Marvels and Pitfalls of the Langevin Algorithm in Noisy High-Dimensional Inference

Stefano Sarao Mannelli[1], Giulio Biroli,[2] Chiara Cammarota,[3] Florent Krzakala,[4]
Pierfrancesco Urbani,[1] and Lenka Zdeborová[1]

[1]*Université Paris-Saclay, CNRS, CEA, Institut de Physique Théorique, 91191, Gif-sur-Yvette, France*
[2]*Laboratoire de Physique de l'Ecole Normale Supérieure ENS, Université PSL, CNRS, Sorbonne Université, Université Paris-Diderot, Sorbonne Paris Cité Paris, France*
[3]*Department of Mathematics, King's College London, Strand London WC2R 2LS, United Kingdom*
[4]*Laboratoire de Physique Statistique, CNRS & Université Pierre & Marie Curie & Ecole Normale Supérieure & PSL Université, 75005 Paris, France*

Gradient-descent-based algorithms and their stochastic versions have widespread applications in machine learning and statistical inference. In this work, we carry out an analytic study of the performance of the algorithm most commonly considered in physics, the Langevin algorithm, in the context of noisy high-dimensional inference. We employ the Langevin algorithm to sample the posterior probability measure for the spiked mixed matrix-tensor model. The typical behavior of this algorithm is described by a system of integrodifferential equations that we call the Langevin state evolution, whose solution is compared with the one of the state evolution of approximate message passing (AMP). Our results show that, remarkably, the algorithmic threshold of the Langevin algorithm is suboptimal with respect to the one given by AMP. This phenomenon is due to the residual glassiness present in that region of parameters. We also present a simple heuristic expression of the transition line, which appears to be in agreement with the numerical results.

Subject Areas: Computational Physics, Statistical Physics

I. MOTIVATION

Algorithms based on noisy variants of gradients descent [1,2] stand at the root of many modern applications of data science and are being used in a wide range of high-dimensional nonconvex optimization problems. The wide-spread use of stochastic gradient descent in deep learning [3] is certainly one of the most prominent examples. For such algorithms, the existing theoretical analysis mostly concentrates on convex functions and convex relaxations or on regimes where spurious local minima become irrelevant. For problems with complicated landscapes where, instead, useful convex relaxations are not known and spurious local minima cannot be ruled out, the theoretical understanding of the behavior of the gradient-descent-based algorithm remains poor and represents a major avenue of research.

The goal of this paper is to contribute to such an understanding in the context of statistical learning and to transfer ideas and techniques developed for glassy dynamics [4] to the analysis of nonconvex high-dimensional inference. In statistical learning, the minimization of a cost function is not the goal *per se* but rather a way to uncover an unknown structure in the data. One common way to model and analyze this situation is to generate data with a hidden structure and to see if the structure can be recovered. This method is easily set up as a teacher-student scenario [5,6]: First, a teacher generates latent variables and uses them as input of a prescribed model to generate a synthetic data set. Then, the student observes the data set and tries to infer the values of the latent variables. The analysis of this setting has been carried out rigorously in a wide range of teacher-student models for high-dimensional inference and learning tasks as diverse as a planted clique [7], generalized linear models such as compressed sensing or phase retrieval [8], factorization of matrices and tensors [9,10], or simple models of neural networks [11]. In these works, the information-theoretic optimal performances—the ones obtained by an ideal Bayes-optimal estimator, not limited in time and memory—have been computed.

The main question is, of course, how practical algorithms—operating in polynomial time with respect to the problem size—compare to these ideal performances. The last decade brought remarkable progress in our understanding of the performances achievable computationally. In particular, many algorithms based on message passing [6,12], spectral methods [13], and semidefinite programs

(SDP) [14] were analyzed. Depending on the signal-to-noise ratio, these algorithms were shown to be very efficient in many of those tasks. Interestingly, all these algorithms fail to reach good performance in the same region of the parameter space, and this striking observation has led to the identification of a well-defined hard phase. This case is a regime of parameters in which the underlying statistical problem can be information theoretically solved, but no efficient algorithms are known, rendering the problem essentially unsolvable for large instances. This stream of ideas is currently gaining momentum and impacting research in statistics, probability, and computer science.

The performance of the noisy-gradient-descent algorithms remains an entirely open question. Do they allow one to reach the same performances as message passing and SDPs? Can they enter the hard phase? Do they stop being efficient at the same moment as the other approaches, or are they worse? The ambition of the present paper is to address these questions by analyzing the performance of the Langevin algorithm in the high-dimensional limit of a particular spiked mixed matrix-tensor model, defined in detail in the next section.

Similar models have played a fundamental role in statistics and random matrix theory [15,16]. Tensor factorization is also an important topic in machine learning and is widely used in data analysis [17–22]. At variance with the pure spiked tensor case [18], this mixed matrix-tensor model has the advantage that the algorithmic threshold appears at the same scale as the information-theoretic one, similarly to what is observed in simple models of neural networks [8,11]. We view the spiked mixed matrix-tensor model as a prototype for the nonconvex high-dimensional landscape. The key virtue of the model is its tractability.

We focus on the Langevin algorithm for two main reasons: First, it is the gradient-based algorithm that is most widely studied in physics. Second, at long times (possibly growing exponentially with the system size), it is known to sample the associated Boltzmann measure, thus evaluating the Bayes-optimal estimator for the inference problem. We evaluate performance of the algorithm at times that are long but not growing with the system size. We explicitly compare the obtained performance to the one of the Bayes-optimal estimator and to the best-known efficient algorithm so far—the approximate message-passing (AMP) algorithm [6,12]. In particular, contrary to what has been anticipated in Refs. [23,24], but as surmised in Ref. [25], we observe that the performance of the Langevin algorithm is hampered by the many spurious metastable states still present in the AMP easy phase. Thus, we shed light on a number of properties of the Langevin algorithm that may seem counterintuitive at first sight (e.g., the performance gets worse as the noise decreases).

The possibility to analytically describe the behavior of the Langevin algorithm in this model is enabled by the existence of the Crisanti-Horner-Sommers-Cugliandolo-Kurchan (CHSCK) equations in spin-glass theory, describing the behavior of the Langevin dynamics in the so-called spherical p-spin model [26,27], where the method can be rigorously justified [28]. These equations were a key development in the field of statistical physics of disordered systems that lead to detailed understanding and predictions about the slow dynamics of glasses [4]. In this paper, we bring these powerful methods and ideas into the realm of statistical learning.

II. SPIKED MATRIX-TENSOR MODEL

We now detail the spiked mixed matrix-tensor problem: A teacher generates an N-dimensional vector x^* by choosing each of its components independently from a normal Gaussian distribution of zero mean and unit variance. In the large-N limit, this method is equivalent to having a flat distribution over the N-dimensional hypersphere $\mathcal{S}_{N-1}$ defined by $|x^*|^2 = N$. In this paper, we use either of these two, as convenient. The teacher then generates a symmetric matrix Y_{ij} and a symmetric order-p tensor $T_{i_1,\ldots,i_p}$ as

$$Y_{ij} = \frac{1}{\sqrt{N}} x_i^* x_j^* + \xi_{ij} \quad \forall \; i < j,$$

$$T_{i_1\ldots i_p} = \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} x_{i_1}^* \ldots x_{i_p}^* + \xi_{i_1\ldots i_p} \quad \forall \; i_1 < \ldots < i_p,$$

$$(1)$$

where ξ_{ij} and $\xi_{i_1,\ldots,i_p}$ are iid Gaussian components of a symmetric random matrix and tensor of zero mean and variance Δ_2 and Δ_p, respectively; ξ_{ij} and $\xi_{i_1,\ldots,i_p}$ correspond to noises corrupting the signal of the teacher. In the limit $\Delta_2 \to 0$ and $\Delta_p \to 0$, the above model reduces to the canonical spiked Wigner model [29] and spiked tensor model [18], respectively. The goal of the student is to infer the vector x^* from the knowledge of the matrix Y, of the tensor T, of the values Δ_2 and Δ_p, and the knowledge of the spherical prior. The scaling with N as specified in Eq. (1) is chosen in such a way that the information-theoretic best-achievable error varies between a perfectly reconstructed spike x^* and a random guess from the flat measure on $\mathcal{S}_{N-1}$. Here and in the rest of the paper, we denote $x \in \mathcal{S}_{N-1}$ the N-dimensional vector and x_i, with $i = 1, \ldots, N$ its components.

This model belongs to the generic direction of study of Gaussian functions on the N-dimensional sphere, known as p-spin spherical spin-glass models in the physics literature and as isotropic models in the Gaussian process literature [30–34]. In statistics and machine learning, these models have appeared following the studies of spiked matrix and tensor models [16,18,29]. Analogous mixed matrix-tensor models—where, next to an order-p tensor, one observes a

matrix created from the same spike—are studied, e.g., in Ref. [17] in the context of topic modeling (also in Ref. [18]). From the optimization-theory point of view, this model is highly nontrivial, as it is high dimensional and nonconvex. For the purpose of the present paper, this model is chosen with the hypothesis that its energy landscape presents properties that will generalize to other nonconvex high-dimensional problems. The following three ingredients are key to the analysis: (a) It is in the class of models for which the Langevin algorithm can be analyzed exactly in the large-N limit. (b) The different phase transitions, both algorithmic and information theoretic, discussed hereafter, all happen at $\Delta_2 = \mathcal{O}(1)$, $\Delta_p = \mathcal{O}(1)$, which means that when the problem becomes algorithmically tractable, it is still in the noisy regime, where the optimal mean squared error (MSE) is bounded away from zero. (c) In this model, the AMP algorithm is conjectured to be optimal among polynomial algorithms. The second and third ingredients are not present in the pure spiked tensor model [18], making it unsuitable for our present study. We note that the Langevin algorithm was recently analyzed for the pure spiked tensor model in Ref. [21] in a regime where the noise variance is very small, $\Delta \sim N^{-p/2}$, but we also note that, in that model, algorithms such as tensor unfolding, semidefinite programming, homotopy methods, or improved message-passing schemes work better, roughly up to $\Delta \sim N^{-p/4}$ [17–19,35,36].

III. BAYES-OPTIMAL ESTIMATION AND MESSAGE-PASSING ALGORITHM

In this section, we present the performance of the Bayes-optimal estimator and of the approximate message-passing algorithm. This theory is based on a straightforward adaptation of analogous results known for the pure spiked matrix model [9,29,37] and for the pure spiked tensor model [10,18].

The Bayes-optimal estimator $\hat{x}$ is defined as the one that, among all estimators, minimizes the MSE with the spike x^*. Starting from the posterior probability distribution

$$P(x|Y,T) = \frac{1}{Z(Y,T)} \left[\prod_{i=1}^{N} e^{-x_i^2/2} \right] \prod_{i<j} e^{-\frac{1}{2\Delta_2}\left(Y_{ij} - \frac{x_i x_j}{\sqrt{N}}\right)^2}$$
$$\times \prod_{i_1<\ldots<i_p} e^{-\frac{1}{2\Delta_p}\left(T_{i_1\ldots i_p} - \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} x_{i_1}\ldots x_{i_p}\right)^2}, \quad (2)$$

the Bayes-optimal estimator reads

$$\hat{x}_i = \mathbb{E}_{P(x|Y,T)}(x_i). \quad (3)$$

To simplify notation, and to make contact with the energy landscape and the statistical physics notations, it is convenient to introduce the energy cost function, or Hamiltonian, as

$$\mathcal{H}(x) = \mathcal{H}_2 + \mathcal{H}_p = -\frac{1}{\Delta_2\sqrt{N}} \sum_{i<j} Y_{ij} x_i x_j$$
$$-\frac{\sqrt{(p-1)!}}{\Delta_p N^{(p-1)/2}} \sum_{i_1<\ldots<i_p} T_{i_1,\ldots,i_p} x_{i_1}\ldots x_{i_p}. \quad (4)$$

Thus, keeping in mind that for $N \to \infty$ the spherical constraint is satisfied, $|x|^2 = N$, the posterior is written as $P(x|Y,T) = \exp[-\mathcal{H}(x)]/\tilde{Z}(Y,T)$, where $\tilde{Z}$ is the normalizing partition function.

With the use of the replica theory and its recent proofs from Refs. [9,10,38], one can rigorously establish that the mean squared error achieved by the Bayes-optimal estimator (2) is given as MMSE $= 1 - m^*$, where $m^* \in \mathbb{R}$ is the global maximizer of the so-called free entropy of the problem:

$$\Phi_{RS}(m) = \frac{1}{2}\log(1-m) + \frac{m}{2} + \frac{m^2}{4\Delta_2} + \frac{m^p}{2p\Delta_p}. \quad (5)$$

This expression is derived, and proven, in the Appendix B 2. We note that the proof applies to the posterior distribution (2) with the Gaussian prior.

We now turn to the AMP algorithm [10,18], which is the best algorithm known so far for this problem. AMP is an iterative algorithm inspired by the work of Thouless-Anderson and Palmer in statistical physics [39]. We give its form explicitly in Appendix B 1. Most remarkably, performance of the AMP can be evaluated by tracking its evolution with the iteration time, and it is given in terms of the (possibly local) maximum of the above free entropy, which is reached as a fixed point of the following iterative process:

$$m^{t+1} = 1 - \frac{1}{1 + m^t/\Delta_2 + (m^t)^{p-1}/\Delta_p}, \quad (6)$$

with the initial condition $m^{t=0} = \epsilon$, with $0 < \epsilon \ll 1$. Equation (6) is called the state evolution (SE) of the AMP, and its validity is proven for closely related models in Ref. [40]. We denote the corresponding fixed point m_{AMP} and the corresponding estimation error $\text{MSE}_{\text{AMP}} = 1 - m_{\text{AMP}}$.

The phase diagram presented in Fig. 1 summarizes this theory for the spiked $2 + 3$-spin model. It is deduced by investigating the local maxima of the scalar function (5). Notably, we observe that the phase diagram in terms of Δ_2 and Δ_p splits into three phases:

(i) Easy, in green for $\Delta_2 < 1$ and any Δ_p: The fixed point of the state evolution (6) is the global maximizer of the free entropy (5), and $m^* = m_{\text{AMP}} > 0$.

(ii) Hard, in orange, for $\Delta_2 > 1$ and low $\Delta_p < \Delta_p^{\text{IT}}(\Delta_2)$: The fixed point of the state evolution (6) is not
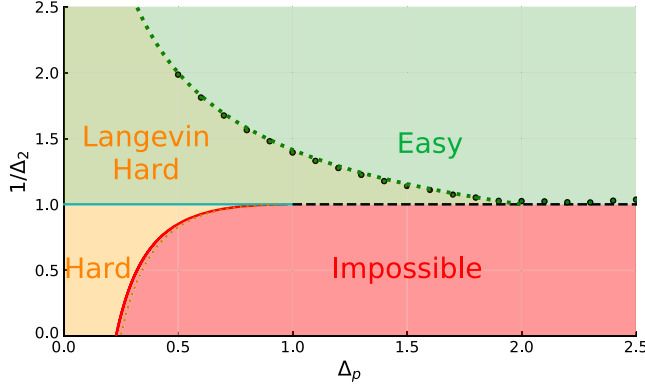
FIG. 1. Phase diagram of the spiked $2 + 3$-spin model (the matrix and order-3 tensor are observed). In the easy (green) region, the AMP achieves an optimal error smaller than the random pick from the prior. In the impossible region (red), the optimal error is as bad as the random pick from the prior, and the AMP achieves it as well. In the hard region (orange), the optimal error is low, but the AMP does not find an estimator better than the random pick from the prior. In the case of the Langevin algorithm, the performance is strictly worse than that for the AMP in the sense that the hard region increases up to the line $1/\Delta_2^* = \max(1, \sqrt{\Delta_3/2})$, depicted by green dots. The green circles are obtained by numerical extrapolation of the Langevin state evolution equations.

the global maximizer of the free entropy (5), and $m^* > m_{\mathrm{AMP}} = 0$.

(iii) Impossible, in red, for $\Delta_2 > 1$ and high $\Delta_p > \Delta_p^{\mathrm{IT}}(\Delta_2)$: The fixed point of the state evolution (6) is the global maximizer of the free entropy (5), and $m^* = m_{\mathrm{AMP}} = 0$.

For the $2 + p$-spin model with $p > 3$, the phase diagram is slightly richer and is presented in Appendix D 4.

IV. LANGEVIN ALGORITHM AND ITS ANALYSIS

We now turn to the core of the paper and the analysis of the Langevin algorithm. In statistics, the most common way to compute the Bayes-optimal estimator (3) is to attempt to sample the posterior distribution (2) and to use several independent samples to compute the expectation in Eq. (3). In order to do perform this process, one needs to set up a stochastic dynamics on x that has a stationary measure at long times given by the posterior measure (2). The Langevin algorithm is one possibility (others include, notably, the Markov chain Monte Carlo method). The common bottleneck is that the time needed to achieve stationarity can, in general, be exponential in the system size, in which case the algorithm is practically useless. However, this result is not always the case, and there are regions in parameter space where one can expect that the relaxation to the posterior measure happens on tractable

timescales. Therefore, it is crucial to understand where this happens and what the associated relaxation timescales are.

The Langevin algorithm on the hypersphere with the Hamiltonian given by Eq. (4) reads

$$\dot{x}_i(t) = -\mu(t)x_i(t) - \frac{\partial \mathcal{H}}{\partial x_i} + \eta_i(t), \qquad (7)$$

where $\eta_i(t)$ is a zero mean noise term, with $\langle \eta_i(t)\eta_j(t') \rangle = 2\delta_{ij}\delta(t - t')$, where the average $\langle \cdot \rangle$ is with respect to the realizations of the noise. The Lagrange multiplier $\mu(t)$ is chosen in such a way that the dynamics remains on the hypersphere. In the large-N limit, one finds $\mu(t) = 1 - 2\mathcal{H}_2(t) - p\mathcal{H}_p(t)$, where the $\mathcal{H}_2(t)$ is the first term from Eq. (4) evaluated at $x(t)$, and $\mathcal{H}_p(t)$ is the value of the second term from Eq. (4).

The presented spiked matrix-tensor model falls into the particular class of spherical $2 + p$-spin glasses [41,42] for which the performance of the Langevin algorithm can be tracked exactly in the large-N limit via a set of integro-partial differential equations [26,27], previously dubbed CHSCK. We call this generalized version of the CHSCK equations the Langevin state evolution (LSE) equations, in analogy to the state evolution of AMP.

In order to write the LSE equations, we define three dynamical correlation functions,

$$C_N(t, t') \equiv \frac{1}{N}\sum_{i=1}^N x_i(t)x_i(t'), \qquad (8)$$

$$\bar{C}_N(t) \equiv \frac{1}{N}\sum_{i=1}^N x_i(t)x_i^*, \qquad (9)$$

$$R_N(t, t') \equiv \frac{1}{N}\sum_{i=1}^N \partial x_i(t)/\partial h_i(t')|_{h_i=0}, \qquad (10)$$

where h_i is a pointwise external field applied at time t' to the Hamiltonian as $\mathcal{H} + \sum_i h_i x_i$. We note that the correlation functions defined above depend on the realization of the thermal history [i.e., of the noise $\eta(t)$] and on the disorder (here, the matrix Y and tensor T). However, in the large-N limit, they all concentrate around their averages. Thus, we define $C(t, t') = \lim_{N\to\infty}\mathbb{E}_{Y,T}\langle C_N(t, t') \rangle_\eta$ and analogously for $\bar{C}(t)$ and $R(t, t')$. Standard field theoretical methods [43] or dynamical cavity method arguments [44] can then be used to obtain a closed set of integrodifferential equations for the averaged dynamical correlation functions, describing the average global evolution of the system under the Langevin algorithm. The resulting LSE equations are (see Appendix C for a complete derivation)

$$\frac{\partial}{\partial t}C(t,t') = 2R(t',t) - \mu(t)C(t,t') + Q'(\bar{C}(t))\bar{C}(t') + \int_0^t dt'' R(t,t'')Q''(C(t,t''))C(t',t'') + \int_0^{t'} dt'' R(t',t'')Q'(C(t,t'')),$$

$$\frac{\partial}{\partial t}R(t,t') = \delta(t-t') - \mu(t)R(t,t') + \int_{t'}^t dt'' R(t,t'')Q''(C(t,t''))R(t'',t'),$$

$$\frac{\partial}{\partial t}\bar{C}(t) = -\mu(t)\bar{C}(t) + Q'(\bar{C}(t)) + \int_0^t dt'' R(t,t'')\bar{C}(t'')Q''(C(t,t'')), \tag{11}$$

where we have defined $Q(x) = x^2/(2\Delta_2) + x^p/(p\Delta_p)$. The Lagrange multiplier $\mu(t)$ is fixed by the spherical constraint, through the condition $C(t,t) = 1 \; \forall \; t$. Furthermore, causality implies that $R(t,t') = 0$ if $t < t'$. Finally, the Ito convention on the stochastic equation (7) gives $\forall \, t \lim_{t' \to t^-} R(t,t') = 1$.

V. BEHAVIOR OF THE LANGEVIN ALGORITHM

In order to assess the performance of the Langevin algorithm and compare it with AMP, we notice that the correlation function $\bar{C}(t)$ is directly related to the accuracy of the algorithm. We solve the differential equations (11) numerically along the lines of Refs. [45,46] (for a detailed procedure, see Appendix C 1; codes are available online [47]). In Fig. 2, we plot the correlation with the spike $\bar{C}(t)$ as a function of the running time t for $p = 3$, fixed $\Delta_2 = 0.7$, and several values of Δ_p, and we use as an initial condition $\bar{C}(t = 0) = 10^{-4}$. In the inset of the plot, we compare it to the same quantity obtained from the state evolution of the AMP algorithm, with the same initial condition. For the Langevin algorithm in Fig. 2, we see a pattern that is striking. One would expect that as the noise Δ_p decreases, the inference problem is easier; the correlation with the signal is larger, and it is reached sooner in the iteration. This case is exactly what we observe for the

AMP algorithm in the inset of Fig. 2. Also for the Langevin algorithm, the plateau reached for long times t becomes higher (better accuracy) as the noise Δ_p is reduced. Furthermore, the height of the plateau coincides with that reached by AMP, thus testifying that the algorithm reached equilibrium. However, contrary to AMP, the relaxation time for the Langevin algorithm increases dramatically with diminishing Δ_p (notice the log scale on the x axis of Fig. 2, as compared to the linear scale of the inset).

We define τ as the time it takes for the correlation to reach a value $\bar{C}_{\text{plateau}}/2$. We then plot the value of this equilibration time in the insets of Fig. 3 as a function of the noise Δ_2 having fixed Δ_p. The data are consistent with a divergence of τ at a certain finite value of Δ_2^*. We found that the divergence points are affected by the initial condition of the dynamics $\bar{C}(t = 0)$; this aspect is discussed in Appendix D 5. In the analysis of the phase diagram, we initialize the dynamics to $\bar{C}(t = 0) = 10^{-40}$ (smaller values have not led to noticeable changes in Δ_2^*). We calculate the divergence time and fit the data with a power law $\tau(\Delta) = |(1/\Delta) - (1/\Delta^*)|^{-\gamma}$; we obtain, in the particular case of fixed $\Delta_p = 1.0$, that $\gamma = 2.24$ and $\Delta_2^* = 0.72$. We are not able to strictly prove that the divergence of the relaxation time truly occurs, but at least our results imply
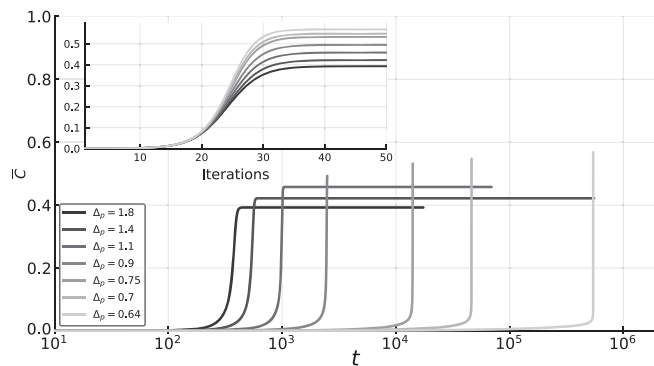

FIG. 2. Evolution of the correlation with the signal $\bar{C}(t)$ starting from $\bar{C}(t = 0) = 10^{-4}$ in the Langevin algorithm at fixed noise on the matrix ($\Delta_2 = 0.7$) and different noises on the tensor (Δ_p). As we decrease Δ_p, the time required to jump to the solution appears to diverge. Inset: The behavior of $\bar{C}(t)$ as a function of the iteration time for the AMP algorithm for the same values of Δ_p and the same initialization.
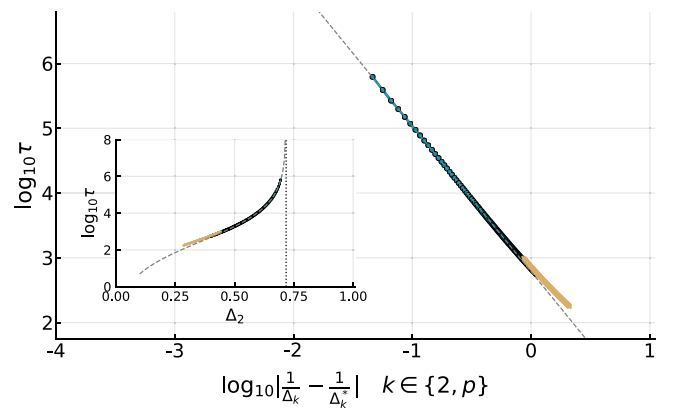

FIG. 3. Extrapolation of the Langevin relaxation time. The inset presents the relaxation time for fixed $\Delta_p = 1$. The main panel presents a fit using a power law consistent with a divergence at $\Delta_2^* \approx 0.72$. The circles are obtained with a numerical solution of LSE that uses the dynamical grid, while crosses are obtained using a fixed grid; the initial condition is $\bar{C}(t = 0) = 10^{-40}$ (details are given in Appendix D5).
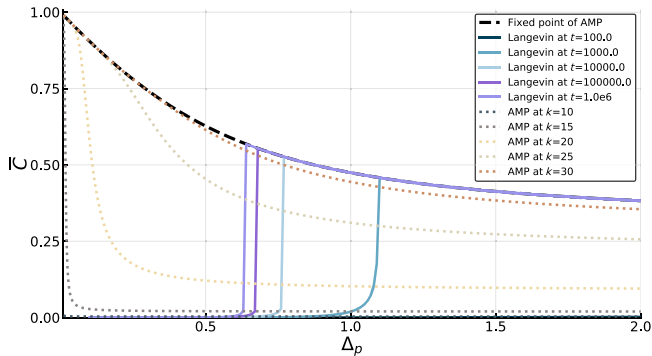
FIG. 4. Correlation with the signal of the AMP and Langevin at the kth iteration (at time t) for fixed $\Delta_2 = 0.7$, where both the evolutions start with initial overlap 10^{-4}.

that for $\Delta_2 > \Delta_2^*$ the Langevin algorithm (7) is not a practical solver for the spiked mixed matrix-tensor problem. We call the region $\Delta_2^* < \Delta_2 < 1$, where the AMP algorithm works optimally without problems yet the Langevin algorithm does not, the Langevin hard region. Then, Δ_2^* is plotted in Fig. 1 with green points, and it delimits the Langevin hard region that extends considerably into the region where the AMP algorithm works optimally in a small number of iterations. Thus, our main conclusion is that the Langevin algorithm designed to sample the posterior measure works efficiently in a considerably smaller region of parameters than the AMP, as quantified in Fig. 1.

Figure 4 presents another way to depict the observed data; the correlation $\bar{C}(t)$ reached after time t is plotted as a function of the tensor noise variance Δ_p. The results of AMP are depicted with dotted lines and, as one would expect, they decrease monotonically as the noise Δ_p increases. The equilibrium value (black dashed line) is reached within a few dozen iterations. On the contrary, the correlation reached by the Langevin algorithm after time t is nonmonotonic and close to zero for small values of noise Δ_p, signaling again a rapidly growing relaxation time when Δ_p is decreased.

VI. GLASSY NATURE OF THE LANGEVIN HARD PHASE

The behavior of the Langevin dynamics as presented in the last section might seem counterintuitive at first sight because one would expect any problem to get simpler when noise Δ_p is decreased. In the present model, instead, as Δ_p is decreased, the tensor part of the cost function (4) becomes more important. As a consequence, the landscape becomes rougher and causes the failure of the Langevin algorithm.

In the presence of the hard (for AMP) phase, it was recently argued in Ref. [25] that sampling-based algorithms are indeed expected to be worse than the approximate message-passing ones. This result is due to residual

glassiness that extends beyond the hard phase. We repeat the analysis of Ref. [25] in the present model (details can be found in Appendix E) and conclude that, while this explanation provides the correct physical picture, the transition line obtained in this way does not agree quantitatively with the numerical extrapolation of the relaxation times we have obtained numerically in the previous section, at least on the timescales on which we were able to solve the LSE equations. The reasons behind this conclusion remain open.

