On the glassy nature of the hard phase in inference problems

Fabrizio Antenucci, Silvio Franz, Pierfrancesco Urbani, Lenka Zdeborová

On the glassy nature of the hard phase in inference problems

Fabrizio Antenucci,[1,2] Silvio Franz,[3] Pierfrancesco Urbani,[1] and Lenka Zdeborová[1]

[1] *Institut de physique théorique, Université Paris Saclay, CNRS, CEA, F-91191 Gif-sur-Yvette, France*
[2] *Soft and Living Matter Lab., Rome Unit of CNR-NANOTEC, Institute of Nanotechnology, Piazzale Aldo Moro 5, I-00185, Rome, Italy*
[3] *LPTMS, Université Paris-Sud 11, UMR 8626 CNRS, Bât. 100, 91405 Orsay Cedex, France*

An algorithmically hard phase was described in a range of inference problems: even if the signal can be reconstructed with a small error from an information theoretic point of view, known algorithms fail unless the noise-to-signal ratio is sufficiently small. This *hard phase* is typically understood as a metastable branch of the dynamical evolution of message passing algorithms. In this work we study the metastable branch for a prototypical inference problem, the low-rank matrix factorization, that presents a hard phase. We show that for noise-to-signal ratios that are below the information theoretic threshold, the posterior measure is composed of an exponential number of metastable glassy states and we compute their entropy, called the complexity. We show that this glassiness extends even slightly below the algorithmic threshold below which the well-known approximate message passing (AMP) algorithm is able to closely reconstruct the signal. Counter-intuitively, we find that the performance of the AMP algorithm is not improved by taking into account the glassy nature of the hard phase.

I. INTRODUCTION

Inference problems are ubiquitous in many scientific areas involving data. They can be summarized as follows: a signal is measured or observed in some way and the inference task is to reconstruct the signal from the set of observations. Many practical applications involving data rely on our ability to solve inference problems fast and efficiently. While from the point of view of computational complexity theory many of the practically important inference problems are algorithmically hard in the worst case, practitioners are solving them every day in many cases of interest. It is hence an important research question to know which types of inference problems can be solved efficiently and which cannot. Formally satisfying answer to this question would lead to an entirely new theory of typical computational complexity, and would likely shed new light on the way we develop algorithms.

For a range of inference problems the Bayesian inference naturally leads to statistical physics of systems with disorder, see e.g. [1]. This connection was explored in a range of recent works and brought a class of models for inference problem in which the Bayes-optimal inference can be analyzed and presents a first order phase transition. As common in physics in high dimension, the first order phase transition is associated to the existence of a metastable region in which known efficient algorithms fail to reach the theoretical optimal performance. This metastable region was coined as the *hard phase*, see e.g. [2]. It has been located in error correcting codes [3, 4], compressed sensing [5], community detection [6], the hidden-dense submatrix problem [7, 8], low-rank estimation problems including data clustering, sparse PCA or tensor factorization [9, 10], learning in neural networks [11]. The nature of the hard phase in all these problems is of the same origin, and therefore it is expected that algorithmic improvement in any of them would lead to improvement in all the others as well.

In the current state-of-the-art (including the references above) the hard phase is located as a performance barrier of a class of message passing algorithms. Message passing algorithms can be seen as spin-offs of the cavity method of spin glasses [12]. In the context of inference on dense graphical models the algorithms is called approximate message passing (AMP) known from the context of compressed sensing [13]. In the limit of large system size, the dynamical evolution of AMP can be tracked by the so-called *state evolution* (SE) [13, 14], whose fixed point equations coincide with the saddle point equations describing the thermodynamic of the system under the *replica symmetric* assumption. The analysis of SE and its comparison to the analysis of the Bayes-optimal performance reveals that there is an interval of noise-to-signal ratio where the signal could be reconstructed by sampling the posterior measure, while AMP is not able to converge to the optimal error. This interval marks the presence of the *hard phase*.

In this paper we want to attract further attention of the physics community towards the existence of this hard phase related to a 1st order phase transition in the optimal performance in inference problems. The following open questions might use the physics-like approach and insights: Could there be a physics-inspired algorithm that is able to overcome the algorithmic barrier the AMP algorithm encounters? Note that in problems where the corresponding graphical model can be designed, such as compressed sensing or error correcting codes, such a strategy related to nucleation indeed exists [5, 15]. But what about the more ubiquitous problems where the graphical model is fixed? Are there some physical principles or laws that can provide further evidence towards the impenetrability of the algorithmic barrier?

The motivation of the present work was to challenge the conjecture about impenetrability of the hard phase for efficient algorithms. We analyze the following physics-motivated strategy: It is known that the metastable

part of the posterior measure in the hard phase is glassy [16, 17]. Yet, the AMP algorithm fails to describe this glassiness properly. In some other contexts where message passing algorithms are successfully used, a correct account of glassiness leads to algorithm that improve over simpler ones. Notably this is the case of random constraint satisfaction problems, where the influential work [18] has shown that *survey propagation*, that takes correctly glassiness into account, beats the performance of *belief propagation*.

We pose therefore the problem whether, in inference tasks, the reconstruction of the signal becomes easier when one uses algorithms in which the glassiness is correctly taken into account. We investigate this strategy thoroughly in the present work. We confirm that the hard phase is glassy in the sense that it consists of an exponential number of local optima at higher free energy than the equilibrium one. However, when it comes to the reconstruction of the signal, our analysis leads us to the remarkable conclusion that, in contrast to constraint satisfaction and optimization problems, in inference problems taking into account the glassiness of the hard phase does not improve upon the performance of the simplest AMP algorithm. We thus provide an additional evidence towards the bold conjecture that in the corresponding inference problems AMP is the best of low-computational-complexity inference algorithms.

