

Glassy Nature of the Hard Phase in Inference Problems

Fabrizio Antenucci,^{1,2} Silvio Franz,³ Pierfrancesco Urbani,¹ and Lenka Zdeborová¹

¹*Institut de physique théorique, Université Paris Saclay, CNRS, CEA, F-91191 Gif-sur-Yvette, France*

²*Soft and Living Matter Laboratory, Rome Unit of CNR-NANOTEC, Institute of Nanotechnology, Piazzale Aldo Moro 5, I-00185, Rome, Italy*

³*LPTMS, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France*



(Received 21 May 2018; revised manuscript received 12 September 2018; published 31 January 2019)

An algorithmically hard phase is 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. Counterintuitively, we find that the performance of the AMP algorithm is not improved by taking into account the glassy nature of the hard phase. This result provides further evidence that the hard phase in inference problems is algorithmically impenetrable for some deep computational reasons that remain to be uncovered.

DOI: [10.1103/PhysRevX.9.011020](https://doi.org/10.1103/PhysRevX.9.011020)

Subject Areas: Complex Systems,
Interdisciplinary Physics,
Statistical Physics

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. A 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., Ref. [1]. This connection is explored in a range of recent works and brings a class of models for the inference problem in which the Bayes-optimal inference can be analyzed and presents a first-order phase transition. As is common in physics in high dimensions, 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 is coined as the *hard phase*; see, e.g., Ref. [2]. It is 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 principle component analysis, or tensor factorization [9,10], and 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

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

message passing (AMP) known from the context of compressed sensing [13]. In the limit of a 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 the 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 first-order phase transition in the optimal performance in inference problems. The following open questions might use the physicslike 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 is to investigate the above questions. We analyze the following physics-motivated strategy: It is known that the metastable part of the posterior measure in the hard phase is glassy [16–18]. 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 algorithms that improve over simpler ones. Notably, this is the case of random constraint-satisfaction problems, where the influential work in Ref. [19] shows that *survey propagation*, that correctly takes 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 additional evidence towards the bold conjecture that in the corresponding inference problems

AMP is the best of the low-computational-complexity inference algorithms.

Note that such a *negative* result is very interesting from both a physics and a computer-science point of view. In physics, a common intuitive narrative tells us that the properties of the energy landscape control the algorithmic difficulty of the problem. Yet a solid and physically intuitive explanation of why an inference algorithm could not penetrate the hard phase remains open. Our results invite researchers to progress in this question, eventually leading to a precise understanding of the interplay between the dynamics and landscape. In computer science, developments that go beyond the traditional worst-case computational-complexity results are rare, and the hard phase provides a unique and a sharply delimited case that might be computationally hard even for a typical instance. Building a theory that would explain the nature of the hard phase might be the next pillar of our understanding of computational complexity.

Our analysis of the glassiness of the hard phase provides new insights on the performance of Monte Carlo or Langevin dynamics. The 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 suggestion aligns with some of the early literature [16], more recent literature [6] suggests, based on numerical evidence, that Monte Carlo sampling is as good as the message-passing algorithm. Based on the conclusion of our work, this question of performance barriers of sampling-based algorithms should be reopened and investigated more thoroughly. A good understanding of the performance of these algorithms is especially important in 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 Sec. 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 is identified. In Sec. III, we recall the basic setting of Bayesian inference. In Sec. IV, we give a summary of the main algorithmic consequences of our work. In Sec. V, we then recall the replica approach to the study of the corresponding posterior measure. Section VA then summarizes the known replica-symmetric diagram and the resulting phase transitions. Section VB 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 Sec. VI.

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 first-order phase transition is identified [7,20,21]. 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 is 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 a 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 $P_X(\underline{x}^{(0)}) = \prod_{i=1}^N P(x_i^{(0)})$. This signal is subjected to noisy measurements of the following form:

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

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 is analyzed in detail in Ref. [21], and this analysis is proven rigorously to be correct in Refs. [22–25]. References [21,22,26] also analyze 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

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

as this is a prototypical yet simple example in which the hard phase appears for sufficiently low ρ [20,21]. Let us mention that the rank-one matrix estimation with the Rademacher-Bernoulli prior has a very natural interpretation in terms of the 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 Refs. [23,27], the results 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 so-called Bayes optimal setting, which means that we know both the prior $P_X(\underline{x})$ and the variance Δ of the noise. The probability distribution of \underline{x} given Y is given by the Bayes formula

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

Since the noise ξ_{ij} is Gaussian, we have

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

In both Eqs. (3) and (4), we omit 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

$$\hat{x}_i = \langle x_i \rangle, \quad (5)$$

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 Ref. [21]. AMP boils down to a set of recursion relations of the form

$$\hat{x}_i^{(t+1)} = \text{AMP}_i(\hat{\underline{x}}^{(t)}, \hat{x}_i^{(t-1)}), \quad (6)$$

whose iterative fixed point is taken as an estimate of the signal. It is known that fixed points of the state evolution of the AMP algorithm are in the thermodynamic limit described by the replica-symmetric (RS) solution of the model [13,14]. AMP follows the RS solution irrespective of the fact whether RS is the physically correct description of the posterior measure or not.