In order to obtain a theoretical estimate that quantitatively agrees with the observed behavior of the LSE, we find the following argument. We first notice that the Langevin dynamics initialized at very small overlap $\bar{C}(t=0)$ remains, for a long time, at small values of the correlation with the signal. We assume that, during this time, the dynamics behaves as it would in the mixed $2 + p$-spin model without the spike. The model without the spike has been studied extensively in the physics literature, precisely with the aim to understand the dynamical properties of glasses [4,27,48]. One of the important results of those studies is that the randomly initialized dynamics converges asymptotically to the so-called threshold states. Indeed, in Fig. 5, this aspect can be observed in the evolution of the energy. It soon approaches a value that can be evaluated [4,27,48], and it corresponds to the threshold state energy E^{th} (horizontal lines),

$$E^{\text{th}} = -Q(1) - \left[\frac{1}{(1-q^{\text{th}})Q'(q^{\text{th}})} - \frac{1}{q^{\text{th}}} \right] Q(q^{\text{th}}). \quad (12)$$

In the above equation, qth represents the correlation, a.k.a. overlap, of two configurations randomly picked from the same threshold state, which can also be evaluated as the solution of

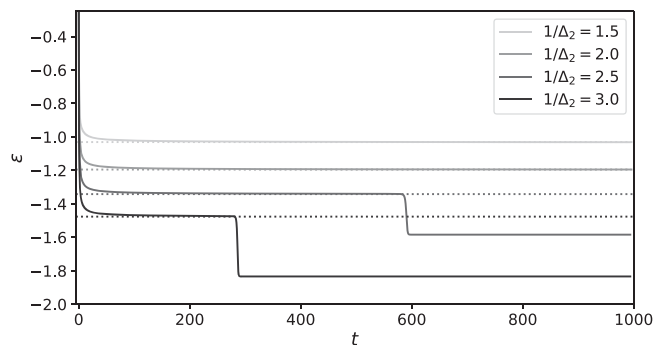$$\frac{1}{1-q^{\text{th}}} = \sqrt{(p-1)\frac{(q^{\text{th}})^{p-2}}{\Delta_p} + \frac{1}{\Delta_2}}. \quad (13)$$

FIG. 5. Dynamical evolution of the energy starting from a configuration with overlap 10^{-40} at $\Delta_p = 1.0$ and for various Δ_2. The system first tends to the threshold energy [Eq. (12), horizontal lines] and then, for sufficiently small Δ_2, finds the direction toward the signal.

The derivation of these expressions can be found in Appendixes E 1 and F. Supported by the numerical results of Fig. 5, we make an approximation that, already on the observed timescales, the algorithm converges to the threshold states [49]. The presence of the signal determines whether the algorithm develops a correlation with the signal. To understand how it occurs, one has to study the statistical properties of the Thouless-Anderson-Palmer (TAP) free-energy landscape, which is the finite-temperature counterpart of the energy landscape, as it has been shown in early results of spin-glass theory [44,50] and in the recent results of the mathematical community [51]. The generic picture, which resulted from several years of studies on spin-glass models, is that threshold states correspond to marginal local minima of the TAP free energy. Critical points of the TAP free-energy functional that are below the threshold states are typically local minima, while those above are saddles with extensively many negative directions. The threshold states lie in between and have just a few very flat directions. In order to obtain an analytical prediction for the Langevin dynamics threshold, one has to find out how the presence of the spike destabilizes the threshold states, which can be achieved by studying the free-energy Hessian at a threshold state.

As shown in Appendix F, such a Hessian reads

$$\frac{\partial^2 F}{\partial m_i \partial m_j} = G_{ij} + \delta_{ij} 2\sigma_F(q_{\text{th}}) - \frac{1}{\Delta_2}\frac{x_i^* x_j^*}{N} + f''(q_{\text{th}})\frac{m_i m_j}{N},$$

(14)

where $f''(q_{\text{th}})$ is positive (and its expression can be found in Appendix F), m_i is the average magnetization of site i in the given threshold state, and G_{ij} can be shown to be statistically equivalent to a random matrix having elements that are i.i.d. Gaussian random variables with mean zero and variance are given as

$$\frac{\sigma_F^2(q_{\text{th}})}{N} = \frac{(p-1)q_{\text{th}}^{p-2}}{\Delta_p} + \frac{1}{\Delta_2}.$$

The free-energy Hessian evaluated at a typical threshold state is therefore a random matrix belonging to the Gaussian orthogonal ensemble plus two rank-one perturbations: One is negative and in the direction of the signal, whereas the other is positive and in the direction of the threshold state. Results from random matrix theory allow us to completely characterize the spectral properties of the Hessian. Its bulk density of eigenvalues is a shifted semicircle whose left edge touches zero, hence leading to the marginality of the threshold states. For a small signal-to-noise ratio, the minimal eigenvalue is zero, whereas when the signal-to-noise ratio exceeds a certain critical value, the rank-one perturbation in the direction of the signal induces

a BBP (Baik, Ben Arous, Péché) transition [15,52], where a negative eigenvalue pops out from the Wigner semicircle, and correspondingly, a downward descent direction toward the spike emerges and makes the threshold states unstable. Note that the last term of the Hessian has no effect on the development of an unstable direction as it is positive and uncorrelated with the signal. By adapting the known formulas for the BBP transition to our case (see Appendix F), we find a landscape-based conjecture for the algorithmic threshold, that is, the larger value of Δ_2^* between $\Delta_2^* = 1$ and the roots of

$$\Delta_p = (p-1)(\Delta_2^*)^2(1-\Delta_2^*)^{p-3}.$$

(15)

This result is the threshold depicted in Fig. 1 for $p = 3$ as the green dotted line. We note a very good agreement with the data points obtained with extrapolation of the relaxation time from the numerical solution of the LSE equations.

In the following, we present a complementary argument that, interestingly, also makes a direct link with AMP state evolution. We again assume that Langevin dynamics approaches the threshold states. This time, we use AMP to determine whether it will remain there or not. If the initial correlation is $0 < m^{t=0} \ll 1$, its evolution follows

$$m^{t+1} = (1-q^{\text{th}})\left[\frac{(m^t)^{p-1}}{\Delta_p} + \frac{m^t}{\Delta_2}\right].$$

(16)

This equation is obtained from the state evolution of the AMP algorithm, where the overlap is fixed to qth, as detailed in Appendix B 3.

The stability condition that decides whether an infinitesimal correlation will grow or decrease under Eq. (16) reads $q^{\text{th}} = 1 - \Delta_2$. Using Eq. (13), this condition then leads to Eq. (15).

VII. DISCUSSION AND PERSPECTIVES

Motivated by the general aim to shed light on the behavior and performance of noisy-gradient-descent algorithms that are widely used in machine learning, we analytically investigate the performance of the Langevin algorithm in the noisy high-dimensional limit of a spiked matrix-tensor model. We compare it to the performance of the approximate message-passing algorithm. While both of these algorithms are designed with the aim to sample the posterior measure of the model, we show that the Langevin algorithm fails to find correlation with the signal in a considerable part of the AMP easy region. Neither of the two algorithms enters the so-called hard phase. Our analysis is based on the Langevin state evolution equations, a generalization of the dynamical theory for mean field spin glasses, that describe the evolution of the algorithm in the large size limit.

The Langevin algorithm performs worse than the AMP due to the underlying glass transition in the corresponding region of parameters. Relying on the result from spin-glass theory, we present a simple heuristic expression of the Langevin-threshold (15) line, which appears to be in agreement with the value obtained from the numerical solution of the LSE equations.

We note that, so far, in our study of the spiked matrix-tensor model with Langevin dynamics, we only accessed the cost function (4) and its derivatives. We did not allow ourselves to split the cost function in the tensor-related $\mathcal{H}_p$ and the matrix-related $\mathcal{H}_2$ parts. If we did, then there would be a simple way to overcome the Langevin hard regime by first considering only the matrix measurements and then slowly turning on the tensor, similarly to how temperature is tuned in simulated annealing. We study this procedure in Appendix D 6. It is interesting to underline that, from the point of view of Bayesian inference, this finding remains somewhat paradoxical. In the setting of this paper, we know perfectly the model that generated the data and all its parameters, yet we see that for the Langevin algorithm, it is computationally advantageous to mismatch the parameters and perform the annealing in the tensor part in order to reach faster convergence to equilibrium. This idea is particularly striking given that, for AMP, it has been proven in Ref. [7] that mismatching the parameters can never improve the performance. In fact, from a physics point of view, the principle (thanks to which the AMP does not share the hurdles of the Langevin algorithm) remains an interesting open question.

We stress that the above annealing procedure is a particularity of the present model and will not generalize to a broad range of inference problems because it is not clear, in general, how to split the cost function into a simple-to-optimize yet informative part and the rest. A formidably interesting direction for future work consists instead in investigating whether the performance of the Langevin algorithm can be improved in a manner that only accesses the cost function or its derivatives.

While here we studied the spiked matrix-tensor model, we expect that our findings, based on the existence of an underlying glass transition, will hold more universally. We expect them to apply to other local sampling dynamics, e.g., to the Markov chain Monte Carlo method, and to a broader range of models, e.g., simple models of neural networks. An interesting extension of this work would be to investigate algorithms closer to stochastic gradient descent and models closer to current neural network architectures.

APPENDIX A: DEFINITION OF THE SPIKED MATRIX-TENSOR MODEL

We consider a teacher-student setting in which the teacher constructs a matrix and a tensor from a randomly sampled signal, and the student is asked to recover the signal from the observation of the matrix and tensor provided by the teacher [6].

The signal x^* is an N-dimensional vector whose entries are real i.i.d. random variables sampled from the normal distribution [i.e., the prior is $P_X \sim \mathcal{N}(0,1)$]. From the signal, the teacher generates a symmetric matrix and a symmetric tensor of order p. Those two objects are then transmitted through two noisy channels with variances Δ_2 and Δ_p, so at the end, one has two noisy observations given by

$$Y_{ij} = \frac{x_i^* x_j^*}{\sqrt{N}} + \xi_{ij}, \tag{A1}$$

$$T_{i_1,\dots,i_p} = \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} x_{i_1}^* \dots x_{i_p}^* + \xi_{i_1,\dots,i_p}, \tag{A2}$$

where, for $i < j$ and $i_1 < \dots < i_p$, ξ_{ij} and $\xi_{i_1,\dots,i_p}$ are i.i.d. random variables distributed according to $\xi_{ij} \sim \mathcal{N}(0, \Delta_2)$ and $\xi_{i_1,\dots,i_p} \sim \mathcal{N}(0, \Delta_p)$. The ξ_{ij} and $\xi_{i_1,\dots,i_p}$ are symmetric random matrix and tensor, respectively. Given Y_{ij} and $T_{i_1,\dots,i_p}$, the inference task is to reconstruct the signal x^*.

In order to solve this problem, we consider the Bayesian approach. This approach starts from the assumption that both the matrix and the tensor have been produced from the same kind of process as the one described by Eqs. (A1) and (A2). Furthermore, we assume to know the statistical properties of the channel, namely, the two variances Δ_2 and Δ_p, and the prior on x. Given these properties, the posterior probability distribution over the signal is obtained through the Bayes formula

$$P(X|Y,T) = \frac{P(Y,T|X)P(X)}{P(Y,T)}, \tag{A3}$$

where

$$P(Y,T|X) = \prod_{i<j} P_Y\left(Y_{ij}\left|\frac{x_i x_j}{\sqrt{N}}\right.\right) \prod_{i_1<...<i_p} P_T\left(T_{i_1...i_p}\left|\frac{\sqrt{(p-1)!}}{N^{(p-1)/2}}x_{i_1}...x_{i_p}\right.\right) =$$

$$\propto \prod_{i<j} e^{-\frac{1}{2\Delta_2}\left(Y_{ij}-\frac{x_i x_j}{\sqrt{N}}\right)^2} \prod_{i_1<...<i_p} e^{-\frac{1}{2\Delta_p}\left(T_{i_1...i_p}-\frac{\sqrt{(p-1)!}}{N^{(p-1)/2}}x_{i_1}...x_{i_p}\right)^2}. \tag{A4}$$

Therefore, we have

$$P(X|Y,T) = \frac{1}{Z(Y,T)}\prod_i e^{-\frac{1}{2}x_i^2}\prod_{i<j} e^{-\frac{1}{2\Delta_2}\left(Y_{ij}-\frac{x_i x_j}{\sqrt{N}}\right)^2} \prod_{i_1<...<i_p} e^{-\frac{1}{2\Delta_p}\left(T_{i_1...i_p}-\frac{\sqrt{(p-1)!}}{N^{(p-1)/2}}x_{i_1}...x_{i_p}\right)^2}, \tag{A5}$$

where $Z(Y,T)$ is a normalization constant.

Plugging Eqs. (A1) and (A2) into Eq. (A5) allows us to rewrite the posterior measure in the form of a Boltzmann distribution of the mixed $2+p$-spin Hamiltonian [41,42,53],

$$\mathcal{H} = -\frac{1}{\Delta_2\sqrt{N}}\sum_{i<j}\xi_{ij}x_i x_j - \frac{\sqrt{(p-1)!}}{\Delta_p N^{\frac{p-1}{2}}}\sum_{i_1<...<i_p}\xi_{i_1...i_p}x_{i_1}...x_{i_p} - \frac{N}{2\Delta_2}\left(\frac{1}{N}\sum_i x_i x_i^*\right)^2$$

$$+ -\frac{N}{p\Delta_p}\left(\frac{1}{N}\sum_i x_i x_i^*\right)^p - \frac{1}{2}\sum_{i=1}^N x_i^2 + \text{const}, \tag{A6}$$

so

$$P(X|Y,T) = \frac{1}{\tilde{Z}(Y,T)}e^{-\mathcal{H}}. \tag{A7}$$

In the following, we refer to $\tilde{Z}(Y,T)$ as the partition function. We note here that in the large-N limit, using a Gaussian prior on the variables x_i is equivalent to considering a flat measure over the N-dimensional hypersphere $\sum_{i=1}^N x_i^2 = N$. This choice will be used when we describe the Langevin algorithm, and in this case, the last term in the Hamiltonian will become an irrelevant constant.

APPENDIX B: APPROXIMATE MESSAGE PASSING, STATE EVOLUTION, AND PHASE DIAGRAMS

Approximate message passing is a powerful iterative algorithm to compute the local magnetizations $\langle x_i \rangle$ given the observed matrix and tensor. It is rooted in the cavity method of statistical physics of disordered systems [39,44], and it has recently been developed in the context of statistical inference [12], where, in the Bayes-optimal case, it has been conjectured to be optimal among all local iterative algorithms. Among the properties that make AMP extremely useful is the fact that its performances can be analyzed in the thermodynamic limit. Indeed, in such a limit, its dynamical evolution is described by the so-called SE equations [12]. In this section, we derive the AMP equations and their SE description for the spiked matrix-tensor model and solve them to obtain the phase diagram of

the model as a function of the variances Δ_2 and Δ_p of the two noisy channels.

1. Approximate message passing and Bethe free entropy

AMP can be obtained as a relaxed Gaussian closure of the belief propagation (BP) algorithm. The derivation that we present follows the same lines of Refs. [10,37]. The posterior probability can be represented as a factor graph where all the variables are represented by circles and are linked to squares representing the interactions [54] (see Fig. 6).

This representation is very convenient to write down the BP equations. In the BP algorithm, we iteratively update until a set of variables converge; these variables are the beliefs of the (cavity) magnetization of the nodes. The intuitive underlying reasoning behind how BP works is as follows. Given the current state of the variable nodes, take a factor node and exclude one node among its neighbors. The remaining neighbors through the factor node express a belief on the state of the excluded node. This belief is mathematically described by a probability distribution called a message, $\tilde{m}^t_{ij\to i}(x_i)$ and $\tilde{t}^t_{ii_2...i_p\to i}(x_i)$, depending on which factor node is selected. At the same time, another belief on the state of the excluded node is given by the rest of the network, but we use the factor node previously taken into account, $m_{i\to ij}(x_i)$ and $t_{i\to ii_2...i_p}(x_i)$, respectively. All of these messages travel in the factor graph carrying partial information on the real magnetization of the single nodes, and they are iterated until convergence. The iterative scheme is described by the following equations:
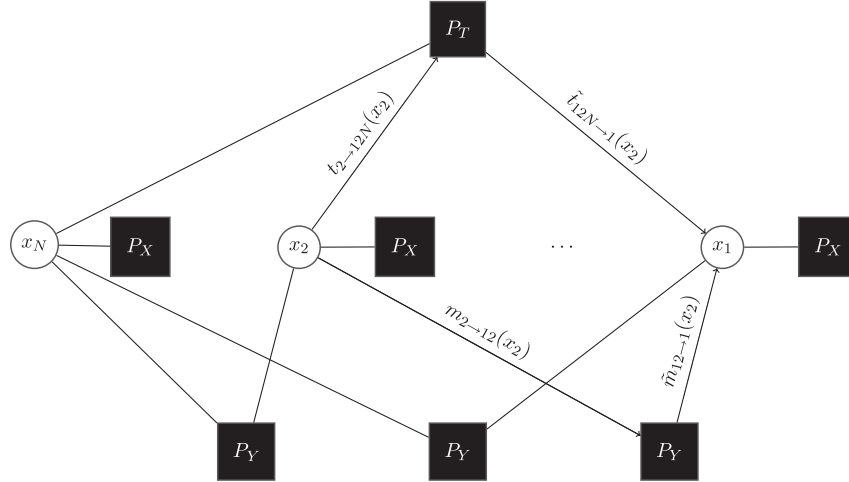
FIG. 6.   The factor graph representation of the posterior measure of the matrix-tensor factorization model. The variable nodes represented with white circles are the components of the signal, while black squares are factor nodes that denote interactions between the variable nodes that appear in the interaction terms of the Boltzmann distribution in Eqs. (A6) and (A7). There are three types of factor nodes: P_X is the prior that depends on a single variable, P_Y is the probability of observing a matrix element Y_{ij} given the values of the variables x_i and x_j, and, finally, P_T is the probability of observing a tensor element $T_{i_1,\dots,i_p}$. The posterior, apart from the normalization factor, is simply given by the product of all the factor nodes.

$$\tilde{m}^t_{ij\to i}(x_i) \propto \int dx_j m^t_{j\to ij}(x_j) P_Y\left(Y_{ij} \middle| \frac{x_i x_j}{\sqrt{N}}\right), \quad \text{(B1)}$$

$$m^{t+1}_{i\to ij}(x_i) \propto P_X(x_i)\prod_{l\neq j}\tilde{m}^t_{il\to i}(x_i) \prod_{i_2<\dots<i_p}\tilde{t}^t_{ii_2\dots i_p\to i}(x_i), \quad \text{(B2)}$$

$$\tilde{t}^t_{ii_2\dots i_p\to i}(x_i) \propto \int \prod_{l=2\dots p}(dx_l t^t_{i_l\to ii_2\dots i_p}(x_l))$$
$$\times P_T\left(T_{ii_2\dots i_p} \middle| \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}}x_i x_{i_2}\dots x_{i_p}\right), \quad \text{(B3)}$$

$$t^{t+1}_{i\to ii_2\dots i_p}(x_i) \propto P_X(x_i)\prod_l\tilde{m}^t_{il\to i}(x_i)$$
$$\times \prod_{k_2<\dots<k_p\neq i_2\dots i_p}\tilde{t}^t_{ik_2\dots k_p\to i}(x_i), \quad \text{(B4)}$$

and we have omitted the normalization constants that guarantee that the messages are probability distributions. When the messages have converged to a fixed point, the estimation of the local magnetizations can be obtained through the computation of the real marginal probability distribution of the variables, given by

$$\mu_i(x_i) = \int\left[\prod_{j(\neq i)}dx_j\right]P(X|Y,T)$$
$$= P_X(x_i)\prod_l\tilde{m}^t_{il\to i}(x_i) \prod_{i_2<\dots<i_p}\tilde{t}^t_{ii_2\dots i_p\to i}(x_i). \quad \text{(B5)}$$

We note that the computational cost to produce an iteration of BP scales as $O(N^p)$. Furthermore,

Eqs. (B1)–(B4) are iterative equations for continuous functions and therefore are extremely hard to solve when dealing with continuous variables. The advantage of AMP is to drastically reduce the computational complexity of the algorithm by closing the equations on a Gaussian ansatz for the messages. This process is justified in the present context since the factor graph is fully connected; therefore, each iteration step of the algorithm involves sums of a large number of independent random variables that give rise to Gaussian distributions. Gaussian random variables are characterized by their mean and covariance, which are readily obtained for $N \gg 1$, expanding the factor nodes for small $\omega_{ij} = x_i x_j/\sqrt{N}$ and $\omega_{i_1\dots i_p} = \sqrt{(p-1)!}x_1\dots x_p/N^{(p-1/2)}$.

Once the BP equations are relaxed on Gaussian messages, the final step to obtain the AMP algorithm is the so-called TAPyfication procedure [37,39], which exploits the fact that the procedure of removing one node or one factor produces only a weak perturbation to the real marginals and therefore can be described in terms of the real marginals of the variable nodes themselves. By applying this scheme, we obtain the AMP equations, which are described by a set of auxiliary variables $A^{(k)}$ and $B^{(k)}_i$ and by the mean $\langle x_i\rangle$ and variance $\sigma_i = \langle x_i^2\rangle$ of the marginals of variable nodes. The AMP iterative equations are

$$B^{(2),t}_i = \frac{1}{\Delta_2\sqrt{N}}\sum_k Y_{ki}\hat{x}^t_k - \frac{1}{\Delta_2}\left(\frac{1}{N}\sum_k\sigma^t_k\right)\hat{x}^{t-1}_i, \quad \text{(B6)}$$

$$A^{(2),t} = \frac{1}{\Delta_2 N}\sum_k(\hat{x}^t_k)^2, \quad \text{(B7)}$$

$$B_i^{(p),t} = \frac{\sqrt{(p-1)!}}{\Delta_p N^{(p-1)/2}} \sum_{k_2 \ldots k_p} T_{i k_2 \ldots k_p}(\hat{x}_{k_2}^t \ldots \hat{x}_{k_p}^t)$$

$$- \frac{p-1}{\Delta_p} \left[\left(\frac{1}{N} \sum_k \sigma_k^t \right) \left[\frac{1}{N} \sum_k \hat{x}_k^t \hat{x}_k^{t-1} \right]^{p-2} \right] \hat{x}_i^{t-1},$$

$$\text{(B8)}$$

$$A^{(p),t} = \frac{1}{\Delta_p} \left[\frac{1}{N} \sum_k (\hat{x}_k^t)^2 \right]^{p-1}, \qquad \text{(B9)}$$

$$\hat{x}_i^{t+1} = f(A^{(2)} + A^{(p)}, B_i^{(2)} + B_i^{(p)}), \qquad \text{(B10)}$$

$$\sigma_i^{t+1} = \frac{\partial}{\partial B} f(A, B) \Big|_{A = A^{(2)} + A^{(p)}, \, B = B_i^{(2)} + B_i^{(p)}}, \qquad \text{(B11)}$$

$$f(A, B) \equiv \int dx \frac{1}{\mathcal{Z}(A,B)} x P_X(x) e^{Bx - \frac{1}{2}Ax^2} = \frac{B}{1+A}. \qquad \text{(B12)}$$

It can be shown that these equations can be obtained as saddle-point equations from the so-called Bethe free entropy defined as $\Phi_{\text{Bethe}} = \log Z^{\text{Bethe}}(Y, T)/N$, where Z^{Bethe} is the Bethe approximation to the partition function, which is defined as the normalization of the posterior measure. The expression of the Bethe free entropy per variable can be computed in a standard way (see Ref. [54]), and it is given by

$$\Phi_{\text{Bethe}} = \frac{1}{N} \left(\sum_i \log Z_i + \sum_{i \leq j} \log Z_{ij} + \sum_{i_1 \leq \ldots \leq i_p} \log Z_{i_1 \ldots i_p} \right.$$

$$\left. - \sum_{i(ij)} \log Z_{i,ij} - \sum_{i(ii_2 \ldots i_p)} \log Z_{i(ii_2 \ldots i_p)} \right), \qquad \text{(B13)}$$

where

$$Z_i = \int dx_i P_X(x_i) \prod_j \tilde{m}_{ij \to i}(x_i) \prod_{(i_2 \ldots i_p)} \tilde{t}_{i i_2 \ldots i_p \to i}(x_i),$$

$$Z_{ij} = \int \prod_{j(\neq i)} [dx_j m_{j \to ij}(x_j)] \prod_{i<j} e^{-\frac{1}{2\Delta_2}(Y_{ij} - \frac{x_i x_j}{\sqrt{N}})^2},$$

$$Z_{i_1 \ldots i_p} = \int \prod_{l=1}^p [dx_{i_l} t_{i_l \to i_1 \ldots i_p}(x_{i_l})]$$

$$\times \prod_{i_1 < \ldots < i_p} e^{-\frac{1}{2\Delta_p}(T_{i_1 \ldots i_p} - \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} x_{i_1} \ldots x_{i_p})^2},$$

$$Z_{i(ij)} = \int dx_i m_{i \to ij}(x) \tilde{m}_{ij \to i}(x_i),$$

$$Z_{i(ii_2 \ldots i_p)} = \int dx_i t_{i \to ii_2 \ldots i_p}(x) \tilde{t}_{ii_2 \ldots i_p \to i}(x_i)$$

are a set of normalization factors. Using the Gaussian approximation for the messages and employing the same TAPyification procedure used to get the AMP equations, we obtain the Bethe free entropy density as

$$\Phi_{\text{Bethe}} = \frac{1}{N} \sum_i \log \mathcal{Z}(A^{(p)} + A^{(2)}, B_i^{(p)} + B_i^{(2)})$$

$$+ \frac{p-1}{p} \frac{1}{N} \sum_i \left[-B_i^{(p)} \hat{x}_i + A_i^{(p)} \frac{\hat{x}_i^2 + \sigma_i}{2} \right]$$

$$+ \frac{p-1}{2p\Delta_p} \left(\frac{\sum_i \hat{x}_i^2}{N} \right)^{p-1} \left(\frac{\sum_i \sigma_i}{N} \right)$$

$$+ \frac{1}{2N} \sum_i \left[-B_i^{(2)} \hat{x}_i + A_i^{(2)} \frac{\hat{x}_i^2 + \sigma_i}{2} \right]$$

$$+ \frac{1}{4\Delta_2} \left(\frac{\sum_i \hat{x}_i^2}{N} \right) \left(\frac{\sum_i \sigma_i}{N} \right), \qquad \text{(B14)}$$

where we use the variables defined in Eqs. (B6)–(B9) for the sake of compactness and $\mathcal{Z}(A, B)$ is defined as

$$\mathcal{Z}(A, B) = \int dx P_X(x) e^{Bx - \frac{Ax^2}{2}} = \frac{1}{\sqrt{A+1}} e^{\frac{B^2}{2(A+1)}}. \qquad \text{(B15)}$$

2. Averaged free entropy and its proof

Equation (B14) represents the Bethe free entropy for a single realization of the factor nodes in the large size limit. Here, we wish to discuss the actual exact value of this free entropy, that is,

$$f_N(Y, T) = \frac{\log Z(Y, T)}{N},$$

where the partition function $Z(Y, T)$ is defined as the normalization of the posterior probability distribution, Eq. (2). The free entropy is a random variable since it depends *a priori* on the planted signal and the noise in the tensor and matrices. However, one expects that, since free entropy is an intensive quantity, we expect from the statistical physics intuition that it should be self-averaging and concentrate around its mean value in the large-N limit [44]. In fact, this case is easily proven. First, since the spherical model has a rotational symmetry, one may assume the planted assignment could be any vector on the hypersphere, and we might as well suppose it is the uniform one $x_i^* = 1 \ \forall \ i$: The true source of fluctuation comes from the noises Y and T. These values can be controlled by noticing that the free entropy is a Lipschitz function of the Gaussian random variables Y and T. Indeed,

$$\partial_{Y_{ij}} f_N(Y, T) = \frac{1}{\Delta_2 N \sqrt{N}} \langle x_i x_j \rangle,$$

so the free energy f_N is a Lipschitz function with respect to Y, with the constant

$$L = \frac{1}{\Delta_2 N \sqrt{N}} \sqrt{\sum_{i<j} \langle x_i x_j \rangle^2} \le \frac{1}{\Delta_2 N \sqrt{N}} \sqrt{\frac{1}{2} \sum_{i,j} \langle x_i x_j \rangle^2}$$

$$= \frac{1}{\Delta_2 N \sqrt{N}} \sqrt{\frac{1}{2} \sum_{i,j} \langle x_i \tilde{x}_i x_j \tilde{x}_j \rangle},$$

where $\tilde{x}$ represents a copy (or replica) of the system. In this case,

$$L \le \frac{1}{\Delta_2 N \sqrt{N}} \sqrt{N^2 \left\langle \left(\frac{\sum_i x_i \tilde{x}_i}{N} \right)^2 \right\rangle} = \frac{\sqrt{\langle q^2 \rangle}}{\Delta_2 \sqrt{N}},$$

where q is the overlap between the two replicas x and $\tilde{x}$, which is bounded by 1 on the sphere; thus, $L \le (1/\Delta_2 \sqrt{N})$. Therefore, by Gaussian concentration of Lipschitz functions (the Tsirelson-Ibragimov-Sudakov inequality [55]), we have, for some constant K,

$$\Pr\left[|f_n - \mathbb{E}_Y f_n| \ge t\right] \le 2 e^{-Nt^2/K}, \qquad (B16)$$

and, in particular, any fluctuation larger than $O(1/\sqrt{N})$ is (exponentially) rare. A similar computation shows that f_N is also concentrated with respect to the tensor T. This result shows that, in the large size limit, we can consider the averaged free entropy:

$$\mathcal{F}_N \equiv \frac{1}{N} \mathbb{E}[\log Z_N].$$

With our (nonrigorous) statistical physics tools, this result can be obtained by averaging Eq. (B14) over the disorder (see, for instance, Ref. [37]), and this process yields an expression for the free energy called the replica symmetric (RS) formula:

$$\Phi_{\text{RS}} = \lim_{N \to \infty} \mathbb{E}_{Y,T} \frac{\log Z(Y,T)}{N}. \qquad (B17)$$

We now state precisely the form of Φ_{RS} and prove the validity of Eq. (B17). The RS free entropy for any prior distribution P_X reads as

$$\Phi_{\text{RS}} \equiv \max_m \tilde{\Phi}_{\text{RS}}(m),$$

$$\tilde{\Phi}_{\text{RS}}(m) = \mathbb{E}_{W,x^*} \left[\log \left[\mathcal{Z}\left(\frac{m}{\Delta_2} + \frac{m^{p-1}}{\Delta_p}, \left(\frac{m}{\Delta_2} + \frac{m^{p-1}}{\Delta_p} \right) x^* \right. \right.$$

$$\left. \left. + \sqrt{\frac{m}{\Delta_2} + \frac{m^{p-1}}{\Delta_p}} W \right) \right] \right] - \frac{1}{4\Delta_2} m^2 - \frac{p-1}{2p\Delta_p} m^p,$$

$$\qquad (B18)$$

where W is a Gaussian random variable of zero mean and unit variance and x^* is a random variable taken from the prior P_X. We remind the reader that the function $\mathcal{Z}(A, B)$ is defined via Eq. (B15).

