with the optimization of a deep architecture at $\mathbf{h} = 0$ [18, 19]. Precisely we see

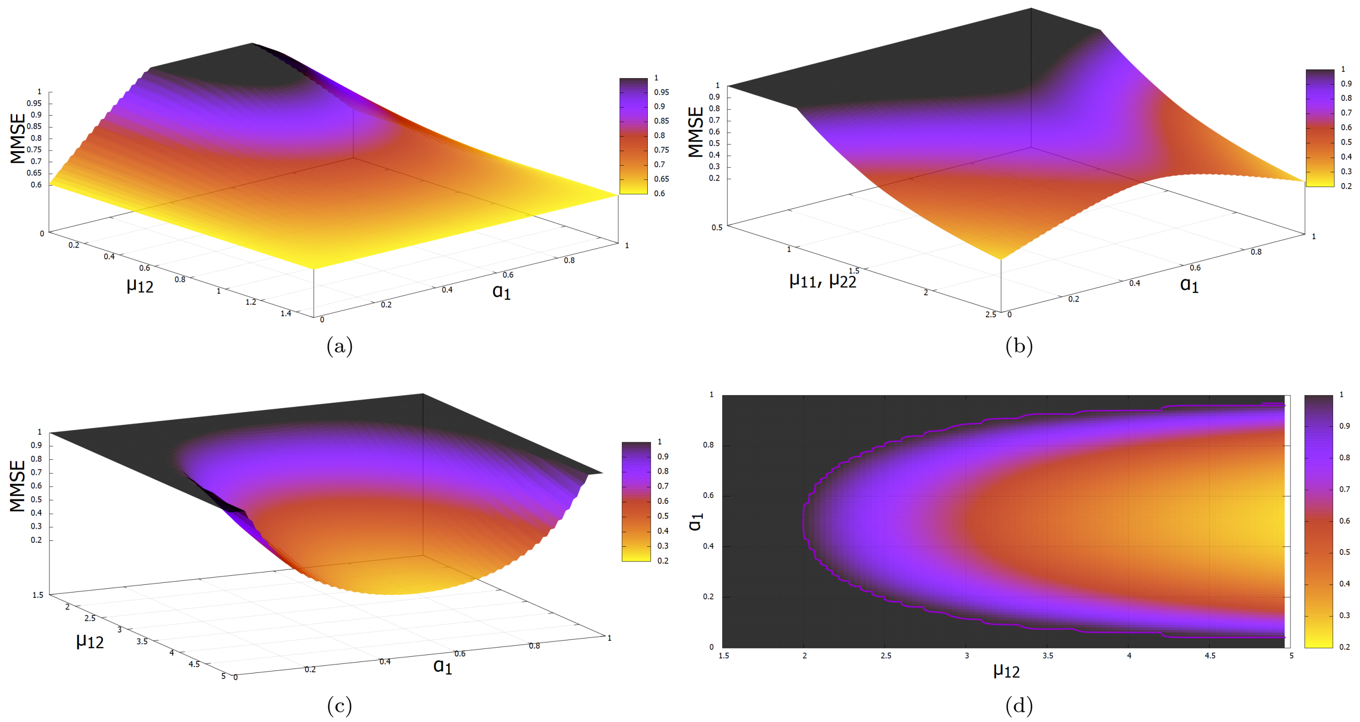

Figure 1: Plots of the asymptotic MMSE for $K = 2$ using the consistency equations (12). Convex architecture in the top row, deep architecture in the bottom one. The critical line between MMSE$= 1$ (dark color) and MMSE< 1(brighter colors) signals a second order phase transition. (a) displays MMSE as a function of $\mu_{12} \in [0, 1.5]$ and α_1 with $\mu_{11} = \mu_{22} = 1.5$. (b) plots MMSE vs α_1, μ_{11} drawn with the constraint $\mu_{11} = \mu_{22} \in [0.5, 2.5]$ and $\mu_{12} = 0.5$. For (a) and (b) the ranges of the parameters keep the SNR matrix μ positive semidefinite. (c) shows the MMSE as a function of the only two free parameters μ_{12}, α_1. (d) is the 2D projection of (c), the contour draws the critical line of the model. (d) shows that in the deep architecture it is impossible to have MMSE< 1 if $\mu_{12} < 2$. Indeed for $K = 2$ the spectral radius equals at most $\mu_{12}^2/4$ (attained at $\alpha_1 = 0.5$) according to Proposition 3.