As shown in Ref. [28], it is possible to derive a generalized AMP, that we call the *approximate survey propagation* (ASP) algorithm, whose state evolution fixed points coincide with the replica equations in the one-step replica symmetry-breaking (1RSB) ansatz. Just as AMP, the ASP algorithm can be also written in the form [28]

$$\hat{x}_i^{(t+1)} = \text{ASP}_i(\hat{\underline{x}}^{(t)}, \hat{x}_i^{(t-1)}, s), \quad (7)$$

depending on one additional free parameter s , corresponding to the Parisi parameter from the spin-glass literature. The special case of $s = 1$ reduces the ASP algorithm back to AMP. The 1RSB solution is known to provide a better description—in many cases, exact—of glassy states. In Sec. V, we hence study the thermodynamics of the above model in the RS and 1RSB ansatz, focusing on its properties in the hard phase.

IV. SUMMARY OF MAIN ALGORITHMIC RESULT

Before going to the technical part of the replica analysis in Sec. V, we briefly summarize the corresponding main algorithmic result. In Sec. V, we then investigate in detail the 1RSB solution of the low-rank matrix estimation model (1) focusing on the glassy properties of the hard phase. Our main interest, however, is in the relation between the 1RSB solution and the associated algorithmic performance. The main question we ask is whether ASP can (for a suitable choice of the Parisi parameter s) improve on AMP. The experience with the survey propagation algorithm applied to constraint-satisfaction problems [29] suggests that this improvement should be possible.

In Fig. 1, we plot the magnetization achieved by the ASP algorithm as a function of the noise Δ for several values of the Parisi parameter s . We observe that, as the noise Δ decreases, the equilibrium value (yellow) is reached first by the $s = 1$ curve, corresponding to the performance of AMP. In Fig. 3, we then plot the mean-squared error (MSE) as a function of the Parisi parameter s for several values of the noise Δ . Again, we see that in all cases the best error is achieved with $s = 1$. Algorithmically, this result means

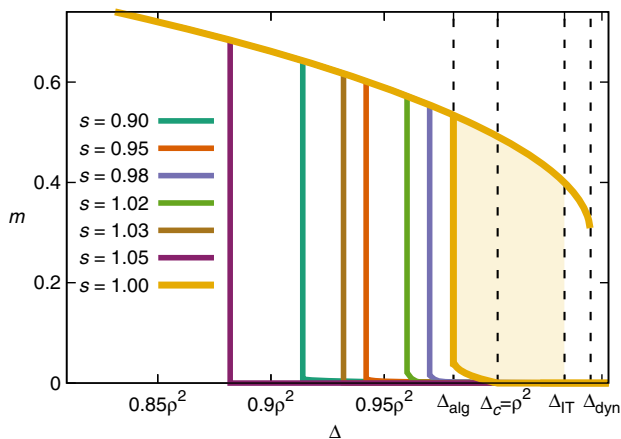


FIG. 1. The magnetization, also known as the overlap, between the signal and the states described by the 1RSB solution at Parisi parameter s , as a function of the noise strength Δ , and sparsity $\rho = 0.08$. The curve that shows a spinodal transition towards the strongly magnetized solution at the 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.

that, in the present setting, ASP never obtains better accuracy than 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 hard region works better as an inference algorithm than an algorithm that correctly describes the metastable states in this region. At the same time, the above result could be anticipated based on a mathematical theorem of Ref. [7] that implies that AMP is optimal among all local algorithms. This theorem applies as long as an iterative algorithm uses only information from nearest neighbors and (nearly) reaches a fixed point after $O(1)$ iterations.

V. THE REPLICA APPROACH TO THE POSTERIOR MEASURE

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

$$f[\Delta; Y] = -\frac{1}{N} \ln \int \left(\prod_{i=1}^N dx_i P_X(x_i) \right) \prod_{i \leq j} \mathcal{G} \left(Y_{ij} \left| \frac{x_i x_j}{\sqrt{N}} \right. \right). \quad (8)$$

This energy 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

$$f(\Delta) = \overline{f[\Delta; Y]} \equiv \int \left[\prod_{i \leq j} dY_{ij} \right] P(Y) f[\Delta; Y], \quad (9)$$