For the Gaussian prior P_X, which is the one of interest here, we obtain

$$\tilde{\Phi}_{\text{RS}}(m) = -\frac{1}{2} \log \left(\frac{m}{\Delta_2} + \frac{m^{p-1}}{\Delta_p} + 1 \right) + \frac{1}{2} \left(\frac{m}{\Delta_2} + \frac{m^{p-1}}{\Delta_p} \right)$$

$$- \frac{1}{4\Delta_2} m^2 - \frac{p-1}{2p\Delta_p} m^p. \qquad (B19)$$

The expression given in the main text is slightly different but can be obtained as follows. First, notice that the extremization condition for $\tilde{\Phi}_{\text{RS}}(m)$ reads

$$m = 1 - \frac{1}{1 + \frac{m}{\Delta_2} + \frac{m^{p-1}}{\Delta_p}}, \qquad (B20)$$

and by plugging this expression into Eq. (B19), we recover the more compact expression $\Phi_{\text{RS}}(m)$ shown in the main text:

$$\Phi_{\text{RS}}(m) = \frac{1}{2} \log(1 - m) + \frac{m}{2} + \frac{m^2}{4\Delta_2} + \frac{m^p}{2p\Delta_p}. \qquad (B21)$$

The two expressions $\Phi_{\text{RS}}(m)$ and $\tilde{\Phi}_{\text{RS}}(m)$ are thus equal for each value of m that satisfies Eq. (B20). The parameter m can be interpreted as the average correlation between the true and the estimated signal,

$$m = \frac{1}{N} \sum_{i=1}^{N} x_i^* \hat{x}_i. \qquad (B22)$$

The average minimal mean squared error (MMSE) can be obtained from the maximizer m of the average Bethe free entropy as

$$\text{MMSE} \equiv \frac{1}{N} \sum_{i=1}^{N} \overline{(x_i^* - \hat{x}_i)^2} = 1 - m^*, \quad \text{where}$$

$$m^* = \arg\max \tilde{\Phi}_{\text{RS}}(m), \qquad (B23)$$

where the overbar stands for the average over the signal x^* and the noise of the two Gaussian channels.

The validity of Eq. (B18) can be proven rigorously for every prior having a bounded second moment. The proof we present is a straightforward generalization of the one presented in Ref. [10] for the pure tensor case and in Ref. [38] for the matrix case, and it is based on two main ingredients. The first one is the Guerra interpolation method applied on the Nishimori line [38,56,57], in which we construct an interpolating Hamiltonian that depends on a parameter $t \in [0; 1]$ that is used to move from the original

Hamiltonian of Eq. (A6) to the one corresponding to a scalar denoising problem whose free entropy is given by the first term in Eq. (B18). The second ingredient is the Aizenman-Sims-Starr method [58], which is the mathematical version of the cavity method (note that other techniques could also be employed to obtain the same results; see Refs. [9,59–61]). The theorem we want to prove is as follows:

Theorem 1 (Replica-symmetric formula for the free energy). Let P_X be a probability distribution over $\mathbb{R}$, with a finite second moment Σ_X. Then, for all $\Delta_2 > 0$ and $\Delta_p > 0$,

$$\mathcal{F}_N \equiv \frac{1}{N}\mathbb{E}[\log Z_N] \underset{[N \to \infty]}{\longrightarrow} \sup_{m \geq 0} \tilde{\Phi}_{\mathrm{RS}}(m) \equiv \Phi_{\mathrm{RS}}(\Delta_2, \Delta_p).$$

(B24)

For almost every $\Delta_2 > 0$ and $\Delta_p > 0$, $\tilde{\Phi}_{\mathrm{RS}}$ admits a unique maximizer m over $\mathbb{R}_+ \times \mathbb{R}_+$ and

$$\text{T-MMSE}_N \underset{[N \to \infty]}{\longrightarrow} \Sigma_X^p - (m^*)^p,$$

$$\text{M-MMSE}_N \underset{[N \to \infty]}{\longrightarrow} \Sigma_X^2 - (m^*)^2.$$

Here, we have defined the tensor-MMSE T-MMSE$_N$ by the error in reconstructing the tensor:

$$\text{T-MMSE}_N(\Delta_2, \Delta_p)$$
$$= \inf_{\hat{\theta}}\left\{ \frac{p!}{N^p} \sum_{i_1 < \ldots < i_p} (x_{i_1}^0 \ldots x_{i_p}^0 - \hat{\theta}(Y)_{i_1 \ldots i_p})^2 \right\},$$

and the matrix-MMSE M-MMSE$_N$ by the error in reconstructing the matrix:

$$\text{M-MMSE}_N(\Delta_2, \Delta_p) = \inf_{\hat{\theta}}\left\{ \frac{2}{N^2} \sum_{i<j} (x_i^0 x_j^0 - \hat{\theta}(Y)_{i,j})^2 \right\},$$

where in both cases the infimum is taken over all measurable functions $\hat{\theta}$ of the observations Y.

The result concerning the MMSE is a simple application of the I-MMSE theorem [62], which relates the derivative of the free energy with respect to the noise variances and the MMSE. The details of the arguments are the same as in the matrix ($p = 2$) case (Ref. [38], Corollary 17) and the tensor one (Ref. [10], Theorem 2). Indeed, as discussed in Refs. [10,38], these M-MMSE and T-MMSE results imply the vector MMSE result of Eq. (B23) when p is odd and thus, in particular, for the $p = 3$ case discussed in the main text.

a. Sketch of proof

In this section, we give a detailed sketch of the proof of Theorem 1. Following the techniques used in many recent works [8–10,38,56,57,59,60], we make a few technical remarks:

(i) We consider only priors with bounded support, $\mathrm{supp}(P_X) = S \subset [-K; K]$. This method allows us to switch integrals and derivatives without problems. This condition can then be relaxed to unbounded distributions with a bounded second moment using the same techniques as the ones that we are going to present, and the proof is therefore valid in this case. This case is detailed, for instance, in Ref. [38], Sec. VI.2.2.

(ii) Another key ingredient is the introduction of a small perturbation in the model that takes the form of a small amount of side information. These kinds of techniques are frequently used in statistical physics, where a small "magnetic field" forces the Gibbs measure to be in a single pure state [63]. It has also been used in the context of coding theory [64] for the same reason. In the context of Bayesian inference, we follow the generic scheme proposed by Montanari in Ref. [65] (see also Ref. [66]) and add a small additional source of information that allows the system to be in a single pure state so that the overlap is concentrated on a single value. This source depends on Bernoulli random variables $L_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(\epsilon)$, $i \in [N]$; if $L_i = 1$, the channel, call it A, transmits the correct information. We can then consider the posterior of this new problem, $P(X|A, Y, T)$ and focus on the associated free-energy density $F_{N,\epsilon}$ defined as the expected value of the average of the logarithm of the normalization constant divided by the number of spins. Then, we can immediately prove that for all $N \geq 1$ and $\epsilon, \epsilon' \in [0; 1]$, it follows that $|F_{N,\epsilon} - F_{N,\epsilon'}| \leq [(K^{2p}/\Delta_p) + (K^4/\Delta_2)]|\epsilon - \epsilon'|$. This result allows us (see, for instance, Ref. [10]) to obtain the concentration of the posterior distribution around the replica parameter ($q = (1/N)\langle x^{(1)} \cdot x^{(2)} \rangle$),

$$\mathbb{E}\left\langle \left(\frac{x^{(1)} \cdot x^{(2)}}{N} - q \right)^2 \right\rangle \overset{N \to \infty}{\longrightarrow} 0, \quad (\text{B25})$$

$$\mathbb{E}\left\langle \left(\frac{x^* \cdot x}{N} - q \right)^2 \right\rangle \overset{N \to \infty}{\longrightarrow} 0, \quad (\text{B26})$$

where $x, x^{(1)}, x^{(2)}$ are sampled from the posterior distribution and the averages $\langle \cdot \rangle$ and $\mathbb{E}[\cdot]$ are, respectively, the average over the posterior measure and the remaining random variables.

(iii) Finally, a fundamental property of inference problems—which is a direct consequence of the Bayes theorem and of the fact that we are in the Bayes-optimal setting where we know the statistical properties of the signal, namely, the prior, and the

statistical properties of the channels, namely, Δ_2 and Δ_p—is the so-called Nishimori symmetry [6,67]: Let (X, Y) be a couple of random variables on a polished space. Let $k \geq 1$ and let $X^{(1)}, \ldots, X^{(k)}$ be k i.i.d. samples (given Y) from the distribution $P(X = \cdot|Y)$, independently of every other random variable. Let us denote $\langle \cdot \rangle$ as the expectation with respect to $P(X = \cdot|Y)$ and $\mathbb{E}$ the expectation with respect to (X, Y). Then, for all continuous bounded functions f,

$$\mathbb{E}\langle f(Y, X^{(1)}, \ldots, X^{(k)})\rangle = \mathbb{E}\langle f(Y, X^{(1)}, \ldots, X^{(k-1)}, X)\rangle.$$

While the consequences of this identity are important, the proof is rather simple: It is equivalent to sampling the couple (X, Y) according to its joint distribution or to sample Y first according to its marginal distribution and then to sample X conditionally to Y from its conditional distribution $P(X = \cdot|Y)$. Thus, the $(k+1)$-tuple $(Y, X^{(1)}, \ldots, X^{(k)})$ is equal to $(Y, X^{(1)}, \ldots, X^{(k-1)}, X)$.

The proof of Theorem 1 is obtained by using the Guerra interpolation technique to prove a lower bound for the free entropy and then applying the Aizenman-Sims-Star scheme to get a matching upper bound.

b. Lower bound

Guerra interpolation. We now move to the core of the proof. The first part combines the Guerra interpolation method [68] developed for matrices in Ref. [57] and tensors in Ref. [10].

Consider the interpolating Hamiltonian depending on $t \in [0, 1]$,

$$\mathcal{H}_{N,t} = -\sum_{i<j}\left[\frac{\sqrt{t}}{\Delta_2\sqrt{N}}Y_{ij}x_ix_j + \frac{t}{2\Delta_2 N}(x_ix_j)^2\right] + -\sum_{i_1<\ldots<i_p}\left[\frac{\sqrt{t(p-1)!}}{\Delta_p N^{\frac{p-1}{2}}}T_{i_1\ldots i_p}x_{i_1}\ldots x_{i_p} + \frac{t(p-1)!}{2\Delta_p N^{p-1}}(x_{i_1}\ldots x_{i_p})^2\right] +$$
$$-\sum_j\left[\sqrt{1-t}\sqrt{\frac{m^{p-1}}{\Delta_p} + \frac{m}{\Delta_2}}W_jx_j + (1-t)\left(\frac{m^{p-1}}{\Delta_p} + \frac{m}{\Delta_2}\right)x_j^*x_j + \frac{1-t}{2}\left(\frac{m^{p-1}}{\Delta_p} + \frac{m}{\Delta_2}\right)x_j^2\right], \tag{B27}$$

where we have, for $t = 1$, the regular Hamiltonian and, for $t = 0$, the first term of Eq. (B18), where W_j are i.i.d. canonical Gaussian variables. More importantly, for all $t \in [0, 1]$, we can show that the Hamiltonian above can be seen as the one emerging for an appropriate inference problem, so the Nishimori property is kept valid for generic $t \in [0, 1]$ [57].

Given the interpolating Hamiltonian, we can write the corresponding Gibbs measure

$$P(x|W, Y, T) = \frac{1}{\mathcal{Z}_{N,t}}P_X(x)e^{\mathcal{H}_{N,t}(x)} \tag{B28}$$

and the interpolating free entropy

$$\psi_N(t) \doteq \frac{1}{N}\mathbb{E}[\log \mathcal{Z}_{N,t}], \tag{B29}$$

whose boundaries are $\psi_N(1) = (1/N)\mathcal{F}_N$ (our target) and $\psi_N(0) = (1/N)\tilde{\Phi}_{\mathrm{RS}} + (1/4\Delta_2)m^2 + (p-1/2p\Delta_p)m^p$. We then use the fundamental theorem of calculus to write

$$\mathcal{F}_N = \psi_N(1) = \psi_N(0) + \frac{1}{N}\underbrace{\mathbb{E}\int_0^1\left(-\frac{\partial\log\mathcal{Z}_{N,t}}{\partial t}\right)dt}_{\doteq\mathcal{R}}.$$
$$\tag{B30}$$

We work with the second term and use Stein's lemma, which, given a well-behaving function g, provides the useful relation for a canonical Gaussian variable Z: $\mathbb{E}_Z[Zg(Z)] = \mathbb{E}_Z[g'(Z)]$. This case yields

$$\mathcal{R} = -\mathbb{E}\int_0^1\left[\frac{1}{\mathcal{Z}_{N,t}}\int dx^N \frac{\partial\mathcal{H}_{N,t}(x)}{\partial t}P_X(x)e^{\mathcal{H}_{N,t}(x)}\right]dt = -\mathbb{E}\int_0^1\left\langle\frac{\partial\mathcal{H}_{N,t}(x)}{\partial t}\right\rangle dt$$

$$= -\mathbb{E}\int_0^1\left\langle\sum_{i<j}\frac{1}{\Delta_2 N}(x_i^*x_ix_j^*x_j) + \sum_{i_1<\ldots<i_p}\frac{(p-1)!}{\Delta_2 N^{p-1}}(x_{i_1}^*x_{i_1}\ldots x_{i_p}^*x_{i_p}) - \sum_i\left(\frac{m}{2\Delta_2} + \frac{m^{p-1}}{2\Delta_p}\right)x_i^*x_i\right\rangle dt$$

$$= \mathbb{E}\int_0^1\left[\frac{1}{4\Delta_2}\left\langle\left(\frac{x\cdot x^*}{N}\right)^2 - 2m\left(\frac{x\cdot x^*}{N}\right)\right\rangle + \frac{1}{2p\Delta_p}\left\langle\left(\frac{x\cdot x^*}{N}\right)^p - pm^{p-1}\left(\frac{x\cdot x^*}{N}\right)\right\rangle\right]dt,$$

where we have used the Nishimori property to replace terms such as $\langle x\rangle^2$ by $\langle xx^*\rangle$. At this point, we can write

$$\mathcal{R} = \mathbb{E}\int_0^1\left[\frac{1}{4\Delta_2}\left\langle\left(\frac{x\cdot x^*}{N}\right)^2 - 2m\left(\frac{x\cdot x^*}{N}\right)\right\rangle\right]dt + \mathbb{E}\int_0^1\left[\frac{1}{2p\Delta_p}\left\langle\left(\frac{x\cdot x^*}{N}\right)^p - pm^{p-1}\left(\frac{x\cdot x^*}{N}\right)\right\rangle\right]dt$$

$$= -\frac{m^2}{4\Delta_2} + \frac{1}{4\Delta_2}\mathbb{E}\int_0^1\frac{1}{4\Delta_2}\left\langle\left(\frac{x\cdot x^*}{N} - m\right)^2\right\rangle dt + \frac{1}{2p\Delta_p}\mathbb{E}\int_0^1\left\langle\left(\frac{x\cdot x^*}{N}\right)^p - pm^{p-1}\left(\frac{x\cdot x^*}{N}\right)\right\rangle dt. \tag{B31}$$

The first integral is clearly positive. However, the second one is more difficult to estimate. We may, however, use a simple convexity argument on the function $f(x) = x^k$. Indeed, observe that $\forall\, a, b \geq 0$ and $p \geq 1$: $a^p - pb^{p-1}a \geq (1-p)b^p$. We would like to use this property, but there is the subtlety that we need $x \cdot x^*$ to be non-negative. To bypass this problem, we can add again a small perturbation that forces $x \cdot x^*$ to concentrate around a non-negative value, without affecting the "interpolating free entropy" $\psi_N(t)$ in the $N \to \infty$ limit. This case is, again, the argument used in Ref. [10] and originally in Ref. [56]. In this way, we can write

$$\mathcal{R} \geq -\frac{m^2}{4\Delta_2} + \mathbb{E}\int_0^1\left[\frac{1}{4\Delta_2}\left\langle\left(\frac{x\cdot x^*}{N}\right)^2 - 2m\left(\frac{x\cdot x^*}{N}\right)\right\rangle\right]dt$$
$$+ \frac{(1-p)m^p}{4\Delta_2}$$
$$\geq -\frac{m^2}{4\Delta_2} - \frac{(p-1)m^p}{4\Delta_2}. \tag{B32}$$

This concludes the proof and yields the lower bound

$$\mathcal{F}_N \geq \psi_N(0) - \frac{1}{4\Delta_2}m^2 - \frac{p-1}{2p\Delta_p}m^p = \frac{1}{N}\tilde{\Phi}_{\mathrm{RS}}(m), \tag{B33}$$

so for all $m \geq 0$,

$$\lim_{N\to\infty}\inf\mathcal{F}_N = \lim_{N\to\infty}\inf\psi_N(1)$$
$$= \lim_{N\to\infty}\inf\left[\psi_N(0) + \int_0^1\psi_N'(t)dt\right] \geq \tilde{\Phi}_{\mathrm{RS}}(m).$$

c. Upper bound: Aizenman-Sims-Starr scheme

The matching upper bound is obtained using the Aizenman-Sims-Starr scheme [58]. This scheme is a particularly effective tool that has already been used for these problems; see, e.g., Refs. [10,38,66]. The method goes as follows. Consider the original system with N variables, $\mathcal{H}_N$, and add a new variable x_0 so that we get a Hamiltonian $\mathcal{H}_{N+1}$. Define the Gibbs measures of the two systems, the first with N variables and the second with $N + 1$ variables, and consider the two relative free entropies. Call $A_N = \mathbb{E}[\log \mathcal{Z}_{N+1}] - \mathbb{E}[\log \mathcal{Z}_N]$ their difference. First, we notice that we have $\limsup_N \mathcal{F}_N \leq \limsup_N A_N$ because

$$\mathcal{F}_N = \mathbb{E}\frac{1}{N}\log Z_N = \frac{1}{N}\mathbb{E}\log\left(\frac{Z_N}{Z_{N-1}}\frac{Z_{N-1}}{Z_{N-2}}\cdots\frac{Z_1}{Z_0}\right)$$
$$= \frac{1}{N}\sum_i A_i \leq \sup_i A_i.$$

Moreover, we can separate the contribution of the additional variable in the Hamiltonian $\mathcal{H}_{N+1}$ so that $\mathcal{H}_{N+1} = \tilde{\mathcal{H}}_N + x_0 z(x) + x_0^2 s(x)$, with $x = (x_1, \ldots, x_N)$, and

$$z(x) = \frac{1}{\sqrt{\Delta_2(N+1)}}\sum_{i=1}^N Z_{0i}x_i + \frac{\sqrt{(p-1)!}}{\sqrt{\Delta_p(N+1)^{(p-1)/2}}}\sum_{1\leq i_1<\ldots<i_{p-1}\leq N} Z_{0i_1\ldots i_{p-1}}x_{i_1}\ldots x_{i_{p-1}}$$

$$+ \frac{1}{\Delta_2(N+1)}\sum_{i=1}^N x_0^*x_i^*x_i + \frac{(p-1)!}{\Delta_p(N+1)^{p-1}}\sum_{1\leq i_1<\ldots<i_{p-1}\leq N} x_0^*x_{i_1}^*x_{i_1}\ldots x_{i_{p-1}}^*x_{i_{p-1}},$$

$$s(x) = -\frac{1}{2\Delta_2(N+1)}\sum_{i=1}^N x_i^2 - \frac{(p-1)!}{2\Delta_p(N+1)^{p-1}}\sum_{1\leq i_1<\ldots<i_{p-1}\leq N}(x_{i_1}\ldots x_{i_{p-1}})^2,$$

and $\mathcal{H}_{N+1}$ is the same expression as Eq. (A6), where the N in the denominators are replaced by $N + 1$. We also rewrite $\mathcal{H}_N(x)$ as a perturbation of $\tilde{\mathcal{H}}_N$: $\mathcal{H}_N(x) = \tilde{\mathcal{H}}_N(x) + y(x) + O(1)$, with

$$y(x) = \frac{1}{\sqrt{\Delta_2 N}}\sum_{i<j}V_{ij}x_ix_j + \sqrt{p-1}\frac{\sqrt{(p-1)!}}{\sqrt{\Delta_p N^{p/2}}}\sum_{i_1<\ldots<i_p}V_{i_1\ldots i_p}x_{i_1}\ldots x_{i_p} + \frac{1}{N^2}\sum_{i<j}\left[x_i^*x_ix_j^*x_j - \frac{1}{2}(x_ix_j)^2\right]$$

$$+ (p-1)!\frac{p-1}{N^p}\sum_{i_1<\ldots<i_p}\left[x_{i_1}^*x_{i_1}\ldots x_{i_p}^*x_{i_p} - \frac{1}{2}(x_{i_1}\ldots x_{i_p})^2\right],$$

where the Zs and the Vs are standard Gaussian random variables.

Finally, we observe that the partition functions Z_N can be interpreted as ensemble averages with respect to $\tilde{\mathcal{H}}_N$. Thus, $A_N = \mathbb{E} \log \langle \int P_X(x_0) e^{x_0 z(x) + x_0^2 s(x)} dx_0 \rangle_{\tilde{\mathcal{H}}_N} - \mathbb{E} \log \langle e^{y(x)} \rangle_{\tilde{\mathcal{H}}_N}$. Now, using the Nishimori property and the concentration of the overlap around a non-negative value—which we denote $m(Y, T)$ since it depends explicitly on the disorder—it yields (see Ref. [38], Sec. IV.3 for details) Eq. (B18) in the thermodynamic limit, with $m(Y, T)$ instead of m. From this result, we can now obtain the upper bound that concludes the proof:

$$\lim_N \sup \mathcal{F}_N \leq \lim_N \sup A_N \leq \lim_N \sup \mathbb{E}_{Y,T} \tilde{\Phi}_{RS}[m(Y, T)]$$

$$\leq \lim_N \sup \sup_m \Phi_{RS}(m) \leq \tilde{\Phi}_{RS}. \qquad (B34)$$

3. State evolution of AMP and its analysis

The dynamical evolution of the AMP algorithm in the large-N limit is described by the so-called SE equations. The derivation of these equations can be straightforwardly done using the same techniques as developed in Ref. [37]. They can be written in terms of two dynamical order parameters, namely, $m^t = \sum_i \hat{x}_i^t x_i^* / N$, encoding for the alignment of the current estimation $\hat{x}_i^t$ of the components of the signal with the signal itself at time t and $q^t = \sum_i \hat{x}_i^t \hat{x}_i^t / N$. Keeping the spherical constraint in mind, we obtain the following SE equations:

$$\frac{m^{t+1}}{1 - q^{t+1}} = \frac{m^t}{\Delta_2} + \frac{(m^t)^{p-1}}{\Delta_p}, \qquad (B35)$$

$$\frac{q^{t+1}}{(1 - q^{t+1})^2} = \left[\frac{m^t}{\Delta_2} + \frac{(m^t)^{p-1}}{\Delta_p}\right]^2 + \left[\frac{q^t}{\Delta_2} + \frac{(q^t)^{p-1}}{\Delta_p}\right]. \qquad (B36)$$

Note that Eq. (B35) at fixed values of q describes the evolution of the parameter m, which is why we use it in the main text to derive the Langevin threshold equation (15). Finally, using the Nishimori symmetry, it can be shown that $m^t = q^t$ at all times (see, e.g., Ref. [6]), and therefore, the evolution of the algorithm is characterized by a single order parameter m^t whose dynamical evolution is given by

$$m^{t+1} = 1 - \frac{1}{1 + \frac{m^t}{\Delta_2} + \frac{(m^t)^{p-1}}{\Delta_p}}. \qquad (B37)$$

Note that AMP satisfies the Nishimori property at all times, while this condition is violated by Langevin dynamics. In that case, the Nishimori symmetry is recovered only when equilibrium is reached, and therefore it is violated when the Langevin algorithm gets trapped in the glass phase (see Appendix E). If we initialize the configuration of the estimator $\hat{x}$ at random, the initial value of m will be equal

to zero on average. However, finite-size fluctuations produce, by chance, a small bias towards the signal, and therefore we consider the initialization to be $m^{t=0} = \epsilon$, with ϵ an arbitrarily small positive number. We call m_{AMP} the fixed point of Eq. (B37), which is reached from this infinitesimal initialization. The MSE reached by AMP after convergence is then given by $MSE_{AMP} = 1 - m_{AMP}$.

We underline that Eq. (B37) can be proven rigorously following Refs. [18,40]. Finally, we note that the fixed point of the SE satisfies the very same Eq. (B20) that gives the replica free entropy. In the rest of this section, we study the fixed points of Eq. (B37). This method will allow us to determine the phase diagram of the spiked matrix-tensor model.

We start by observing that $m = 0$ is a fixed point of Eq. (B37). However, in order to understand whether it is a possible attractor of the AMP dynamics, we need to understand its local stability, which can be obtained perturbatively by expanding Eq. (B37) around $m = 0$,

$$m^{t+1} = \frac{m^t}{\Delta_2} + \left(\frac{m^t}{\Delta_2}\right)^2 - \frac{(m^t)^{p-1}}{\Delta_p} + O((m^t)^3). \qquad (B38)$$

It is clear that the noninformative fixed point $m = 0$ is stable as long as $\Delta_2 > 1$. We call $\Delta_2 = 1$ the stability threshold.

When $p = 3$, the SE equations are particularly simple, and the fixed points are written explicitly as

$$m_0 = 0; \quad m_\pm = \frac{1}{2}\left[1 - \frac{\Delta_3}{\Delta_2} \pm \sqrt{\left(1 + \frac{\Delta_3}{\Delta_2}\right)^2 - 4\Delta_3}\right]. \qquad (B39)$$

In the regime where $\Delta_2 > 1$, m_0 and m_+ are stable, while m_- is unstable. When Δ_2 becomes smaller than 1, m_+ becomes the only non-negative stable solution, and therefore $\Delta_2 = 1$ is also known as the algorithmic spinodal since it corresponds to the point where the AMP algorithm converges to the informative fixed point. The informative solution m_+ exists as long as $\Delta_2 \leq \Delta_2^{dyn}$, where we have defined the dynamical spinodal by

$$\Delta_2^{dyn} = \frac{\Delta_3}{2\sqrt{\Delta_3} - 1}. \qquad (B40)$$

For a generic p, we cannot determine the values of the informative fixed points explicitly, but we can easily study Eq. (B37) numerically to get the full phase diagram.