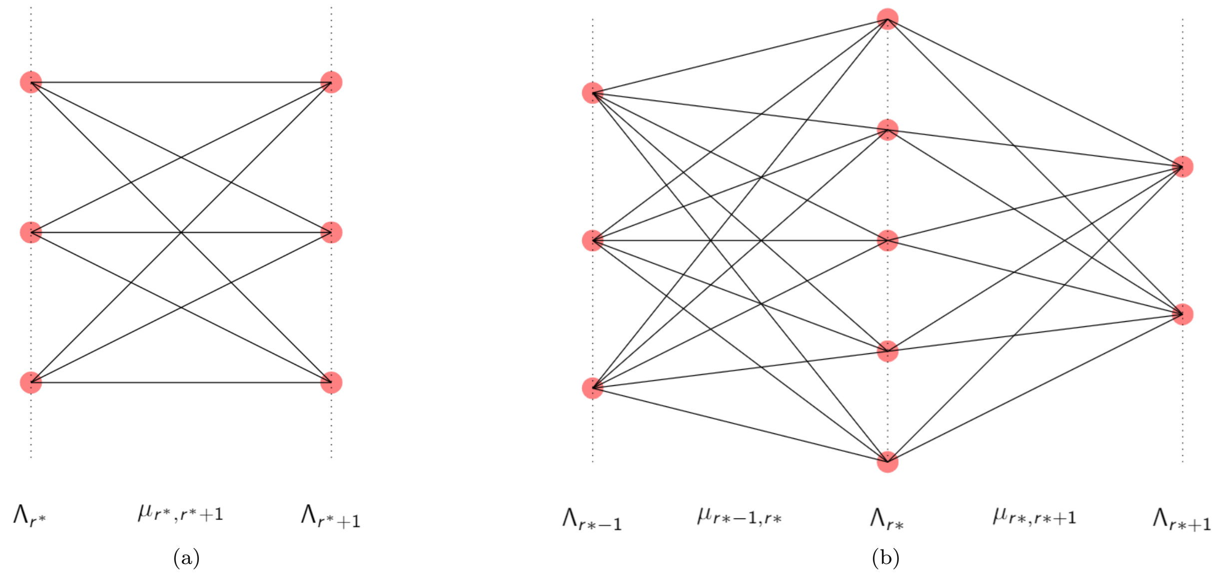

Figure 2: Graphical representation of the two possible cases (a) and (b), with reference to Proposition 3, in which the spectral radius attains its maximum value. The non extensive layers are not displayed.

how the signal components can be partitioned in different species (layers) in order to minimize the set of those signal-to-noise ratios μ which do not allow for signal reconstruction, namely when all the $\mathbb{E}\langle m_r\rangle$'s in (9) asymptotically vanish. Indeed in that case the MMSE reaches its maximum value.

Proposition 3 (Deep architecture optimization [13]). *Let μ be the tridiagonal matrix* (17) *with K even. The spectral radius of the $K/2 \times K/2$ matrix obtained from $(\mu\alpha)^2$ by erasing the even rows and columns equals at most $\frac{1}{4}\max_r \mu_{r,r+1}^2$. Moreover this maximum is attained if and only if one of the following conditions is verified:*

(a) there are only two extensive layers, which have equal size and have the highest interaction strength. Namely:

$$\alpha_{r^*} = \alpha_{r^*+1} = \frac{1}{2} \quad , \quad \mu_{r^*,\,r^*+1} = \max_r \mu_{r,r+1} \tag{20}$$

for some $r^* \in \{1, \dots, K-1\}$;

(b) there are three consecutive extensive layers with the highest interaction strengths. One half of the volume is in the middle layer and the remaining half is arbitrarily shared by the other two. Namely:

$$\begin{aligned} \alpha_{r^*} = \alpha_{r^*-1} + \alpha_{r^*+1} = \frac{1}{2}\,, \\ \mu_{r^*-1,\,r^*} = \mu_{r^*,\,r^*+1} = \max_r \mu_{r,r+1} \end{aligned} \tag{21}$$

for some $r^* \in \{2, \dots, K-1\}$.

In this paper we have outlined the solution of a multichannel Wigner spiked model from a Statistical Mechanics perspective. Two special architectures are explicitly studied, the convex and the deep case, corresponding respectively to a non-restricted Boltzmann Machine and a restricted one. The inference problem that we present here is in the Bayesian Optimal Setting, when the receiver knows the signal-to-noise ratios and the distribution of the ground truth signal. In this setting one is allowed to use the Nishimori identities which remarkably simplify the treatment. When instead the receiver ignores the prior or the signal-to-noise ratio, only a few results are available [20–24]. In such settings, called *mismatched*, the Nishimori identities do not hold true and a replica symmetry breaking variational principle may arise [25]. It is worth noticing that the Statistical Mechanics model described by the Hamiltonian (8) with centered interactions (which for $K = 1$ reduces to the standard Sherrington-Kirkpatrick model [26–29]) is still unsolved for non-convex architectures such as the deep one, see nevertheless [19, 30–32].