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

$$P(Y) \propto \int d\mathbf{x}^{(0)} \underline{P}_X(\mathbf{x}^{(0)}) \prod_{i \leq j} \mathcal{G} \left(Y_{ij} \left| \frac{x_i^{(0)} x_j^{(0)}}{\sqrt{N}} \right. \right). \quad (10)$$

In order to perform the average defined in Eq. (9), we use the replica method [12]. Introducing

$$\mathcal{Z} = \int \left(\prod_{i=1}^N dx_i P_X(x_i) \right) \prod_{i \leq j} \mathcal{G} \left(Y_{ij} \left| \frac{x_i x_j}{\sqrt{N}} \right. \right), \quad (11)$$

we get

$$f(\Delta) = -\frac{1}{N} \lim_{n \rightarrow 0} \partial_n \int \left[\prod_{i \leq j} dY_{ij} \right] P(Y) \mathcal{Z}^n. \quad (12)$$

For integer n , we can represent \mathcal{Z}^n as an n -dimensional integral over n replicas $\mathbf{x}^{(a)}$ with $a = 1, \dots, n$. 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., Ref. [12]), we arrive at a closed expression for $f(\Delta)$ that is

$$f(\Delta) = -\frac{1}{N} \ln \int \mathcal{D}q \mathcal{D}\hat{q} \exp [N\mathcal{S}(q, \hat{q})], \quad (13)$$

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. (13) 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

$$q_{ab} = \frac{1}{N} \sum_{i=1}^N \overline{\langle x_i^{(a)} x_i^{(b)} \rangle}, \quad (14)$$

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

$$m \equiv q_{0a} = q_{a0} = \frac{1}{N} \sum_{i=1}^N \overline{\langle x_i^{(0)} x_i^{(a)} \rangle}, \quad a > 0. \quad (15)$$

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 $n \rightarrow 0$. One needs an appropriate scheme from which one can take the replica limit. Here, we consider two schemes: the RS and the 1RSB one. We refer here to symmetry under permutations of the n replicas with index $a = 1, \dots, n$.

A. Reminder of the replica-symmetric solution

The RS scheme boils down to considering

$$\begin{aligned} q_{ab} &= (q_d - q_0)\delta_{ab} + q_0, & a, b \geq 1, \\ \hat{q}_{ab} &= (\hat{q}_d - \hat{q}_0)\delta_{ab} + \hat{q}_0, & a, b \geq 1, \\ q_{0a} &= q_{a0} = m, & a \geq 1, \\ \hat{q}_{0a} &= \hat{q}_{a0} = \hat{m}, & a \geq 1. \end{aligned} \quad (16)$$

From the point of view of the inference, the relevant quantity to look at is the MSE:

$$\begin{aligned} \text{MSE} &= \frac{1}{N} \sum_{i=1}^N \overline{\langle (x_i) - x_i^{(0)} \rangle^2} \\ &= \rho - 2m + q_0, \end{aligned} \quad (17)$$

where $\rho \equiv \overline{\langle x^{(0)} \rangle^2}$. 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 the 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 one to describe metastable states where the Nishimori identities might not hold. Plugging this ansatz inside the expression for \mathcal{S} and taking the saddle-point equations with respect to all these parameters, one gets the replica-symmetric solution as reported in Ref. [21] and proven to give the equilibrium solution in Refs. [24,25]. The RS free energy can be expressed as

$$f_{\text{RS}}(\Delta) = \min_m \{ \phi_{\text{RS}}(m, \Delta) \} \quad (18)$$

with

$$\phi_{\text{RS}}(m, \Delta) = \frac{m^2}{4\Delta} - \mathbb{E}_{x^{(0)}, W} \left[f \left(\frac{m}{\Delta}, \frac{m}{\Delta} x^{(0)} + \sqrt{\frac{m}{\Delta}} W \right) \right], \quad (19)$$

where

$$f(A, B) = \ln \left[\int dx P_X(x) e^{-(1/2)Ax^2 + Bx} \right] \quad (20)$$

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

$$m = \mathbb{E}_{x^{(0)}, W} \left[x^{(0)} \frac{\partial f}{\partial B} \left(\frac{m}{\Delta}, \frac{m}{\Delta} x^{(0)} + \sqrt{\frac{m}{\Delta}} W \right) \right]. \quad (21)$$

