Solving the inverse Ising problem by mean-field methods in a clustered phase space with many states

Aurélien Decelle[1,*] and Federico Ricci-Tersenghi[2,3,4]

[1]*Laboratoire de Recherche en Informatique, TAO - INRIA, CNRS and Université Paris-Sud 11, Bâtiment 660, 91190 Gif-sur-Yvette, France*
[2]*Dipartimento di Fisica, Università La Sapienza, Piazzale Aldo Moro 5, I-00185 Roma, Italy*
[3]*Sezione di Roma 1, INFN, I-00185 Roma, Italy*
[4]*Unità di Roma, NANOTEC, CNR, I-00185 Roma, Italy*

In this work we explain how to properly use mean-field methods to solve the inverse Ising problem when the phase space is clustered, that is, many states are present. The clustering of the phase space can occur for many reasons, e.g., when a system undergoes a phase transition, but also when data are collected in different regimes (e.g., quiescent and spiking regimes in neural networks). Mean-field methods for the inverse Ising problem are typically used without taking into account the eventual clustered structure of the input configurations and may lead to very poor inference (e.g., in the low-temperature phase of the Curie-Weiss model). In this work we explain how to modify mean-field approaches when the phase space is clustered and we illustrate the effectiveness of our method on different clustered structures (low-temperature phases of Curie-Weiss and Hopfield models).

I. INTRODUCTION

The Ising inverse problem has been the subject of a large amount of works recently [1–5]. Although this problem is known since many decades under the name of Boltzmann machine learning (BML), many recent applications and developments in different fields (e.g., biology [6–8], computer science [9], and physics [10–12]) have renewed the interest in studying such problems. The BML can be investigated under two very different approaches. In the first one, which concerns this work, a set of data is generated according to the Gibbs-Boltzmann measure of a generic Ising model. The input data for the inverse problem are therefore independent and distributed accordingly to the Boltzmann distribution of the system [13]. In a second case, the data are generated according to a stochastic dynamical process which correlates configurations close in time, and these correlations in the input data are exploited in solving the inverse problem [14]. In both cases, the traditional Bayesian approach consists of maximizing the likelihood function of the data. In this work, we focus on the first case which is commonly named "static inverse Ising problem" and is harder than the second case.

In the static case, maximizing the likelihood is a complicated task, because it directly depends on the partition function which is impossible to compute efficiently (in the general case, its complexity grows exponentially with the system size). However, it is still possible to maximize the likelihood by the expectation-maximization method using a Monte Carlo (MC) algorithm and doing a Boltzmann learning procedure [13]. The MC algorithm is used to evaluate the average value of the observables of the system (here the magnetizations and the correlations) and to update the value of the magnetic fields and the couplings by doing a gradient ascent. Yet, it is known that MC estimates do not converge quickly in many cases and may require many steps to obtain accurate mean values. It means that the MC algorithm should be run for a long time at each step of the BML procedure making the method quite slow. For this

reason, faster methods based on mean-field approximations are commonly used in practical applications.

In a recent work [11] Nguyen and Berg have revisited the problem of finding a good mean-field (MF) approximation for the inverse Ising problem. It was already known that MF methods fail to provide a good couplings reconstruction at low temperatures even for ferromagnetic systems (see Fig. 1 for an example on a ferromagnet and [15] for an example on a MF spin glass). Worse than that, this problem in coupling reconstruction occurs also in cases where the MF approximation is exact in the thermodynamical limit (e.g., the Curie-Weiss model). This failure in reconstructing couplings in ferromagnetic systems can be understood by looking at the input configurations at low temperatures: below the ferromagnetic transition, indeed, configurations are clustered in two groups of respectively positive and negative magnetization. The naive MF (nMF) approximation is based on the self-consistency equations for the magnetizations, $m_i = \tanh(\beta \sum_j J_{ij} m_j)$, with β being the inverse temperature, which have three solutions for $\beta > \beta_c$: it is well known that the $m_i = 0$ solution is unphysical, while the two solutions with $m_i \neq 0$ are thermodynamically stable. However, considering all the input configurations together, the average magnetizations are zero by symmetry. Therefore, a naive use of MF equations infer the couplings using the unphysical $m_i = 0$ fixed point, and lead to a very poor result. Please notice that the same problem arises if one computes correlations in a naive way: using all input data connected correlations would not decay at long distance. Therefore, in order to use properly the nMF equations, it is mandatory to look at the two other solutions characterized by nonzero magnetization. These solutions arise naturally when considering the decomposition of the Gibbs-Boltzmann measure in the configuration space.