References

[1] El Alaoui A. and Krzakala F., *Estimation in the spiked wigner model: A short proof of the replica formula* 2018 pp. 1874–1878.

[2] Barbier J. and Macris N., *Probability Theory and Related Fields*, **174** (2019) .

[3] Lelarge M. and Miolane L., *Fundamental limits of symmetric low-rank matrix estimation* in proc. of *Proceedings of the 2017 Conference on Learning Theory* Vol. 65 of *Proceedings of Machine Learning Research* (PMLR) 2017 pp. 1297–1301.
`http://proceedings.mlr.press/v65/lelarge17a.html`

[4] Agliari E. and Marzo G. D., *The European Physics Journal Plus*, **135** (2020) 883.

[5] Agliari E., Alemanno F., Barra A., Centonze M. and Fachechi A., *Physical Review Letters*, **124** (2020) 028301.

[6] Sompolinsky H., *Physical Review A*, **34** (1986) 2571.

[7] Barbier J., Dia M., Macris N., Krzakala F., Lesieur T. and Zdeborová L., *Advances in Neural Information Processing Systems*, **29** (2016) .

[8] Mézard M. and Montanari A., *Information, Physics, and Computation* (Oxford University Press, Inc., USA) 2009.

[9] Nishimori H., *Statistical Physics of Spin Glasses and Information Processing: an Introduction* (Oxford University Press, Oxford; New York) 2001.

[10] Barra A., Contucci P., Mingione E. and Tantari D., *Annales Institut Henri Poincaré*, **16** (2013) .

[11] Panchenko D., *Annals of Probability*, **43** (2015) .

[12] Alberici D., Camilli F., Contucci P. and Mingione E., *Journal of Statistical Physics*, **182** (2020) .

[13] Alberici D., Camilli F., Contucci P. and Mingione E., *Communications in Mathematical Physics*, (2021) .

[14] Salakhutdinov R. and Hinton G., *Deep boltzmann machines* in proc. of *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, edited by van Dyk D. and Welling M., Vol. 5 of *Proceedings of Machine Learning Research* (PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA) 2009 pp. 448–455.
`https://proceedings.mlr.press/v5/salakhutdinov09a.html`

[15] Contucci P. and Giardinà C., *Perspectives on Spin Glasses* (Cambridge University Press) 2012.

[16] Barbier J., *Information and Inference: A Journal of the IMA*, **10** (2020) 597.

[17] El Alaoui A., Krzakala F. and Jordan M. I., *arXiv e-prints*, (2017) .

[18] Alberici D., Barra A., Contucci P. and Mingione E., *Journal of Statistical Physics*, **180** (2020) .

[19] Alberici D., Contucci P. and Mingione E., *Annales Institut Henri Poincaré*, (2021) .

[20] Verdú S., *IEEE Transactions on Information Theory*, **56** (2010) 3712.

[21] Barbier J., Panchenko D. and Sáenz M., *Strong replica symmetry for high-dimensional disordered log-concave gibbs measures* (2020).

[22] Barbier J., Chen W.-K., Panchenko D. and Sáenz M., *Performance of bayesian linear regression in a model with mismatch* (2021).

[23] Mukherjee S. and Sen S., *Variational inference in high-dimensional linear regression* (2021).

[24] Pourkamali F. and Macris N., *Mismatched estimation of rank-one symmetric matrices under gaussian noise* (2021).

[25] Camilli F., Contucci P. and Mingione E., *arXiv e-prints*, (2021) .

[26] Mezard M., Parisi G. and Virasoro M., *Spin Glass Theory And Beyond: An Introduction To The Replica Method And Its Applications* World Scientific Lecture Notes In Physics (World Scientific Publishing Company) 1987.
https://books.google.it/books?id=DwY8DQAAQBAJ

[27] Guerra F., *Communications in Mathematical Physics*, **233** (2003) .

[28] Talagrand M., *Mean Field Models for Spin Glasses: Volume I: Basic Examples* (Springer) 2010.

[29] Panchenko D., *The Sherrington-Kirkpatrick Model* (Springer) 2015.

[30] Baik J. and Lee J. O., *Annales Institut Henri Poincaré*, **56** (2020) .

[31] Auffinger A. and Chen W.-K., *Journal of Statistical Physics*, **157** (2014) .

[32] Mourrat J., *Prob. Math. Phys.*, (2021) .