Optimization of the Sherrington-Kirkpatrick Hamiltonian

Andrea Montanari
Department of Electrical Engineering and Department of Statistics
Stanford University
Stanford, CA, USA
montanari@stanford.edu

***Abstract*—Let A be a symmetric random matrix with independent and identically distributed Gaussian entries above the diagonal. We consider the problem of maximizing the quadratic form associated to A over binary vectors. In the language of statistical physics, this amounts to finding the ground state of the Sherrington-Kirkpatrick model of spin glasses. The asymptotic value of this optimization problem was characterized by Parisi via a celebrated variational principle, subsequently proved by Talagrand. We give an algorithm that, for any $\varepsilon > 0$, outputs a feasible solution whose value is at least $(1-\varepsilon)$ of the optimum, with probability converging to one as the dimension n of the matrix diverges. The algorithm's time complexity is of order n^2. It is a message-passing algorithm, but the specific structure of its update rules is new.**

As a side result, we prove that, at (low) non-zero temperature, the algorithm constructs approximate solutions of the Thouless-Anderson-Palmer equations.

***Keywords*-Sherrington-Kirkpatrick; spin glasses; replica symmetry breaking; message passing algorithms**

I. Introduction and main result

Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a random matrix from the $\mathsf{GOE}(n)$ ensemble. Namely, $\boldsymbol{A} = \boldsymbol{A}^{\mathsf{T}}$ and $(A_{ij})_{i\le j\le n}$ is a collection of independent random variables with $A_{ii} \sim \mathsf{N}(0, 2/n)$ and $A_{ij} = \mathsf{N}(0, 1/n)$ for $i < j$. We are concerned with the following optimization problem (here $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \sum_{i\le n} u_i v_i$ is the standard scalar product)

$$\begin{aligned} &\text{maximize } \ \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle \,, \\ &\text{subject to } \ \boldsymbol{\sigma} \in \{+1,-1\}^n \,. \end{aligned} \tag{1}$$

From a worst-case perspective, this problem is NP-hard and indeed hard to approximate within a sublogarithmic factor [3]. For random data $\boldsymbol{A}$, the energy function $H_n(\boldsymbol{\sigma}) = \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle/2$ is also known as the Sherrington-Kirkpatrick model [41]. Its properties have been intensely studied in statistical physics and probability theory for over 40 years as a prototypical example of complex energy landscape and a mean field model for spin glasses [28], [45], [37]. Generalizations of this model have been used to understand structural glasses, random combinatorial problems, neural networks, and a number of other systems [19], [29], [48], [33], [27].

In this paper we consider the computational problem of finding a vector $\boldsymbol{\sigma}_* \in \{+1,-1\}^n$ that is a near optimum, namely such that $H_n(\boldsymbol{\sigma}_*) \ge (1-\varepsilon)\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} H_n(\boldsymbol{\sigma})$. Under a widely believed assumption about the structure of the associated Gibbs measure (more precisely, on the support of the asymptotic overlap distribution) we prove that, for any $\varepsilon > 0$ there exists an algorithm with complexity $O(n^2)$ that –with high probability– outputs such a vector.

In order to state our assumption, we need to take a detour and introduce Parisi's variational formula for the value of the optimization problem (1). Let $\mathscr{P}([0,1])$ be the space of probability measures on the interval $[0,1]$ endowed with the topology of weak convergence. For $\mu \in \mathscr{P}([0,1])$, we will write (with a slight abuse of notation) $\mu(t) = \mu([0,t])$ for its distribution function. For $\beta \in \mathbb{R}_{\ge 0}$, consider the following parabolic partial differential equation (PDE) on $(t,x) \in [0,1]\times\mathbb{R}$

$$\begin{aligned} &\partial_t \Phi(t,x) + \frac{1}{2}\beta^2 \partial_{xx}\Phi(t,x) + \frac{1}{2}\beta^2 \mu(t)\big(\partial_x \Phi(t,x)\big)^2 = 0 \,, \\ &\Phi(1,x) = \log 2\cosh x \,. \end{aligned} \tag{2}$$

It is understood that this is to be solved backward in time with the given final condition at $t = 1$. Existence and uniqueness where proved in [23]. We will also write Φ_μ to emphasize the dependence of the solution on the measure μ. The Parisi functional is then defined as

$$\mathsf{P}_\beta(\mu) \equiv \Phi_\mu(0,0) - \frac{1}{2}\beta^2 \int_0^1 t\,\mu(t)\,\mathrm{d}t \,. \tag{3}$$

The relation between this functional and the original optimization problem is given by a remarkable variational principle, first proposed by Parisi [38] and established rigorously, more than twenty-five years later, by Talagrand [44], and Panchenko [36].

Theorem 1 (Talagrand [44])**.** *Consider the partition function* $Z_n(\beta) = \sum_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \exp\{\beta H_n(\boldsymbol{\sigma})\}$. *Then*

2575-8454/19/$31.00 ©2019 IEEE
DOI 10.1109/FOCS.2019.00087

we have, almost surely (and in L^1)

$$\lim_{n\to\infty}\frac{1}{n}\log Z_n(\beta) = \min_{\mu\in\mathscr{P}([0,1])}\mathsf{P}_\beta(\mu)\,. \qquad (4)$$

The following consequence for the optimization problem (1) is elementary, see e.g. [17].