Equilibrium properties of the inference problem are given by the global minima of the free energy Eq. (19). 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 a hard phase [21]. 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 with respect to the noise, and for $\Delta = \Delta_{\text{dyn}} \sim 1.041\rho^2$ the system undergoes a *dynamical transition*. On the one hand, one can see that the free energy (19) 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 [30,31]. The corresponding RS free energy ceases to describe a paramagnetic state, and it describes a nonergodic phase with an exponential number $\exp[N\Sigma(\Delta)]$ of metastable states—also known as 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 complexity $f(m > 0) - f(m = 0) = \Sigma(\Delta)$. Moreover, as we 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 $\Delta = \Delta_{\text{IT}} \sim 1.0295\rho^2$, where there is the information-theoretic phase transition and $\Sigma(\Delta_{\text{IT}}) = 0$. 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 a 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 $\Delta = \Delta_{\text{alg}} \sim 0.9805\rho^2$, this metastable minimum disappears with a spinodal transition. In the interval $[\Delta_{\text{alg}}, \Delta_{\text{IT}}]$, 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 nonphysical 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 nonphysical below Δ_{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

$$q_{ab} = \begin{cases} q_d & a = b, \\ q_1 & a, b \text{ in the same block,} \\ q_0 & a, b \text{ in different blocks} \end{cases} \quad (22)$$

and analogous for \hat{q} . For s strictly equal to one, we get back the replica-symmetric ansatz Eq. (16). Note that, for $s \neq 1$, m and q_0 are, in general, different in the solution: This difference is crucial when evaluating the MSE Eq. (17), as the minimum of the MSE does not correspond, in general, to the maximum of m .

The 1RSB free energy takes the form

$$f_{\text{1RSB}}(\Delta, s) = \mathbf{extr}\{\phi_{\text{1RSB}}(m, q_0, q_1, \Delta, s)\}, \quad (23)$$

with

$$\begin{aligned} \phi_{\text{1RSB}}(m, q_0, q_1, \Delta, s) &= \frac{m^2}{2\Delta} - s \frac{q_0^2}{4\Delta} - (1-s) \frac{q_1^2}{4\Delta} \\ &\quad - \frac{1}{s} \mathbb{E}_{x^{(0)}, W} \left[f \left(\frac{q_1}{\Delta}, \frac{m}{\Delta} x^{(0)} + \sqrt{\frac{q_0}{\Delta}} W, \frac{q_1 - q_0}{\Delta} \right) \right], \end{aligned} \quad (24)$$

where

$$\begin{aligned} f(A, B, C) &= \ln \int dh \sqrt{\frac{C}{2\pi}} e^{-(1/2)Ch^2} \\ &\quad \cdot \left[\int dx P_X(x) e^{-(1/2)Ax^2 + (B+Ch)x} \right]^s. \end{aligned} \quad (25)$$

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

$$\begin{aligned} m &= \frac{1}{s} \mathbb{E}_{x^{(0)}, W} \left[x^{(0)} \frac{\partial f}{\partial B} \right], \\ q_0 &= \frac{1}{s^2} \mathbb{E}_{x^{(0)}, W} \left[\left(\frac{\partial f}{\partial B} \right)^2 \right], \\ q_1 &= \frac{2}{s(s-1)} \mathbb{E}_{x^{(0)}, W} \left[\frac{\partial f}{\partial A} + \frac{\partial f}{\partial C} \right], \end{aligned} \quad (26)$$

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

We 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 [28]. In particular, the expression (17) exactly gives the MSE of such an algorithm with m and q_0 being the solution of Eq. (26).

For high enough Δ , the 1RSB solution collapses to the RS one, meaning that $q_0 = q_1 = m = 0$. At Δ_{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 Δ_{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 [32].

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 one to select families of metastable states of given (internal) free energy [33]. Their corresponding complexity Σ (defined as the log of their number) is obtained by deriving Eq. (24) with respect to s [33] and multiplying the result by s^2 , i.e.,

$$\Sigma(\Delta, s) = \frac{s^2}{4\Delta} (q_1^2 - q_0^2) - s^2 \frac{\partial}{\partial s} \mathbb{E}_{x^{(0)}, W} \left[\frac{1}{s} f \left(\frac{q_1}{\Delta}, \frac{m}{\Delta} x^{(0)} + \sqrt{\frac{q_0}{\Delta}} W, \frac{q_1 - q_0}{\Delta} \right) \right]. \quad (27)$$

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 a 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 a nonphysical one where $\Sigma < 0$. Note that the physical region with positive complexity continues not only below Δ_{IT} but even well below Δ_{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 stability requires that all the eigenvalues of the Hessian of the free energy should be positive in the solution. The 1RSB solutions can lose stability in two possible ways, associated to negative values of the following eigenvalues [34–36]:

$$\begin{aligned} \lambda_{\text{I}} &= 1 - \frac{1}{\Delta} \int_{-\infty}^{\infty} dh P(s, h) [f''(s, h)]^2, \\ \lambda_{\text{II}} &= 1 - \frac{1}{\Delta} \int_{-\infty}^{\infty} dh P(1, h) [f''(1, h)]^2, \end{aligned} \quad (28)$$