Our analysis of the glassiness of the hard phase provides new insights on the performance of Monte Carlo or Langevin dynamics. Presence of the glassiness suggests that these sampling-based algorithms are slowed-down and thus their commonly used versions may not be able to match the performance of AMP. While this aligns with some of the the early literature [16], more recent literature [6] suggested, based on numerical evidence, that Monte Carlo sampling is as good as the message passing algorithm. Based on conclusion of our work, this question of performance barriers of sampling-based algorithms should be re-opened and investigated more thoroughly. Good understanding of performance of these algorithms is especially important in the view of the fact that some of the most performing systems currently use stochastic gradient descent, that can be seen as a variant of the Langevin dynamics.

This paper is organized as follows. In Section II we introduce the model on which we illustrate the main findings of this paper, we expect this picture to be generic and apply to all the models where the hard phase related to a first order phase transition in the performance of the Bayesian inference was identified. In Section III we remind the basic setting of Bayesian inference. In Section IV we then remind the replica approach to the study of the corresponding posterior measure. Section IV A then summarized the known replica symmetric diagram and the resulting phase transitions. Section IV B then includes the main technical results of the paper where we quantitatively analyze the glassiness of the hard phase, giving rise to our conclusions in section V.

II. MODEL

In order to be concrete we concentrate on a prototypical example of an inference problem with a hard phase - the constrained rank-one matrix estimation. This problem is representative of the whole class of inference problems where the hard phase related to a 1st order phase transition was identified [7, 19, 20]. We choose this example because it is very close to the Sherrington-Kirkpatrick model for which the study of glassy states is the most advanced [12]. Glassiness was also studied in detail in the spherical or Ising p-spin model, corresponding to spiked tensor estimation [9]. However, in that model the hard phase spans the full low-noise phase and the transition towards the easy phase, on which we aim focus here, happens for noise-to-signal-ratio too low to be straightforwardly investigated within the replica method.

In the rank-one matrix estimation problem the signal, denoted by $\underline{x}^{(0)} \in \mathbb{R}^N$, is extracted from some separable prior probability distribution given by $\underline{P}_X(\underline{x}^{(0)}) = \prod_{i=1}^N P(x_i^{(0)})$. This signal is subjected to noisy measurements of the following form

$$Y_{ij} = \frac{1}{\sqrt{N}} x_i^{(0)} x_j^{(0)} + \xi_{ij} \,, \qquad \forall\ i \leq j \tag{1}$$

where ξ_{ij} are Gaussian random variables with zero mean and variance Δ. Therefore one observes the signal through the matrix Y. The inference problem is to reconstruct the signal $\underline{x}^{(0)}$ given the observation of the matrix Y. The informational-theoretically optimal performance in this problem was analyzed in detail in [20] and this analysis was proven rigorously to be correct in [21–24]. Refs. [20, 21, 25] also analyzed the performance of the AMP algorithm.

While the theoretical part of this paper is for a generic prior P_X, the results section focuses on the Rademacher-Bernoulli prior

$$P_X(x) = (1-\rho)\,\delta(x) + \frac{\rho}{2}\left[\delta(x-1) + \delta(x+1)\right] \tag{2}$$

as this is a prototypical yet simple example in which the hard phase appears for sufficiently low ρ [19, 20]. Let us mention that the rank-one matrix estimation with the Rademacher-Bernoulli prior has a very natural interpretation in terms of community detection problem. Keeping this interpretation in mind can help the reader to get intuition about the problem. Nodes are of three types: $x^{(0)} = 1$ belong to one community, $x^{(0)} = -1$ to a second community, and $x^{(0)} = 0$ does not belong to any community. The observations Y_{ij} (1) can be interpreted as weights on edges of a graph that are on average larger for nodes that are either both in community one or both in community two, they are on average smaller if one of the nodes is in community one and the other in community two, and they are independent and unbiased when one of the nodes does not belong to any community. Thanks to the output universality result of [22, 26] the result

presented in this paper also hold for a model where the observations $Y_{ij} \in \{0,1\}$ correspond to the adjacency matrix of an unweighted graph with Fisher information corresponding to the inverse of the variance Δ.

III. BAYESIAN INFERENCE AND APPROXIMATE MESSAGE PASSING

We study the the so-called Bayes optimal setting, which means that we know both the prior $\underline{P}_X(\underline{x})$ and the variance Δ of the noise. The probability distribution of $\underline{x}$ given Y is given by Bayes formula

$$P(\underline{x}|Y) \propto P_X(\underline{x})P(Y|\underline{x})\,. \tag{3}$$

Since the noise ξ_{ij} is Gaussian we have

$$\begin{aligned} P(Y|\underline{x}) &\propto \prod_{i\leq j} \exp\left[-\frac{1}{2\Delta}\left(Y_{ij}-\frac{x_i x_j}{\sqrt{N}}\right)^2\right] \\ &\equiv \prod_{i\leq j} \mathcal{G}\left(Y_{ij}|\frac{x_i x_j}{\sqrt{N}}\right)\,. \end{aligned} \tag{4}$$

Both in Eq. (3) and (4) we have omitted the normalization constants. An estimate of the components of the signal that minimize the mean-squared-error with the ground truth signal $\underline{x}^{(0)}$ is computed as

$$\hat{x}_i = \langle x_i \rangle \tag{5}$$

where the brackets stand for the average over the posterior measure Eq. (3). Therefore in order to solve the inference problem we need to compute the local magnetizations $\{\hat{x}_i\}$. The AMP algorithm is aiming to do precisely that, its derivation can be found e.g. in [20]. AMP boils down to a set of recursion relations of the form

$$\hat{x}_i^{(t+1)} = \mathrm{AMP}_i\left(\underline{\hat{x}}^{(t)}, \hat{x}_i^{(t-1)}\right)\,, \tag{6}$$

whose iterative fixed point is taken as an estimate of the signal. In this work we do not investigate the properties of AMP algorithm but we focus on the statistical properties of the posterior measure that AMP is trying to describe, analyzing the free energy of the system within the replica method. It is known that the replica symmetric (RS) solution describes the fixed points of the state evolution for AMP in the thermodynamic limit. In fact, it is possible to derive a generalized AMP, that we call *Approximate Survey Propagation* (ASP) algorithm, whose state evolution fixed points coincide with the replica equations in the one-step replica symmetry breaking (1RSB) ansatz, which is known to provide a better description -in many case exact- of glassy states [27]. In the next sections we hence study the thermodynamics of the above model in the RS and 1RSB ansatz, focusing on its properties in the hard phase.