Furthermore, we can obtain the spinodal transition lines as follows. The key observation is that the two spinodals are critical points of the equation $\Delta_p(m; \Delta_2)$, where Δ_2 is fixed, or analogously $\Delta_2(m; \Delta_p)$, where Δ_p is fixed (to have
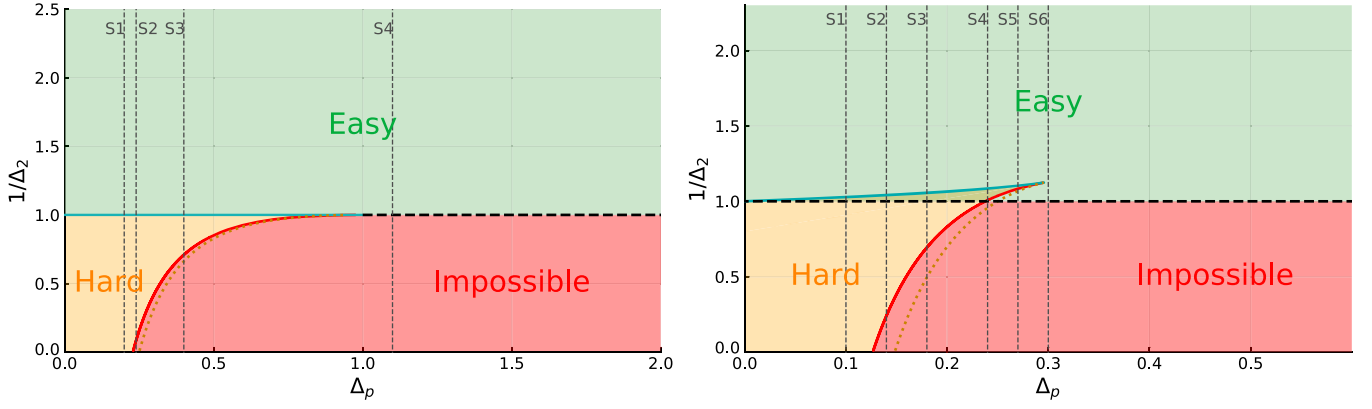
FIG. 7.   Left panel: Phase diagram of the spiked matrix-tensor model for $p = 3$. The phase diagram identifies four regions: easy (green), impossible (red), and hard (orange). The lines correspond to different phase transitions, namely, the stability threshold (dashed black), the information-theoretic threshold (solid red), the algorithmic threshold (solid cyan), and the dynamical threshold (dotted orange). The vertical cuts represent the section along which the magnetization is plotted in Fig. 9. Right panel: Phase diagram of the spiked matrix-tensor model for $p = 4$. The main difference with respect to the case $p = 3$ is that the algorithmic spinodal (solid cyan) is strictly above the stability threshold (dashed black). The hybrid-hard phase appears between these two lines (combined green and orange color). The vertical cuts represent the section along which the magnetization is plotted in Fig. 10.
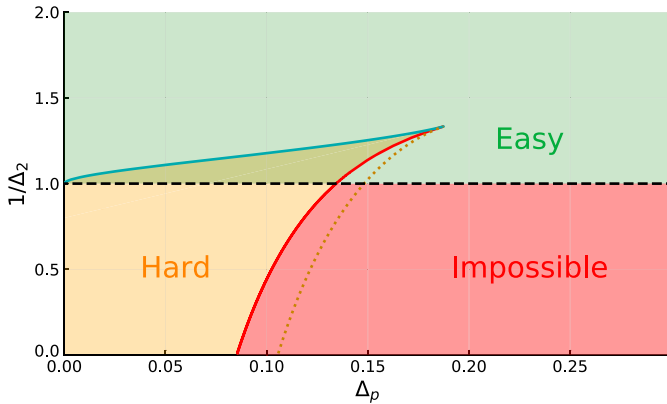
a pictorial representation of the idea, see Fig. 10). We find $x = m/\Delta_2 + m^{p-1}/\Delta_p$, and $f_{SE}(x) \equiv 1 - (1/1 + x)$; then,

$$\Delta_p \equiv \Delta_p(x; \Delta_2) = \frac{(f_{SE}(x))^{p-1}}{x - \frac{f_{SE}(x)}{\Delta_2}}. \quad (B41)$$

The stationary points are implicitly defined by

$$0 = \frac{d \log \Delta_p}{dm} = \frac{\partial \log \Delta_p}{\partial x}(1+x)^2$$

$$\propto (p-1)\frac{f'_{SE}(x)}{f_{SE}(x)} - \frac{1 - \frac{f'_{SE}(x)}{\Delta_2}}{x - \frac{f_{SE}(x)}{\Delta_2}} = \frac{\frac{2-p}{\Delta_2} + (1+x)(p-x-2)}{x(1+x)[x+1-\frac{1}{\Delta_p}]},$$

giving

$$x_\pm(\Delta_2) = \frac{1}{2}\left[p - 3 \pm \sqrt{(p-1)^2 - \frac{4}{\Delta_2}(p-2)}\right]. \quad (B42)$$

Finally, $\Delta_p(x_\pm(\Delta_2); \Delta_2)$ describes the two spinodals. We can also derive the tricritical point, when the two spinodals meet, which is given by the zero discriminant condition on (B42),

$$(\Delta_p^{tri}; 1/\Delta_2^{tri}) = \left(\frac{4(p-2)(\frac{p-3}{p-1})^{p-1}}{(p-3)^2}; \frac{(p-1)^2}{4(p-2)}\right). \quad (B43)$$

4. Phase diagrams of spiked matrix-tensor model

In this section, we present the phase diagrams for the spiked matrix-tensor model as a function of the two noise
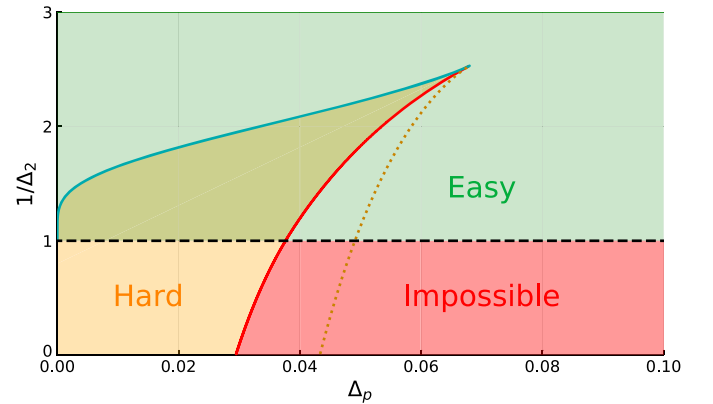

FIG. 8.   Left panel: Phase diagram of the spiked matrix-tensor model for $p = 5$. Right panel: Phase diagram of the spiked matrix-tensor model for $p = 10$. In both cases, we observe qualitatively the same scenario found in the right panel of Fig. 7.
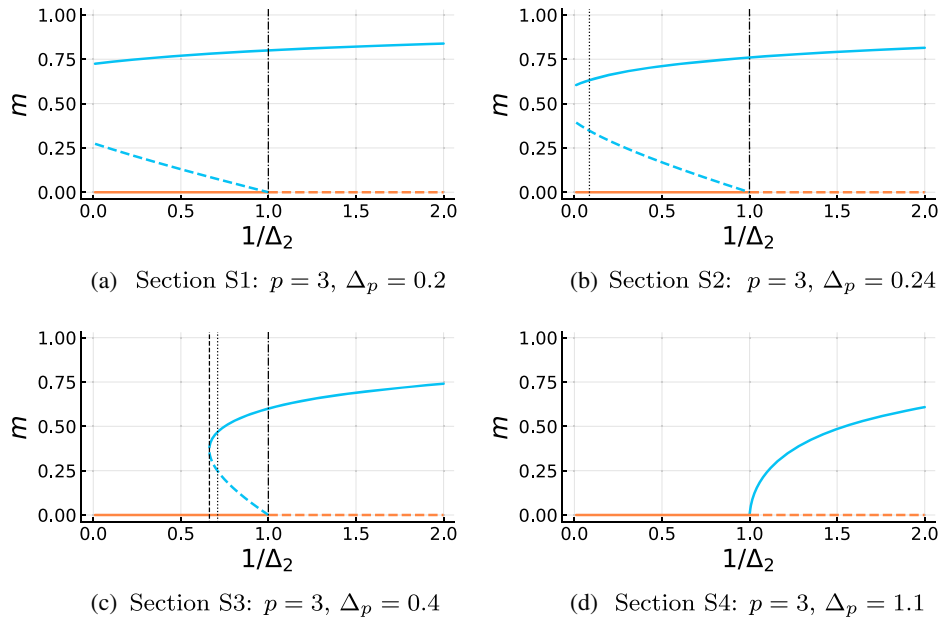
FIG. 9. Fixed points of Eq. (B37) as a function of Δ_2 for $p = 3$ and several fixed values of Δ_p. The values of Δ_p correspond to the vertical cuts in the left panel of Fig. 7. Solid lines are stable fixed points; dashed lines are unstable fixed points. The blue line represents informative fixed points with positive overlap with the signal, while the orange line represents uninformative fixed points with no overlap with the signal. Starting from high Δ_2, an informative fixed point appears at the dynamical threshold (vertical dashed line) but is energetically disfavored until it reaches the information-theoretic threshold (vertical dotted line); finally, it becomes the only stable solution crossing the algorithmic threshold (vertical dotted-dashed line). When the transition is continuous, the three vertical threshold lines merge, and we have a single second-order phase transition, which occurs here at $\Delta_p \geq 1$.
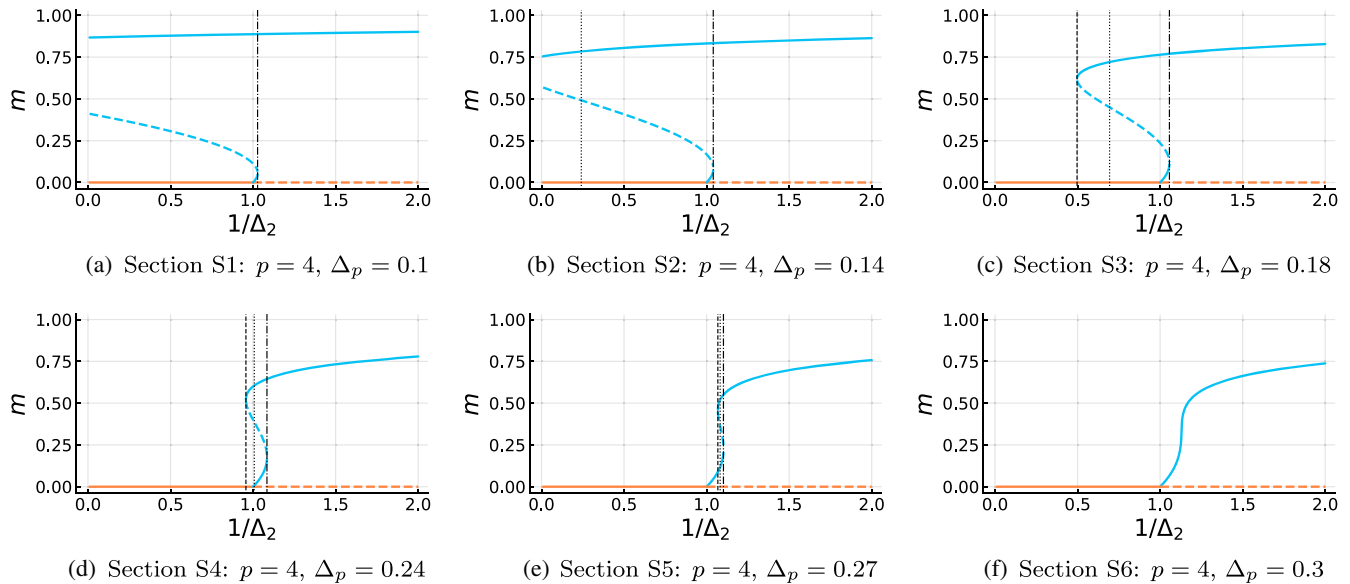

FIG. 10. Fixed points, Eq. (B37), as a function of Δ_2 for $p = 4$ and several fixed values of Δ_p. The values of Δ_p correspond to the vertical cuts of the right panel of Fig. 7. The situation is qualitatively similar to Fig. 9, the difference being only the presence of the hybrid-hard phase. We can observe that when the transition is discontinuous, from (a) to (e), for $1/\Delta_2 > 1.0$, the uninformative solution becomes unstable and continuously goes to a stable-informative solution, which is not the optimal one.

levels Δ_2 and Δ_p and for several values of p. These phase diagrams are plotted in Figs. 7 and 8.

Generically, we can have four regions:

(i) Easy phase (green), where the MSE obtained through AMP coincides with the MMSE, which is better than random sampling of the prior.

(ii) Impossible phase (red), where the MMSE and MSE of AMP coincide and are equal to 1 (meaning that $m^* = m_{\mathrm{AMP}} = 0$).

(iii) Hard phase (orange), where the MMSE is smaller than the MSE obtained from AMP and $m^* > m_{\mathrm{AMP}} \geq 0$.

(iv) Hybrid-hard phase [69] (mix of green and orange), which is part of the hard phase where the AMP performance is strictly better than random sampling from the prior but the MSE obtained this way still does not match the MMSE, i.e., $m^* > m_{\mathrm{AMP}} > 0$. The hybrid-hard phase can be found for $p \geq 4$.

All of these phases are separated by the following transition lines:

(i) The stability threshold (dashed black line) at $\Delta_2 = 1$ for all p. This threshold corresponds to the point where the uninformative fixed point $m = 0$ loses its local stability.

(ii) The information-theoretic threshold (solid red line). Here, $m^* > 0$, and the MMSE jumps to a value strictly smaller than 1.

(iii) The algorithmic threshold (solid cyan line). This threshold is where the fixed point of AMP jumps to the MMSE < 1. For $p = 3$, this line coincides with a segment of the stability threshold, while for $p \geq 4$, it is strictly above the threshold.

(iv) The dynamic threshold (dotted orange line). Here, the most informative fixed point (the one with the largest m_{AMP}) disappears.

In Figs. 9 and 10, we plot the evolution of the magnetization m, as found through the fixed points of the SE equation, for several fixed values of Δ_p and $p = 3$ and $p = 4$, respectively. The values of Δ_p are identified by the vertical cuts in the phase diagrams of Fig. 7.

APPENDIX C: LANGEVIN ALGORITHM AND ITS STATE EVOLUTION

The main goal of our analysis is to compare AMP with the performance of the Langevin dynamics. The advantage of the spiked matrix-tensor model is that, in this case, the Langevin dynamics can be studied in the large-N limit through integrodifferential equations for the correlation function $C(t, t') = \lim_{N \to \infty} \sum_i \langle x_i(t) x_i(t') \rangle / N$, the response function $R(t, t') = \lim_{N \to \infty} (1/N) \sum_i \frac{d\langle x_i(t) \rangle}{d\eta_i(t')}$, and the magnetization $\bar{C}(t) = \lim_{N \to \infty} \sum_i \langle x_i(t) x_i^* \rangle / N$.

To obtain these equations, we use the techniques developed in the context of mean-field spin-glass systems [44,70]. We call $\eta_i(t)$ a time-dependent noise, and we indicate by $\langle \cdot \rangle$ the average with respect to it. The noise is Gaussian and characterized by $\langle \eta_i(t) \rangle = 0$ for all t and $i = 1, \ldots, N$, and $\langle \eta_i(t) \eta_j(t') \rangle = 2\delta_{ij}\delta(t - t')$. As before, we denote by $\mathbb{E}[\ldots]$ the average with respect to the realization of disorder, which, in this case, goes back to the specific realization of the signal.

Before proceeding, it is useful to introduce a set of auxiliary variables that will help in the following. For $k \in \{2, p\}$, we define $r_k \equiv r_k(t) = 2/(kT_k(t)\Delta_k)$, $f_k(x) = x^k/2$, and $m(t) \doteq (1/N) \sum_i x_i(t) x_i^*$, and the random variable $\tilde{\xi}_{i_1 \ldots i_k} \equiv (1/\Delta_k)\xi_{i_1 \ldots i_k} \sim \mathcal{N}(0, 1/\Delta_k)$. We employ the time dependence in T_k in the tensor-annealing protocol, which will be used to avoid part of the Langevin hard phase. We introduce a time-dependent Hamiltonian

$$
\begin{aligned}
\mathcal{H}(t) = & -\frac{1}{T_2(t)\sqrt{N}} \sum_{i<j} \tilde{\xi}_{ij} x_i(t) x_j(t) \\
& - \frac{\sqrt{(p-1)!}}{T_p(t)N^{\frac{p-1}{2}}} \sum_{i_1 < \ldots < i_p} \tilde{\xi}_{i_1 \ldots i_p} x_{i_1}(t) \ldots x_{i_p}(t) \\
& - Nr_2(t)f_2(m(t)) - Nr_p(t)f_p(m(t)),
\end{aligned}
$$

and the associated Langevin dynamics

$$
\begin{aligned}
\dot{x}_i(t) = & -\mu(t)x_i(t) - \frac{\partial \mathcal{H}}{\partial x_i}(t) - \eta_i(t) = -\mu(t)x_i(t) - \frac{1}{T_2(t)\sqrt{N}} \sum_{j(\neq i)} \tilde{\xi}_{ij} x_j(t) \\
& + r_2(t)f_2'(m(t)) - \frac{\sqrt{(p-1)!}}{T_p(t)N^{\frac{p-1}{2}}} \sum_{(i,i_1,\ldots,i_{p-1})\backslash i} \tilde{\xi}_{ii_1 \ldots i_{p-1}} x_{i_1}(t) \cdots x_{i_{p-1}}(t) + r_p(t)f_p'(m(t)) - \eta_i(t),
\end{aligned} \tag{C1}
$$

with μ a Langrange multiplier that enforces the spherical constraint $\sum_{i=1}^{N} x_i^2(t) = N$. If $T_k(t) = 1$ for all $k = 2, p$, the stationary equilibrium distribution for the Langevin dynamics is given by the posterior measure. Using Ito's lemma, one finds

$$
\frac{1}{N}\frac{d}{dt}\sum_i x_i^2(t) = \frac{2}{N}\sum_i x_i(t)\dot{x}_i(t) + 2.
$$

Since the spherical constraint imposes the left-hand side to be zero, one obtains a condition on the right-hand side. By plugging in the expression (C1), one gets that, in the large-N limit,

$$\mu(t) = 1 - 2\mathcal{H}_2(t) - p\mathcal{H}_p(t), \tag{C2}$$

where

$$\mathcal{H}_k = -\frac{\sqrt{(k-1)!}}{T_k(t)N^{\frac{k-1}{2}}} \sum_{i_1 < ... < i_k} \tilde{\xi}_{i_1...i_k} x_{i_1}(t)...x_{i_k}(t)$$
$$- N r_k(t) f_k(m(t)), \qquad k = 2, p, \tag{C3}$$

are the parts of the Hamiltonian defined in Eq. (C) relative to the matrix ($k = 2$) and to the tensor ($k = p$).

Note that we have not specified any initial condition for the variables $x_i(t = 0)$. Therefore, since we always employ the spherical constraint, the initial condition for the dynamics is a point on the N-dimensional hypersphere $|x|^2 = N$ extracted with the flat measure.

In order to analyze the Langevin dynamics in the large-N limit, we use the dynamical cavity method [44,71,72]. We consider a system of N variables, with $N \gg 1$, and add a new one. This new variable will be considered as a small

perturbation to the original system, but at the same time, it will be treated self-consistently.

1. Dynamical mean-field equations

In the following, we drop the time dependence for simplicity, restoring it only when it is needed. Given the system with N variables, $i = 1, ..., N$, we add a new one, say, $i = 0$, and define $\tilde{m} = (1/N + 1) \sum_{i=0}^{N} x_i x_i^* \simeq (1/N) \sum_{i=0}^{N} x_i x_i^*$ (henceforth, we use the symbol $\simeq$ to denote two quantities that are equal up to terms that vanish in the large-N limit). The Langevin equation associated with the new variable is

$$\dot{x}_0 = -\mu x_0 - \frac{1}{T_2(t)\sqrt{N}} \sum_{j(\neq 0)} \tilde{\xi}_{0j} x_j + r_2 f_2'(\tilde{m})$$
$$- \frac{\sqrt{(p-1)!}}{T_p(t)N^{\frac{p-1}{2}}} \sum_{(0,i_1,...,i_{p-1})\backslash 0} \tilde{\xi}_{0i_1...i_{p-1}} x_{i_1} \cdots x_{i_{p-1}}$$
$$+ r_p f_p'(\tilde{m}) - \eta_0, \tag{C4}$$

where we use that $N \simeq N + 1$ for $N \gg 1$. We consider the contribution of the new variable on the others in perturbation theory. In the dynamical equations for the variables $i = 1, ..., N$, we can isolate the variable $i = 0$ and write

$$\dot{x}_i = -\mu x_i - \frac{1}{T_2(t)\sqrt{N}} \sum_{j(\neq i,0)} \tilde{\xi}_{ij} x_j + r_2 f_2'(m) - \frac{\sqrt{(p-1)!}}{T_p(t)N^{\frac{p-1}{2}}} \sum_{(i,i_1,...,i_{p-1})\backslash i,0} \tilde{\xi}_{ii_1...i_{p-1}} x_{i_1} \cdots x_{i_{p-1}} + r_p f_p'(m) - \eta_i + H_i, \tag{C5}$$

with

$$H_i(t) = (r_2 f_2''(m) + r_p f_p''(m)) \frac{1}{N} x_0 - \frac{1}{T_2(t)\sqrt{N}} \tilde{\xi}_{0i} x_0 - \frac{\sqrt{(p-1)!}}{T_p(t)N^{\frac{p-1}{2}}} \sum_{(i,0,i_1,...,i_{p-2})\backslash i,0} \tilde{\xi}_{i0i_1...i_{p-2}} x_0 x_{i_1} \cdots x_{i_{p-2}}. \tag{C6}$$

Consider the unperturbed variables $x_i^0 = x_i|_{H_i=0}$. At leading order in N, we can write

$$x_i \simeq x_i^0 + \int_{t_o}^{t} dt' \frac{\delta x_i(t)}{\delta H_i(t')} \bigg|_{H_i=0} H_i(t'). \tag{C7}$$

In the dynamical equation for the variable 0, we can identify a piece associated with the unperturbed variables x_i^0. This term can be thought of collectively as a stochastic term $\Xi(t)$,

$$\dot{x}_0 = -\mu x_0 \overbrace{- \frac{1}{T_2(t)\sqrt{N}} \sum_{j(\neq 0)} \tilde{\xi}_{0j} x_j^0 - \frac{\sqrt{(p-1)!}}{T_p(t)N^{\frac{p-1}{2}}} \sum_{(0,i_1,...,i_{p-1})\backslash 0} \tilde{\xi}_{0i_1...i_{p-1}} x_{i_1}^0 \cdots x_{i_{p-1}}^0 - \eta_0}^{\doteq \Xi(t)}$$
$$+ r_2 f_2'(m) + r_p f_p'(m) + (r_2 f_2''(m) + r_p f_p''(m)) \frac{1}{N} x_0 - \frac{1}{T_2(t)\sqrt{N}} \sum_{j(\neq 0)} \tilde{\xi}_{0j} \int_{t_o}^{t} dt' \frac{\delta x_j(t)}{\delta H_j(t')} \bigg|_{H_j=0} H_j(t') +$$
$$- \left[\frac{\sqrt{(p-1)!}}{T_p(t)N^{\frac{p-1}{2}}} \sum_{(0,i_1,...,i_{p-1})\backslash 0} \tilde{\xi}_{0i_1...i_{p-1}} \int_{t_o}^{t} dt' \frac{\delta x_{i_1}(t)}{\delta H_{i_1}(t')} \bigg|_{H_{i_1}=0} H_{i_1}(t') x_{i_2}^0 \cdots x_{i_{p-1}}^0 + \text{permutations} \right]. \tag{C8}$$

Indeed, $\Xi(t)$ encodes the effect of a kind of bath made of the unperturbed variables $i = 1, \ldots, N$ on the new one. We can show that at leading order in N, $\Xi(t)$ is a Gaussian noise with zero mean and variance given by

$$\mathbb{E}\langle \Xi(t)\Xi(t')\rangle = 2\delta(t - t') - \mathbb{E}\left[\frac{1}{T_2(t)T_2(t')N}\sum_{j(\neq 0)}\sum_{l(\neq 0)}\tilde{\xi}_{0j}\tilde{\xi}_{0l}x_j^0(t)x_l^0(t')\right] +$$

$$- \mathbb{E}\left[\frac{(p-1)!}{T_p(t)T_p(t')N^{p-1}}\sum_{(0,i_1,\ldots,i_{p-1})\backslash 0}\sum_{(0,j_1,\ldots,j_{p-1})\backslash 0}\tilde{\xi}_{0i_1\ldots i_{p-1}}\tilde{\xi}_{0j_1\ldots j_{p-1}}x_{i_1}^0\cdots x_{i_{p-1}}^0 x_{j_1}^0\cdots x_{j_{p-1}}^0\right],$$

and the second term can be simplified as

$$\mathbb{E}\left[\frac{(p-1)!}{T_p(t)T_p(t')N^{p-1}}\sum_{(0,i_1,\ldots,i_{p-1})\backslash 0}\sum_{(0,j_1,\ldots,j_{p-1})\backslash 0}\tilde{\xi}_{0i_1\ldots i_{p-1}}\tilde{\xi}_{0j_1\ldots j_{p-1}}x_{i_1}^0\cdots x_{i_{p-1}}^0 x_{j_1}^0\cdots x_{j_{p-1}}^0\right] =$$

$$\simeq \frac{(p-1)!}{N^{p-1}}\frac{1}{T_p(t)T_p(t')\Delta_p}\sum_{(0,i_1,\ldots,i_{p-1})\backslash 0}\langle x_{i_1}^0(t)x_{i_1}^0(t')\cdots x_{i_{p-1}}^0(t)x_{i_{p-1}}^0(t')\rangle = \frac{1}{T_p(t)T_p(t')\Delta_p}C^{p-1}(t,t'),$$

where we used $\sum_{(i_1,\ldots,i_k)} = (1/k!)\sum_{1\leq i_1,\ldots,i_k\leq N}$, neglected terms subleading in N, and used the definition of the dynamical correlation function

$$C(t, t') = \frac{1}{N}\sum_{i=1}^{N}\langle x_i(t)x_i(t')\rangle.$$

Therefore, we have

$$\mathbb{E}\langle\Xi(t)\rangle = 0; \tag{C9}$$

$$\mathbb{E}\langle\Xi(t)\Xi(t')\rangle = 2\delta(t - t') + \frac{1}{T_2(t)T_2(t')}C(t,t') + \frac{1}{T_p(t)T_p(t')\Delta_p}C^{p-1}(t,t'). \tag{C10}$$

Now, we can focus of the deterministic term coming from the first-order perturbation in Eq. (C8). Consider just the integral for the p-body term; the other will be given by setting $p = 2$,

$$\frac{\sqrt{(p-1)!}}{T_p(t)N^{\frac{p-1}{2}}}\sum_{(0,i_1,\ldots,i_{p-1})\backslash 0}\tilde{\xi}_{0i_1\ldots i_{p-1}}\int_{t_o}^{t}dt'\left.\frac{\delta x_{i_1}(t)}{\delta H_{i_1}(t')}\right|_{H_{i_1}=0}H_{i_1}(t')x_{i_2}^0\cdots x_{i_{p-1}}^0 + \text{permutations} =$$

$$\simeq \frac{(p-1)!}{T_p(t)N^{p-1}}\sum_{(0,i_1,\ldots,i_{p-1})\backslash 0}\tilde{\xi}_{0i_1\ldots i_{p-1}}^2\int_{t_o}^{t}dt'\frac{1}{T_p(t')}\left.\frac{\delta x_{i_1}(t)}{\delta H_{i_1}(t')}\right|_{H_{i_1}=0}x_{i_1}^0(t)x_{i_1}^0(t')\cdots x_{i_{p-2}}^0(t)x_{i_{p-2}}^0(t')x_0(t')$$

$$+ \text{permutations} \simeq -\frac{p-1}{T_p(t)\Delta_p}\int_{t_o}^{t}dt'\frac{1}{T_p(t')}R(t,t')C^{p-2}(t,t')x_0(t'), \tag{C11}$$

where we have used the definition of the response function

$$R(t, t') = \frac{1}{N}\sum_{i=1}^{N}\left\langle\frac{\delta x_i(t)}{\delta H_i(t')}\right\rangle.$$

Plugging Eq. (C11) into Eq. (C8), we obtain an effective dynamical equation for the new variable in terms of the correlation and response function of the system with N variables,

$$\dot{x}_0(t) = -\mu(t)x_0(t) + \Xi(t) + r_pf_p'(\bar{C}(t)) + r_2f_2'(\bar{C}(t)) + \frac{p-1}{T_p(t)\Delta_p}\int_{t_o}^{t}dt''\frac{1}{T_p(t'')}R(t,t'')C^{p-2}(t,t'')x_0(t'')$$

$$+ \frac{1}{T_2(t)\Delta_2}\int_{t_o}^{t}dt''\frac{1}{T_p(t'')}R(t,t'')x_0(t''). \tag{C12}$$

In order to solve Eq. (C12), we need to give the recipe to compute the correlation and response function.