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

$$\begin{aligned} f(1, h) &= \ln \int dx P(x) \exp \left[-\frac{A}{2} x^2 + hx \right], \\ f(s, h) &= \frac{1}{s} \ln \int \frac{dz}{\sqrt{2\pi C}} e^{-(z^2/2C)} e^{sf(1, h-z)}, \end{aligned} \quad (29)$$

$$\begin{aligned} P(s, h) &= \mathbb{E}_{x^{(0)}} \left[\sqrt{\frac{\Delta}{2\pi q_0}} \exp \left(-\frac{\Delta}{2q_0} B^2 \right) \right], \\ P(1, h) &= e^{sf(1, h)} \int \frac{dz e^{-(z^2/2C)}}{\sqrt{2\pi C}} \cdot P(s, h-z) e^{-sf(s, h-z)}. \end{aligned} \quad (30)$$

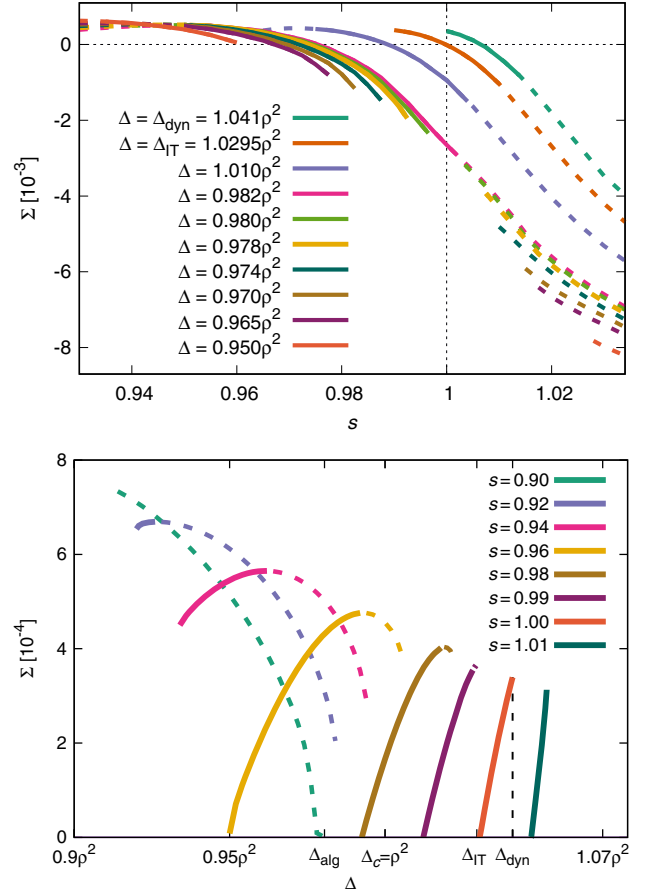


FIG. 2. The complexity of metastable states Σ as a function of the Parisi parameter s and the noise Δ , for prior (2) with sparsity $\rho = 0.08$. Upper: Complexity at fixed s in the whole domain of existence of a nontrivial fixed point. Lower: The physical region of positive Σ as a function of Δ . We draw the stable solutions with a solid line and the unstable, with respect to the eigenvalues (28), with a dashed line. For each value of $\Delta \in [\Delta_{\text{IT}}, \Delta_{\text{dyn}}]$, the value of $\Sigma(\Delta, s = 1)$ represents the complexity of the family of thermodynamically dominating states. Below Δ_{IT} , the $s = 1$ solution is nonphysical 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.

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 when the glassy states are unstable against a *Gardner transition* to further RSB [34,35]: Each metastable state splits into a hierarchy of new states (type-II instability) [36]. In Fig. 2, we mark with full lines the stable region and with dashed lines the unstable ones. Type-I instability is found for large s in the nonphysical region of negative complexity. Type-II instability is found in the physical region at small values of s , and it has been found also in spin-glass models [36–38].

Let us now discuss in detail the glassy solutions that one finds for $\Delta < \Delta_{\text{IT}}$ 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 the lowest total free energy turn out to be the ones with zero complexity Σ . For different fixed values of the parameter s , the complexity curves reach a 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 $\Delta_{\text{IRSB, equil}} < \Delta_{\text{alg}}$. Stable solutions of positive complexity exist down to $\Delta_{\text{IRSB, stable}} < \Delta_{\text{IRSB, equil}}$ and solutions with positive complexity (irrespective of the stability) down to $\Delta_{\text{IRSB, all}} < \Delta_{\text{IRSB, stable}}$. Example of specific values for $\rho = 0.08$ in Fig. 2 are $\Delta_{\text{alg}} \sim 0.9805\rho^2$, $\Delta_{\text{IRSB, equil}} \sim 0.951\rho^2$, $\Delta_{\text{IRSB, stable}} \sim 0.918\rho^2$, and $\Delta_{\text{IRSB, all}} \sim 0.903\rho^2$. This result notably means that for $\Delta < \Delta_{\text{alg}}$, 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 the principles of physical dynamics. On the one hand, one could conjecture that Monte Carlo dynamics gets trapped by glassy states even below Δ_{alg} . On the other hand, the dynamics is expected to fall out of equilibrium for all $\Delta < \Delta_{\text{dyn}}$, and it is not *a priori* clear in which states it should get trapped. While AMP clearly works for $\Delta < \Delta_{\text{alg}}$ and does not work for $\Delta > \Delta_{\text{alg}}$, our analysis does not provide any reason why the threshold Δ_{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 the corresponding algorithmic threshold is.