IV. THE REPLICA APPROACH TO THE POSTERIOR MEASURE

In order to study the posterior measure, we define the corresponding free energy as

$$\mathrm{f}\left[\Delta; Y\right] = -\frac{1}{N}\ln\int\left(\prod_{i=1}^{N}\mathrm{d}x_i P_X(x_i)\right)\prod_{i\leq j}\mathcal{G}\left(Y_{ij}\left|\frac{x_i x_j}{\sqrt{N}}\right.\right)\,. \tag{7}$$

This is a random object since it depends on the matrix Y. Furthermore it depends on Δ through the function $\mathcal{G}$. Indeed, we want to study the typical behavior of this sample-dependent free energy. Therefore we define

$$\mathrm{f}(\Delta) = \overline{\mathrm{f}\left[\Delta; Y\right]} \equiv \int\left[\prod_{i\leq j}\mathrm{d}Y_{ij}\right] P(Y)\mathrm{f}\left[\Delta; Y\right]\,, \tag{8}$$

where Y is obtained as in Eq. (1), so that $P(Y)$ is given by

$$P(Y) \propto \int \mathrm{d}\underline{x}^{(0)}\,\underline{P}_X\left(\underline{x}^{(0)}\right)\prod_{i\leq j}\mathcal{G}\left(Y_{ij}\left|\frac{x_i^{(0)}x_j^{(0)}}{\sqrt{N}}\right.\right)\,. \tag{9}$$

In order to perform the average defined in Eq. (8) we use the replica method [12]. Introducing

$$\mathcal{Z} = \int\left(\prod_{i=1}^{N}\mathrm{d}x_i P_X(x_i)\right)\prod_{i\leq j}\mathcal{G}\left(Y_{ij}\left|\frac{x_i x_j}{\sqrt{N}}\right.\right)\,, \tag{10}$$

we get

$$\mathrm{f}(\Delta) = -\frac{1}{N}\lim_{n\to 0}\partial_n\int\left[\prod_{i\leq j}\mathrm{d}Y_{ij}\right]P(Y)\mathcal{Z}^n\,. \tag{11}$$

For integer n we can represent $\mathcal{Z}^n$ as an n-dimensional integral over n replicas $\underline{x}^{(a)}$ with $a = 1,\ldots,n$. Stated in this way the problem is obviously symmetric under the exchange of the n replicas among themselves. Moreover since we need to integrate over the signal distribution $P(Y)$ we end up with a system of $n+1$ replicas, that, in the Bayes optimal case, is symmetric under the permutation among *all* the $n+1$ replicas. Performing standard manipulations, see e.g. [12], we arrive at a closed expression for $\mathrm{f}(\Delta)$ that is

$$\mathrm{f}(\Delta) = -\frac{1}{N}\ln\int\mathcal{D}q\mathcal{D}\hat{q}\exp\left[N\mathcal{S}\left(q,\hat{q}\right)\right]\,, \tag{12}$$

where $\mathcal{S}$ is a function that can be computed explicitly and q and $\hat{q}$ are $(n+1)\times(n+1)$ overlap matrices. In the large N limit, the integral in Eq. (12) can be evaluated using the saddle point method. At the saddle point level the physical meaning of the overlap matrix q is given in terms of

$$q_{ab} = \frac{1}{N}\sum_{i=1}^{N}\overline{\left\langle x_i^{(a)}x_i^{(b)}\right\rangle}\,, \tag{13}$$

while the matrix $\hat{q}$ is just a Lagrange multiplier. We denote m the magnetization of the system, meaning

$$m \equiv q_{0a} = q_{a0} = \frac{1}{N}\sum_{i=1}^{N} \overline{\left\langle x_i^{(0)} x_i^{(a)} \right\rangle} \quad a > 0\,. \tag{14}$$

The saddle point equations for q and $\hat{q}$ can be written in complete generality for any n but then one needs to take the analytic continuation down to $n \to 0$. One needs an appropriate scheme from which one can take the replica limit. Here we consider two schemes: the replica symmetric (RS) and the 1-step replica symmetry breaking (1RSB) one. We refer here to symmetry under permutations of the n replicas with index $a = 1, \ldots, n$.

A. Reminder of the replica symmetric solution

The RS scheme boils down to consider

$$\begin{aligned} q_{ab} &= (q_d - q_0)\,\delta_{ab} + q_0 \quad a, b \geq 1\,, \\ \hat{q}_{ab} &= (\hat{q}_d - \hat{q}_0)\,\delta_{ab} + \hat{q}_0 \quad a, b \geq 1\,, \\ q_{0a} &= q_{a0} = m \quad a \geq 1\,, \\ \hat{q}_{0a} &= \hat{q}_{a0} = \hat{m} \quad a \geq 1\,. \end{aligned} \tag{15}$$