Corollary I.1. *We have, almost surely*

$$\lim_{n\to\infty}\frac{1}{2n}\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n}\langle\boldsymbol{\sigma},\boldsymbol{A}\boldsymbol{\sigma}\rangle = \lim_{\beta\to\infty}\frac{1}{\beta}\min_{\mu\in\mathscr{P}([0,1])}\mathsf{P}_\beta(\mu)\,. \qquad (5)$$

Remark I.1. The limit $\beta\to\infty$ on the right-hand side of Eq. (5) can be removed by defining a new variational principle directly 'at $\beta=\infty$'. Namely, the right-hand side of Eq. (5) can be replaced by $\min_\gamma \hat{\mathsf{P}}(\gamma)$ where $\hat{\mathsf{P}}$ is a modification of P and the minimum is taked over a suitable functional space [5]. In this paper we use the $\beta<\infty$ formulation, but it should be possible to work directly at $\beta=\infty$: we defer such extensions to future work.

We also note that while we stated Theorem 1 and Corollary I.1 for simplicity in the case of $\boldsymbol{A}\sim\mathsf{GOE}(n)$, these results holds more generally for symmetric matrices $\boldsymbol{A}$ with independent entries above the diagonal, provided $\mathbb{E}\{A_{ij}\}=0$, $\mathbb{E}\{A_{ij}^2\}=1/n$ and $\mathbb{E}\{|A_{ij}|^3\}\le C/n^{3/2}$ [12]. (Indeed even weaker conditions are sufficient [17].

Existence and uniqueness of the minimizer of $\mathsf{P}_\beta(\,\cdot\,)$ were proved in [4] and [23], which also proved that $\mu\mapsto\mathsf{P}_\beta(\mu)$ is strongly convex. We will denote by μ_β the unique minimizer, and refer to it as the 'Parisi measure' or 'overlap distribution' at inverse temperature β. Our key assumption will be that –at large enough β– the support of μ_β is an interval $[0,q_*(\beta)]$.

Assumption 1 (No overlap gap)**.** *There exist $\beta_0<\infty$ such that, for any $\beta>\beta_0$, the function $t\mapsto\mu_\beta([0,t])$ is strictly increasing on $[0,q_*]$, where $q_*=q_*(\beta)$ and $\mu_\beta([0,q_*])=1$.*

This assumption is sometimes referred to as 'continuous replica symmetry breaking' or 'full replica symmetry breaking' and is widely believed to be true (with $\beta_0=1$) within statistical physics [28]. In particular, this conjecture is supported by high precision numerical solutions of the variational problem for P_β[15], [35], [40]. Rigorous evidence was recently obtained in [6]. Addressing this conjecture goes beyond the scope of the present paper.

Let us emphasize that the expression 'no overlap gap' captures the content of this assumption better than 'continuous' or 'full replica symmetry breaking.' Indeed, the latter are generally used whenever the support of the probability measure μ_β, $\mathrm{supp}(\mu_\beta)$, has infinite cardinality. In contrast, here we are requiring the stronger condition $\mathrm{supp}(\mu_\beta)=[0,q_*]$ (which implies $q_*>0$ for all $\beta>1$ [46]).

We are now in position to state our main result.

Theorem 2. *Under Assumption 1, for any $\varepsilon>0$ there exists an algorithm that takes as input the matrix $\boldsymbol{A}\in\mathbb{R}^{n\times n}$, and outputs $\boldsymbol{\sigma}_*=\boldsymbol{\sigma}_*(\boldsymbol{A})\in\{+1,-1\}^n$, such that the following hold:*

- *(i) The complexity (floating point operations) of the algorithm is at most $C(\varepsilon)n^2$.*
- *(ii) We have $\langle\boldsymbol{\sigma}_*,\boldsymbol{A}\boldsymbol{\sigma}_*\rangle\ge(1-\varepsilon)\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n}\langle\boldsymbol{\sigma},\boldsymbol{A}\boldsymbol{\sigma}\rangle$, with high probability (with respect to $\boldsymbol{A}\sim\mathsf{GOE}(n)$).*

The same result holds when $\boldsymbol{A}=\boldsymbol{A}(n)$ is symmetric with $A_{ii}=0$ and $(A_{ij})_{1\le i<j\le n}$ a collection of independent random variables, satisfying: $\mathbb{E}\{A_{ij}\}=0$, $\mathbb{E}\{A_{ij}^2\}=1/n$ and $\mathbb{E}\{\exp(\lambda A_{ij})\}\le\exp(C_*\lambda^2/2n)$ for some constant C_* and all $i<j\le n$ (in words, entries are subgaussian with common subgaussian parameter C_*/n).

In other words, on average, the optimization problem (1) is much easier than in worst case. Of course, this is far from being the only example of this phenomenon (a gap between worst case and average case complexity). However, it is a rather surprising example given the complexity of the energy landscape $H_n(\boldsymbol{\sigma})$. Its proof uses in a crucial way a fine property of the associated Gibbs measure, namely the support overlap distribution.