So far, we focus on glassy states of positive complexity (i.e., existing with probability one for a typical instance). There are also solutions of the 1RSB equations having negative complexity. We 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 [28] 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 $\Delta_{\text{alg}}(s)$ the ghost-glassy state disappears, 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 $\Delta_{\text{alg}} \equiv \Delta_{\text{alg}}(s = 1)$ of the usual AMP ($s = 1$) and if we could have an s for which $\Delta_{\text{alg}}(s) > \Delta_{\text{alg}}(1)$. With this question in mind, in Fig. 3, we plot the MSE with the ground truth signal given by Eq. (17) as a function of s for various values of Δ . We initialize the 1RSB fixed-point equations at infinitesimal

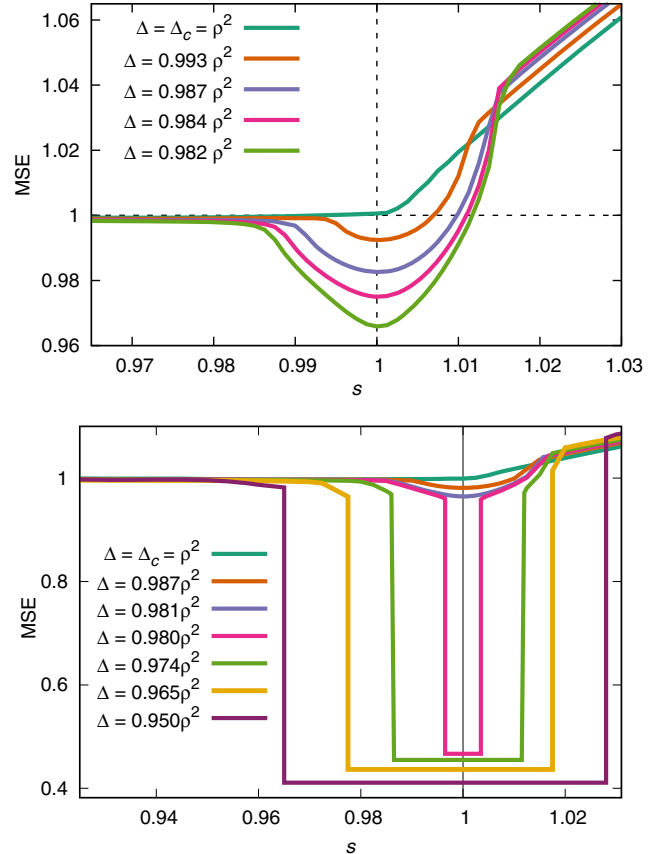


FIG. 3. The MSE as a function of the Parisi parameter s for different values of the noise strength Δ . The smallest MSE is always reached for $s = 1$, corresponding to the performance of the AMP algorithm, with a threshold at $\Delta_{\text{alg}} = 0.9805\rho^2$.

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.

VI. 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 hard phase. This finding reopens the natural question of performance limits of Monte Carlo-based sampling. While some recent works [6] anticipated numerically that Monte Carlo and message passing will share the same algorithmic threshold, our results do not provide any evidence of this result. 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. In order to validate this proposition, one needs to study a different model than the present one. The present model is

dense and thus not suitable for large-scale simulations; also, the analytically tractable description of sampling-based dynamics for the present model is a major open question. One possibility is to perform a large-scale numerical study with Monte Carlo-based dynamics in diluted models such as those studied in Ref. [39]. Another possibility is to aim at an analytical description of the Langevin dynamics that is known in a tractable form so far only for mixtures of spherical p -spin models.

While we anticipate that the performance of the usual sampling-based algorithms will be hampered by the glassiness, it is an interesting open question to investigate whether other algorithms are able to match the performance of AMP. We have in mind, for instance, the algorithms based on the robust ensemble as introduced in Ref. [40].

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 conclusion 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 [29]. We note that this observation is surprising, and we are missing a physically intuitive explanation for why taking glassiness into account improves performance in optimization problems but not in Bayes-optimal inference. We stress that our results provide strong evidence towards the conjecture that the hard phase is impenetrable for some computationally fundamental reasons. Further investigation of this conjecture is an exciting direction for both physics and theoretical computer science.