The authors of Ref. [11] consider the nMF equations for both states (of positive and negative magnetizations) at the same time. In this way they obtain an overconstrained system of linear equations to be solved. They manage to find a solution by using the pseudoinverse of a matrix (see [11] for further details). We will see that this approach can be considerably simplified in the case of the Curie-Weiss (CW) model, and

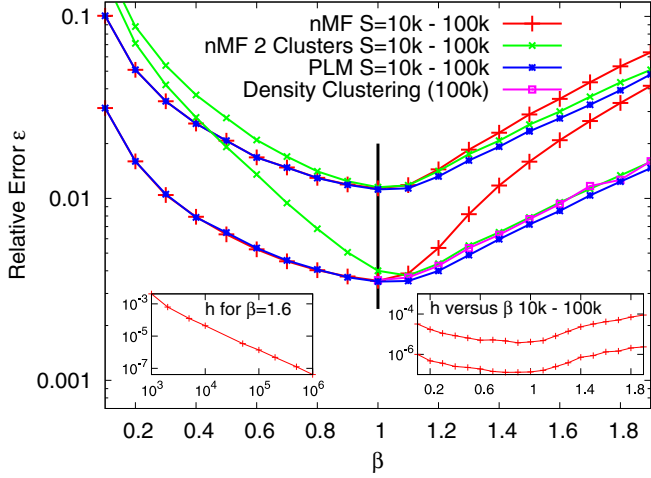*Corresponding author: aurelien.decelle@lri.fr

FIG. 1. Inference of couplings in the CW model with $N = 100$ and two different values of the number M of input configurations. We see that the nMF method with all the input data is good only for $\beta < \beta_c = 1$. For $\beta > \beta_c$ the phase space separates into two states and the nMF method with two clusters gives much better performance (although it fails badly at high temperature). Inference methods, like PLM and nMF with density clustering, that take correctly into account the clustering of input configurations provide the best estimate in the entire temperature range, both above and below the transition temperature. In the left inset, we show how the inferred magnetic field at $\beta = 1.6$ decreases when increasing the number of samples ($M \in [10^3, 10^6]$) used for the inference process via nMF with K-means clustering and $K = 2$. In the right inset the same inferred magnetic field is plotted vs β, for $M = 10^4, 10^5$.

then generalized to models with many free-energy minima. In Ref. [11] also the case of the Sherrington-Kirpatrick (SK) model is considered as a case study with a clustered phase space at low temperatures. We would like to emphasize, however, that the division in metastable states of the SK model is somehow problematic for this approach. The metastable states of the SK model in the glassy phase are highly nontrivial and therefore it is very difficult even to define them properly in a system of limited size. Therefore we claim that the inference algorithm of Ref. [11] as well as the one presented in the present work are not suitable for this kind of model, for which more elaborate techniques (such as the pseudolikelihood method [4,16]) are required.

In the present work we show that couplings can be well inferred using nMF equations also in the low-temperature phase if input configurations are previously clustered and the nMF inference algorithm is applied separately to data in each cluster. We show that our inference procedure based on solving the nMF or Thouless Anderson Palmer (TAP) equations inside each cluster *separately* is much simpler than the method proposed in Ref. [11], where self-consistency equations for each cluster need to be solved *simultaneously*. Therefore the use of complicated numerical algorithms such as the pseudoinverse is not necessary. In addition, we show that, at variance to what is claimed in Ref. [11], using the present inference procedure one does not estimate wrongly the magnetic fields. It is worth mentioning that, when using one of the MF fixed points with $m_i \neq 0$, a spurious magnetic field unavoidably appears due to errors on the inferred couplings.