From the point of view of the inference, the relevant quantity to look at is the Mean Square Error (MSE)

$$\begin{aligned} \text{MSE} =& \frac{1}{N}\sum_{i=1}^{N} \overline{\left(\langle x_i \rangle - x_i^{(0)}\right)^2} \\ =& \rho - 2m + q_0\,, \end{aligned} \tag{16}$$

where $\rho \equiv \overline{\left\langle x^{(0)} \right\rangle^2}$. Replica symmetry among all the $n+1$ replicas is obtained for $m = q_0$. It is well known that, as a direct consequence of Bayes optimality (also called Nishimori condition [2]), this fully replica symmetric solution is the one that describes thermodynamically dominant states. The more general ansatz is, however, important as it allows to describes metastable states where the Nishimori identities might not hold. Plugging this ansatz inside the expression for $\mathcal{S}$ and taking the saddle point equations w.r.t. all these parameters one gets the replica symmetric solution as reported in [20], and proven to give the equilibrium solution in [23, 24]. The RS free energy can be expressed as

$$\mathrm{f}_{\mathrm{RS}}(\Delta) = \min_m \left\{\phi_{\mathrm{RS}}\left(m, \Delta\right)\right\} \tag{17}$$

with

$$\phi_{\mathrm{RS}}\left(m, \Delta\right) = \frac{m^2}{4\Delta} - \mathbb{E}_{x^{(0)},W}\left[f\left(\frac{m}{\Delta}, \frac{m}{\Delta}x^{(0)} + \sqrt{\frac{m}{\Delta}}W\right)\right] \tag{18}$$

where

$$f\left(A, B\right) = \ln\left[\int dx\, P_X(x)\, e^{-\frac{1}{2}Ax^2 + Bx}\right], \tag{19}$$

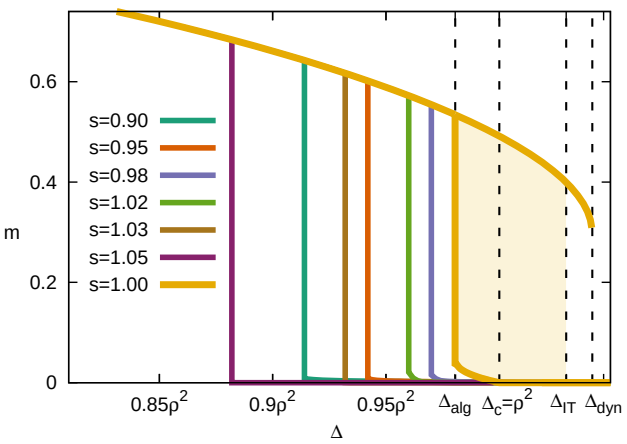

FIG. 1: The magnetization, *aka* the overlap, between the signal and the states described by the 1RSB solution at Parisi parameter s, as a function of the noise strength Δ, and sparsity $\rho = 0.08$. The curve that show a spinodal transition towards the strongly magnetized solution at largest values of Δ is the one for $s = 1$. The same curve represents also the performance of the AMP algorithm. Taking the glassiness of the metastable branch into account does not improve upon AMP.

and $x^{(0)}$ and W are random variables distributed according $P_X\left(x^{(0)}\right)$ and a standard normal distribution, respectively. The values of m for which ϕ_{RS} is stationary are the solution of

$$m = \mathbb{E}_{x^{(0)},W}\left[x^{(0)}\frac{\partial f}{\partial B}\left(\frac{m}{\Delta}, \frac{m}{\Delta}x^{(0)} + \sqrt{\frac{m}{\Delta}}W\right)\right]. \tag{20}$$

Equilibrium properties of the inference problem are given by the global minima of the free energy Eq. (18). Local minima of the free energy that do not correspond to the equilibrium solution are called *metastable*.

For illustration, we consider the case of the Rademacher-Bernoulli prior (2) and we set $\rho = 0.08$ so that the inference problem has an hard phase [20]. The replica symmetric phase diagram is represented in Fig. 1 (yellow curve).

At high Δ the noise is so strong that the signal cannot be recovered and therefore $m = 0$. Upon decreasing Δ the signal is relatively stronger w.r.t the noise and for $\Delta = \Delta_{\mathrm{dyn}} \sim 1.041\rho^2$ the system undergoes a *dynamical transition*. On the one hand one can see that the free energy (18) develops a local metastable minimum with $m > 0$. On the other hand, the $m = 0$ state undergoes a clustering transition according to the pattern familiar in the physics of spin glasses [28, 29]. The corresponding RS free energy ceases to describe a paramagnetic state and it describes a non-ergodic phase with an exponential number $\exp(N\Sigma(\Delta))$ of metastable states - aka clusters - with zero overlap among each other and identical energy and internal entropy. Both the zero m dominating branch and the metastable $m > 0$ branch have identical energy and internal entropy. Their free energy difference is the com-

plexity $f(m>0)-f(m=0)=\Sigma(\Delta)$. Moreover, as we will see in the next section, the typical overlap q_1 between configurations in these states coincides with the value of m of the magnetized solution. For that reason the magnetized state corresponds just to one cluster among the exponential multiplicity dominating the thermodynamics. The complexity (i.e. log of their number) of the thermodynamic states decreases with Δ, until it vanishes at a value $\Delta=\Delta_{\rm IT}\sim 1.0295\rho^2$ where there is the information theoretic phase transition and $\Sigma(\Delta_{\rm IT}=0)$. The signal is here strong enough so that a first order phase transition happens where the minimum with positive magnetization becomes the global minimum of the free energy. The complexity of the $m=0$ solution becomes negative, the solution is non physical and consequently RSB is necessary to describe the metastable branch. Despite this fact, this RS metastable branch cannot be just dismissed as unphysical: it continues to be relevant algorithmically as dynamical attractor of the AMP algorithm. Decreasing the intensity of the noise further, another phase transition happens in this RS branch. At $\Delta=\Delta_c=\rho^2$ the metastable minimum develops a small magnetization. Decreasing even further Δ, at $\Delta=\Delta_{\rm alg}\sim 0.9805\rho^2$ this metastable minimum disappears with a spinodal transition. In the interval $[\Delta_{\rm alg},\Delta_{\rm IT}]$ one finds the hard phase defined by the property that the AMP algorithm is suboptimal (the shaded yellow region in Fig. 1): the global minimum of the free energy has a high m (low MSE), but the small m non-physical local minimum continues to describe the attractor of the AMP. The state evolution describing the AMP algorithm starting from random conditions converges to the local minimum of lowest magnetization.