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.

We also note that our above conclusions apply to the case of Bayes-optimal inference, where the generative model is matched to the inference model. In case the hyperparameters are not known or mismatched, the message-passing algorithm that takes glassiness into account can provide better error and robustness, which is investigated in detail in Ref. [28].

Finally, we mention that the results shown here may be compelling also beyond inference problems. In particular, the instabilities of the RS solution at Δ_{alg} and Δ_c can be related to a similar phenomenon occurring in the mean-field theory of liquids and glasses [41,42]. 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 nonergodic extensive complexity phase at higher density. As 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 [43,44]. Our analysis suggests that within inference models not only the nonphysical 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 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 Program Grant Agreement No. 714608-SMiLe. The work of S.F. was supported by a grant from the Simons Foundation (No. 454941).

-
- [1] P. Grassberger and J.-P. Nadal, *From Statistical Physics to Statistical Inference and Back*, NATO ASI, Series C (Springer, Dordrecht, 1992).
 - [2] L. Zdeborová and F. Krzakala, *Statistical Physics of Inference: Thresholds and Algorithms*, *Adv. Phys.* **65**, 453 (2016).
 - [3] T. Richardson and R. Urbanke, *Modern Coding Theory* (Cambridge University Press, Cambridge, England, 2008).
 - [4] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University, New York, 2009).
 - [5] F. Krzakala, M. Mézard, F. Sausset, Y.F. Sun, and L. Zdeborová, *Statistical-Physics-Based Reconstruction in Compressed Sensing*, *Phys. Rev. X* **2**, 021005 (2012).
 - [6] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Asymptotic Analysis of the Stochastic Block Model for Modular Networks and Its Algorithmic Applications*, *Phys. Rev. E* **84**, 066106 (2011).
 - [7] Y. Deshpande and A. Montanari, *Finding Hidden Cliques of Size \sqrt{N}/e in Nearly Linear Time*, *Found. Comput. Math.* **15**, 1069 (2015).
 - [8] A. Montanari, *Finding One Community in a Sparse Graph*, *J. Stat. Phys.* **161**, 273 (2015).
 - [9] E. Richard and A. Montanari, *A Statistical Model for Tensor PCA*, in *Advances in Neural Information Processing Systems*, 2014, pp. 2897–2905, <https://papers.nips.cc/paper/5616-a-statistical-model-for-tensor-pca>.
 - [10] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová, *Statistical and Computational Phase Transitions in Spiked Tensor Estimation*, in *Proceedings of the*