2. Integrodifferential equations

In order to obtain the final equations for dynamical order parameters, we assume that the new variable x_0 is a typical one; namely, it has the same statistical nature as all of the others. Therefore, we can assume that

$$C(t, t') \doteq \mathbb{E}\langle x_0(t) x_0(t') \rangle,$$

$$R(t, t') \doteq \mathbb{E}\left\langle \frac{\delta x_0(t)}{\delta \Xi(t')} \right\rangle,$$

$$\bar{C}(t) \doteq \mathbb{E}\langle x_0(t) x_0^* \rangle. \tag{C13}$$

Equations (C13) give us a way to obtain the equation for all the correlation functions. Indeed, we can consider Eq. (C12), multiply it by $x_0(t')$, differentiate it with respect to an external field $h_0(t')$, or multiply it by x_0^*, and we can average the results over the disorder and thermal noise. Using the following identity,

$$\mathbb{E}\langle \Xi(t) x_0(t') \rangle = \int \mathcal{D}\Xi(t) \Xi(t) x_0(t') e^{-\int d\tilde{t} d\tilde{\tilde{t}} \Xi(\tilde{t}) \mathbb{K}^{-1}(\tilde{t}, \tilde{\tilde{t}}) \Xi(\tilde{\tilde{t}})}$$

$$= -\int dt'' \int \mathcal{D}\Xi(t) x_0(t') \frac{\delta}{\delta \Xi(t'')} e^{-\int d\tilde{t} d\tilde{\tilde{t}} \Xi(\tilde{t}) \mathbb{K}^{-1}(\tilde{t}, \tilde{\tilde{t}}) \Xi(\tilde{\tilde{t}})} \mathbb{K}(t, t'')$$

$$= \int dt'' \mathbb{E}\left\langle \frac{\delta x_0(t')}{\delta \Xi(t'')} \mathbb{K}(t, t'') \right\rangle = \int dt'' R(t', t'') \mathbb{K}(t, t'')$$

$$= 2R(t', t) + \frac{1}{T_p(t)\Delta_p} \int_{t_o}^{t'} dt'' \frac{1}{T_p(t'')} R(t', t'') C^{p-1}(t, t'') + \frac{1}{T_2(t)\Delta_2} \int_{t_o}^{t'} dt'' \frac{1}{T_2(t'')} R(t', t'') C(t, t''), \tag{C14}$$

we get the following LSE equations:

$$\frac{\partial}{\partial t} C(t, t') = \mathbb{E}\langle \dot{x}_0(t) x_0(t') \rangle = 2R(t', t) - \mu(t) C(t, t') + r_p(t) f'_p(\bar{C}(t)) \bar{C}(t') + r_2(t) f'_2(\bar{C}(t)) \bar{C}(t')$$

$$+ (p - 1) \frac{1}{T_p(t)\Delta_p} \int_{t_o}^{t} dt'' \frac{1}{T_p(t'')} R(t, t'') C^{p-2}(t, t'') C(t', t'') + \frac{1}{T_p(t)\Delta_p} \int_{t_o}^{t'} dt'' \frac{1}{T_p(t'')} R(t', t'') C^{p-1}(t, t'')$$

$$+ \frac{1}{T_2(t)\Delta_2} \int_{t_o}^{t} dt'' \frac{1}{T_2(t'')} R(t, t'') C(t', t'') + \frac{1}{T_2(t)\Delta_2} \int_{t_o}^{t'} dt'' \frac{1}{T_2(t'')} R(t', t'') C(t, t''); \tag{C15}$$

$$\frac{\partial}{\partial t} R(t, t') = \mathbb{E}\left\langle \frac{\delta \dot{x}_0(t)}{\delta \Xi(t')} \right\rangle$$

$$= \delta(t - t') - \mu(t) R(t, t') + (p - 1) \frac{1}{T_p(t)\Delta_p} \int_{t'}^{t} dt'' \frac{1}{T_p(t'')} R(t, t'') R(t'', t') C^{p-2}(t, t'')$$

$$+ \frac{1}{T_2(t)\Delta_2} \int_{t'}^{t} dt'' \frac{1}{T_2(t'')} R(t, t'') R(t'', t'); \tag{C16}$$

$$\frac{\partial}{\partial t} \bar{C}(t) = \mathbb{E}\langle \dot{x}_0(t) x_0^* \rangle$$

$$= -\mu(t) \bar{C}(t) + r_p(t) f'_p(\bar{C}(t)) + r_2(t) f'_2(\bar{C}(t))$$

$$+ (p - 1) \frac{1}{T_p(t)\Delta_p} \int_{t_o}^{t} dt'' \frac{1}{T_p(t'')} R(t, t'') C^{p-2}(t, t'') \bar{C}(t'') + \frac{1}{T_2(t)\Delta_2} \int_{t_o}^{t} dt'' \frac{1}{T_2(t'')} R(t, t'') \bar{C}(t''); \tag{C17}$$

$$\mu(t) = 1 + r_p(t) f'_p(\bar{C}(t)) \bar{C}(t) + r_2(t) f_2'(\bar{C}(t)) \bar{C}(t)$$

$$+ p \frac{1}{T_p(t)\Delta_p} \int_{t_o}^{t} dt'' \frac{1}{T_p(t'')} R(t, t'') C^{p-1}(t, t'') + 2 \frac{1}{T_2(t)\Delta_2} \int_{t_o}^{t} dt'' \frac{1}{T_2(t'')} R(t, t'') C(t, t''). \tag{C18}$$

Note that the last equation for $\mu(t)$ is obtained by imposing the spherical constraint $C(t,t) = 1 \; \forall \; t$ using the fact that $0 = [dC(t,t)/dt] = [\partial C(t,t')/\partial t]|_{t'=t} + [\partial C(t',t)/\partial t]|_{t'=t}$. The boundary conditions of this equations are as follows: $C(t,t) = 1$, which is the spherical constraint; $R(t,t) = 0$, which comes from causality in the Itô approach; and $R(t, t' \to t^-) = 1$. The initial condition for $\bar{C}(0) = \bar{C}_0$ is the overlap with the initial configuration with the true signal. If the initial configuration is random, $\bar{C}_0 = 0$, but it will have finite-size fluctuations, as in the case of AMP. Therefore, we find $\bar{C}_0 = \epsilon$, with ϵ an arbitrary small positive number.

APPENDIX D: NUMERICAL SOLUTION OF LSE EQUATIONS

The dynamical equations (C15)–(C18) were integrated numerically using two schemes:

(i) Fixed time grid: The derivatives were discretized and integrated according to their causal structure. This method is suitable only for short times (up to 500 time units).

(ii) Dynamic time grid: The step size is doubled after a given number of steps, and the equations are solved self-consistently for each waiting time. This approach is proposed in Ref. [45] and described in Appendix C of Ref. [73]. It allows integration up to very long times (up to 10^6 time units).

The results of these algorithms are concisely reported in the phase diagram shown in the main part of the paper. In what follows, we present the algorithms and a series of investigations that we carried out to check their stability; we explain the procedure to delimit the Langevin hard region, and we discuss how we can enter into part of that region by choosing a proper annealing protocol. The codes are available online [47].

1. Fixed time-grid $(2+p)$ spin

In this approach, time derivatives and integrals are discretized using $(\partial/\partial t)f(t,t') \simeq (1/\Delta t)[f(t+\Delta t, t') - f(t,t')]$ and the trapezoidal rule for integration, $\int_0^t f(t)dt \simeq (\Delta t/2)\sum_{l=0}^{t/\Delta t - 1}[f(l\Delta t) + f((l+1)\Delta t)]$. For instance, we define a function for computing the update in the response function, Eq. (C16), as follows:

$$R(t + \Delta t, t') = R(t, t') - \Delta t \mu(t) R(t, t') + \frac{1}{2}\frac{\Delta t^2}{\Delta_2}\sum_{l=t'/\Delta t}^{t/\Delta t - 1}[R(t, l\Delta t)R(l\Delta t, t') + R(t, (l+1)\Delta t)R((l+1)\Delta t, t')]$$

$$+ (p-1)\frac{\Delta t^2}{\Delta_p}\sum_{l=t'/\Delta t}^{t/\Delta t - 1}[C^{p-2}(t, l\Delta t)R(t, l\Delta t)R(l\Delta t, t') + C^{p-2}(t, (l+1)\Delta t)R(t, (l+1)\Delta t)R((l+1)\Delta t, t')].$$

Analogously, we define the other integrators. A simple causal integration scheme, being careful with the Itô prescription, gives the pseudocode below:

```
C(0,0) ← 1; R(0,0) ← 0; C̄(0) ← C̄₀;
for t ≤ t_max do
    C(t + Δt, t + Δt) ← 1; R(t + Δt, t + Δt) ← 0;
    μ(t) ← compute_mu(C, R, C̄, t);
    C̄(t + Δt) ← compute_mag(μ, C, R, C̄, t);
    for t' ≤ t do
        C(t + Δt, t') ← compute_C(μ, C, R, C̄, t);
        R(t + Δt, t') ← compute_R(μ, C, R, C̄, t);
    end for
    R(t + Δt, t) ← 1;
end for
```

2. Dynamical time-grid $(2+p)$ spin

The numerical scheme we discuss is presented in the Bayes-optimal case, where $T_2(t) \equiv T_p(t) \equiv 1$. However, the derivation that we propose can be easily generalized to the case where the Ts assume different values but are constants. Therefore, we do not employ this algorithm to

solve the LSE equations in the annealing protocol, for which instead we use the fixed time-grid algorithm. It is convenient to manipulate the equations to obtain an equivalent set of equations for the functions $C(t,t')$, $Q(t,t') \doteq 1 - C(t,t') - \int_{t'}^t R(t,t'')dt''$, $\bar{C}(t)$, where $Q(t,t')$ represents the deviation from the fluctuation dissipation theorem (FDT) at time t starting from time t'. Indeed, when the FDT theorem holds, it states that $R(t,t') = -\partial_t C(t,t')$.

We briefly anticipate the strategy that the algorithm uses to solve the equations. The algorithm discretizes the times into N_t intervals—starting from the boundary conditions $C(t,t) = 1$, $Q(t,t) = 0$, and $\bar{C} = \bar{C}_0 \in [0; 1]$; it fills the grid for small times (or small time differences $\tau = t - t' \ll 1$) using linear propagation. Given a time t and the initial guess for the Lagrange multiplier obtained by the linear propagator, the integrals are discretized and evaluated; then, the results are used to update the value of the Lagrange multiplier. This procedure is repeated iteratively until convergence. Once the first grid is filled, it follows a coarse-graining procedure, where the sizes of the time intervals are doubled and only half of the information is retained. This procedure is repeated a fixed number of

doubling of the original grid. The doubling scheme allows us to explore exponentially long times at the cost of losing part of the information; the direct consequence of this is the loss of stability for very long times [especially when the functions $C(t, t')$, $R(t, t')$, $\bar{C}(t)$ undergo fast changes at long times].

a. Dynamical equations in the algorithm

We recall the function $f_k(x) = (x^k/2)$ and its derivatives, $f'_k(x) = (kx^{k-1}/2)$ and $f''_k(x) = [k(k-1)x^{k-2}/2]$. For simplicity in the notation, we also introduce $f_k(t, t') \doteq f_k(C(t, t'))$,

$$(\partial_t + \mu(t))C(t, t') = 2R(t', t) + r_2\bar{C}(t')f'_2(\bar{C}(t)) + r_p\bar{C}(t')f'_p(\bar{C}(t)) + \frac{1}{\Delta_2}\int_0^{t'} dt'' f'_2(t, t'')R(t', t'')$$

$$+ \frac{1}{\Delta_2}\int_0^{t} dt'' f''_2(t, t'')R(t, t'')C(t', t'') + \frac{2}{p\Delta_p}\int_0^{t'} dt'' f'_p(t, t'')R(t', t'')$$

$$+ \frac{2}{p\Delta_p}\int_0^{t} dt'' f''_p(t, t'')R(t, t'')C(t', t''),$$

$$(\partial_t + \mu(t))R(t, t') = \delta(t - t') + \frac{1}{\Delta_2}\int_{t'}^{t} dt'' f''_2(t, t'')R(t, t'')R(t'', t') + \frac{2}{p\Delta_p}\int_{t'}^{t} dt'' f''_p(t, t'')R(t, t'')R(t'', t'),$$

$$(\partial_t + \mu(t))\bar{C}(t) = r_2 f'_2(\bar{C}(t)) + r_p f'_p(\bar{C}(t)) + \frac{1}{\Delta_2}\int_0^{t} dt'' f''_2(t, t'')R(t, t'')\bar{C}(t'') + \frac{2}{p\Delta_p}\int_0^{t} dt'' f''_p(t, t'')R(t, t'')\bar{C}(t''),$$

$$\mu(t) = 1 + r_2\bar{C}(t)f'_2(\bar{C}(t)) + r_p\bar{C}(t)f'_p(\bar{C}(t)) + \frac{2}{\Delta_2}\int_0^{t} dt'' f'_2(t, t'')R(t, t'') + \frac{2}{\Delta_p}\int_0^{t} dt'' f'_p(t, t'')R(t, t'').$$

Following the lines of Ref. [73], we introduce the FDT violation function $Q(t, t')$, and after some manipulation, the system becomes

$$(\partial_t + \mu(t))C(t, t') = \bar{C}(t')[r_2 f'_2(\bar{C}(t)) + r_p f'_p(\bar{C}(t))] + \frac{1}{\Delta_2}\left\{\int_0^{t'} dt''\left[f'_2(t, t'')\frac{\partial Q(t', t'')}{\partial t''} + f''_2(t, t'')\frac{\partial Q(t, t'')}{\partial t''}C(t', t'')\right] + \right.$$

$$- \int_{t'}^{t} dt''\left[f'_2(t, t'')\frac{\partial C(t'', t')}{\partial t''} - f''_2(t, t'')\frac{\partial Q(t, t'')}{\partial t''}C(t'', t')\right] + f'_2(1)C(t, t') - f'_2(t, 0)C(t', 0)\right\}$$

$$+ \frac{2}{p\Delta_p}\left\{\int_0^{t'} dt''\left[f'_p(t, t'')\frac{\partial Q(t', t'')}{\partial t''} + f''_p(t, t'')\frac{\partial Q(t, t'')}{\partial t''}C(t', t'')\right] + \right.$$

$$- \int_{t'}^{t} dt''\left[f'_p(t, t'')\frac{\partial C(t'', t')}{\partial t''} - f''_p(t, t'')\frac{\partial Q(t, t'')}{\partial t''}C(t'', t')\right] + f'_p(1)C(t, t') - f'_p(t, 0)C(t', 0)\right\},$$

$$\tag{D1}$$

$$(\partial_t + \mu(t))Q(t, t') = \mu(t) - 1 + \frac{1}{\Delta_2}\left\{-\int_{t'}^{t} dt'' f'_2(t, t'')\frac{\partial Q(t'', t')}{\partial t''} + \int_{t'}^{t} dt'' f''_2(t, t'')\frac{\partial Q(t, t'')}{\partial t''}[Q(t'', t') - 1]\right.$$

$$+ f'_2(1)[Q(t, t') - 1] + f'_2(t, 0)C(t', 0) - \int_0^{t'} dt''\left[f'_2(t, t'')\frac{\partial Q(t', t'')}{\partial t''} + f''_2(t, t'')\frac{\partial Q(t, t'')}{\partial t''}C(t', t'')\right]\right\}$$

$$+ \frac{2}{p\Delta_p}\left\{-\int_{t'}^{t} dt'' f'_p(t, t'')\frac{\partial Q(t'', t')}{\partial t''} + \int_{t'}^{t} dt'' f''_p(t, t'')\frac{\partial Q(t, t'')}{\partial t''}[Q(t'', t') - 1]\right.$$

$$+ f'_p(1)[Q(t, t') - 1] + f'_p(t, 0)C(t', 0) - \int_0^{t'} dt''\left[f'_p(t, t'')\frac{\partial Q(t', t'')}{\partial t''} + f''_p(t, t'')\frac{\partial Q(t, t'')}{\partial t''}C(t', t'')\right]\right\}$$

$$- \bar{C}(t')[r_2 f'_2(\bar{C}(t)) + r_p f'_p(\bar{C}(t))], \tag{D2}$$

$$(\partial_t + \mu(t))\bar{C}(t) = r_2 f_2'(\bar{C}(t)) + r_p f_p'(\bar{C}(t)) + \frac{1}{\Delta_2}\left\{ f_2'(1)\bar{C}(t) - f_2'(t,0)\bar{C}(0) - \int_0^t dt'' f_2'(t,t'') \frac{d}{dt''}\bar{C}(t'') \right.$$

$$+ \int_0^t dt'' f_2''(t,t'') \frac{\partial Q(t,t'')}{\partial t''}\bar{C}(t'') \Bigg\} + \frac{2}{p\Delta_p}\left\{ f_p'(1)\bar{C}(t) - f_p'(t,0)\bar{C}(0) - \int_0^t dt'' f_p'(t,t'') \frac{d}{dt''}\bar{C}(t'') \right.$$

$$+ \int_0^t dt'' f_p''(t,t'') \frac{\partial Q(t,t'')}{\partial t''}\bar{C}(t'') \Bigg\}, \tag{D3}$$

$$\mu(t) = 1 + r_2 \bar{C}(t) f_2{}'(\bar{C}(t)) + r_p \bar{C}(t) f_p'(\bar{C}(t)) + \frac{2}{\Delta_2}[f_2(1) - f_2(t,0)] + \frac{2}{\Delta_p}[f_p(1) - f_p(t,0)]$$

$$+ \int_0^t dt'' \left[\frac{2}{\Delta_2} f_2{}'(t,t'') + \frac{2}{\Delta_p} f_p'(t,t'') \right] \frac{\partial Q(t,t'')}{\partial t''}. \tag{D4}$$

Further simplifications can be obtained by introducing $\mu'(t) = \mu(t) - (2/\Delta_2)f_2(1) - (2/\Delta_p)f_p(1)$,

$$\mu'(t) = 1 + r_2 \bar{C}(t) f_2'(\bar{C}(t)) + r_p \bar{C}(t) f_p'(\bar{C}(t)) - \frac{2}{\Delta_2} f_2(t,0) - \frac{2}{\Delta_p} f_p(t,0)$$

$$+ \int_0^t dt'' \left[\frac{2}{\Delta_2} f_2'(t,t'') + \frac{2}{\Delta_p} f_p'(t,t'') \right] \frac{\partial Q(t,t'')}{\partial t''}, \tag{D5}$$

$$(\partial_t + \mu'(t))C(t,t') = \bar{C}(t')[r_2 f_2'(\bar{C}(t)) + r_p f_p'(\bar{C}(t))] + \frac{1}{\Delta_2}\left\{ \int_0^{t'} dt'' \left[f_2'(t,t'') \frac{\partial Q(t',t'')}{\partial t''} + f_2''(t,t'') \frac{\partial Q(t,t'')}{\partial t''} C(t',t'') \right] + \right.$$

$$- \int_{t'}^t dt'' \left[f_2'(t,t'') \frac{\partial C(t'',t')}{\partial t''} - f_2''(t,t'') \frac{\partial Q(t,t'')}{\partial t''} C(t'',t') \right] - f_2'(t,0)C(t',0) \Bigg\}$$

$$+ \frac{2}{p\Delta_p}\left\{ \int_0^{t'} dt'' \left[f_p'(t,t'') \frac{\partial Q(t',t'')}{\partial t''} + f_p''(t,t'') \frac{\partial Q(t,t'')}{\partial t''} C(t',t'') \right] + \right.$$

$$- \int_{t'}^t dt'' \left[f_p'(t,t'') \frac{\partial C(t'',t')}{\partial t''} - f_p''(t,t'') \frac{\partial Q(t,t'')}{\partial t''} C(t'',t') \right] - f_p'(t,0)C(t',0) \Bigg\}, \tag{D6}$$

$$(\partial_t + \mu'(t))Q(t,t') = \mu'(t) - 1 - \bar{C}(t')[r_2 f_2'(\bar{C}(t)) + r_p f_p'(\bar{C}(t))]$$

$$+ \frac{1}{\Delta_2}\left\{ - \int_{t'}^t dt'' f_2'(t,t'') \frac{\partial Q(t'',t')}{\partial t''} + \int_{t'}^t dt'' f_2''(t,t'') \frac{\partial Q(t,t'')}{\partial t''}[Q(t'',t') - 1] \right.$$

$$+ f_2'(t,0)C(t',0) - \int_0^{t'} dt'' \left[f_2'(t,t'') \frac{\partial Q(t',t'')}{\partial t''} + f_2''(t,t'') \frac{\partial Q(t,t'')}{\partial t''} C(t',t'') \right] \Bigg\}$$

$$+ \frac{2}{p\Delta_p}\left\{ - \int_{t'}^t dt'' f_p'(t,t'') \frac{\partial Q(t'',t')}{\partial t''} + \int_{t'}^t dt'' f_p''(t,t'') \frac{\partial Q(t,t'')}{\partial t''}[Q(t'',t') - 1] \right.$$

$$+ f_p'(t,0)C(t',0) - \int_0^{t'} dt'' \left[f_p'(t,t'') \frac{\partial Q(t',t'')}{\partial t''} + f_p''(t,t'') \frac{\partial Q(t,t'')}{\partial t''} C(t',t'') \right] \Bigg\}, \tag{D7}$$

$$(\partial_t + \mu'(t))\bar{C}(t) = r_2 f_2'(\bar{C}(t)) + r_p f_p'(\bar{C}(t)) + \frac{1}{\Delta_2}\left\{ -f_2'(t,0)\bar{C}(0) + \right.$$

$$- \int_0^t dt'' f_2'(t,t'') \frac{d}{dt''}\bar{C}(t'') + \int_0^t dt'' f_2''(t,t'') \frac{\partial Q(t,t'')}{\partial t''}\bar{C}(t'') \Bigg\} + \frac{2}{p\Delta_p}\left\{ -f_p'(t,0)\bar{C}(0) + \right.$$

$$- \int_0^t dt'' f_p'(t,t'') \frac{d}{dt''}\bar{C}(t'') + \int_0^t dt'' f_p''(t,t'') \frac{\partial Q(t,t'')}{\partial t''}\bar{C}(t'') \Bigg\}. \tag{D8}$$

b. First-order expansion coefficients

In the numerics, we initialize the grid by a linear propagation of the initial conditions. To determine which coefficients to use, we can expand the functions up to the second term for small values of τ (and in the last equation of t),

$$C(t'+\tau,t') = C(t',t') + C^{(1,0)}(t',t')\tau$$
$$+ \frac{1}{2}C^{(2,0)}(t',t') + O(\tau^3),$$
$$Q(t'+\tau,t') = Q(t',t') + Q^{(1,0)}(t',t')\tau$$
$$+ \frac{1}{2}Q^{(2,0)}(t',t') + O(\tau^3),$$
$$\bar{C}(t) = \bar{C}(0) + \bar{C}^{(1)}(0)\tau + \frac{1}{2}\bar{C}^{(2)}(0) + O(t^3). \quad \text{(D9)}$$

This expansion gives the following coefficients: $C(t,t) = 1$, $C^{(1,0)}(t,t) = -1$, $Q(t,t) = 0$, $Q^{(1,0)}(t,t) = 0$, $\bar{C}(0) = \bar{C}_0$, and $\bar{C}^{(1)}(0) = [r_2 f_2'(\bar{C}_0) + r_p f_p'(\bar{C}_0)](1 - (\bar{C}_0)^2) - \bar{C}_0$, where $\bar{C}_0$ is the initial value of the overlap with the signal.

c. Numerical integration and derivation

The set of equations derived above presents six types of integrals,

$$I_{ij}^{(1AB)} = \int_{t_j}^{t_i} dt'' A(t_i, t'') \frac{\partial B(t'', t_j)}{\partial t''},$$

$$I_{ij}^{(2ABC)} = \int_{t_j}^{t_i} dt'' A(t_i, t'') \frac{\partial B(t_i, t'')}{\partial t''} C(t'', t_j),$$

$$I_{ij}^{(3AB)} = \int_0^{t_j} dt'' A(t_i, t'') \frac{\partial B(t_i, t'')}{\partial t''},$$

$$I_{ij}^{(4ABC)} = \int_0^{t_j} dt'' A(t_i, t'') \frac{\partial B(t_i, t'')}{\partial t''} C(t_j, t''),$$

$$I_i^{(5AB)} = \int_0^{t_i} dt'' A(t_i, t'') \frac{\partial B(t'')}{\partial t''},$$

$$I_i^{(6ABC)} = \int_0^{t_i} dt'' A(t_i, t'') \frac{\partial B(t_i, t'')}{\partial t''} C(t'').$$

The integrals can be easily discretized,

$$I_{ij}^{(2ABC)} = \sum_{t_l=t_j+\delta t}^{t_i} \int_{t_l-\delta t}^{t_l} dt'' A(t_i, t'') \frac{\partial B(t_i, t'')}{\partial t''} C(t'', t_j)$$

$$\simeq \sum_{t_l=t_j+\delta t}^{t_i} \int_{t_l-\delta t}^{t_l} dt_1 A(t_i, t_1) \int_{t_l-\delta t}^{t_l} dt_2 \frac{\partial B(t_i, t_2)}{\partial t_2} \int_{t_l-\delta t}^{t_l} dt_3 C(t_3, t_j)$$

$$\simeq \sum_{t_l=t_j+\delta t}^{t_i} \frac{1}{2}[A(t_i, t_l) + A(t_i, t_l - \delta t)][B(t_i, t_l) - B(t_i, t_l - \delta t)]\frac{1}{2}[C(t_l, t_j) + C(t_l - \delta t, t_j)].$$

In particular, the six integrals become

$$I_{ij}^{(1AB)} = A_{im}B_{mj} - A_{ij}B_{jj} + \sum_{l=m+1}^{i} \frac{1}{2}(A_{il} + A_{i(l-1)})(B_{lj} - B_{(l-1)j}) + - \sum_{l=j+1}^{m} \frac{1}{2}(B_{lj} + B_{(l-1)j})(A_{il} - A_{i(l-1)})$$

$$= A_{im}B_{mj} - A_{ij}B_{jj} + \sum_{l=m+1}^{i} dA_{il}^{(v)}(B_{lj} - B_{(l-1)j}) - \sum_{l=j+1}^{m} (A_{il} - A_{i(l-1)})dB_{lj}^{(h)}, \quad \text{(D10)}$$

$$I_{ij}^{(2ABC)} = \sum_{l=j+1}^{i} \frac{1}{2}(A_{il} + A_{i(l-1)})(B_{lj} - B_{(l-1)j})\frac{1}{2}(C_{lj} + C_{(l-1)j})$$

$$= \sum_{l=m+1}^{i} dA_{il}^{(h)}(B_{il} - B_{i(l-1)})\frac{1}{2}(C_{lj} + C_{(l-1)j}) + \sum_{l=j+1}^{m} \frac{1}{2}(A_{il} + A_{i(l-1)})(B_{il} - B_{i(l-1)})dC_{lj}^{(v)}, \quad \text{(D11)}$$

$$I_{ij}^{(3AB)} = A_{ij}B_{jj} - A_{i0}B_{j0} - \sum_{l=1}^{j}(A_{il} - A_{i(l-1)})dB_{jl}^{(v)}, \quad \text{(D12)}$$

$$I_{ij}^{(4ABC)} = \sum_{l=1}^{j} \frac{1}{2}(A_{il} + A_{i(l-1)})(B_{il} - B_{i(l-1)})dC_{jl}^{(v)}, \quad \text{(D13)}$$
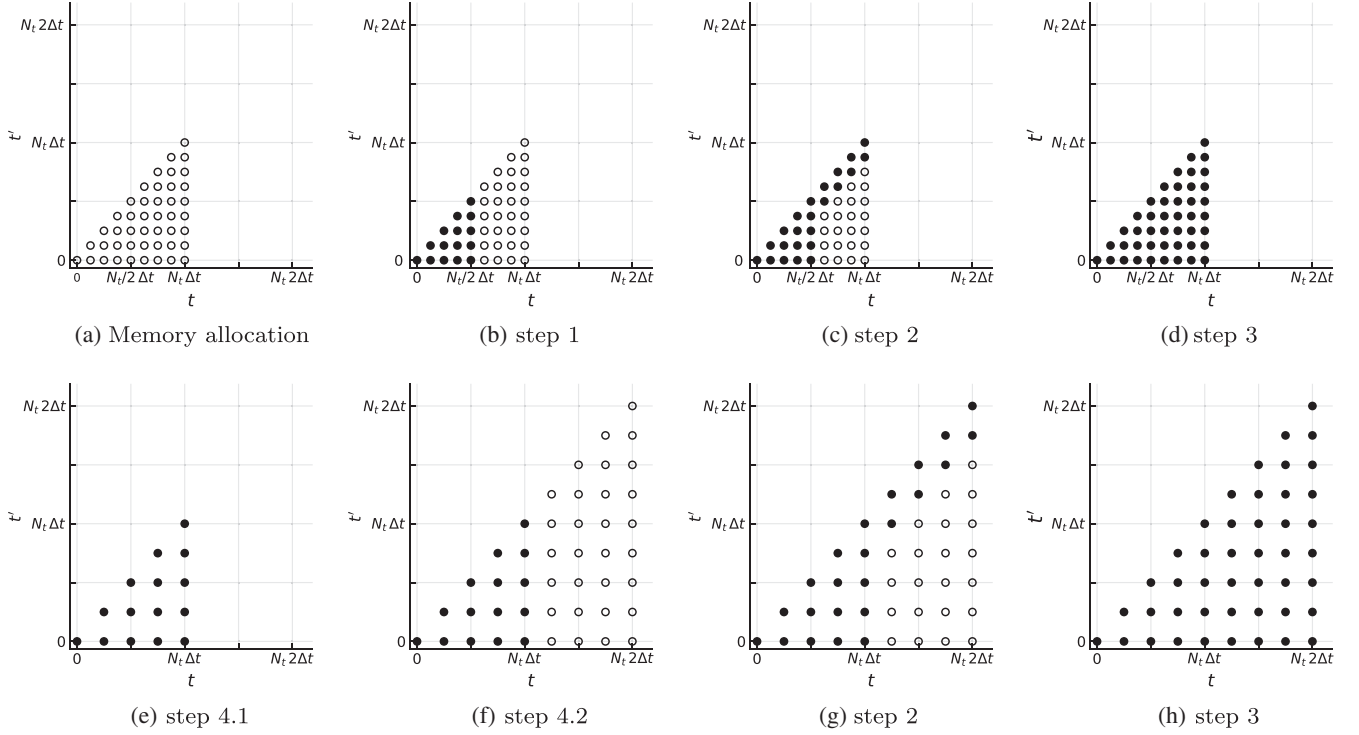
FIG. 11.   Representation of the initialization and the first two iterations for the evaluation of a two-times observable using the dynamic-grid algorithm. The empty circles represent slots allocated in memory but not associated with any specific value, while the full circles are memory slots already associated with a specific value. For any two-times function, it first allocates the memory (a), and then it fills half of the grid by linear propagation (b). Still using linear propagation, it fills the slots with $t - t' \ll 1$ (c), and it sets the other values by imposing self-consistency (d). Finally, it halves the grid (e), doubles the time step, and it allocates the memory (f). Then, the algorithm loops follow the same scheme as in panels (b)–(e).

$$I_i^{(5AB)} = \sum_{l=1}^{i} dA_{il}^{(v)}(B_l - B_{l-1}), \qquad (D14)$$

$$I_i^{(6ABC)} = \sum_{l=1}^{i} \frac{1}{2}(A_{il} + A_{i(l-1)})(B_{il} - B_{i(l-1)})dC_l, \qquad (D15)$$

where the superscripts (v) and (h) represent the vertical (t') and horizontal (t) derivatives in the discretized times; see Fig. 11 for an intuitive understanding.