B. Glassy phase and complexity

The low branch RS solution is non-physical below $\Delta_{\rm IT}$, its existence, however, suggests that metastable states exist that should be described with RSB. We therefore consider the 1RSB ansatz. We divide the n replicas $a=1,\dots,n$ into n/s blocks, where s is the so-called Parisi parameter [12]. The overlap matrix becomes

$$q_{ab}=\begin{cases} q_d & a=b \\ q_1 & a,b \text{ in the same block} \\ q_0 & a,b \text{ in different blocks}\end{cases} \tag{21}$$

and analogous for $\hat{q}$. For s strictly equal to one we get back the replica symmetric ansatz Eq. (15). Note that for $s\neq 1$, m and q_0 are in general different in the solution: this is crucial when evaluating the MSE Eq. (16) as the minimum of the MSE does not correspond in general to the maximum of m.

The 1RSB free energy takes the form

$$\mathrm{f}_{\rm 1RSB}(\Delta,s)=\mathbf{extr}\Big\{\phi_{\rm 1RSB}\left(m,q_0,q_1,\Delta,s\right)\Big\}, \tag{22}$$

with

$$\phi_{\rm 1RSB}\left(m,q_0,q_1,\Delta,s\right)=\frac{m^2}{2\Delta}-s\frac{q_0^2}{4\Delta}-(1-s)\frac{q_1^2}{4\Delta}+ -\frac{1}{s}\mathbb{E}_{x^{(0)},W}\left[f\left(\frac{q_1}{\Delta},\frac{m}{\Delta}x^{(0)}+\sqrt{\frac{q_0}{\Delta}}W,\frac{q_1-q_0}{\Delta}\right)\right], \tag{23}$$

where

$$f\left(A,B,C\right)=\ln\int dh\,\sqrt{\frac{C}{2\pi}}\,e^{-\frac{1}{2}Ch^2}\cdot \cdot\left[\int dx\,P_X(x)\,e^{-\frac{1}{2}Ax^2+(B+Ch)x}\right]^s. \tag{24}$$

The stationary points of the 1RSB free energy are now obtained by the fixed points of

$$\begin{aligned} m&=\frac{1}{s}\mathbb{E}_{x^{(0)},W}\left[x^{(0)}\frac{\partial f}{\partial B}\right] \\ q_0&=\frac{1}{s^2}\mathbb{E}_{x^{(0)},W}\left[\left(\frac{\partial f}{\partial B}\right)^2\right] \\ q_1&=\frac{2}{s(s-1)}\mathbb{E}_{x^{(0)},W}\left[\frac{\partial f}{\partial A}+\frac{\partial f}{\partial C}\right]\end{aligned} \tag{25}$$

where $A=q_1/\Delta$, $B=mx^{(0)}/\Delta+W\sqrt{q_0/\Delta}$ and $C=(q_1-q_0)/\Delta$ and the extremum is a minimum in m and a maximum in the other parameters.

We would like to reiterate here the observation that in the same way that the stationary points of the RS free energy correspond to state evolution fixed points of the AMP algorithm, the stationary points of the 1RSB free energy correspond to the fixed points of the state evolution of an approximate survey propagation algorithm that depends on s [27]. In particular, the expression (16) exactly gives the MSE of such algorithm with m and q_0 being the solution of (25).

For high enough Δ the 1RSB solution collapses to the RS one, meaning that $q_0=q_1=m=0$. At $\Delta_{\rm dyn}$ the saddle point equations for $s=1$ admit a solution with $m=q_0=0$, $q_1>0$. The value of q_1 in this solution coincides with the value of m in the high magnetization RS branch discussed in the previous section. At $\Delta_{\rm IT}$ the metastable states undergo an entropy crisis transition. Although the thermodynamically dominant state becomes the state with high correlation with the ground truth signal, glassy states continue to exist. In fact as far as these states are concerned - if we neglect the high magnetization state - the system undergoes there a Kauzmann transition where the dominant glassy states have zero complexity and a value of the Parisi parameter s is determined by the condition that complexity $\Sigma(\Delta,s)$ (defined below) is equal to zero[1].

[1] Notice the analogy of the high-magnetization state here with the crystal state in the physics of glasses.

Let us now discuss $s \neq 1$ solutions. It is well known that the Parisi parameter s can be interpreted as an effective temperature that enables to select families of metastable states of given (internal) free energy [30]. Their corresponding complexity Σ (defined as the log of their number) is obtained by deriving (23) w.r.t s [30], and multiplying the result by s^2, i.e.

$$\Sigma(\Delta, s) = \frac{s^2}{4\Delta}\left(q_1^2 - q_0^2\right) - s^2 \frac{\partial}{\partial s}\mathbb{E}_{x^{(0)},W}\left[\frac{1}{s} f\left(\frac{q_1}{\Delta}, \frac{m}{\Delta}x^{(0)} + \sqrt{\frac{q_0}{\Delta}}W, \frac{q_1 - q_0}{\Delta}\right)\right] \tag{26}$$

As expected this complexity for $s = 1$ coincides with the free energy difference between the two RS branches discussed in the previous section.

In Fig. 2 we plot the complexity as function of both s and of the noise variance Δ. For each value of s we find two regions: a physical region where Σ is positive, and an non-physical one where $\Sigma < 0$. Note as the physical region with positive complexity continues not only below $\Delta_{\rm IT}$, but even well below $\Delta_{\rm alg}$.