However, this magnetic field is very small and decreases when increasing the number of input data.

In order to prove that our method is very efficient we apply it to different kind of models. First we show that in the CW model the results are as good as those from more elaborate methods, like the pseudolikelihood method. Then we focus on the Hopfield model where the number of different free-energy minima can be controlled and made larger. We show that it is possible to improve the results on the inference process by clustering the set of input configurations and to infer the right number of clusters to be used. We should mention that a previous attempt to infer the couplings in the (sparse) Hopfield model from data collected in a single state was done in [17]. However, in that work, the interaction network was assumed to be known and only coupling intensities were inferred, so a direct comparison with our results is not possible.

II. PROBLEM DEFINITION AND INFERENCE ALGORITHMS

In the static inverse Ising problem one aims at inferring the value of the couplings between the variables and the eventual magnetic fields, given a set of M equilibrium configurations. More precisely we consider an Ising model with N spins defined by the Hamiltonian

$$\mathcal{H}(\underline{s}) = -\sum_{i<j} J_{ij} s_i s_j - \sum_i h_i s_i, \qquad (1)$$

where $i, j = 1, \ldots, N$. In the static case, the inference process is done by using input data distributed according to the Gibbs-Boltzmann measure

$$P_{\mathrm{GB}}(\underline{s}) = \frac{e^{-\beta \mathcal{H}(\underline{s})}}{Z} \quad \text{where} \quad Z = \sum_{\underline{s}} e^{-\beta \mathcal{H}(\underline{s})}. \qquad (2)$$

We remind here that the M sampled configurations are assumed to be independent.

In the following we will consider two different families of inference methods. For mean-field methods, we shall consider the average magnetizations and correlations of the data,

$$\bar{m}_i = \frac{1}{M} \sum_{a=1}^{M} s_i^a, \qquad (3)$$

$$\bar{c}_{ij} = \frac{1}{M} \sum_{a=1}^{M} s_i^a s_j^a - \bar{m}_i \bar{m}_j. \qquad (4)$$

These observables are the only information needed to infer the parameters of the models when using mean-field methods. We will also consider the pseudolikelihood methods for which the entire sampled configurations $\{s_i^a\}$ are needed. Let us now describe how these methods work and how we will use them in the context of a clustered phase space.

We first consider the naive mean-field approach where the equations can be simply derived by considering the solution of the Curie-Weiss model (where $J_{ij} = 1/N$). For this model, the magnetizations and the correlations are given by

$$m_i = \tanh\left(\beta \sum_{j \neq i} J_{ij} m_j + \beta h_i\right), \qquad (5)$$

$$c_{ik} \equiv \frac{\partial m_i}{\partial h_k} = \beta\left(1 - m_i^2\right)\left(\sum_{j \neq i} J_{ij} c_{jk} + \delta_{ik}\right). \quad (6)$$

By inverting Eq. (6) we can reconstruct directly the couplings J_{ij}^*. Then, by using the J^* and Eq. (5) we can infer the magnetic fields h_i^*,

$$J_{ij}^* = -(c^{-1})_{ij} + \frac{\delta_{ij}}{\beta\left(1 - m_i^2\right)}, \quad (7)$$

$$h_i^* = \beta^{-1}\left[\operatorname{atanh}(m^i) - \sum_{j \neq i} J_{ij}^* m_j\right]. \quad (8)$$

We refer to this method as nMF in the rest of the article.

A second approximation commonly used is to consider the pseudolikelihood method (PLM). PLM is based on the maximization of the marginals probability of one spin s_i given that the rest of the spins are fixed: $p(s_i|s_{j\setminus i})$. The PLM consists of maximizing the sum of all the log-pseudolikelihood [4,16],

$$\mathcal{PL} = \frac{1}{NM}\sum_{i,a}\ln\left[p\left(s_i^a|s_{j\setminus i}^a\right)\right]. \quad (9)$$

In this method, we need to have access to all the configurations $\{s_i^a\}$. The advantage of this method is that it deals also with high-order correlations and thus provides much better performances on finite dimensional systems [12,18], but it also can handle directly clustered phase space. Moreover it has a polynomial complexity at variance to using the true likelihood of the data.