Remark I.2 (Computation model)**.** For the sake of simplicity, we measure complexity in floating point operations. However, all operations in our algorithm appear to be stable and it should be possible to translate this result to weaker computation models.

We also assume that we can choose one value of the inverse temperature β, and query the distribution $\mu_\beta(t)$ and the PDE solution $\Phi(t,x)$ as well as its derivatives $\partial_x\Phi(t,x)$, $\partial_{xx}\Phi(t,x)$ at specified points (t,x), with each query costing $O(1)$ operations.

This is a reasonable model for two reasons: (i) The PDE (2) is independent of the instance, and can be solved to a desired degree of accuracy only once. This solution can be used every time a new instance of the problem is presented. (ii) The function $\mu\mapsto\mathsf{P}_\beta(\mu)$ is uniformly continuous [22] and strongly convex [4], [23]. Further the PDE solution Φ is continuous in μ and can

be characterized as fixed point of a certain contraction [23]. Because of these reasons we expect that an oracle to compute $\Phi(t,x)$, $\partial_x\Phi(t,x)$, $\partial_{xx}\Phi(t,x)$ to accuracy η can be implemented efficiently. We defer to future work a more detailed study of the complexity of this oracle.

Beyond Theorem 2, our general analysis allows us to prove an additional fact that is of independent interest. Namely, for any $\beta > \beta_0$, our message passing iteration constructs an approximate solution of the celebrated Thouless, Anderson, Palmer (TAP) equations [28], [45].

In order to avoid inessential technical complications, the bulk of this paper is devoted to proving Theorem 2 for the case of Gaussian matrices $\boldsymbol{A}$, However, the class of algorithms we use enjoys certain universality properties, first established in [7]. These properties can be used to establish the last part of Theorem 2 which addresses the case of symmetric matrices with independent subgaussian entries. Section V contains such generalization.

As a special case of random matrices $\boldsymbol{A}$ with independent subgaussian entries, we can consider (centered) adjacency matrices of dense Erdös-Renyi random graphs. As a consequence of Theorem 2 we obtain an algorithm to approximate the MAXCUT of such a graph.

Let $G_n = ([n], E_n) \sim \mathcal{G}(n,p)$ be an Erdös-Renyi random graph with edge probability $\mathbb{P}\{(i,j) \in E_n\} = p = \Omega(1)$. A random balanced partition of the vertices (which we encode as a vector $\boldsymbol{\sigma} \in \{+1,-1\}^n$) achieves a cut $\mathsf{CUT}_G(\boldsymbol{\sigma}) = |E_n|/2 + O(n) = n^2p/4 + O(n)$, and simple concentration argument implies that the MAXCUT has size $\max_{\boldsymbol{\sigma}\in\{+1,-1\}} \mathsf{CUT}_G(\boldsymbol{\sigma}) = |E_n|/2 + O(n^{3/2}p^{1/2})$. In fact, it follows from [17] that[1] $\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \mathsf{CUT}_G(\boldsymbol{\sigma}) = |E_n|/2 + (n^3p(1-p)/2)^{1/2}\mathsf{P}_* + o(n^{3/2})$, where P_* is the prediction of Parisi's formula (i.e. the right-hand side of ((4))). In other words, MAXCUT on dense Erdös-Renyi random graphs is non-trivial only once we subtract the baseline value $|E_n|/2$. Once this baseline is subtracted, the problem lies in the universality class of the Sherrington-Kirkpatrick model. As a corollary of Theorem 2 we can approximate this subtracted value arbitrarily well.

Corollary I.2. *Under Assumption 1, for any $\varepsilon > 0$ there exists an algorithm (with complexity at most $C(\varepsilon)\, n^2$), that takes as input an Erdös-Renyi random graph $G_n = ([n], E_n) \sim \mathcal{G}(n,p)$ (with p bounded away from 0 and 1 as $n \to \infty$), and outputs $\boldsymbol{\sigma}_* = \boldsymbol{\sigma}_*(G) \in \{+1,-1\}^n$, such that*

$$\left(\mathsf{CUT}_G(\boldsymbol{\sigma}_*) - \frac{|E_n|}{2}\right) \geq (1-\varepsilon) \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \left(\mathsf{CUT}_G(\boldsymbol{\sigma}_*) - \frac{|E_n|}{2}\right).$$

The rest of this section provides further background. In Section II we describe and analyze a general message passing algorithm, which we call *incremental approximate message passing* (IAMP). We believe this algorithm is of independent interest and can be applied beyond the Sherrington-Kirkpatrick model. In Section III we use this approach to prove Theorem 2. In Section IV we show that the same message passing algorithm of Section II produces approximate solutions of the TAP equations. Finally, Section V discusses a generalization of Theorem 2 using universality. The impatient reader, who is interested in a succinct description of the algorithm (with some technical bells and whistles removed), is urged to read Appendix A.

A. Further background

As mentioned above –under suitable complexity theory assumptions– there is no polynomial-time algorithm that approximates the quadratic program (1) better than within a factor $O((\log n)^c)$, for some $c > 0$ [3]. Little is known on average-case hardness, when $\boldsymbol{A}$ is drawn from one of the random matrix distributions considered here. As an exception, Gamarnik [20] proved that exact computation of the partition function $Z_n(\beta)$ is hard on average.