The 1RSB solution is not guaranteed to give the exact description of the glassy states. It is well known that in the replica solutions should be stable against (further) breaking of the replica symmetry. This requires that all the eigenvalues of the Hessian of the free energy should be positive in the solution. The 1RSB solutions can loose stability in two possible ways, associated, to negative values of the following eigenvalues [31–33]:

$$\begin{aligned}\lambda_I &= 1 - \frac{1}{\Delta}\int_{-\infty}^{\infty} dh P(s,h)\left(f''(s,h)\right)^2 \\ \lambda_{II} &= 1 - \frac{1}{\Delta}\int_{-\infty}^{\infty} dh P(1,h)\left(f''(1,h)\right)^2 \ .\end{aligned} \tag{27}$$

where ($A = q_1/\Delta$, $B = \frac{m}{\Delta}x^{(0)} + h$ and $C = (q_1 - q_0)/\Delta$)

$$\begin{aligned} f(1,h) &= \ln \int dx P(x) \exp\left[-\frac{A}{2}x^2 + hx\right] , \\ f(s,h) &= \frac{1}{s}\ln \int \frac{dz}{\sqrt{2\pi C}} e^{-\frac{z^2}{2C}} e^{s f(1,h-z)} , \end{aligned} \tag{28}$$

$$\begin{aligned} P(s,h) &= \mathbb{E}_{x^{(0)}}\left[\sqrt{\frac{\Delta}{2\pi q_0}}\exp\left(-\frac{\Delta}{2q_0}B^2\right)\right] , \\ P(1,h) &= e^{s f(1,h)} \int \frac{dz\, e^{-\frac{z^2}{2C}}}{\sqrt{2\pi C}} \cdot P(s, h-z)\, e^{-s f(s,h-z)} . \end{aligned} \tag{29}$$

A negative λ_I (type I instability) signals the appearance of new scales of distance between states. A negative λ_{II} on the other hand is met for small Δ and small s, where the glassy states are unstable against a *Gardner transition* to further RSB [31, 32]. Each metastable state splits into a hierarchy of new states (type II instability) [33]. In Fig. 2 we mark with full lines the stable region, with dashed lines the unstable ones. Type I instability is found in the physical region at low value of s for sufficiently large Δ and it has been found also in spin glass models [33–35]. Type II instability is found for large s in the non-physical region of negative complexity.

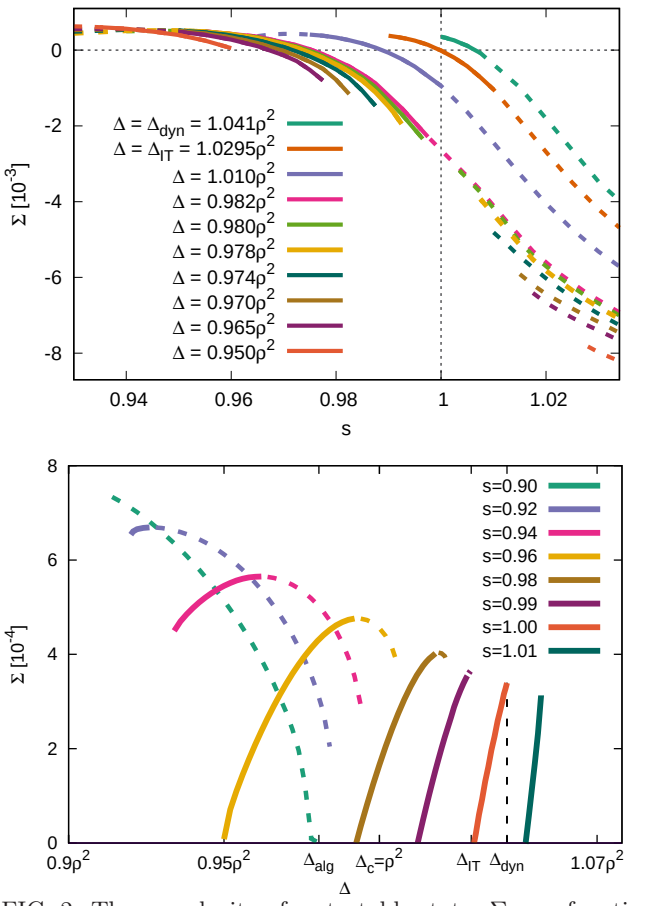

FIG. 2: The complexity of metastable states Σ as a function of the Parisi parameter s and the noise Δ, for prior (2) with sparsity $\rho = 0.08$. Upper panel, complexity at fixed s in the whole domain of existence of a non-trivial fixed point. Lower panel, the physical region of positive Σ as function of Δ. We draw the stable solutions with a solid line and the unstable, wrt the eigenvalues (27), with a dashed line. For each value of $\Delta \in [\Delta_{\rm IT}, \Delta_{\rm dyn}]$ the value of $\Sigma(\Delta, s = 1)$ represents the complexity of the family of thermodynamically dominating states. Below $\Delta_{\rm IT}$ the $s = 1$ solution in non-physical and $\Sigma(\Delta, s = 1) < 0$. The algorithmic threshold of AMP occurs when the ghost-glassy states at $s = 1$ have a spinodal transition towards the signal.

Let's now discuss in detail the glassy solutions that one finds for $\Delta < \Delta_{\rm IT}$ representing metastable states with higher free energy than the high-magnetization solution. These solutions have zero or low magnetization (overlap with the signal). As already remarked, for a given Δ, among all the glassy states the ones with lowest total free energy turn out to be the ones with zero complexity Σ. For different fixed values of the parame-

ter s, the complexity curves reach zero value at different values of Δ. Remarkably, as illustrated in Fig. 2 a stable (towards higher levels of RSB) zero-complexity solution is found down to a value of noise $\Delta_{\rm 1RSB,equil} < \Delta_{\rm alg}$. Stable solutions of positive complexity exists down to $\Delta_{\rm 1RSB,stable} < \Delta_{\rm 1RSB,equil}$, and solutions with positive complexity (irrespective of the stability) down to $\Delta_{\rm 1RSB,all} < \Delta_{\rm 1RSB,stable}$. Example of specific values for $\rho = 0.08$ in Fig. 2 are $\Delta_{\rm alg} \sim 0.9805\rho^2$, $\Delta_{\rm 1RSB,equil} \sim 0.951\rho^2$, $\Delta_{\rm 1RSB,stable} \sim 0.918\rho^2$, $\Delta_{\rm 1RSB,all} \sim 0.903\rho^2$. This notably means that for $\Delta < \Delta_{\rm alg}$, namely in the easy phase where AMP converges close to the signal, families of metastable states continue to exist, some of them being stable with extensive complexity.