Clustering methods and inference with clustered phase space

Here we describe the clustering algorithms that we use to divide configurations in clusters before applying the nMF method. These clustering algorithms group configurations together based on their distances: configurations are put in the same group if they are "close" enough and "far" from the other clusters, where the concepts of close and far usually need to be determined in a self-consistent way. We use the Hamming distance defined by $d_{ab} = 1/(4N)\sum_i(s_i^a - s_i^b)^2$. In the present work we use two different clustering methods. First we consider the soft K-means clustering [19]. This method clusterizes the space of configurations by assigning each configuration to the closest of the k centers "softly" (a configuration is assign to a center with a given probability). Then the position of the k centers is updated accordingly to the position of the configurations inside each cluster. The procedure is repeated until convergence. This method is very fast, the complexity scale as $O(M)$, but the results can depend strongly on the initial conditions (i.e., on how the k centers are chosen at the beginning).

A second method is based on density clustering. The density clustering algorithm we consider [20] first defines the density around each point. In our case the density is the number of configurations within a given range. Then, each data point is associated to its closest neighbor with higher density. This process naturally separates the phase space into a number of clusters which depend on the range used for defining the neighborhoods. Therefore by using this algorithm we do not need to specify the number of clusters. Thus this second

clustering algorithm has the advantage of finding by itself the number of clusters. It suffers however of a larger complexity, scaling as $O(M^2)$.

After clustering the configurations we have to use them properly to infer the parameters of the model. We define the observables of the kth cluster by

$$\bar{m}_i^{(k)} = \frac{1}{M_k}\sum_{a \in \mathcal{C}_k} s_i^a, \quad (10)$$

$$\bar{c}_{ij}^{(k)} = \frac{1}{M_k}\sum_{a \in \mathcal{C}_k} s_i^a s_j^a - \bar{m}_i^{(k)}\bar{m}_j^{(k)}, \quad (11)$$

where $\mathcal{C}_k$ is the set of indices of configurations belonging to the kth cluster and $M_k = |\mathcal{C}_k|$. We now apply the nMF equations separately for each cluster and obtain a different estimate of the parameters for each cluster $J_{ij}^{(k)}$. Finally, to obtain the best estimate for the couplings we take the weighted average of all the different estimates,

$$J_{ij}^* = \frac{1}{M}\sum_k M_k J_{ij}^{(k)}. \quad (12)$$

To estimate the magnetic field, we first compute them within each cluster: $h_i^{(k)}$ is obtained from Eq. (8) with the estimates $J_{ij}^{(k)}$. The final estimate for the magnetic fields is again given by the weighted average over the clusters,

$$h_i^* = \frac{1}{M}\sum_k M_k h_i^{(k)}. \quad (13)$$

III. RESULTS ON THE CURIE-WEISS MODEL

The Curie-Weiss (CW) model is a fully connected ferromagnet with $J_{ij} = 1/N$, $\forall i \neq j$. The model has a paramagnetic phase ($m_i = 0$) at high temperature $\beta < \beta_c = 1$ and a ferromagnetic phase ($m_i \neq 0$) above β_c. In the ferromagnetic phase, two states of positive and negative magnetizations coexist. In the limit of very large system sizes ($N \to \infty$) magnetizations and correlations can be computed analytically by Eqs. (5) and (6), which are exact up to $O(1/N)$ corrections. It means that, by using Eqs. (7) and (8) one should obtain the best possible estimate of the parameters J_{ij} and h_i, but in the ferromagnetic phase, only the solution with nonzero magnetization of Eq. (5) should be considered (as discussed in the Introduction). We evaluate now how the following three inference algorithms perform in the estimate of couplings in the CW model: (i) the nMF method used naively, without clustering the configurations; (ii) the nMF method on configurations clustered using two clusters; (iii) the PLM on the original configurations.