A natural approach to the quadratic program (1) would be to use a convex relaxation. A spectral relaxation yields $\max_{\boldsymbol{\sigma}\in\{+1,-1\}} H_n(\boldsymbol{\sigma})/n \leq \lambda_1(\boldsymbol{A})/2 = 1 + o_n(1)$, and hence is not tight for large n. This can be compared to a numerical evaluation of Parisi's formula which yields $\mathsf{P}_* \approx 0.763166$ [15], [39]. Rounding the spectral solution yields a $H_n(\boldsymbol{\sigma}_{\text{sp}}) = 2/\pi + o_n(1) \approx 0.636619$. Somewhat surprisingly, the simplest semidefinite programming relaxation (degree 2 of the sum-of-squares hierarchy), does not yield any improvement (for large n) over the spectral one [32]. After a preprint of this paper was posted, [26] proved that the degree 4 sum-of-squares relaxation has asymptotically the same value as well. Theorem 2 was conjectured by the author in 2016 [31], based on insights from statistical physics [16], [11]. The same presentation also outlined the basic strategy followed in the present paper, which uses an iterative 'approximate message passing' (AMP) algorithm. This type of algorithms were first proposed in the context of signal processing and compressed sensing

[1] In [17], the same result is shown to hold for sparser graphs, as long as the average degree diverges: $np_n \to \infty$.

[25], [18]. Their rigorous analysis was developed by Bolthausen [10] and subsequently generalized in several papers [8], [24], [7], [9]. In this paper we introduce a specific class of AMP algorithms ('incremental AMP') whose specific properties allow us to match the result predicted by Parisi's formula.

The fundamental phenomenon studied here is expected to be quite general. Namely objective functions with overlap distribution having support of the form $[0, q_*]$ are expected to be easy to optimize. In contrast, if the support has a gap (for instance, has the form $[0, q_1] \cup [q_2, q_*]$ for some $q_1 < q_2$), this is considered as an indication of average case hardness. This intuition originates within spin glass theory [28]. Roughly speaking, the structure of the overlap distribution should reflect the connectivity properties of the level sets $\mathcal{L}_n(\varepsilon) \equiv \{\boldsymbol{\sigma} : \ H_n(\boldsymbol{\sigma}) \geq (1-\varepsilon) \max_{\boldsymbol{\sigma}'} H_n(\boldsymbol{\sigma}')\}$. This intuition was exploited in some cases to prove the failure of certain classes of algorithms in problems with a gap in the overlap distribution, see e.g. [21].

Important progress towards clarifying this connection was achieved recently in two remarkable papers [1], [42].

Addario-Berry and Maillard [1] study an abstract optimization problem that is thought to capture some key features of the the energy landscape of the Sherrington-Kirkpatrick model, the so-called 'continuous random energy model.' They prove that an approximate optimum can be found in time polynomial in the problem dimensions. From an optimization perspective, the random energy model is somewhat un-natural, in that specifying an instance requires memory that is exponential in the problem dimensions.

Subag [42] considers the p-spin spherical spin glass. Roughly speaking, this can be described as the problem of optimizing a random smooth function (which can be taken to be a low-degree polynomial) over the unit sphere. Subag relaxes this problem by extending the optimization over the unit ball, and proves that this objective function can be optimized efficiently by following the positive directions of the Hessian. The solution thus constructed lies on the unit sphere and thus solves the un-relaxed problem. The mathematical insight of [42] is beautifully simple, but uses in a crucial way the spherical geometry. While it might be possible to generalize the same argument to the hypercube case (e.g., using the generalized TAP free energy of [30], [14]) this extension is far from obvious. In particular, uniform control of the Hessian is not as straightforward as in [42].

The algorithm presented here is partially inspired by [42] (in particular, a key role is played by approximate orthogonality of the updates), but its specific structure is dictated by the message passing viewpoint. Thanks to the technique of [10], [8], [24], [9], its analysis does not require uniform control and is relatively simple.

B. Notations

Given vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we denote by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ their scalar product and by $\|\boldsymbol{x}\| \equiv \langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2}$ the associated ℓ_2 norm. Given a function $f : \mathbb{R}^k \to \mathbb{R}$, and k vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^n$ we write $f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ for the vector in $\mathbb{R}^n$ with components $f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)_i = f(x_{1,i}, \dots, x_{k,i})$. The empirical distribution of the coordinates of a vector of vectors $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in (\mathbb{R}^n)^k$ is the probability measure on $\mathbb{R}^k$ defined by

$$\hat{p}_{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k} \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_{1,i}, \dots, x_{k,i})} \, .$$

In other words, if we arrange the vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ in a matrix in $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_k] \in \mathbb{R}^{n \times k}$, $\hat{p}_{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k}$ denotes the probability distribution of a uniformly random row of $\boldsymbol{X}$. In the case of a single vector $\boldsymbol{x} \in \mathbb{R}^n$ (i.e. for $k = 1$), this reduces to the standard empirical distribution of the entries of $\boldsymbol{x}$. We say that a function $f : \mathbb{R}^d \to \mathbb{R}$ is pseudo-Lipschitz if $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq C(1 + \|\boldsymbol{x}\| + \|\boldsymbol{y}\|)\|\boldsymbol{x} - \boldsymbol{y}\|$.