We also discretize the derivative using the last two time steps,

$$\frac{d}{dt}g(t) = \frac{3}{2\delta t}g(t) - \frac{2}{\delta t}g(t - \delta t) + \frac{1}{2\delta t}g(t - 2\delta t) + O(\delta t^3). \qquad (D16)$$

Given the time indices i and j, we define and evaluate the following quantities

$$\{C_{ij}, Q_{ij}, M2_{ij}, N2_{ij}, Mp_{ij}, Np_{ij}, \text{Cbar}_i, P2_i, Pp_i, mu_i\}$$
$$= \{C(t_i, t_j), Q(t_i, t_j), f'_2(C(t_i, t_j)), f''_2(C(t_i, t_j)), f'_p(C(t_i, t_j)), f''_p(C(t_i, t_j)), \bar{C}(t_i), f'_2(\bar{C}(t_i)), f'_p(\bar{C}(t_i)), \mu(t_i)\}$$

plus the respective vertical and horizontal derivatives.
Calling $D_i = (3/2dt) + \mu'_i - (1/\Delta_2)M2_{ii} - (2/p\Delta_p)Mp_{ii}$, the original dynamical equations are integrated as follows:

$$C_{ij}D_i = \frac{2}{dt}C_{(i-1)j} - \frac{1}{2dt}C_{(i-2)j} + \text{Cbar}_j(r_2 P2_i + r_p Pp_i) + \frac{1}{\Delta_2}(-\dot{I}_{ij}^{(1f'_2 C)} + I_{ij}^{(2f''_2 QC)} + I_{ij}^{(3f'_2 Q)} + I_{ij}^{(4f''_2 QC)} - M2_{i0}C_{j0})$$

$$+ \frac{2}{p\Delta_p}(-\dot{I}_{ij}^{(1f'_p C)} + I_{ij}^{(2f''_p QC)} + I_{ij}^{(3f'_p Q)} + I_{ij}^{(4f''_p QC)} - Mp_{i0}C_{j0}), \qquad (D17)$$

$$Q_{ij}D_i = \mu'_i - 1 + \frac{2}{dt}Q_{(i-1)j} - \frac{1}{2dt}Q_{(i-2)j} + \mathrm{Cbar}_j(r_2 P2_i + r_p Pp_i)$$
$$+ \frac{1}{\Delta_2}(-\dot{I}_{ij}^{(1f'_2 Q)} + I_{ij}^{(2f''_2 Q(Q-1))} - I_{ij}^{(3f'_2 Q)} - I_{ij}^{(4f''_2 QC)} - M2_{i0}C_{i0})$$
$$+ \frac{2}{p\Delta_p}(-\dot{I}_{ij}^{(1f'_p Q)} + I_{ij}^{(2f''_p Q(Q-1))} - I_{ij}^{(3f'_p Q)} - I_{ij}^{(4f''_p QC)} - Mp_{i0}C_{i0}), \tag{D18}$$

$$\mathrm{Cbar}_i D_i = \frac{2}{dt}\mathrm{Cbar}_{i-1} - \frac{1}{2dt}\mathrm{Cbar}_{i-2} + r_2 P2_i + r_p Pp_i + \frac{1}{\Delta_2}(-\dot{I}_i^{(5f'_2 \mathrm{Cbar})} + I_i^{(6f''_2 Q\mathrm{Cbar})} - M2_{i0}\mathrm{Cbar}_0)$$
$$+ \frac{2}{p\Delta_p}(-\dot{I}_i^{(5f'_p \mathrm{Cbar})} + I_i^{(6f''_p Q\mathrm{Cbar})} - Mp_{i0}\mathrm{Cbar}_0). \tag{D19}$$

In the systems, we used $\dot{I}$ to characterize the integrals, where we remove from the sum the term present in the left-hand side [e.g., for C_{ij} in Eq. (D17)]. Using Simpson's integration formula, we define the increments

$$\Delta_{il} = \frac{1}{12}(Q_{il} - Q_{i(l-1)})\{W_2^2[-(M2_{i(l+1)} + N2_{i(l+1)}C_{i(l+1)}) + 8(M2_{il} + N2_{il}C_{il}) + 5(M2_{i(l-1)} + N2_{i(l-1)}C_{i(l-1)})]$$
$$+ W_p^2[-(Mp_{i(l+1)} + Np_{i(l+1)}C_{i(l+1)}) + 8(Mp_{il} + Np_{il}C_{il}) + 5(Mp_{i(l-1)} + Np_{i(l-1)}C_{i(l-1)})]\},$$

and we determine μ' as

$$\mu' = 1 + r_2 P2_i + r_p Pp_i + \delta\mu'$$
$$+ \sum_{l=1}^{i-N_t/4} \Delta_{il} - (W_2^2 M2_{i0} + W_p^2 Mp_{i0})C_{i0}, \tag{D20}$$

with $\delta\mu'$ initially set to 0.

d. Algorithm

Here, we describe the main steps of the algorithm, pictorially represented Fig. 11.

Discretize the time (t, t') in N_t (even) intervals; the results shown use $N_t = 1024$.

(1) *Initialization.* Fill the first $N_t/2$ times by linear propagation of the value obtained from the perturbative analysis,

$$C_{ij} = 1 - (i - j)dt, \tag{D21}$$

$$Q_{ij} = 0, \tag{D22}$$

$$\mathrm{Cbar}_i = \bar{C}_0 + \{[r_2 f'_2(\bar{C}_0) + r_p f'_p(\bar{C}_0)]$$
$$\times (1 + (\bar{C}_0)^2) - \bar{C}_0\}dt, \tag{D23}$$

$$M2_{ij} = f'_2(C_{ij}), \tag{D24}$$

$$N2_{ij} = f''_2(C_{ij}), \tag{D25}$$

$$Mp_{ij} = f'_p(C_{ij}), \tag{D26}$$

$$Np_{ij} = f''_p(C_{ij}). \tag{D27}$$

(2) *Fill the grid (small τ).* Continue to propagate the values for small time differences $\tau = t - t' \ll 1$. In terms of the algorithm, we have some elements of the grid, N_c of them, close to the diagonal that will be updated by linear propagation because the approximation of small τ is still valid. In our simulation, the first Δt is of the order 10^{-7} and $N_c = 2$.

(3) *Fill the grid (larger τ).* The rest of the values will be copied from the previous t ($A_{t+\Delta t,t'} = A_{t,t'}$). These values are the initial guesses for solving the self-consistent equations (D17)–(D19) and (D20); in this procedure, the derivatives are updated using the second-order discretization.

(4) *Half the grid and expand.* The grid is decimated, which means that each observable is contracted, $A_{i,j} \leftarrow A_{2i,2j}$, and the derivates are updated as follows: $dA_{i,j}^{(h)} \leftarrow \frac{1}{2}(dA_{2i,2j}^{(h)} + dA_{2i-1,2j}^{(h)})$, $dA_{i,j}^{(v)} \leftarrow \frac{1}{2}(dA_{2i,2j}^{(v)} + dA_{2i,2j-1}^{(v)})$. The new time step is now $\Delta t \leftarrow 2\Delta t$.

(5) *Start over from step 2.*

3. Numerical checks on the dynamical algorithm

The dynamic-grid algorithm has been checked in a variety of ways.

(a) *Cross-checking using the fixed-grid algorithm.* For short times, the dynamical equations were solved using the fixed-grid algorithm and compared with the outcome of the dynamic-grid algorithm, obtaining the same results; see Fig. 12. In the figure, we used the fixed grid with $t_{\max} = 100$ and $\Delta t = 6.25 \times 10^{-3}$.

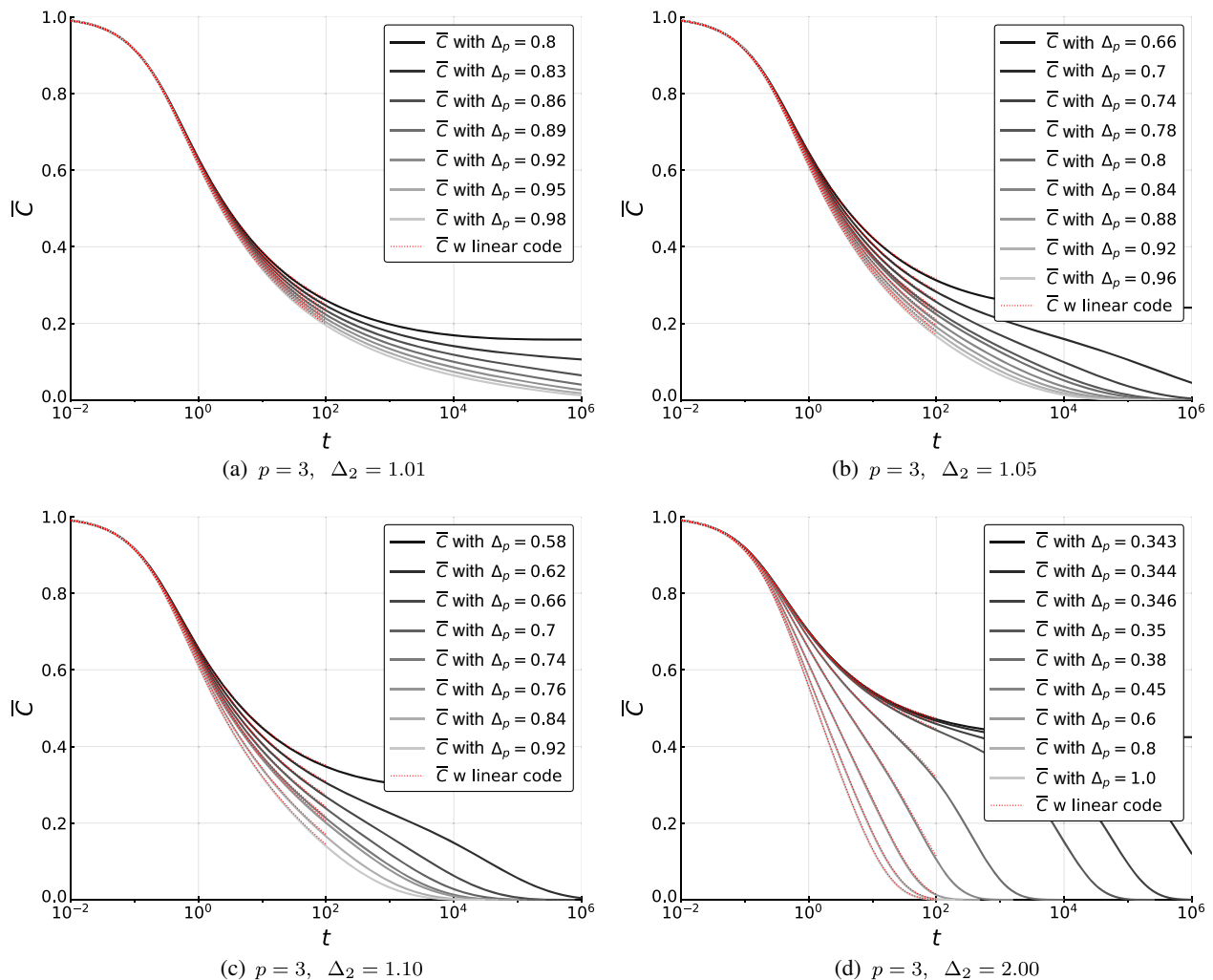(b) *Same magnetization in the easy region.* In the impossible and easy regions, the overlap with the signal of

(a) $p = 3$, $\Delta_2 = 1.01$

(b) $p = 3$, $\Delta_2 = 1.05$

(c) $p = 3$, $\Delta_2 = 1.10$

(d) $p = 3$, $\Delta_2 = 2.00$

FIG. 12. Evolution of the correlation with the signal starting from the solution, $\bar{C}_0 = 1.0$ at fixed Δ_2 for different Δ_p. The dotted red line overlapping with other lines is the same quantity evaluated using the fixed grid algorithm up to time 100. We have started the LSE from an informative initial condition.

both AMP and dynamic-grid integration converges to the same value. In Fig. 13, we show the overlap obtained with AMP (black dashed line) and the overlap achieved by the integration scheme at a given time. We can see that the overlap with the signal as obtained by solving the LSE equations converges to the same value of the fixed point of AMP. Given a fixed Δ_2, we can observe that the time to convergence increases very rapidly as we decrease Δ_p. We fit this increase of the relaxation time to get the boundary of the Langevin hard region.

(c) *Dynamical transition*. The dynamical transition where the finite magnetization fixed point disappears can be regarded as a clustering or dynamical glass transition. Indeed, starting at the impossible phase and going towards the hard phase, at the dynamical transition the free-energy landscape changes and the unique ergodic paramagnetic minimum of the impossible phase gets

clustered into an exponential number of metastable glassy states (see Appendix E). Correspondingly, the relaxation time of the Langevin algorithm diverges. Fitting this divergence with a power law, we obtain an alternative estimation of the dynamical line. In the right panel of Fig. 14, we plot, with yellow points, the dynamical transition line as extracted from the fit of the relaxation time of the Langevin algorithm coming from the impossible phase and entering the hard phase.

4. Extrapolation procedure

In order to determine the Langevin hard region, given a fixed value of Δ_p (Δ_2), we measure the time that it takes to relax to equilibrium. On approaching the Langevin hard region, this relaxation time increases, and we extrapolate the growth to obtain the critical Δ_p^* (Δ_2^* respectively) where the relaxation time appears to diverge. The extrapolation
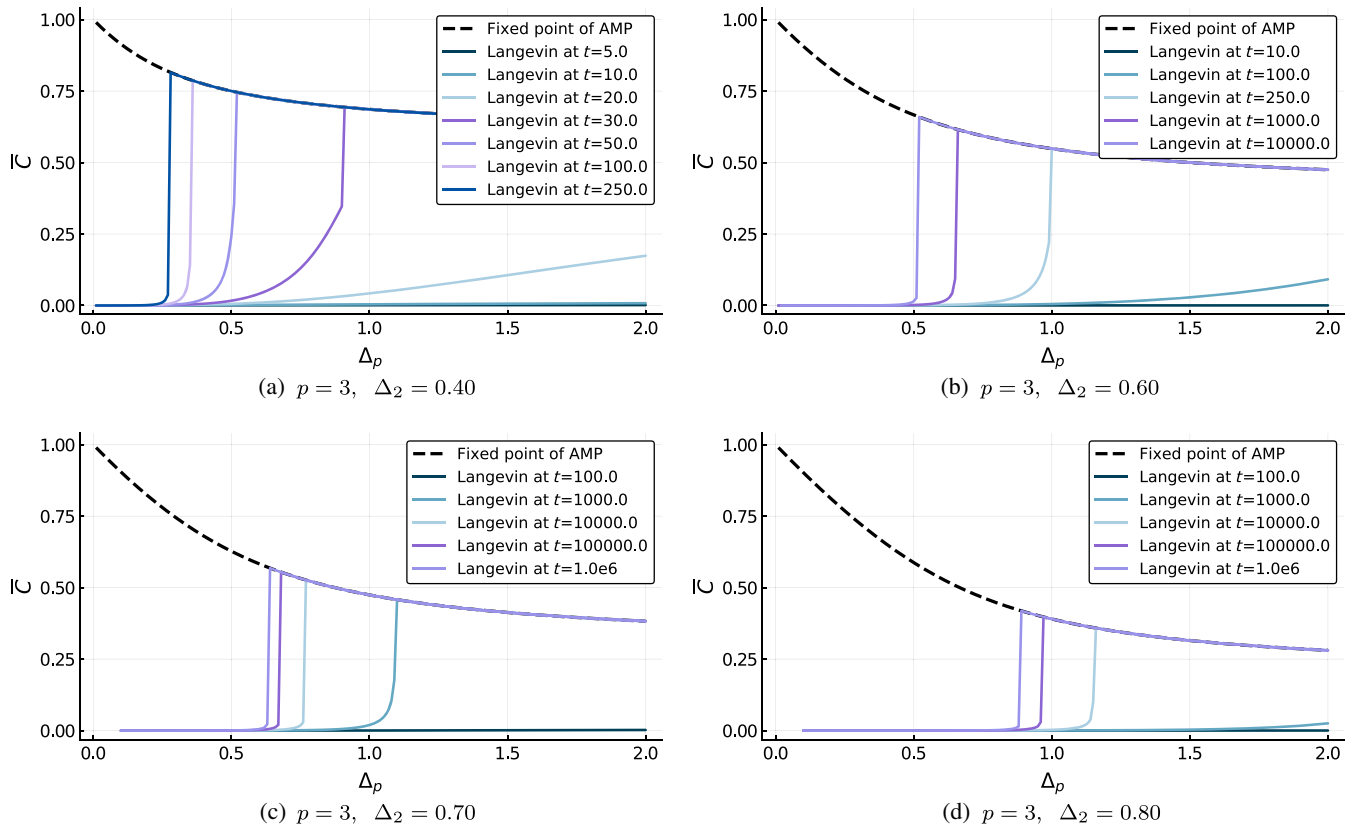
FIG. 13. Correlation with the signal of AMP (dotted lines) and Langevin (solid lines) at the kth iteration and t time, respectively, starting in both cases with an initial overlap of 10^{-4}. The black dashed line is the asymptotic value predicted with AMP. In the easy region, provided there is enough running time, Langevin dynamics finds the same alignment as AMP. The figures show qualitatively the same behavior for different values of Δ_2.

is performed starting from $\bar{C}_0 = 10^{-40}$ and assuming a power-law divergence. Figure 14 shows the results of this procedure for the cases $2 + 3$ and $2 + 4$. We remark that the divergence times increase as p increases. Therefore, because of the instability of the code for long times, it becomes difficult to extrapolate the threshold accurately. In particular, in the right panel of Fig. 14 when estimating the threshold for $2 + 4$, we consider horizontal sections, and the points extrapolated for Δ_2 close to the threshold $\Delta_2 = 1.0$ are very hard to estimate because of these instabilities.

a. Numerical checks on the extrapolation procedure

To test the quality of the fits, we use a similar numerical procedure to locate the spinodal of the informative solution, which is given by the points where the informative solution ceases to exist. This spinodal must be the same for both the AMP and the Langevin algorithm [6].

Since we aim at studying the spinodal of the informative solution, we initialize the LSE with $\bar{C}_0 = 1$ and let it relax, measuring the time it takes to equilibrate at the value of $\bar{C}$ given by the informative fixed point of AMP. We perform this procedure by fixing Δ_2 and changing Δ_p. As we

approach the critical $\Delta_{p,\text{dyn}}$, the relaxation time will diverge, and we can fit this divergence with a power law. The dynamic threshold extracted in this way is finally compared with the one obtained from AMP. In Fig. 17, we show how this scheme has been applied for $\Delta_2 \in \{1.01, 1.05, 1.10, 2.00\}$. As we get closer to the critical line $\Delta_2 = 1$, the relaxation time increases (and the height of the plateau decreases), making the fit harder. All in all, we observe a very good agreement between the points found with these extrapolation procedures and the prediction obtained with AMP, as shown in Fig. 14.

5. Initial conditions

The LSE equations show a rather strong dependence on the initial condition $\bar{C}_0$. A low initial magnetization will give a low initial momentum in the direction of the signal, as can be observed in the linear expansion of $\bar{C}$, Eq. (D9), and consequently, the system will not be able to cross even very small barriers. The direct consequence is that by reducing the initial magnetization, the estimated threshold will get worse, in the sense that a larger signal-to-noise ratio will be required to find the solution. This finding can be
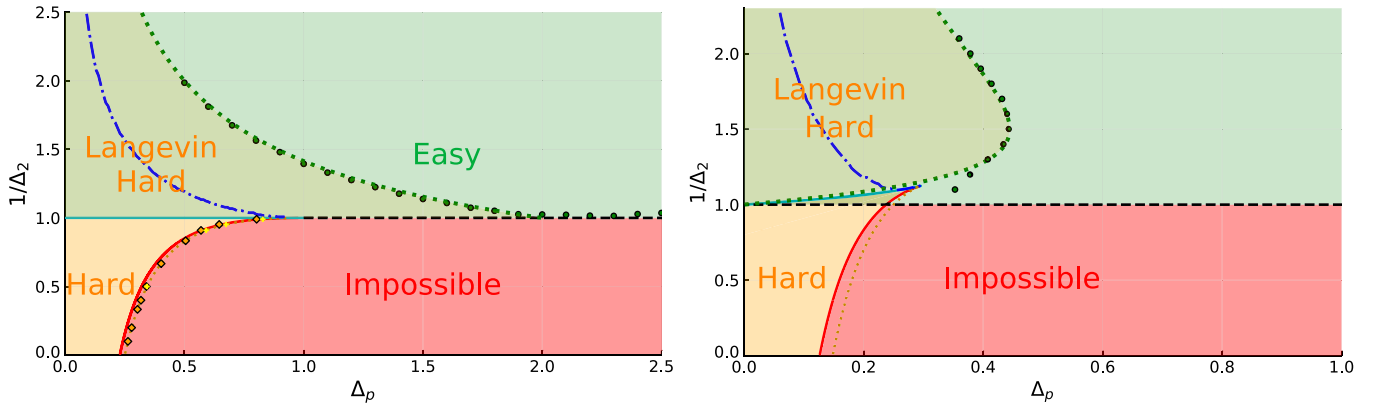
FIG. 14.   Left panel: Phase diagram of the spiked matrix-tensor model for $p = 3$ as presented in the left panel of Fig. 7, with the additional boundary of the Langevin hard phase (green circles and dotted green line). The data points (circles) have been obtained numerically by fitting the relaxation time at fixed Δ_p and increasing Δ_2. The green dotted line shows fixed points of expression (15). The blue dashed-dotted line marks a region above which we no longer observe a stable positive 1RSB complexity. Finally, we plot with orange and yellow dots on the dynamical transition line as extracted from the relaxation time of the Langevin algorithm coming from the hard and impossible phases, respectively. Right panel: Phase diagram of the spiked matrix-tensor model for $p = 4$ as presented in the right panel of Fig. 7 with the additional Langevin hard phase boundary. The data points are obtained by fixing Δ_2 and decreasing Δ_p. Also in this case, we observe that the Langevin hard phase extends to the AMP easy phase. Interestingly, the Langevin hard phase here folds and presents a reentrant behavior; investigating the precise character of this reentrance is hampered by the vicinity of the critical point and is left for future work. The blue dashed line marks a region above which we no longer observe a stable positive 1RSB complexity.

observed in Fig. 15, where different initial conditions are compared on the section $\Delta_p = 0.9$.