In Fig. 1 we report the error achieved by different methods in the temperature range $\beta \in [0.1,2]$ with $M = 10^4$ and $M = 10^5$ in inferring the couplings using the following definition:

$$\epsilon^2 = \frac{2}{N(N-1)}\sum_{i<j}(J_{ij} - J_{ij}^*)^2. \quad (14)$$

For $\beta < \beta_c$ the paramagnetic fixed point is correct and therefore the reconstruction achieved by nMF is the best possible. However, for $\beta > \beta_c$ the nMF error (red curves) suddenly rises, because the $m_i = 0$ fixed point is no longer

the physical one. On the contrary, using the nMF method on the data clustered with exactly two clusters (green curves), provides a small error in the ferromagnetic phase, but fails badly in the paramagnetic phase. The inference methods that provide the best estimate in the whole temperature range are the PLM (blue curves) and the nMF with data clustered via density clustering (purple curve), that automatically split the input data in one or two clusters, depending on symmetries in the input data. It is worth stressing that these two methods have essentially the same error at any temperature: that is even the nMF approximation provides the best possible estimates if applied to properly clustered data.

In Fig. 1 we show results obtained with $M = 10^4$ and $M = 10^5$ in order to make evident whether the uncertainties in the couplings estimates are due to the noise in the input data or to an intrinsic limitation of the inference algorithm. For example deep in the ferromagnetic phase the nMF method has an error decreasing only slightly when M increases, because the error is mainly due to a limitation of the method. On the contrary, PLM and nMF with properly clustered data provide a result whose uncertainty is mainly due to noise in the input data: indeed the error decreases as $1/\sqrt{M}$.

To confirm the correctness of the inference algorithm based on data clustering and nMF equations, we also looked at the inferred value of the magnetic field by using Eqs. (8) and (13). We see clearly in the insets of Fig. 1 that, in the low-temperature phase, the clustering + nMF method does not predict any anomalously large magnetic field, thanks to the fact that, clustering the input data, we are actually using the magnetized solutions of Eq. (8). In our numerical experiments, we have found too large inferred magnetic fields only if either system size was too small or the input data were too noisy: in the former case the problem resides in the fact Eq. (8) is crudely approximate, while in the latter case it is a consequence of large errors in couplings reconstruction.

IV. RESULTS ON THE HOPFIELD MODEL

We now extend our analysis to a more complicated case by considering the Hopfield model. The Hopfield model has been introduced long time ago [21] to model neural networks: it is a fully connected Ising model, whose couplings can be chosen such that the model free energy has $2P$ different minima (that act has attractors for the pattern recovery dynamics). In some sense, the Hopfield model can be seen as a generalization of the Curie-Weiss model, which is indeed equivalent to the $P = 1$ case. We are interested in studying the inverse Ising problem in the Hopfield model, because configurations sampled at low temperature in the Hopfield model are typically clustered around the $2P$ free-energy minima: consequently naive MF methods face even more severe limitations than in the low-temperature phase of the CW model, and we want to study how much MF methods for the inverse Ising problem can be improved by clustering input configurations.

The Hamiltonian of the Hopfield model reads

$$\mathcal{H}(\underline{s}) = -\frac{1}{N} \sum_{ij} \frac{1}{P} \sum_{\alpha=1}^{P} \xi_i^\alpha \xi_j^\alpha s_i s_j, \qquad (15)$$

where the P patterns ξ^α identify the directions of the free-energy minima. In the standard Hopfield model, the ξs are drawn from the bimodal distribution, that is $\xi_i^\alpha = \pm 1$ with probability $1/2$ independently. In our study we also consider the case where the pattern ξ are correlated by setting 10% of their components equal ($\xi_i^\alpha = \xi_i^\beta \quad \forall \alpha, \beta$), and anticorrelated (only when $P = 2$) by setting 10% of their components in an opposite way ($\xi_i^1 = -\xi_i^2$). This model presents a paramagnetic phase at high temperature, and an ordered phase at low temperature defined by the states around the patterns $\{\xi\}$ if the number of patterns is not too high [22]. The ordered phase is characterized by a Gibbs-Boltzmann measure clustered around one of the $2P$ available states (for a given P there will be $2P$ stable states in the low-temperature region due to the spin-flip symmetry).