Given two probability measures μ, ν on $\mathbb{R}^d$, we recall that their Wasserstein W_2 distance is defined as

$$W_2(\mu, \nu) \equiv \left\{ \inf_{\gamma \in \mathcal{C}(\mu,\nu)} \int \|\boldsymbol{x} - \boldsymbol{y}\|^2 \gamma(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{y}) \right\}^{1/2} ,$$

where the infimum is taken over all the couplings of μ and ν (i.e. joint distributions on $\mathbb{R}^d \times \mathbb{R}^d$ whose first marginal coincides with μ, and second with ν. For a sequence of probability measures $(\mu_n)_{n \geq 1}$, and μ on $\mathbb{R}^d$, we say that μ_n converges in Wasserstein distance to μ (and write $\mu_n \xrightarrow{W_2} \mu$) if $\lim_{n \to \infty} W_2(\mu_n, \mu) = 0$. It is well known that $\mu_n \xrightarrow{W_2} \mu$ if and only if $\lim_{n \to \infty} \int \psi(\boldsymbol{x}) \mu_n(\mathrm{d}\boldsymbol{x}) = \int \psi(\boldsymbol{x}) \mu(\mathrm{d}\boldsymbol{x})$ for all bounded Lipschitz functions ψ, and for $\psi(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ [47, Theorem 6.9]. Given a sequence of random variables X_n, we write $X_n \xrightarrow{p} X_\infty$ or $\text{p-lim}_{n \to \infty} X_n = X_\infty$ to state that X_n converge in probability to X_∞.

We will sometimes be interested in double limits of sequences of random variables. If $X_{n,M}$ is a sequence indexed by n, M and x_* is a constant,

$$\lim_{M \to \infty} \text{p-lim}_{n \to \infty} X_{n,M} = x_* \, ,$$

whenever $X_{n,M}$ converges in probability to a non-random quantity x_M as $n \to \infty$, and $\lim_{M\to\infty} x_M = x_*$.

II. A GENERAL MESSAGE PASSING ALGORITHM

Our algorithm is based on the following approximate message passing (AMP) iteration.

AMP iteration. Consider a sequence of (weakly differentiable) functions $f_k : \mathbb{R}^{k+2} \to \mathbb{R}$, and a non-random initialization $\boldsymbol{u}^0 \in \mathbb{R}^n$ and additional vector $\boldsymbol{y} \in \mathbb{R}^n$ with $\hat{p}_{\boldsymbol{u}^0,\boldsymbol{y}} \xrightarrow{W_2} p_{U_0,Y}$ (where $p_{U_0,Y}$ is any probability distribution on $\mathbb{R}^2$ with finite second moment $\int (u_0^2 + y^2) p_{U_0,Y}(\mathrm{d}u_0, \mathrm{d}y) < \infty$). The AMP iteration is defined by letting, for $k \geq 0$,

$$\begin{aligned} \boldsymbol{u}^{k+1} &= \boldsymbol{A}\, f_k(\boldsymbol{u}^0, \dots, \boldsymbol{u}^k; \boldsymbol{y}) \\ &\quad - \sum_{j=1}^{k} \mathsf{b}_{k,j} f_{j-1}(\boldsymbol{u}^0, \dots, \boldsymbol{u}^{j-1}; \boldsymbol{y}) \,, \\ \mathsf{b}_{k,j} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f_k}{\partial u_i^j}(u_i^0, \dots, u_i^k; y_i) \,. \end{aligned} \tag{6}$$

It will be understood throughout that $f_j = 0$ for $j < 0$.

Proposition II.1. *Consider the AMP iteration (6), and assume $f_k : \mathbb{R}^{k+2} \to \mathbb{R}$ to be Lipschitz continuous. Then for any $k \in \mathbb{N}$, and any pseudo-Lipschitz function $\psi : \mathbb{R}^{k+2} \to \mathbb{R}$, we have*

$$\frac{1}{n} \sum_{i=1}^{n} \psi(u_i^0, \dots, u_i^k; y_i) \xrightarrow{p} \mathbb{E}\psi(U_0, \dots, U_k; Y) \,. \tag{7}$$

Here $(U_j)_{j\geq 1}$ is a centered Gaussian process independent of (U_0, Y) with covariance $\widehat{\boldsymbol{Q}} = (\widehat{Q}_{kj})_{k,j\geq 1}$ determined recursively via

$$\widehat{Q}_{k+1,j+1} = \mathbb{E}\big\{ f_k(U_0, \dots, U_k; Y) f_j(U_0, \dots, U_j; Y) \big\} \,. \tag{8}$$

This proposition follows immediately from the general analysis of AMP algorithms developed in [24], [9], cf. Appendix A.

We next consider a special case of the general AMP setting.