- 2017 *IEEE International Symposium on Information Theory (ISIT)* (IEEE, New York, 2017), pp. 511–515.
- [11] G. Györfyi, *First-Order Transition to Perfect Generalization in a Neural Network with Binary Synapses*, *Phys. Rev. A* **41**, 7097 (1990).
- [12] M. Mézard, G. Parisi, and M. Á. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
- [13] D. L. Donoho, A. Maleki, and A. Montanari, *Message-Passing Algorithms for Compressed Sensing*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18914 (2009).
- [14] M. Bayati and A. Montanari, *The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing*, *IEEE Trans. Inf. Theory* **57**, 764 (2011).
- [15] S. Kudekar, T. J. Richardson, and R. L. Urbanke, *Threshold Saturation via Spatial Coupling: Why Convolutional LDPC Ensembles Perform So Well over the BEC*, *IEEE Trans. Inf. Theory* **57**, 803 (2011).
- [16] H. Sompolinsky, N. Tishby, and H. S. Seung, *Learning from Examples in Large Neural Networks*, *Phys. Rev. Lett.* **65**, 1683 (1990).
- [17] S. Franz, M. Mézard, F. Ricci-Tersenghi, M. Weigt, and R. Zecchina, *A Ferromagnet with a Glass Transition*, *Europhys. Lett.* **55**, 465 (2001).
- [18] F. Krzakala and L. Zdeborová, *Hiding Quiet Solutions in Random Constraint Satisfaction Problems*, *Phys. Rev. Lett.* **102**, 238701 (2009).
- [19] M. Mézard, G. Parisi, and R. Zecchina, *Analytic and Algorithmic Solution of Random Satisfiability Problems*, *Science* **297**, 812 (2002).
- [20] T. Lesieur, F. Krzakala, and L. Zdeborová, *Phase Transitions in Sparse PCA*, in *Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT)* (IEEE, New York, 2015), pp. 1635–1639.
- [21] T. Lesieur, F. Krzakala, and L. Zdeborová, *Constrained Low-Rank Matrix Estimation: Phase Transitions, Approximate Message Passing and Applications*, *J. Stat. Mech.* (2017) 073403.
- [22] Y. Deshpande and A. Montanari, *Information-Theoretically Optimal Sparse PCA*, in *Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT)* (IEEE, New York, 2014), pp. 2197–2201.
- [23] F. Krzakala, J. Xu, and L. Zdeborová, *Mutual Information in Rank-One Matrix Estimation*, in *Proceedings of the 2016 IEEE Information Theory Workshop (ITW)* (IEEE, New York, 2016), pp. 71–75.
- [24] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, *Mutual Information for Symmetric Rank-One Matrix Estimation: A Proof of the Replica Formula*, in *Advances in Neural Information Processing Systems*, 2016, pp. 424–432, <https://papers.nips.cc/paper/6380-mutual-information-for-symmetric-rank-one-matrix-estimation-a-proof-of-the-replica-formula>.
- [25] M. Lelarge and L. Miolane, *Fundamental Limits of Symmetric Low-Rank Matrix Estimation*, *Probab. Theory Relat. Fields* **1** (2018).
- [26] S. Rangan and A. K. Fletcher, *Iterative Estimation of Constrained Rank-One Matrices in Noise*, in *Proceedings of the 2012 IEEE International Symposium on Information Theory Proceedings (ISIT)* (IEEE, New York, 2012), pp. 1246–1250.
- [27] T. Lesieur, F. Krzakala, and L. Zdeborová, *MMSE of Probabilistic Low-Rank Matrix Estimation: Universality with Respect to the Output Channel*, in *Proceedings of the 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (IEEE, New York, 2015), pp. 680–687.
- [28] F. Antenucci, F. Krzakala, P. Urbani, and L. Zdeborová, *Approximate Survey Propagation for Statistical Inference*, [arXiv:1807.01296](https://arxiv.org/abs/1807.01296).
- [29] A. Braunstein, M. Mézard, and R. Zecchina, *Survey Propagation: An Algorithm for Satisfiability*, *Random Struct. Algorithms* **27**, 201 (2005).
- [30] S. Franz and G. Parisi, *Recipes for Metastable States in Spin Glasses*, *J. Phys. I (France)* **5**, 1401 (1995).
- [31] T. Castellani and A. Cavagna, *Spin Glass Theory for Pedestrians*, *J. Stat. Mech.* (2005) P05012.
- [32] Notice the analogy of the high-magnetization state here with the crystal state in the physics of glasses.
- [33] R. Monasson, *Structural Glass Transition and the Entropy of the Metastable States*, *Phys. Rev. Lett.* **75**, 2847 (1995).
- [34] E. Gardner, *Spin Glasses with p -Spin Interactions*, *Nucl. Phys. B* **257**, 747 (1985).
- [35] D. J. Gross, I. Kanter, and H. Sompolinsky, *Mean-Field Theory of the Potts Glass*, *Phys. Rev. Lett.* **55**, 304 (1985).
- [36] A. Montanari, G. Parisi, and F. Ricci-Tersenghi, *Instability of One-Step Replica-Symmetry-Broken Phase in Satisfiability Problems*, *J. Phys. A* **37**, 2073 (2004).
- [37] A. Montanari and F. Ricci-Tersenghi, *On the Nature of the Low-Temperature Phase in Discontinuous Mean-Field Spin Glasses*, *Eur. Phys. J. B* **33**, 339 (2003).
- [38] A. Crisanti, L. Leuzzi, and T. Rizzo, *Complexity in Mean-Field Spin-Glass Models: Ising p -Spin*, *Phys. Rev. B* **71**, 094202 (2005).
- [39] F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, *Typology of Phase Transitions in Bayesian Inference Problems*, [arXiv:1806.11013](https://arxiv.org/abs/1806.11013).
- [40] C. Baldassi, C. Borgs, J. T. Chayes, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina, *Unreasonable Effectiveness of Learning Neural Networks: From Accessible States and Robust Ensembles to Basic Algorithmic Schemes*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7655 (2016).
- [41] G. Parisi and F. Zamponi, *Mean-Field Theory of Hard Sphere Glasses and Jamming*, *Rev. Mod. Phys.* **82**, 789 (2010).
- [42] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani, and F. Zamponi, *Glass and Jamming Transitions: From Exact Results to Finite-Dimensional Descriptions*, *Annu. Rev. Condens. Matter Phys.* **8**, 265 (2017).
- [43] H. L. Frisch and J. K. Percus, *High Dimensionality as an Organizing Device for Classical Fluids*, *Phys. Rev. E* **60**, 2942 (1999).
- [44] R. Mari and J. Kurchan, *Dynamical Transition of Glasses: From Exact to Approximate*, *J. Chem. Phys.* **135**, 124504 (2011).