We show now our results on inferring the Hopfield couplings by using MF methods on clustered data. In Fig. 2 we consider systems with $N = 100$ spins, $P = 2$ (therefore four states) in all the three possible cases (standard, correlated, and anticorrelated patterns). We observed that MF methods with the right number of clusters perform similarly to the PLM, which is at present the best possible algorithm to solve the inverse Ising problem. The right number of clusters can be obtained either by density clustering or by maximizing the likelihood of the clustering obtained by K means [see Fig. 2(a)].

As in the CW model, also for the Hopfield model the magnetic fields inferred by MF methods on clustered data are very small, and independent on the eventual long-range order present in the model [see inset in Fig. 2(c)].

In Fig. 3 we show the results on inferring couplings of Hopfield models with $P = 3$ patterns (and thus six free-energy minima). Again MF methods applied on input data clustered with the right number of clusters perform very similarly to PLM, and much better than standard MF methods applied directly to all input data. It is worth noticing that the best result by the clustering + nMF algorithm has been obtained by running the clustering procedure several times with different initial conditions (data labeled "many IC" in Fig. 3) and then picking the clustering having the largest likelihood. This is expected since a clustering algorithm as K means is not very stable for large K and its outcome strongly depends on the initial condition.

Let us finally discuss the time complexity of the three algorithms we have used: PLM, K means + nMF and dens.clus. + nMF. Regarding the system size dependence, all three algorithms have a time complexity $O(N^3)$, either because of the inversion of a $N \times N$ matrix in nMF methods, or because of the computation of the gradient of the pseudolikelihood (PL), which is $O(N)$, in a space of $O(N^2)$ variables. Their dependence on the number M of input configurations is different: PLM is linear in M, but the search for the maximum of the PL, requires us to compute PL and its derivatives many times; K means is linear in M, but often a search for the optimal clustering requires to run it with many different initial conditions; density clustering is $O(M^2)$, so, although it provides a robust result, it is impracticable when the number of samples is very high (however, we are aware that the authors of Ref. [20] are developing a faster version of the density clustering algorithm). Therefore, nMF methods are
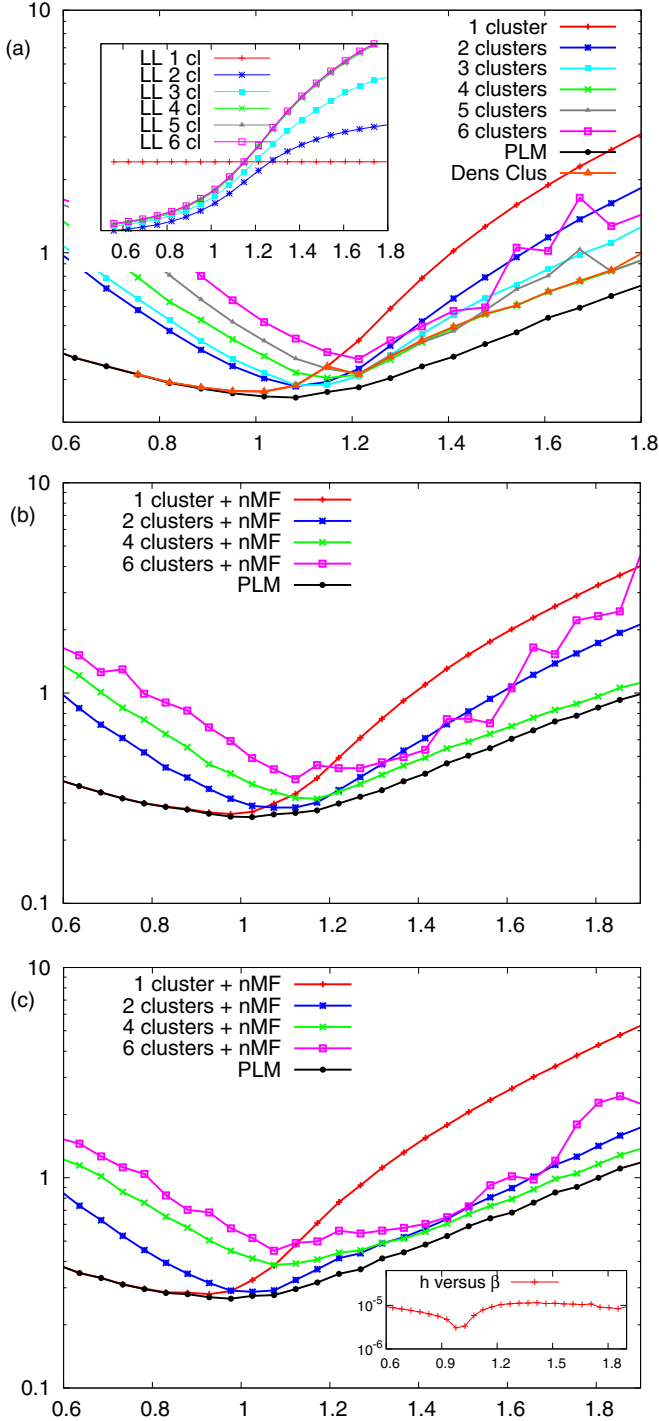
FIG. 2. Main panels: errors in inferring couplings in Hopfield models with $P = 2$ uncorrelated (top), correlated (center), and anticorrelated (bottom) patterns. The comparison is between MF methods with clustered data (either K-means or density clustering) and PLM. In the inset of the top panel, we show that the likelihood of the clustering algorithm suggests to take one cluster below $\beta_c \approx 1.1$ and four clusters above β_c. In the inset of the bottom panel we show the magnetic field inferred by nMF + clustering, which is very small in both phases.