Incremental AMP (IAMP). Fix $\delta, M > 0$, and functions $\widehat{g}_k : \mathbb{R} \to \mathbb{R}$, $k \in \mathbb{N}$, $s, v : \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$. We consider the general iteration (6), with the following choice of functions f_k (independent of y):

$$f_k(u_0, \dots, u_k) = \widehat{g}_k\left(x_{k-1}\right) \cdot [u_k]_M \,, \tag{9}$$

$$\begin{aligned} x_k &= x_{k-1} + v(x_{k-1}, k\delta)\, \delta + \\ &\quad + s(x_{k-1}, k\delta)\, [u_k]_M \sqrt{\delta} \,, \quad x_0 = 1 \,, \end{aligned} \tag{10}$$

where, for $u \in \mathbb{R}$, $[u]_M = \max(-M, \min(u, M))$. Following our convention for f_j, we set $\widehat{g}_j = 0$ for $j < 0$.

We note that, by Eq. (9), x_k is indeed a function of $u_0, \dots, u_k$, and therefore f_k is a function of $u_0, \dots, u_k$ as stated.

Lemma II.2 (State evolution for Incremental AMP)**.** *Consider the incremental AMP iteration, and assume $s, v : \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ to be Lipschitz continuous and bounded, and $\widehat{g}_k : \mathbb{R} \to \mathbb{R}$ to be Lipschitz continuous and bounded for each k. Then for any $k \in \mathbb{N}$, and any pseudo-Lipschitz function $\psi : \mathbb{R}^{k+2} \to \mathbb{R}$, we have*

$$\lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \psi(u_i^0, \dots, u_i^k) = \mathbb{E}\psi(U_0^\delta, \dots, U_k^\delta) \,.$$

(The double limit is to be interpreted as defined in the Notations section.) Here $(U_j^\delta)_{j\geq 1}$ is a centered Gaussian process independent of $U_0^\delta = U_0$, with independent entries, with variance $\mathrm{Var}(U_k^\delta) = \widehat{q}_k$ given recursively by

$$\begin{aligned} \widehat{q}_{k+1} &= \mathbb{E}\{\widehat{g}_k(X_{k-1}^\delta)^2\} \cdot \widehat{q}_k \,, \\ X_k^\delta &= X_{k-1}^\delta + v(X_{k-1}^\delta; k\delta)\, \delta + s(X_{k-1}; k\delta) U_k^\delta \sqrt{\delta} \,. \end{aligned} \tag{11}$$

Proof: Consider Eqs. (9), (10), and note that, for any k, x_{k-1} is a bounded Lipschitz function of $u_0, \dots, u_{k-1}$ (because bounded Lipschitz functions are closed under sum, product, and composition). Hence f_k defined in (9) is Lipschitz continuous and we can therefore apply Proposition II.1 to get

$$\frac{1}{n} \sum_{i=1}^{n} \psi(u_i^0, \dots, u_i^k) \xrightarrow{p} \mathbb{E}\psi(U_0^{\delta,M}, \dots, U_k^{\delta,M}) \,.$$

Here $(U_j^{\delta,M})_{j\geq 1}$ is a Gaussian process with covariance $\widehat{\boldsymbol{Q}}^M$ determined by Eq. (8). We next claim the following:

1) $\widehat{Q}_{j,k}^M = 0$ for $k \neq j$ (and we set $\widehat{q}_k^M \equiv \widehat{Q}_{k,k}^M$).
2) $\widehat{q}_k^M \to \widehat{q}_k$ for each k as $M \to \infty$.

With these two claims, the statement of the lemma follows by dominated convergence.

To prove claim 1 note that, by symmetry we only have to consider the case $j < k$. The proof is by induction over k. For $k = 1$ there is nothing to prove. Assume next that the claim holds up to a certain k, and consider $\widehat{Q}_{j,k+1}^M$ for $1 \leq j \leq k$. By Eq. (8) we have (dropping for simplicity the superscripts δ, M from the random

variables)

$$\widehat{Q}^M_{j,k+1} = \mathbb{E}\{\widehat{g}_{j-1}(X_{j-2})[U_{j-1}]_M\, \widehat{g}_k(X_{k-1})[U_k]_M\} \\ = \mathbb{E}\{\widehat{g}_{j-1}(X_{j-2})[U_{j-1}]_M\, \widehat{g}_k(X_{k-1})\}\mathbb{E}\{[U_k]_M\} = 0\,.$$

Here the second equality follows from the induction hypothesis.