One can discuss how do these states influence Monte-Carlo dynamics, that explore the space of configuration according to principles of physical dynamics. On the one hand, one could conjecture that Monte-Carlo dynamics gets trapped by glassy states even below $\Delta_{\rm alg}$. On the other hand, the dynamics is expected to fall out of equilibrium for all $\Delta < \Delta_{\rm dyn}$ and it is not a priori clear in which states it should get trapped. While AMP clearly works for $\Delta < \Delta_{\rm alg}$ and does not work for $\Delta > \Delta_{\rm alg}$, our analysis does not provide any reason why the threshold $\Delta_{\rm alg}$ should be relevant for Monte Carlo or other sampling-based algorithms. For such physical dynamics, numerical simulations and analytic studies in suitable models are necessary to clarify the question of what is the corresponding algorithmic threshold.

So far we focused on glassy states of positive complexity (i.e. existing with probability one for typical instance). There are also solutions of the 1RSB equations having negative complexity. We will call the negative-complexity solution the *ghost-glassy* states. From the physics point of view those solutions do not correspond to physical states for typical instances. Yet, from the algorithmic point of view they do correspond to the fixed points of the ASP algorithm [27] run for a given value of Parisi parameter s, as such they can be reached algorithmically. At this point it becomes relevant to understand for which value $\Delta_{\rm alg}(s)$ do the ghost-glassy state disappear, developing a spinodal instability towards the high-magnetization state. In particular we can ask the natural question if with a suitable choice of the Parisi parameter s the ASP improves over the algorithmic threshold $\Delta_{\rm alg} \equiv \Delta_{\rm alg}(s=1)$ of the usual AMP ($s=1$) and if we could have an s for which $\Delta_{\rm alg}(s) > \Delta_{\rm alg}(1)$. With this question in mind in Fig. 3 we plot the mean-squared error (MSE) with the ground truth signal given by Eq. (16) as a function of s for various values of Δ. We initialize the 1RSB fixed point equations at infinitesimal magnetization and iterate them till a fixed point. We observe that for all values of Δ the MSE is minimized for $s = 1$, i.e. by the canonical AMP algorithm.

The fact that among all the values of s the lowest MSE is reached by the $s = 1$ states for all Δ is unexpected from the physics point of view. It implies that the AMP that neglects glassiness and wrongly describes the region

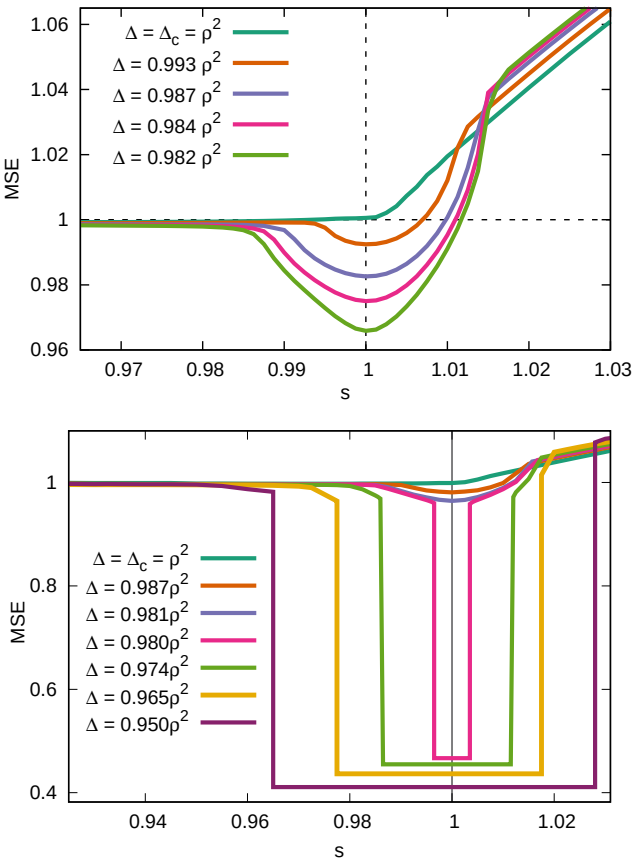

FIG. 3: The MSE as a function of the Parisi parameter s for different values of the noise strength Δ. The smallest MSE is always reached for $s = 1$, corresponding to the performance of the AMP algorithm, with a threshold at $\Delta_{\rm alg} = 0.9805\rho^2$.

$\Delta < \Delta_{\rm IT}$ works better as an inference algorithm than an algorithm that correctly describes the metastable states in this region. However, such a result could be anticipated based on mathematical theorem of [7] that implies that AMP is optimal among all local algorithms. This theorem applies as long as an iterative algorithm only uses information from nearest neighbours and (nearly) reaches a fixed point after $O(1)$ iterations.

V. CONCLUSION

In conclusion, we studied the glassy nature of the hard phase in inference problems. Our results imply that indeed the corresponding metastable state is glassy, i.e. composed of exponentially many states. We evaluate their number (complexity) as a function of their internal free energy to conclude that this glassiness extends to a range of the noise parameter Δ even larger than the extent of the the hard phase. This finding re-opens the natural question of performance limits of Monte-Carlo based sampling. While some recent works [6] anticipated numerically that Monte-Carlo and message pass-

ing will share the same algorithmic threshold, our results do not provide any evidence of this. Instead they suggest that since glassiness is present also below the algorithmic threshold of AMP the performance of sampling-based algorithms will be different in general. Concerning the AMP algorithm, we conclude that, despite the fact that it assumes the hard-phase not to be glassy, the improved description in terms of one-step replica symmetry breaking, that takes glassiness into account, does not provide algorithmic improvement. This is at variance with the situation in random constraint satisfaction problems, where the knowledge of the organization space of solutions provided by 1RSB leads to algorithmic improvement [36].

In this paper we use the example of low-rank matrix estimation with spins 0 and ± 1 as a prototypical example in which the hard phase exists. We checked that the resulting picture applies in a range of parameters and also for some other models (such as planted mixed p-spin model) where the hard phase was identified. We expect the picture presented here to be generic in all the problems where the hard phase related to a first order phase transition was identified.