always faster with a total complexity of $O(KMN + N^3)$ whereas PLM is $O(MN^3)$. In practice, we observe it is better
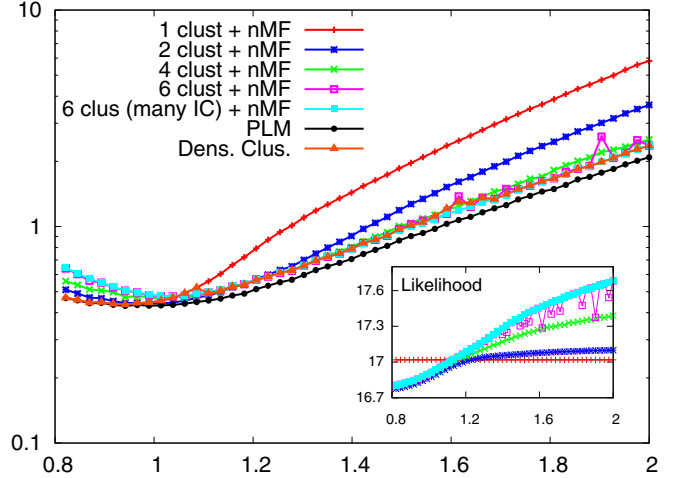

FIG. 3. Errors on inferring couplings in the Hopfield model with three patterns (and thus six minima). We observe again that our algorithm, based on MF methods applied to clustered data, achieves its best performance when input data are split in six clusters. We also put for comparison the results obtained when the clustering is done many times with different initial conditions (labeled "many IC") and then we picked the clustering having the largest likelihood. In this case, the error matches the error obtained when putting each configuration in the correct cluster. We can see that our method performs its best at almost any β value, but at few points where it is particularly difficult to find the best clustering. In the inset we see that likelihood maximization suggests to use one cluster for $\beta < \beta_c$ and six clusters for $\beta > \beta_c$.

to use PLM when the number M of input configurations is small since it gives in general better estimates of the reconstructed couplings. When M becomes large, nMF with K-means clustering is clearly recommended since PLM would be affected by the large number of samples.

V. CONCLUSIONS

In this work we have presented a very simple way to make mean-field approximations to the inverse Ising problem effective also in the low-temperature phase, where symmetries get usually broken and, correspondingly, input data get clustered. The idea is to cluster the input data and to apply mean-field methods to each data cluster. We have tested this clustering + nMF algorithm on the Curie-Weiss and Hopfield models, comparing results with the most sophisticated and state-of-the-art pseudolikelihood method.