To prove claim 2, note that $\widehat{q}^M_k$ satisfies the recursion that follows from Eq. (8), namely

$$\widehat{q}^M_{k+1} = \mathbb{E}\{\widehat{g}_k(X^{\delta,M}_{k-1})^2\}\cdot \mathbb{E}\{[U^{\delta,M}_k]^2_M\}\,, \\ X^{\delta,M}_k = X^{\delta,M}_{k-1} + v(X^{\delta,M}_{k-1}; k\delta)\,\delta \\ + s(X^{\delta,M}_{k-1}; k\delta)\,[U^{\delta,M}_k]_M\sqrt{\delta}\,. \tag{12}$$

Also note that

$$|X^{\delta,M}_k| \le C_{\delta,k}\big(1 + |U_0|^k + |U^{\delta,M}_1|^k + \cdots + |U^{\delta,M}_k|^k\big)\,.$$

For some constants $C_{\delta,k}$ independent of M. This follows by induction over k using Eq. (12), and the fact that s, v are Lipschitz continuous. Since $\widehat{g}_k$ is Lipschitz continuous as well, we obtain $|\widehat{g}_k(X^{\delta,M}_{k-1})| \le F_k(U_0, U^{\delta,M}_1, \dots, U^{\delta,M}_{k-1})$ for some polynomial F_k independent of M. Hence the claim follows by applying recursively dominated convergence. ∎

Remark II.1. The use of truncation $[u_k]_M$ in the definition (9) is dictated by the need to ensure that f_k is Lipschitz, and to be able to apply Proposition II.1. We believe that the conclusion of Proposition II.1 holds under weaker assumptions (e.g. f_k locally Lipschitz with polynomial growth). Such a generalization would allow to replace $[u_k]_M$ by u_k in Eq. (9), and hence get rid of the parameter M in our algorithm.

We are now in position of defining our candidate for a near optimum of the problem (1). We fix $\overline{q} > 0$ and define (recalling the definition of f_k in Eqs. (9), (10))

$$\boldsymbol{z} = \sqrt{\delta}\sum_{k=1}^{\lfloor \overline{q}/\delta\rfloor} f_k(\boldsymbol{u}_0, \dots, \boldsymbol{u}_k)\,. \tag{13}$$

Note that this vector depends on parameters $\delta, M, \overline{q}$, and on the functions g, s, v. Parameters δ and M must be taken (respectively) small enough and large enough (but independent of n). The next section will be devoted to choosing $\overline{q}$ and the functions g, s, v. In this section we will establish some general properties of $\boldsymbol{z}$ (for small δ and large M).

Lemma II.3. *Consider the incremental AMP iteration, and assume $g, s, v : \mathbb{R}\times\mathbb{R}_{\ge 0} \to \mathbb{R}$ to be Lipschitz continuous and bounded. Further assume $\partial_x \widehat{g}_k(x)$, $\partial_x s(x,t)$, $\partial_x v(x,t)$ to exist and be Lipschitz continuous.*

Define the random variable

$$Z^\delta \equiv \sqrt{\delta}\sum_{k=1}^{\lfloor \overline{q}/\delta\rfloor} \widehat{g}_k(X_{k-1})\, U^\delta_k\,.$$

Then we have, for any pseudo-Lipschitz function $\psi : \mathbb{R}\to\mathbb{R}$,

$$\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \psi(z_i) = \mathbb{E}\{\psi(Z^\delta)\}\,, \tag{14}$$

$$\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{2n}\langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{z}\rangle = \tag{15}$$
$$= \delta\sum_{k=1}^{\lfloor \overline{q}/\delta\rfloor - 1} \mathbb{E}\{(U^\delta_k)^2\}\,\mathbb{E}\{\widehat{g}_k(X^\delta_{k-1})\}\,\mathbb{E}\{\widehat{g}_k(X^\delta_{k-1})^2\}\,.$$

Proof: Equation (14) follows immediately from Lemma II.2 upon noticing that $\psi(z_i)$ is a pseudo-Lipschitz function of $u_{0,i}, \dots, u_{k,i}$.

In order to prove Eq. (15), we will write $\boldsymbol{f}_k = f_k(\boldsymbol{u}_0, \dots, \boldsymbol{u}_k)$, and $K = \lfloor \overline{q}/\delta\rfloor$. We further notice that, for $j < k$,

$$\operatorname*{p-lim}_{n\to\infty} \mathsf{b}_{k,j} = \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \frac{\partial g}{\partial u_j}(x_i^{k-1}; k\delta)[u_i^k]_M \\ = \mathbb{E}\left\{\frac{\partial g}{\partial u_j}(X^{\delta,M}_{k-1}; k\delta)[U^{\delta,M}_k]_M\right\} \\ = \mathbb{E}\left\{\frac{\partial g}{\partial u_j}(X^{\delta,M}_{k-1}; k\delta)\right\}\mathbb{E}\left\{[U^{\delta,M}_k]_M\right\} = 0\,. \tag{16}$$

Here and below the random variables $U^{\delta,M}_k, X^{\delta,M}_k$ are defined as in the proof of Lemma II.2. On the other hand

$$\operatorname*{p-lim}_{n\to\infty} \mathsf{b}_{k,k} = \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \widehat{g}_k(x_i^{k-1})\mathbf{1}_{u_i^k\in[-M,M]} \\ = \mathbb{E}\{\widehat{g}_k(X^{\delta,M}_{k-1})\}\,\mathbb{P}(U^{\delta,M}_k \in [-M, M])\,.$$

Note that we applied Lemma II.2 to a non-Lipschitz function. The limit holds nevertheless by a standard weak convergence argument (namely, using upper and lower Lipschitz approximations of the indicator function). We therefore conclude that (using $U^{\delta,M}_k \sim \widehat{q}^M_k$, $\widehat{q}^M_k \to \widehat{q}_k$ as $M\to\infty$):

$$\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty} \mathsf{b}_{k,k} = \mathbb{E}\{\widehat{g}_k(X^\delta_{k-1})\}\,. \tag{17}$$