Finally, we want to mention that the results shown here may be compelling also beyond inference problems. In particular, we notice that the instabilities of the RS solution at $\Delta_{\rm alg}$ and $\Delta_{\rm c}$ can be related to a similar phenomenon occurring in the mean field theory of liquids and glasses [37, 38]. A phase structure similar to the one presented in this paper is found in that case, if we identify Δ as analogue to an (inverse) density parameter and the reconstruction phase as the crystal. Also in that case, the RS solution representing the liquid at low density describes a non-ergodic extensive complexity phase at higher density. As it is the case here, there is a density where complexity vanishes, but the solution can be continued below this point. Finally, there is a maximum density where the solution undergoes an instability - called Kirkwood instability - and ceases to exist [39, 40]. Our analysis suggests that within inference models not only the non-physical negative complexity RS solution could undergo this instability, but also the glassy ones. Whether this phenomenon could be relevant for other glassy systems is an intriguing question.

Acknowledgments

We would like to thank Giulio Biroli, Florent Krzakala, and Guilhem Semerjian for fruitful discussions. This work is supported by "Investissements d'Avenir" LabEx PALM (ANR-10-LABX-0039-PALM) (SaMURai and StatPhysDisSys projects), and from the ERC under the European Unions Horizon 2020 Research and Innovation Programme Grant Agreement 714608-SMiLe. The work of SF was supported by a grant from the Simons Foundation (No. 454941, Silvio Franz).

[1] P. Grassberger and J.-P. Nadal, NATO ASI series. Series C: mathematical and physical sciences (1992).
[2] L. Zdeborová and F. Krzakala, Advances in Physics **65**, 453 (2016).
[3] T. Richardson and R. Urbanke, *Modern coding theory* (Cambridge university press, 2008).
[4] M. Mézard and A. Montanari, *Information, physics, and computation* (Oxford University Press, 2009).
[5] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, Physical Review X **2**, 021005 (2012).
[6] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Physical Review E **84**, 066106 (2011).
[7] Y. Deshpande and A. Montanari, Foundations of Computational Mathematics **15**, 1069 (2015).
[8] A. Montanari, Journal of Statistical Physics **161**, 273 (2015).
[9] E. Richard and A. Montanari, in *Advances in Neural Information Processing Systems* (2014), pp. 2897–2905.
[10] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová, in *Information Theory (ISIT), 2017 IEEE International Symposium on* (IEEE, 2017), pp. 511–515.
[11] G. Györgyi, Physical Review A **41**, 7097 (1990).
[12] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond* (World Scientific, Singapore, 1987).
[13] D. L. Donoho, A. Maleki, and A. Montanari, Proceedings of the National Academy of Sciences **106**, 18914 (2009).
[14] M. Bayati and A. Montanari, IEEE Transactions on Information Theory **57**, 764 (2011).
[15] S. Kudekar, T. J. Richardson, and R. L. Urbanke, IEEE Transactions on Information Theory **57**, 803 (2011).
[16] H. Sompolinsky, N. Tishby, and H. S. Seung, Physical Review Letters **65**, 1683 (1990).
[17] F. Krzakala and L. Zdeborová, Physical Review Letters **102**, 238701 (2009).
[18] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[19] T. Lesieur, F. Krzakala, and L. Zdeborová, in *Information Theory (ISIT), 2015 IEEE International Symposium on* (IEEE, 2015), pp. 1635–1639.
[20] T. Lesieur, F. Krzakala, and L. Zdeborová, Journal of Statistical Mechanics: Theory and Experiment **2017**, 073403 (2017).
[21] Y. Deshpande and A. Montanari, in *Information Theory (ISIT), 2014 IEEE International Symposium on* (IEEE, 2014), pp. 2197–2201.
[22] F. Krzakala, J. Xu, and L. Zdeborová, in *Information Theory Workshop (ITW), 2016 IEEE* (IEEE, 2016), pp. 71–75.
[23] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, in *Advances in Neural Information Processing Systems* (2016), pp. 424–432.
[24] M. Lelarge and L. Miolane, arXiv preprint arXiv:1611.03888 (2016).
[25] S. Rangan and A. K. Fletcher, in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on* (IEEE, 2012), pp. 1246–1250.

[26] T. Lesieur, F. Krzakala, and L. Zdeborová, in *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on* (IEEE, 2015), pp. 680–687.
[27] F. Antenucci, F. Krzakala, P. Urbani, and L. Zdeborová, To appear (2018).
[28] S. Franz and G. Parisi, Journal de Physique I **5**, 1401 (1995).
[29] T. Castellani and A. Cavagna, Journal of Statistical Mechanics: Theory and Experiment **2005**, P05012 (2005).
[30] R. Monasson, Phys. Rev. Lett. **75**, 2847 (1995).
[31] E. Gardner, Nuclear Physics B **257**, 747 (1985).
[32] D. Gross, I. Kanter, and H. Sompolinsky, Physical Review Letters **55**, 304 (1985).
[33] A. Montanari, G. Parisi, and F. Ricci-Tersenghi, Journal of Physics A: Mathematical and General **37**, 2073 (2004).
[34] A. Montanari and F. Ricci-Tersenghi, The European Physical Journal B-Condensed Matter and Complex Systems **33**, 339 (2003).
[35] A. Crisanti, L. Leuzzi, and T. Rizzo, Physical Review B **71**, 094202 (2005).
[36] A. Braunstein, M. Mézard, and R. Zecchina, Random Structures & Algorithms **27**, 201 (2005).
[37] G. Parisi and F. Zamponi, Rev. Mod. Phys. **82**, 789 (2010).
[38] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani, and F. Zamponi, Annual Review of Condensed Matter Physics **8**, 265 (2017).
[39] H. Frisch and J. Percus, Physical Review E **60**, 2942 (1999).
[40] R. Mari and J. Kurchan, The Journal of chemical physics **135**, 124504 (2011).