Finally, we remark how the different initial conditions affect the phase diagram. In Fig. 16, we compare the Langevin hard-easy threshold evaluated starting from different initial conditions, showing that the region gets larger as the initial condition decreases, up to convergence.
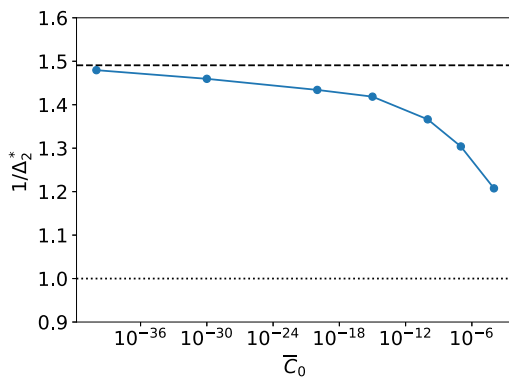
6. Annealing protocol

In this section, we show that by using specific protocols that separate the matrix and tensor parts of the cost function (something we do not allow in the main part of this paper for the purpose of having a model as generic as possible), we are able to enter in the Langevin hard region. A generic annealing scheme would lower the noises of both channels simultaneously, and we will not be able to avoid the Langevin hard region. Instead, we can use the following protocol:

FIG. 15.   Estimated divergence point Δ_2^* at fixed $\Delta_p = 0.9$, with $p = 3$ a function of the initial condition $\bar{C}$, ranging from 10^{-40} to 10^{-4}. The vertical axis is in the linear scale, while the horizontal axis is in the log scale. We observe that the dependence of estimated divergence points on the initial condition is consistent with the asymptotic value $1/\Delta_2^* = \sqrt{2/\Delta_p} \approx 1.491$ following from Eq. (15) and depicted by the dashed line. The dotted line instead represents the AMP threshold, for comparison.
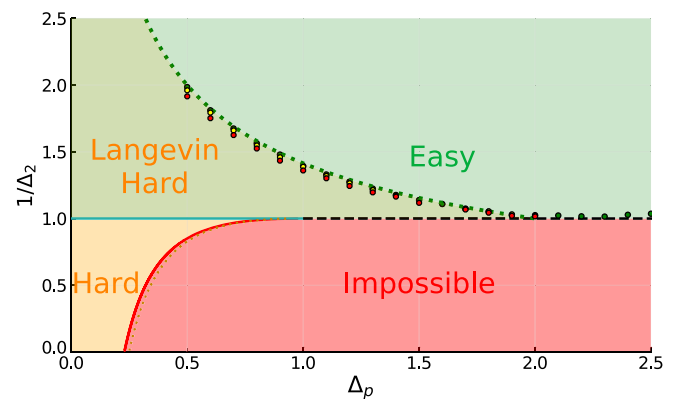

FIG. 16.   Phase diagram of the $p = 3$ model, where we compare the threshold of the Langevin hard region using different values of the initial conditions, respectively: green circles $\bar{C}_0 = 10^{-40}$, yellow circles $\bar{C}_0 = 10^{-30}$, red circles $\bar{C}_0 = 10^{-20}$.
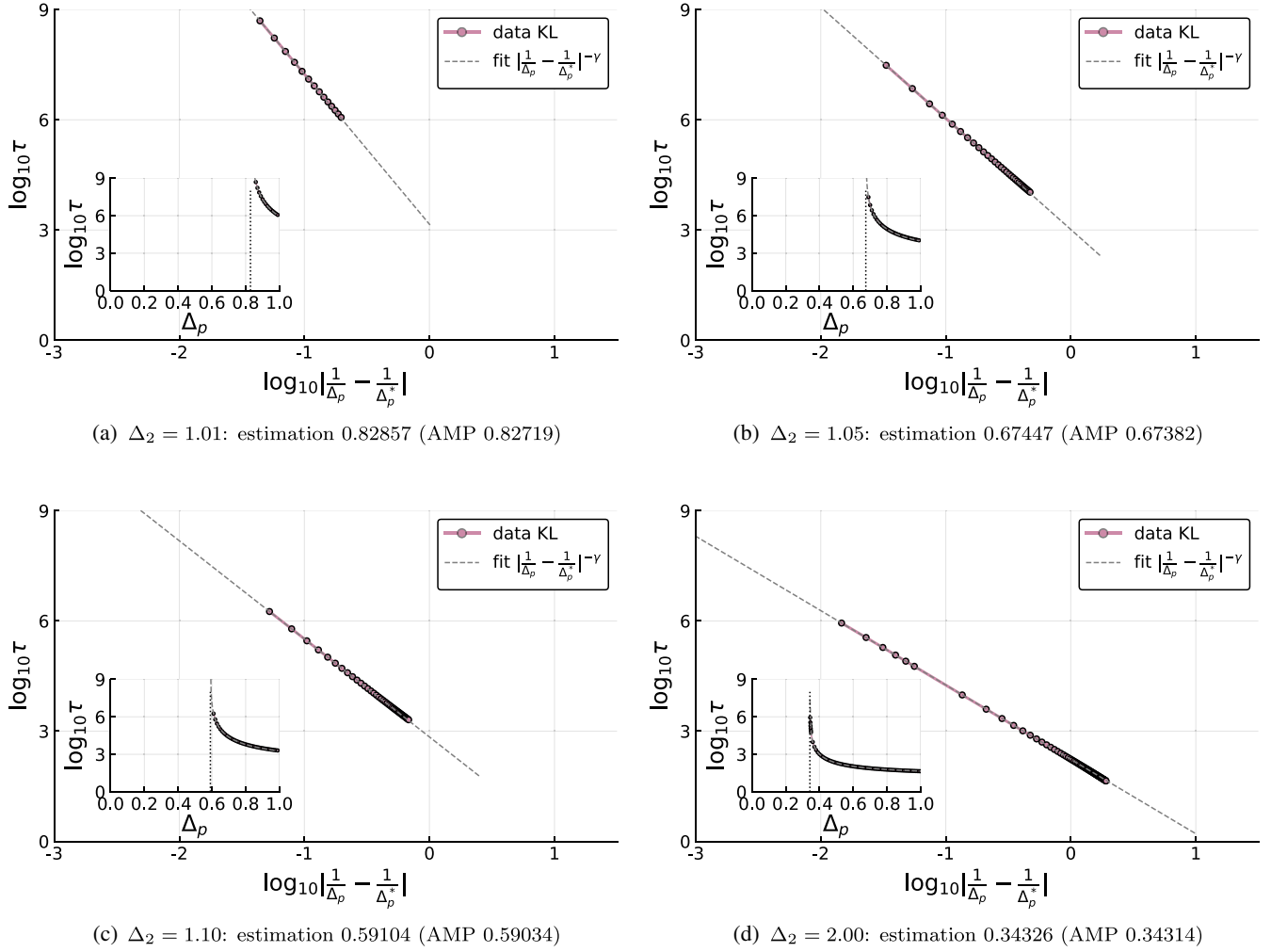
(a) $\Delta_2 = 1.01$: estimation 0.82857 (AMP 0.82719)

(b) $\Delta_2 = 1.05$: estimation 0.67447 (AMP 0.67382)

(c) $\Delta_2 = 1.10$: estimation 0.59104 (AMP 0.59034)

(d) $\Delta_2 = 2.00$: estimation 0.34326 (AMP 0.34314)

FIG. 17. Relaxation time obtained from the LSE starting from an informative initial condition $\bar{C}_0 = 1$. The four cases refer to the $2 + 3$ model and are fitted with a power law, and the relaxation time appears to diverge very close to the point predicted by AMP (the AMP prediction is given in the captions).

$$T_2 \equiv 1,$$

$$T_p \equiv T_p(t) = 1 + \frac{C}{\Delta_p} e^{-\frac{t}{\tau_{\mathrm{ann}}}}. \tag{D28}$$

The constant C allows us to select, at the initial time, the desired effective $\Delta_{p,\mathrm{eff}} = \Delta_p + Ce^{-(t/\tau_{\mathrm{ann}})}$ far from (and much larger than) the original one. Instead, τ_{ann} chooses the speed of the annealing protocol. Figure 18 shows that using this protocol, we are able to enter in the Langevin hard region even with Δ_2 close to the AMP threshold. To this purpose, we initiate the effective $\Delta_{p,\mathrm{eff}}$ close to 100 (i.e., $C = 100$), very far from the Langevin hard region, and we use different speeds for the annealing of Δ_p (different colors in the figures). In the figures, we can observe that when approaching $\Delta_2 = 1$, we need slower and slower protocols (larger and larger τ_{ann}). This behavior is due to the fact that when approaching $\Delta_2 = 1$ with $\Delta_p = 100$, a longer time is required to gain a nontrivial overlap with the

solution. Evidence of this growing timescale at $\Delta_p = 100$ is given in Fig. 19, where we show the relaxation time for magnetizing the solution by varying Δ_2. In particular, we can observe that at $\Delta_2 = 0.70$, the relaxation time is of the order of 100 time units.

For the protocol to be successful, it is therefore crucial that the annealing time τ_{ann} is large enough to give the possibility of magnetizing the solution before $\Delta_{p,\mathrm{eff}}$ has significantly decreased towards Δ_p. According to this analysis, it is not surprising that in Fig. 18, for $\Delta_2 = 0.90$, the proposed protocol seems to be unsuccessful. For this value of the parameter Δ_2, the time to find a solution even with $\Delta_p = 100$ should be larger than 1000 time units, which is much larger than the τ_{ann} used and out of the time window of our numerical solution. However, with an annealing time large enough, it would, in principle, be possible to recover exactly the same boundaries of the AMP easy region.
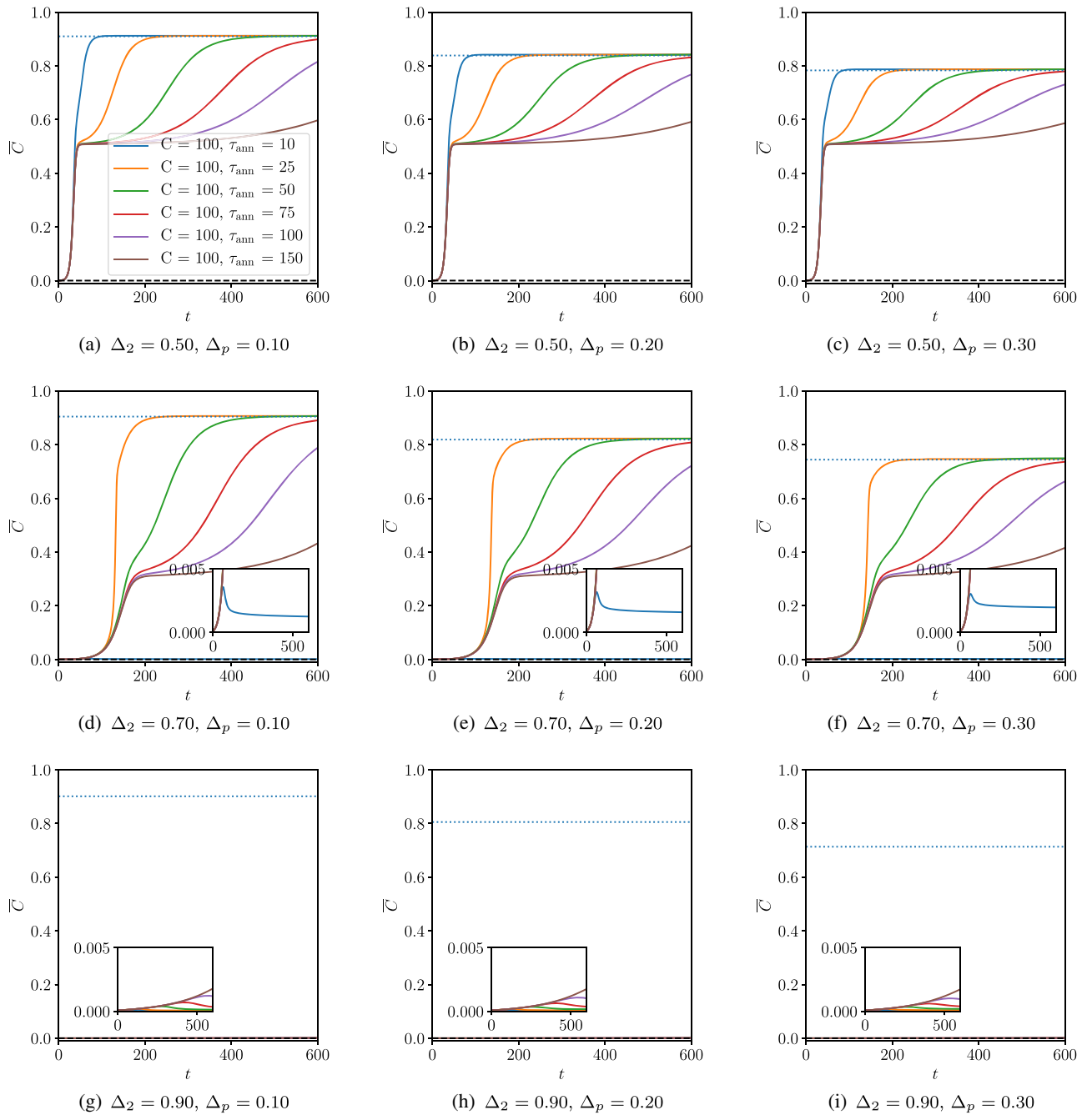
FIG. 18. The figures show the correlation with the signal in time obtained using different annealing protocols, whose details are reported in the legend of the first figure. All of the protocols have $C = 100$, which means that all the dynamics start with close effective $\Delta_p \sim 100$, $T_p(0)\Delta_p \simeq 100$ and an initial overlap $\bar{C}_0 = 10^{-4}$. What changes among the different lines is the relaxation speed, from the fastest (drawn in blue) to the slowest (drawn in brown). They are compared with the asymptotic value of AMP (dotted line) and the Langevin dynamics without tensor annealing (dashed line).

APPENDIX E: GLASSY NATURE OF THE LANGEVIN HARD PHASE: THE REPLICA APPROACH

In this section, we study the landscape of the spiked matrix-tensor problem following the approach of Ref. [25]. We underline here that we are interested in studying the free-energy landscape problem rather than the energy landscape since the former is the relevant quantity for finite temperatures ($\beta = 1$ in our case, as discussed in Appendix A). The results of Ref. [25] suggest that the AMP hard phase and part of the AMP easy phase are glassy. Therefore, we could expect that low-magnetization glassy states trap the Langevin algorithm and forbid the relaxation to the equilibrium configurations that surround the signal.
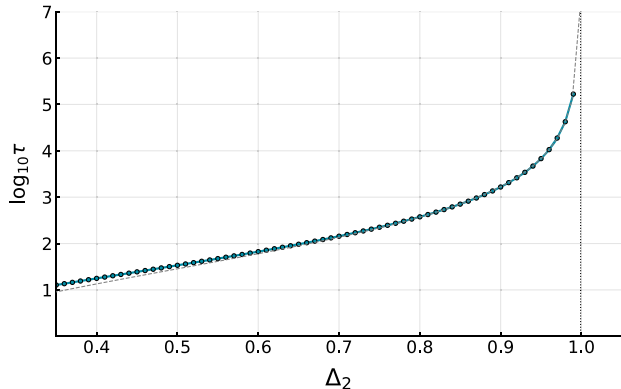
FIG. 19.   Relaxation times of Langevin dynamics at $\Delta_p = 100$ without using protocols starting from $\bar{C}_0 = 10^{-4}$.

This result may also occur in a region where AMP, instead, is perfectly fine producing configurations strongly correlated with the signal. In order to check this hypothesis, we compute the logarithm of the number of glassy states, called the complexity, by using the replica method [25,74]. The goal of this analysis is to trace an additional line in the phase diagram that delimits the region where stable one-step-replica symmetry-breaking (1RSB) metastable states exist. We conjecture that this method provides a physical lower bound to the Langevin hard phase in the $(\Delta_p, 1/\Delta_2)$ phase diagram.

1. Computation of the complexity through the replica method

The replica trick is based on the simple identity $\mathbb{E}\log x = \lim_{n\to 0}\frac{\partial}{\partial n}\mathbb{E}x^n$. Using this observation, we can compute the expected value of the free energy, $\Phi = -(\log Z)/N$, averaging the Z^n and taking the limit $n \to 0$. This process is, in general, as difficult as the initial problem; however, if we consider only integer n and extrapolate to 0, the computation becomes much less involved due to the fact that, for integer n, the average $\mathbb{E}x^n$ can sometimes be performed analytically. Indeed, in this case, the replicated partition function Z^n can be regarded as the partition function of n identical uncoupled systems or replicas. Averaging over the disorder, we obtain a clean system of interacting replicas. The Hamiltonian of this system displays an emerging replica symmetry since it is left unchanged by a permutation of replicas. This symmetry can be spontaneously broken in certain disordered models where frustration is sufficiently strong [44].

In mean field models characterized by fully connected factor graphs, the resulting Hamiltonian of interacting replicas depends on the configuration of the system only through a simple order parameter, the overlap $\tilde{Q}$ between them, which is an $n \times n$ matrix that describes the similarities of the configurations of different replicas in phase space. Furthermore, the Hamiltonian is proportional to N,

which means that in the thermodynamic limit $N \to \infty$, the model can be solved using the saddle-point method. In this case, one needs to consider a simple ansatz for the saddle-point structure of the matrix $\tilde{Q}$ that allows us to take the analytic continuation for $n \to 0$. The solution to this problem comes from spin-glass theory, and general details can be found in Ref. [44]. The saddle-point solutions for $\tilde{Q}$ can be classified according to the replica symmetry-breaking level, going from the replica symmetric solution where replica symmetry is not spontaneously broken to various degrees of spontaneous replica symmetry breaking (including full-replica symmetry breaking). Here, we will not review this subject, but the interested reader can find details in Ref. [44]. The model we are analyzing can be studied in full generality at any degree of RSB (see, e.g., Refs. [41,42,53], where the same models have been studied in the absence of a signal). However, here we limit ourselves to considering saddle-point solutions up to a 1RSB level.

The complexity of the landscape can be directly related to replica symmetry breaking. A replica symmetric solution implies an ergodic free-energy landscape characterized by a single pure state. When replica symmetry is broken instead, a large number of pure states arise, and the phase space gets clustered in a hierarchical way [44]. Making a 1RSB approximation means looking for a situation in which the hierarchical organization contains just one level: The phase space gets clustered into an exponential number of pure states with no further internal structure.

If we assume a 1RSB glassy landscape, we can compute the complexity of metastable states using a recipe from Monasson [74] (see also Ref. [75] for a pedagogical introduction). The argument goes as follows.

Let us consider a system with x *real* replicas infinitesimally coupled. If the free-energy landscape is clustered into an exponential number of metastable states, the replicated partition function, namely, the partition function of the system of x real replicas, can be written as

$$Z^x \simeq e^{N[\Sigma(f^*)-x\beta f^*]},$$

where f^* is the internal free energy of the dominant metastable states that is determined by the saddle-point condition $(d\Sigma/df)(f^*) = \beta x$ and β is the inverse temperature. Note that since we are interested in the Bayes-optimal case, this corresponds to setting $\beta = 1$. In the analysis, we consider a generic β before taking the limits in order to derive the average energy by taking its derivative. The function $\Sigma(f)$ is the complexity of metastable states having internal entropy f. Therefore, using the free parameter x, we can reconstruct the form of $\Sigma(f)$ from the replicated free energy. In order to compute the replicated free energy, we need to apply the replica trick on the replicated system, $\overline{\log Z^x} = \lim_{n\to 0}(\partial/\partial n)\overline{(Z^x)^n}$. Calling the replicated free energy $\Phi = -(1/N)\overline{\log Z^x}$, we get the complexity $\Sigma = x(\partial\Phi/\partial x) - \Phi$.

We can now specify the computation to our case, where the partition function is the normalization of the posterior measure. With simple manipulations of the equations [71], the partition function can be expressed as the integral over the overlap matrix,

$$\overline{(Z^x)^n} = \overline{Z_x^n} \propto \int \prod_{ab} dQ_{ab} e^{\beta NnxS(Q)} \simeq \lim_{Q} \sup e^{\beta NnxS(Q)}, \tag{E1}$$

where the overlap Q is a $(nx+1) \times (nx+1)$ matrix

$$Q = \begin{pmatrix} 1 & m \cdots m \\ \hline m & \\ \vdots & \tilde{Q} \\ m & \end{pmatrix}$$

that contains a special row and column that encode the overlap between different replicas with the signal; therefore, the corresponding overlap is the magnetization m.

The 1RSB structure for the matrix $\tilde{Q}$ can be obtained by defining the following $nx \times nx$ matrices: the identity matrix $\mathbb{I}_{ij} = \delta_{ij}$, the full matrix $\mathbb{J}^{(0)}_{nx,ij} = 1$, and a block diagonal matrix $\mathbb{J}^{(1)}_{nx} = \mathrm{diag}(J_x^{(0)}, \ldots, J_x^{(0)})$, where the diagonal blocks $J_x^{(0)}$ have size $x \times x$ and are matrices full of 1s. In this case, the 1RSB ansatz for $\tilde{Q}$ reads

$$\tilde{Q} = (1 - q_M)\mathbb{I}_{nx} + (q_M - q_m)\mathbb{J}^{(1)}_{nx} + q_m\mathbb{J}^{(0)}_{nx}.$$

Using this ansatz, we can compute $S(Q)$, which is given by

$$\begin{aligned} \beta S(Q) &= \frac{1}{nx}\left[\frac{1}{2}\log \det Q + \frac{\beta^2}{2p\Delta_p}\sum_{a,b=1}^n Q_{ab}^p + \frac{\beta^2}{4\Delta_2}\sum_{a,b=1}^n Q_{ab}^2 + \frac{\beta}{p\Delta_p}\sum_{a=1}^n Q_{0a}^p + \frac{\beta}{2\Delta_2}\sum_{a=1}^n Q_{0a}^2\right] \\ &= \frac{1}{2}\log(1 - q_M) + \frac{1}{2x}\log\frac{1 - q_M + x(q_M - q_m)}{1 - q_M} + \frac{1}{2}\frac{q_m - m^2}{1 - q_M + x(q_M - q_m)} \\ &\quad + \frac{\beta^2}{2p\Delta_p}\left(1 - q_M^p + x(q_M^p - q_m^p) + \frac{2}{\beta}m^p\right) + \frac{\beta^2}{4\Delta_2}\left(1 - q_M^2 + x(q_M^2 - q_m^2) + \frac{2}{\beta}m^2\right). \end{aligned} \tag{E2}$$

From Eq. (E2), we obtain the saddle-point equations

$$0 = 2\frac{\partial S}{\partial q_M} = (x-1)\left[\frac{1}{x}\left(\frac{1}{1 - q_M + x(q_M - q_m)} - \frac{1}{1 - q_M}\right) - \frac{q_m - m^2}{[1 - q_M + x(q_M - q_m)]^2} + \beta^2\left(\frac{q_M^{p-1}}{\Delta_p} + \frac{q_M}{\Delta_2}\right)\right],$$

$$0 = 2\frac{\partial S}{\partial q_m} = x\left[\frac{q_m - m^2}{[1 - q_M + x(q_M - q_m)]^2} - \beta^2\left(\frac{q_m^{p-1}}{\Delta_p} + \frac{q_m}{\Delta_2}\right)\right],$$

$$0 = \frac{\partial S}{\partial m} = \frac{-m}{1 - q_M + x(q_M - q_m)} + \beta^2\left(\frac{m^{p-1}}{\Delta_p} + \frac{m}{\Delta_2}\right). \tag{E3}$$

The above 1RSB fixed-point equations can be used to derive the de Almeida-Thouless instability of the RS solution towards 1RSB. This stability condition, sometimes called the replicon, also determines the overlap of the marginal threshold states. The stability analysis is performed by expanding Eqs. (E3) in small parameters $q_M - q_m = \varepsilon \ll 1$ and investigating whether, under iterations, such a small difference grows or decreases. This process leads directly to the threshold condition on the overlap,

$$\frac{1}{\beta^2(1 - q^{\mathrm{th}})^2} = (p - 1)\frac{(q^{\mathrm{th}})^{(p-2)}}{\Delta_p} + \frac{1}{\Delta_2}. \tag{E4}$$

This condition is then used in the derivation of the Langevin threshold (15) in the main text.

From Eq. (E2), we also obtain the averaged energy

$$\begin{aligned} E &= \frac{\partial \overline{\log Z^x}}{\partial \beta}\bigg|_{\beta=1} = \frac{1 - q_M^p + x(q_M^p - q_m^p) + m^p}{p\Delta_p} \\ &\quad + \frac{1 - q_M^2 + x(q_M^2 - q_m^2) + m^2}{2\Delta_2}. \end{aligned} \tag{E5}$$

In particular, the threshold states are characterized by $q_M = q^{\mathrm{th}}$ [fixed by Eq. (E4)], $q_m = 0$, and $m = 0$. Imposing these values, we can use the saddle-point equation for q_M, Eq. (E3), to fix the Parisi parameters x,

$$x(q^{\mathrm{th}}) = \frac{1}{(1 - q^{\mathrm{th}})[\frac{(q^{\mathrm{th}})^{p-1}}{\Delta_p} + \frac{q^{\mathrm{th}}}{\Delta_2}]} - \frac{1}{q^{\mathrm{th}}} + 1. \tag{E6}$$

These pieces together give Eq. (12) in the main text.

Having obtained the energy, we can consider $\beta = 1$ fixed for the rest of the analysis. We can observe that starting from this expression, we can derive the RS free energy, Eq. (B18), $q_M = q_m$, or equivalently in the limit $x \to 1$. The low-magnetization solution to these equations gives the complexity of the metastable branch of the posterior measure, which is given by

$$-\Sigma(x; Q^*) = -\frac{1}{2} \log \frac{1 - q_M + x(q_M - q_m)}{1 - q_M}$$
$$+ \frac{x}{2} \frac{q_M - q_m}{1 - q_M + x(q_M - q_m)}$$
$$- \frac{x^2}{2} \frac{(q_m - m^2)(q_M - q_m)}{[1 - q_M + x(q_M - q_m)]^2}$$
$$+ \frac{x^2}{2} \frac{q_M^p - q_m^p}{p \Delta_p} + \frac{x^2}{2} \frac{q_M^2 - q_m^2}{2 \Delta_2}. \quad (E7)$$

The free parameter x allows us to tune the free energy of the states for which we compute the complexity. Thus, we can characterize the part of the phase diagram where an exponential number of states is present.

To complete the 1RSB analysis, we compute the stability of the 1RSB saddle-point solution for Q. This computation is done analogously to the derivation of the replicon condition (E4), analyzing the stability of the 1RSB towards further replica symmetry breaking. Following Refs. [53,76], we obtain two replicon eigenvalues given by

$$\lambda_I = 1 - (1 - q_M + x(q_M - q_m))^2 \left[(p-1) \frac{q_m^{p-2}}{\Delta_p} + \frac{1}{\Delta_2} \right],$$
$$(E8)$$

$$\lambda_{II} = 1 - (1 - q_M)^2 \left[(p-1) \frac{q_M^{p-2}}{\Delta_p} + \frac{1}{\Delta_2} \right]. \quad (E9)$$

We can analyze what happens to the landscape when we fix $\Delta_p < 1$, start from a large value of $\Delta_2 < \Delta_{2,\text{dyn}}(\Delta_p)$, and decrease Δ_2. In this case, for sufficiently high Δ_2 and large enough Δ_p, the system is in a paramagnetic phase, and no glassy states are present. At the dynamical transition line, instead, we find a positive complexity as plotted in Fig. 20. At this point, the equilibrium states that dominate the posterior measure are the so-called threshold states for which the complexity is maximal. For those states, the eigenvalue $\lambda_{II} = 0$, which confirms that these states are marginally stable [27]. Decreasing Δ_2, one crosses the information-theoretic phase transition where the relevant metastable states that dominate the posterior measure have zero complexity. This process corresponds to a freezing-condensation-Kauzmann transition. Below the information-theoretic phase transition, the thermodynamics of the posterior measure is dominated by the state containing the signal. However, one can neglect the high-magnetization solution of the 1RSB equations to get the properties of the metastable branch and compute the complexity of states that have zero overlap with the signal. The complexity curves as a function of the Parisi parameter x for decreasing values of Δ_2 are plotted in Fig. 20 for fixed $\Delta_p = 0.5$ and several Δ_2. The curves contain a stable 1RSB part and an unstable one where λ_{II} is negative. The 1RSB line shown in Fig. 14 is obtained by looking at when the states with positive complexity and $\lambda_{II} = 0$ disappear. This method gives us the point where the 1RSB marginally stable states disappear, and therefore, it is expected to be a lower bound for the disappearance of glassiness in the phase diagram. The important outcome of this analysis is that, for $\Delta_2 < 1$ but not sufficiently small, namely, in part of the AMP easy phase, the replica analysis predicts the existence of 1RSB marginally stable glassy states that may stop the Langevin algorithm from relaxing towards the signal [25]; therefore, it supports the existence of the Langevin hard phase. This approach, however, does not quantitatively correctly
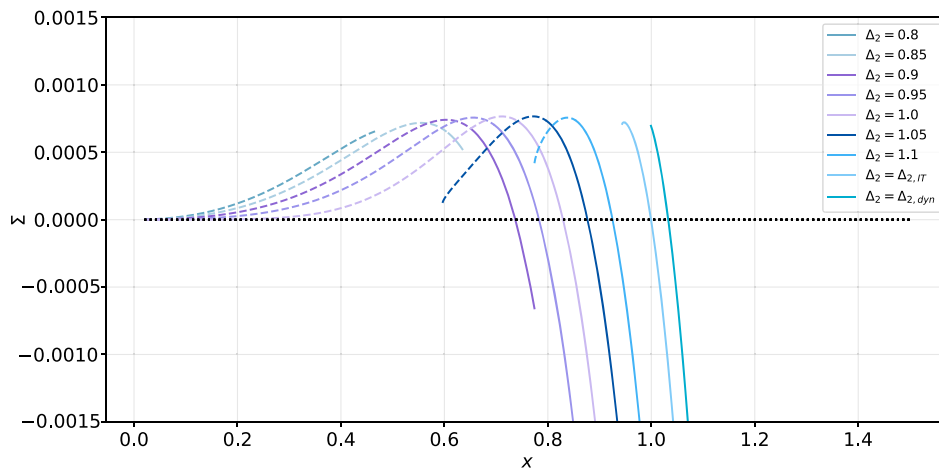

FIG. 20.   Complexity as a function of the Parisi parameter x for $p = 3$ on the line $\Delta_p = 0.5$. The solid line characterizes the stable part of the complexity, while the dashed line is the unstable one.