Results are very promising and redeem mean-field approximations to inverse problems, even in those cases where the structure of the input data is such that a straightforward application of mean-field methods would be ineffective.

The natural followup to this work is application of clustering + nMF methods to inverse problems based on real data. It is worth remembering that often in solving inverse problems based on real and noisy data, the robustness of simple MF methods is more valuable than the putative higher accuracy of more sophisticated methods; see, e.g., the case of inferring protein contacts [6]. It is also worth mentioning cases where the data can be naturally divided in two or more classes,

exhibiting different statistical properties, but this is usually not taken into account when estimating model parameters. For example Ref. [23] presents an impressively detailed analysis of neuronal spiking patterns. Nonetheless the data belonging to two different regimes (quiescent and spiking) are merged together before doing the analysis according to mean-field approximation. The application of the method presented in this work is likely to improve inference and reduce errors. Finally, the numerous recent studies on pattern recognition using neural networks might also benefit from an approach

dealing with clusters. In those systems it is quite common to deal with many basins of attraction that are used to improve the neural network efficiency. Mean-field techniques would be more than welcome since methods such as PLM cannot deal with the large dataset size (particularly since the average over all samples has to be done at each step of the algorithm).

From this point of view, enlarging the range of applicability of MF methods by data clustering is certainly very useful and may be better than developing higher-order approximations (that strongly depends on the model used to describe the data).

[1] S. Cocco and R. Monasson, J. Stat. Phys. **147**, 252 (2012).

[2] H. Huang and Y. Kabashima, Phys. Rev. E **87**, 062129 (2013).

[3] I. Mastromatteo, J. Stat. Phys. **150**, 658 (2013).

[4] P. Ravikumar, M. J. Wainwright, and J. D. Lafferty, Ann. Stat. **38**, 1287 (2010).

[5] Y. Roudi, J. Tyrcha, and J. Hertz, Phys. Rev. E **79**, 051915 (2009).

[6] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, Proc. Natl. Acad. Sci. USA **108**, E1293 (2011).

[7] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Proc. Natl. Acad. Sci. USA **109**, 4786 (2012).

[8] J. Barton and S. Cocco, J. Stat. Mech. (2013) P03002.

[9] C.-J. Hsieh, I. S. Dhillon, P. K. Ravikumar, and M. A. Sustik, in *Advances in Neural Information Processing Systems 24*, edited by J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger (Curran Associates, Inc., Red Hook, NY, 2011), pp. 2330–2338.

[10] F. Ricci-Tersenghi, J. Stat. Mech. (2012) P08015.

[11] H. C. Nguyen and J. Berg, Phys. Rev. Lett. **109**, 050602 (2012).

[12] A. Decelle and F. Ricci-Tersenghi, Phys. Rev. Lett. **112**, 070603 (2014).

[13] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, Cognit. Sci. **9**, 147 (1985).

[14] S. Cocco, S. Leibler, and R. Monasson, Proc. Natl. Acad. Sci. USA **106**, 14058 (2009).

[15] M. Mézard and T. Mora, J. Physiol. (Paris) **103**, 107 (2009).

[16] J. Besag, The Statist. **24**, 179 (1975).

[17] A. Braunstein, A. Ramezanpour, R. Zecchina, and P. Zhang, Phys. Rev. E **83**, 056114 (2011).

[18] M. Ekeberg, C. Lövkvist, Y. Lan, M. Weigt, and E. Aurell, Phys. Rev. E **87**, 012707 (2013).

[19] D. J. MacKay, *Information Theory, Inference, and Learning Algorithms* (Citeseer, Cambridge, 2003), Vol. 7.

[20] A. Rodriguez and A. Laio, Science **344**, 1492 (2014).

[21] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[22] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).

[23] D. Panas, H. Amin, A. Maccione, O. Muthmann, M. van Rossum, L. Berdondini, and M. H. Hennig, J. Neurosci. **35**, 8480 (2015).