(a) $p = 3; \Delta_p = 0.50; \Delta_2 = 0.85$

(b) $p = 3; \Delta_p = 0.50; \Delta_2 = 0.90$

(c) $p = 3; \Delta_p = 0.50; \Delta_2 = 0.95$
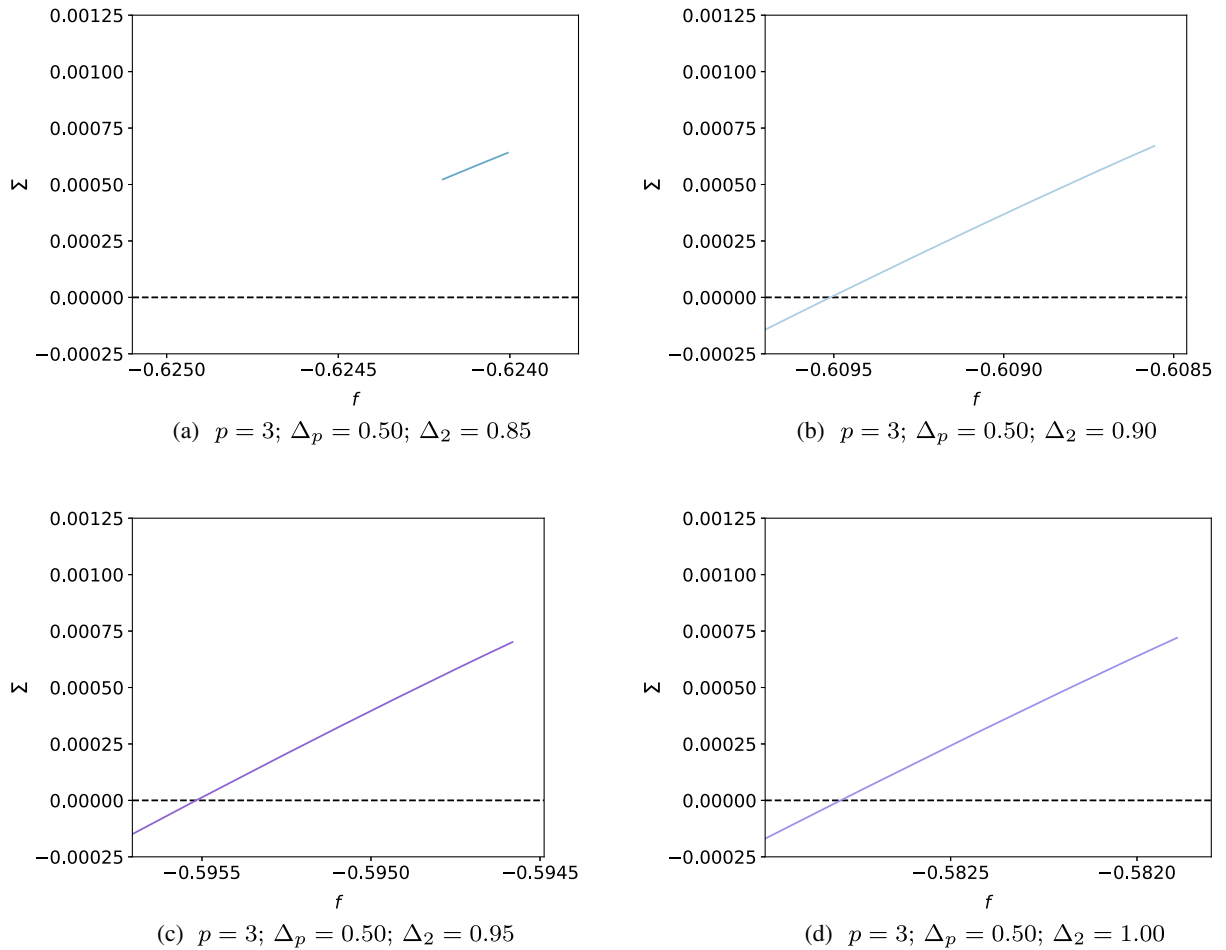
(d) $p = 3; \Delta_p = 0.50; \Delta_2 = 1.00$

FIG. 21.   The stable part of the 1RSB complexity as a function of the free energy for $p = 3$ and $\Delta_p = 0.5$.

predict the extent on the Langevin hard phase for reasons that remain obscure and should be investigated further.

Finally, in Fig. 21, we plot the complexity as a function of the internal free energy of the metastable states for some values of Δ_2 and Δ_p.

2. Breakdown of the fluctuation-dissipation theorem in the Langevin hard phase

When the Langevin algorithm is able to reach equilibrium, since it is the signal or the paramagnetic state, it should satisfy the fluctuation-dissipation theorem (FDT), according to which the response function is related to the correlation function through $R(t, t') = -[\partial C(t, t')/\partial t]$. Furthermore, time translational invariance (TTI) should arise, implying that both correlation and response functions should be functions of only the time difference, meaning that $R(t, t') = R(t - t')$ and $C(t, t') = C(t - t')$; note that all one-time quantities are constant in equilibrium. When the dynamics is run in the glass phase, metastable states may forbid equilibration. In this case, time translational invariance is never reached at long times; it should be reached only on exponential timescales in the system size,

and the dynamics displays aging violating at the same time as the FDT relation. The analysis of the asymptotic aging dynamics has been given by Cugliandolo and Kurchan in Refs. [27,48] (see also Ref. [70] for a pedagogical review) in the simplest spin-glass model (see also Ref. [77] for a much more complex situation), where no signal is present. The outcome of this work is that when the that dynamics started from random initial conditions is run in the glass phase, it drives the system to surf on the threshold states. In the model analyzed in Ref. [27], these states correspond to the 1RSB marginally stable glassy states that maximize the complexity. In this section, we analyze the Cugliandolo-Kurchan scenario by contrasting the numerical solution of the dynamical equations with the replica analysis of the complexity. According to Ref. [27], the long time Langevin dynamics, but still for times that are not exponentially large in the system size N, can be characterized by two time regimes. For short time differences $t - t' \sim \mathcal{O}(1)$ and $t' \to \infty$, the system obeys the FDT theorem and TTI; this regime can be understood as a first fast local equilibration in the nearest metastable state available. On a longer timescale $t - t' \to \infty$ and $t/t' < \infty$, the dynamics surfs on threshold states, and FDT and TTI are both violated. In
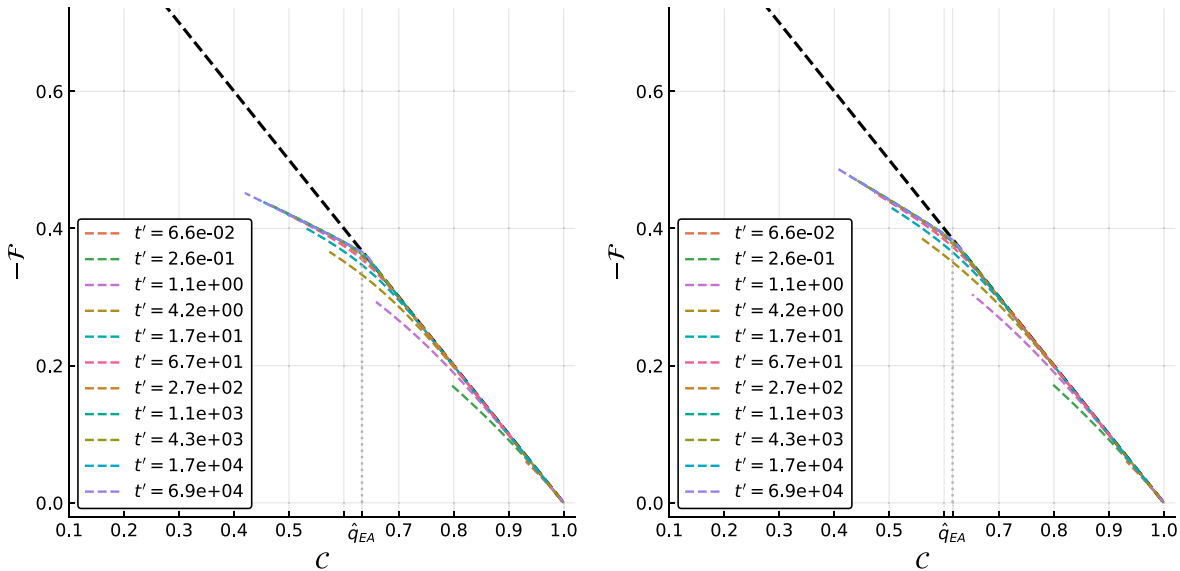
FIG. 22. Left panel: Parametric plot of the integrated response function with respect to the correlation function for $p = 3$, $\Delta_2 = 0.8$, and $\Delta_p = 0.2$. The different lines represent different waiting times, t'. The black dashed line corresponds to the FDT prediction $x = 1$. The vertical dotted line is the point where we observe a kink, which we denote by $\mathcal{C} = \hat{q}_{EA}$ and should be equal to the saddle-point value of q_M as extracted from the 1RSB threshold states in the replica computation [27]: $\hat{q}_{EA} = 0.633$ and $q_M = 0.638$. For $\mathcal{C}$ smaller than q_{EA}, the FDT is violated and is replaced by a generalized version as in Eq. (E10). We can obtain the value of the FDT ratio from a fit of the slope of the asymptotic curves for $\mathcal{C} < \hat{q}_{EA}$. We obtain $\hat{x} = 0.397$, which should be compared with the Parisi parameter that corresponds to 1RSB marginally stable states obtained from the replica computation, that is, $x = 0.408$. Right panel: Parametric plot of the integrated response as a function of the correlation for $p = 3$ and $\Delta_2 = 1.4$ and $\Delta_p = 0.2$. In this case, the value of the FDT ratio extracted from fitting the data is $\hat{x} = 0.397$, to be compared with the value of the Parisi parameter for the 1RSB threshold states, that is, $x = 0.408$. At the same time, data give $\hat{q}_{EA} = 0.633$, while the replica computation gives $q_M = 0.638$.

this time window, both the response and the correlation functions become functions of $\lambda = h(t)/h(t')$, with $h(t)$ an arbitrary reparametrization of the time variable. The function $h(t)$ must be a monotonously increasing function. The asymptotic reparametrization invariance is a key property of the dynamical equations [27]. By defining $\mathcal{C}(\lambda) = C(t, t')$ and $\mathcal{R}(\lambda) = tR(t, t')$, the Cugliandolo-Kurchan solution implies that, in this aging regime, the FDT relation can be generalized to

$$\mathcal{R}(\lambda) = x\,\mathcal{C}'(\lambda), \qquad (E10)$$

with x an effective FDT ratio that controls how much the FDT is violated. In the scenario of Ref. [27], the value of x coincides with the 1RSB Parisi parameter that corresponds to threshold states computed within the replica approach. In order to test this picture, we follow Cugliandolo and Kurchan [78], and we plot the integrated response $\mathcal{F}(t, t') = -\int_{t'}^{t} R(t, t'')dt''$ as a function of $C(t, t')$ in a parametric way. This process is given in Fig. 22.

If FDT holds at all timescales, one should see a straight line with slope -1. Instead, what we see in the Langevin hard phase is that for large values of t', the curves approach two straight lines asymptotically for $t' \gg 1$. For high values of $\mathcal{C}$, meaning for short time differences $t - t' \sim \mathcal{O}(1)$, the slope of the straight line is -1, which means that

$\mathcal{F} = 1 - \mathcal{C}$, as implied by the short time FDT relation. On longer timescales, FDT is violated, confirming the glassiness of the Langevin hard phase. By doing a linear fit, we can use the data plotted in Fig. 22 to estimate the FDT ratio x appearing in Eq. (E10). This estimate can be compared with the Parisi parameter x for which we have marginally stable 1RSB states. Overall, we find very good agreement (data coming from the fit are reported in the caption of Fig. 22). The small discrepancy between the two values of x can be either due to the numerical accuracy in solving the dynamical equations or the possibility that the 1RSB threshold is not exactly the one that characterizes the long time dynamics. Further investigations are needed to clarify this point. Finally, according to Ref. [27], the value of $\mathcal{C}$ at which the two straight lines cross should coincide with the value of q_M computed for the threshold states within the 1RSB solution. Again, we find very good agreement.

APPENDIX F: FREE-ENERGY HESSIAN, BBP TRANSITION, AND LANGEVIN THRESHOLD

In the following, we present the derivation and the analysis of the Langevin threshold based on the study of the free-energy Hessian. The starting point of the analysis is the so-called TAP free energy, i.e., the free energy as a function of the local magnetizations. The TAP free energy was introduced in the early days of spin-glass theory [44,50]

and is now receiving a lot of attention in the mathematical community; see, e.g., Ref. [51]. A straightforward generalization of the results of Ref. [79] allows one to obtain the TAP free energy for the model considered in this work, i.e., for the Hamiltonian (4):

$$F(\{m_i\}) = -\frac{\sqrt{(p-1)!}}{\Delta_p N^{(p-1)/2}} \sum_{i_1 < \ldots < i_p} T_{i_1 \ldots i_p} m_{i_1} \ldots m_{i_p}$$
$$- \frac{1}{\Delta_2 \sqrt{N}} \sum_{i<j} Y_{ij} m_i m_j + f(q)N,$$

where we set the temperature to 1, $q = \sum_i m_i^2 / N$, and $f(q)$ reads

$$f(q) = -\frac{1}{2}\log(1-q) - \frac{1}{2p\Delta_p}[1 + (p-1)q^p - pq^{p-1}]$$
$$- \frac{1}{4\Delta_2}[1 + q^2 - 2q].$$

The so-called TAP states are local minima of $F(\{m_i\})$. We are interested in the free-energy Hessian evaluated at the TAP states having zero overlap with the signal:

$$\frac{\partial^2 F}{\partial m_i \partial m_j} = G_{ij} + \delta_{ij}f'(q) - \frac{1}{\Delta_2}\frac{x_i^* x_j^*}{N} + f''(q)\frac{m_i m_j}{N}, \quad \text{(F1)}$$

where the matrix G_{ij} is defined as

$$G_{ij} = -\frac{1}{\Delta_2 \sqrt{N}}\xi_{ij}$$
$$- \frac{\sqrt{(p-1)!}}{\Delta_p N^{(p-1)/2}(p-2)!} \sum_{i_1, \ldots, i_{p-2}} \xi_{iji_1 \ldots i_{p-2}} m_{i_1} \ldots m_{i_{p-2}}.$$

As shown originally in the spin-glass literature [80], and recently put on a firmer basis by the Kac-Rice method [22,32], the matrix G_{ij} is statistically equivalent to a random matrix belonging to the Gaussian orthogonal ensemble (GOE). In our case, the corresponding GOE matrix has elements that are i.i.d. Gaussian random variables with mean zero and variance σ_F^2/N, where

$$\sigma_F^2(q) = \frac{(p-1)q^{p-2}}{\Delta_p} + \frac{1}{\Delta_2}.$$

Neglecting for the moment the last two terms in Eq. (14), one finds that the free-energy Hessian is the sum of a GOE matrix and the identity multiplied by $f'(q)$. The corresponding density of eigenvalues is therefore the Wigner semicircle with support $[-2\sigma_F(q)+f'(q), 2\sigma_F(q)+f'(q)]$. This result is valid for any TAP state. The threshold states, which are the ones trapping the Langevin dynamics, are characterized by a vanishing fraction of zero modes; i.e., the left edge of the

support of the Wigner semicircle is zero. Their overlap is therefore fixed by the equation

$$2\sigma_F(q_{th}) = f'(q_{th}) \rightarrow \frac{1}{1 - q_{th}} = \sqrt{\frac{(p-1)q_{th}^{p-2}}{\Delta_p} + \frac{1}{\Delta_2}}.$$
$$\text{(F2)}$$

Let us consider now the role of the last two terms in Eq. (14). Both are rank-one perturbations and hence can lead to a BBP transition [15], i.e., an eigenvalue that pops out of the Wigner semicircle with an eigenvector having a finite overlap in the direction of the perturbation. It can be easily checked that $f''(q_{th}) \geq 0$; therefore, the last term cannot lead to any negative eigenvalue and does not play any role in determining the stability of the threshold states. It is the other term that is responsible for the instability in the direction of the signal. In fact, it is the contribution due to the spike; it becomes larger when the signal-to-noise ratio, $1/\Delta_2$, increases.

The condition for the BBP transition for a GOE matrix having elements with variance σ_F^2/N, which is perturbed by a rank-one perturbation of strength $1/\Delta_2$, is $(1/\Delta_2) = \sigma_F$. The equation for the Langevin threshold is

$$\frac{1}{\Delta_2} = \sqrt{\frac{(p-1)q_{th}^{p-2}}{\Delta_p} + \frac{1}{\Delta_2}}.$$

Together with Eq. (F2), this equation leads to Eq. (15) presented in the main text and implies $q_{th} = 1 - \Delta_2^*$ at the Langevin threshold.

---

[1] L. Bottou, *Large-Scale Machine Learning with Stochastic Gradient Descent*, in *Proceedings of COMPSTAT'2010* (Springer, 2010), pp. 177–186.

[2] M. Welling and Y. W. Teh, *Bayesian Learning via Stochastic Gradient Langevin Dynamics*, in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (2011), pp. 681–688.

[3] Y. LeCun, Y. Bengio, and G. Hinton, *Deep Learning*, Nature (London) **521**, 436 (2015).

[4] J.-P. Bouchaud, L. F. Cugliandolo, J. Kurchan, and M. Mézard, *Out of Equilibrium Dynamics in Spin-Glasses and Other Glassy Systems*, Spin Glasses and Random Fields 161 (1998).

[5] H. S. Seung, H. Sompolinsky, and N. Tishby, *Statistical Mechanics of Learning from Examples*, Phys. Rev. A **45**, 6056 (1992).

[6] L. Zdeborová and F. Krzakala, *Statistical Physics of Inference: Thresholds and Algorithms*, Adv. Phys. **65**, 453 (2016).

[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] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, *Optimal Errors and Phase Transitions in High-Dimensional Generalized Linear Models*, Proc. Natl. Acad. Sci. U.S.A. **116**, 5451 (2019).

[9] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborov, *Mutual Information for Symmetric Rank-One Matrix Estimation: A Proof of the Replica Formula*, in *Advances in Neural Information Processing Systems 29* (2016), pp. 424–432.

[10] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová, *Statistical and Computational Phase Transitions in Spiked Tensor Estimation*, in *2017 IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2017), pp. 511–515.

[11] B. Aubin, A. Maillard, J. Barbier, F. Krzakala, N. Macris, and L. Zdeborová, *The Committee Machine: Computational to Statistical Gaps in Learning a Two-Layers Neural Network*, in *Advances in Neural Information Processing Systems* (2018).

[12] D. L. Donoho, A. Maleki, and A. Montanari, *Message-Passing Algorithms for Compressed Sensing*, Proc. Natl. Acad. Sci. U.S.A. **106**, 18914 (2009).

[13] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, *Spectral Redemption in Clustering Sparse Networks*, Proc. Natl. Acad. Sci. U.S.A. **110**, 20935 (2013).

[14] S. B. Hopkins and D. Steurer, *Bayesian Estimation from Few Samples: Community Detection and Related Problems*, arXiv:1710.00264.

[15] J. Baik, G. B. Arous, S. Péché *et al.*, *Phase Transition of the Largest Eigenvalue for Nonnull Complex Sample Covariance Matrices*, Ann. Probability **33**, 1643 (2005).

[16] I. M. Johnstone and A. Y. Lu, *On Consistency and Sparsity for Principal Components Analysis in High Dimensions*, J. Am. Stat. Assoc. **104**, 682 (2009).

[17] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky, *Tensor Decompositions for Learning Latent Variable Models*, J. Mach. Learn. Res. **15**, 2773 (2014).

[18] E. Richard and A. Montanari, *A Statistical Model for Tensor PCA*, in *Advances in Neural Information Processing Systems* (2014), pp. 2897–2905.

[19] S. B. Hopkins, J. Shi, and D. Steurer, *Tensor Principal Component Analysis via Sum-of-Square Proofs*, in *Conference on Learning Theory* (2015), pp. 956–1006.

[20] R. Ge and T. Ma, *On the Optimization Landscape of Tensor Decompositions*, in *Adv. Neural Inf. Process. Syst.* (2017), pp. 3653–3663.

[21] G. B. Arous, R. Gheissari, and A. Jagannath, *Algorithmic Thresholds for Tensor PCA*, arXiv:1808.00921.

[22] V. Ros, G. B. Arous, G. Biroli, and C. Cammarota, *Complex Energy Landscapes in Spiked-Tensor and Simple Glassy Models: Ruggedness, Arrangements of Local Minima, and Phase Transitions*, Phys. Rev. X **9**, 011003 (2019).

[23] F. Krzakala and L. Zdeborová, *Hiding Quiet Solutions in Random Constraint Satisfaction Problems*, Phys. Rev. Lett. **102**, 238701 (2009).

[24] 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).

[25] F. Antenucci, S. Franz, P. Urbani, and L. Zdeborová, *Glassy Nature of the Hard Phase in Inference Problems*, Phys. Rev. X **9**, 011020 (2019).

[26] A. Crisanti, H. Horner, and H.-J. Sommers, *The Spherical p-spin Interaction Spin-Glass Model*, Z. Phys. B **92**, 257 (1993).

[27] L. F. Cugliandolo and J. Kurchan, *Analytical Solution of the Off-Equilibrium Dynamics of a Long-Range Spin-Glass Model*, Phys. Rev. Lett. **71**, 173 (1993).

[28] G. B. Arous, A. Dembo, and A. Guionnet, *Cugliandolo-Kurchan Equations for Dynamics of Spin-Glasses*, Probab. Theory Relat. Fields **136**, 619 (2006).

[29] Y. Deshpande and A. Montanari, *Information-Theoretically Optimal Sparse PCA*, in *2014 IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2014), pp. 2197–2201.

[30] D. J. Gross and M. Mézard, *The Simplest Spin Glass*, Nucl. Phys. **B240**, 431 (1984).

[31] Y. V. Fyodorov, *Complexity of Random Energy Landscapes, Glass Transition, and Absolute Value of the Spectral Determinant of Random Matrices*, Phys. Rev. Lett. **92**, 240601 (2004).

[32] A. Auffinger, G. B. Arous, and J. Černý, *Random Matrices and Complexity of Spin Glasses*, Commun. Pure Appl. Math. **66**, 165 (2013).

[33] L. Sagun, V. U. Guney, G. B. Arous, and Y. LeCun, *Explorations on High Dimensional Landscapes*, arXiv:1412.6615.

[34] G. B. Arous, S. Mei, A. Montanari, and M. Nica, *The Landscape of the Spiked Tensor Model*, Commun. Pure Appl. Math. **72**, 2282 (2019).

[35] A. Anandkumar, Y. Deng, R. Ge, and H. Mobahi, *Homotopy Analysis for Tensor PCA*, arXiv:1610.09322.

[36] A. S. Wein, A. E. Alaoui, and C. Moore, *The Kikuchi Hierarchy and Tensor PCA*, arXiv:1904.03858.

[37] T. Lesieur, F. Krzakala, and L. Zdeborová, *Constrained Low-Rank Matrix Estimation: Phase Transitions, Approximate Message Passing and Applications*, J. Stat. Mech. (2017) 073403.

[38] M. Lelarge and L. Miolane, *Fundamental limits of symmetric low-rank matrix estimation*, Probab. Theory Relat. Fields **173**, 1 (2016).

[39] D. J. Thouless, P. W. Anderson, and R. G. Palmer, *Solution of 'Solvable Model of a Spin Glass,'* Philos. Mag. **35**, 593 (1977).

[40] A. Javanmard and A. Montanari, *State Evolution for General Approximate Message Passing Algorithms, with Applications to Spatial Coupling*, Inf. Inference **2**, 115 (2013).

[41] A. Crisanti and L. Leuzzi, *Spherical $2 + p$ Spin-Glass Model: An Exactly Solvable Model for Glass to Spin-Glass Transition*, Phys. Rev. Lett. **93**, 217203 (2004).

[42] A. Crisanti and L. Leuzzi, *Spherical $2 + p$ Spin-Glass Model: An Analytically Solvable Model with a Glass-to-Glass Transition*, Phys. Rev. B **73**, 014412 (2006).

[43] P. C. Martin, E. Siggia, and H. Rose, *Statistical Dynamics of Classical Systems*, Phys. Rev. A **8**, 423 (1973).

[44] M. Mézard, G. Parisi, and M.-A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[45] B. Kim and A. Latz, *The Dynamics of the Spherical p-spin Model: From Microscopic to Asymptotic*, Europhys. Lett. **53**, 660 (2001).

[46] L. Berthier, G. Biroli, J.-P. Bouchaud, W. Kob, K. Miyazaki, and D. Reichman, *Spontaneous and Induced Dynamic Fluctuations in Glass Formers. I. General Results and Dependence on Ensemble and Dynamics*, J. Chem. Phys. **126**, 184503 (2007).

[47] S. S. Mannelli, G. Biroli, C. Cammarota, F. Krzakala, P. Urbani, and L. Zdeborová, *Langevin State Evolution Integrators* (2018), available at https://github.com/sphinxteam/spiked_matrix-tensor.

[48] L. Cugliandolo and J. Kurchan, *Weak Ergodicity Breaking in Mean-Field Spin-Glass Models*, Philos. Mag. B **71**, 501 (1995).

[49] This is just an approximation because the relaxation to the threshold state is power law and only asymptotic. Therefore, our assumption is expected to provide a coarse-grained description of the short time dynamics. Whether it provides an exact description of what happens on long timescales remains an open problem.

[50] D. Thouless, P. Anderson, and R. Palmer, *Solution of a Solvable Model of a Spin Glass*, Philos. Mag. **35**, 593 (1977).

[51] W.-K. Chen and D. Panchenko, *On the Tap Free Energy in the Mixed p-spin Models*, Commun. Math. Phys. **362**, 219 (2018).

[52] S. F. Edwards and R. C. Jones, *The Eigenvalue Spectrum of a Large Symmetric Random Matrix*, J. Phys. A **9**, 1595 (1976).

[53] A. Crisanti and L. Leuzzi, *Exactly Solvable Spin–Glass Models with Ferromagnetic Couplings: The Spherical Multi-p-spin Model in a Self-Induced Field*, Nucl. Phys. **B870**, 176 (2013).

[54] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University, New York, 2009).

[55] S. Boucheron, G. Lugosi, and O. Bousquet, *Concentration Inequalities*, in *Advanced Lectures on Machine Learning* (Springer, New York, 2004), pp. 208–240.

[56] S. B. Korada and N. Macris, *Exact Solution of the Gauge Symmetric p-spin Glass Model on a Complete Graph*, J. Stat. Phys. **136**, 205 (2009).

[57] F. Krzakala, J. Xu, and L. Zdeborová, *Mutual Information in Rank-One Matrix Estimation*, in *2016 IEEE Information Theory Workshop (ITW)* (2016), pp. 71–75.

[58] M. Aizenman, R. Sims, and S. L. Starr, *Extended Variational Principle for the Sherrington-Kirkpatrick Spin-Glass Model*, Phys. Rev. B **68**, 214403 (2003).

[59] J. Barbier and N. Macris, *The Adaptive Interpolation Method: A Simple Scheme to Prove Replica Formulas in Bayesian Inference*, Probab. Theory Relat. Fields **174**, 1133 (2018).

[60] A. E. Alaoui and F. Krzakala, *Estimation in the Spiked Wigner Model: A Short Proof of the Replica Formula*, in *2018 IEEE International Symposium on Information Theory (ISIT)* (2018), pp. 1874–1878.

[61] J.-C. Mourrat, *Hamilton-Jacobi Equations for Mean-Field Disordered Systems*, arXiv:1811.01432.

[62] D. Guo, S. Shamai, and S. Verdú, *Mutual Information and Minimum Mean-Square Error in Gaussian Channels*, IEEE Trans. Inf. Theory **51**, 1261 (2005).

[63] H.-O. Georgii, *Gibbs Measures and Phase Transitions* (Walter de Gruyter, 2011), Vol. 9.

[64] N. Macris, *Griffith-Kelly-Sherman Correlation Inequalities: A Useful Tool in the Theory of Error Correcting Codes*, IEEE Trans. Inf. Theory **53**, 664 (2007).

[65] A. Montanari, *Estimating Random Variables from Random Sparse Observations*, European Transactions on Telecommunications and Related Technologies **19**, 385 (2008).

[66] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborova, *Information-Theoretic Thresholds from the Cavity Method*, in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)* (Oxford Publications, 2017), pp. 146–157.

[67] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Clarendon, Oxford, 2001), Vol. 111.

[68] F. Guerra and F. L. Toninelli, *The Thermodynamic Limit in Mean Field Spin Glass Models*, Commun. Math. Phys. **230**, 71 (2002).

[69] F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, *Typology of Phase Transitions in Bayesian Inference Problems*, Phys. Rev. E **99**, 042109 (2019).

[70] L. F. Cugliandolo, *Course 7: Dynamics of Glassy Systems*, in *Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter* (Springer, New York, 2003), pp. 367–521.

[71] T. Castellani and A. Cavagna, *Spin Glass Theory for Pedestrians*, J. Stat. Mech. (2005) P05012.

[72] E. Agoritsas, G. Biroli, P. Urbani, and F. Zamponi, *Out-of-Equilibrium Dynamical Mean-Field Equations for the Perceptron Model*, J. Phys. A **51**, 085002 (2018).

[73] L. Berthier, G. Biroli, J.-P. Bouchaud, W. Kob, K. Miyazaki, and D. R. Reichman, *Spontaneous and Induced Dynamic Correlations in Glass Formers. II. Model Calculations and Comparison to Numerical Simulations*, J. Chem. Phys. **126**, 184504 (2007).

[74] R. Monasson, *Structural Glass Transition and the Entropy of the Metastable States*, Phys. Rev. Lett. **75**, 2847 (1995).

[75] F. Zamponi, *Mean Field Theory of Spin Glasses*, arXiv:1008.4844.

[76] A. Crisanti and H.-J. Sommers, *The Spherical p-spin Interaction Spin Glass Model: The Statics*, Z. Phys. B **87**, 341 (1992).

[77] L. Cugliandolo and J. Kurchan, *On the Out-of-Equilibrium Relaxation of the Sherrington-Kirkpatrick Model*, J. Phys. A **27**, 5749 (1994).

[78] L. F. Cugliandolo and J. Kurchan, *Aging and Effective Temperatures in the Low Temperature Mode-Coupling Equations*, Prog. Theor. Phys. Suppl. **126**, 407 (1997).

[79] A. Crisanti and H.-J. Sommers, *Thouless-Anderson-Palmer Approach to the Spherical p-Spin Spin Glass Model*, J. Phys. I **5**, 805 (1995).

[80] A. Bray and M. Moore, *Evidence for Massless Modes in the 'Solvable Model' of a Spin Glass*, J. Phys. C **12**, L441 (1979).