Next notice that, for $j < k$,

$$\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle \boldsymbol{f}_j, \boldsymbol{f}_k\rangle \\ = \mathbb{E}\{\widehat{g}_j(X^{M,\delta}_{j-1})\,[U^{M,\delta}_j]_M \widehat{g}_k(X^{M,\delta}_{k-1})\,[U^{M,\delta}_k]_M\} \\ = \mathbb{E}\{\widehat{g}_j(X^{M,\delta}_{j-1})\,[U^{M,\delta}_j]_M \widehat{g}_k(X^{M,\delta}_{k-1})\}\mathbb{E}\{[U^{M,\delta}_k]_M\} = 0\,.$$

By a similar argument, always for $j < k$,

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\langle \boldsymbol{f}_j, \boldsymbol{u}^k\rangle = 0\,.$$

On the other hand

$$\begin{aligned}\lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} &\frac{1}{n}\|\boldsymbol{f}_k\|^2 \\ &= \lim_{M\to\infty} \mathbb{E}\big\{\widehat{g}_k(X_{k-1}^{M,\delta})^2\,[U_k^{M,\delta}]_M^2\big\} \\ &= \lim_{M\to\infty} \mathbb{E}\big\{\widehat{g}_k(X_{k-1}^{M,\delta})^2\big\}\,\mathbb{E}\big\{[U_k^{M,\delta}]_M^2\big\} \\ &= \mathbb{E}\big\{\widehat{g}_k(X_{k-1}^{\delta})^2\big\}\,\mathbb{E}\big\{(U_k^{\delta})^2\big\}\,.\end{aligned}$$

By the AMP iteration, we know that $\boldsymbol{A}\boldsymbol{f}_k = \boldsymbol{u}^{k+1} + \sum_{\ell=1}^k \mathsf{b}_{k,\ell}\boldsymbol{f}_{\ell-1}$. Hence, using the above limits, for $j \le k$,

$$\begin{aligned}&\lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\langle \boldsymbol{f}_j, \boldsymbol{A}\boldsymbol{f}_k\rangle \\ &= \lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\langle \boldsymbol{f}_j, \boldsymbol{u}^{k+1}\rangle \\ &\qquad + \sum_{\ell=1}^{k} \lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \mathsf{b}_{k,\ell}\langle \boldsymbol{f}_j, \boldsymbol{f}_{\ell-1}\rangle \\ &= \lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \mathsf{b}_{k,k}\langle \boldsymbol{f}_j, \boldsymbol{f}_{k-1}\rangle \\ &= \mathbf{1}_{\{k=j+1\}}\mathbb{E}\big\{\widehat{g}_j(X_{j-1}^{\delta})^2\big\}\,\mathbb{E}\big\{(U_j^{\delta})^2\big\}\mathbb{E}\{\widehat{g}_j(X_{j-1}^{\delta})\}\,.\end{aligned}$$

We finally can compute

$$\begin{aligned}&\lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{2n}\langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{z}\rangle \\ &= \delta\sum_{j=1}^{K} \lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{2n}\langle \boldsymbol{f}_j, \boldsymbol{A}\boldsymbol{f}_j\rangle \\ &\qquad + \delta \sum_{1\le j<k\le K} \lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\langle \boldsymbol{f}_j, \boldsymbol{A}\boldsymbol{f}_k\rangle \\ &= \delta\sum_{j=1}^{K-1} \lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\langle \boldsymbol{f}_j, \boldsymbol{A}\boldsymbol{f}_{j+1}\rangle \\ &= \delta\sum_{j=1}^{K-1} \mathbb{E}\big\{\widehat{g}_j(X_{j-1}^{\delta})^2\big\}\,\mathbb{E}\big\{(U_j^{\delta})^2\big\}\mathbb{E}\{\widehat{g}_j(X_{j-1}^{\delta})\}\,.\end{aligned}$$

■

In the case of models with full replica symmetry breaking, it is natural to consider the limit of small step size $\delta \to 0$. This limit is described by a stochastic differential equation (SDE) described below.

SDE description. Consider Lipschitz functions $g, s, v : \mathbb{R}\times\mathbb{R}_{\ge 0}\to\mathbb{R}$, with $|s(x,t)| + |v(x,t)| \le C(1+|x|)$. Let $(B_t)_{t\ge 0}$ be a standard Brownian motion. We define the process $(X_t, Z_t)_{t\ge 0}$ via

$$\mathrm{d}X_t = v(X_t,t)\,\mathrm{d}t + s(X_t,t)\,\mathrm{d}B_t\,,\ \mathrm{d}Z_t = g(X_t,t)\,\mathrm{d}B_t\,, \tag{18}$$

with initial condition $X_0 = Z_0 = 0$. Equivalently

$$\begin{aligned}X_t &= \int_0^t v(X_r,r)\,\mathrm{d}r + \int_0^t s(X_r,r)\,\mathrm{d}B_r\,, \\ Z_t &= \int_0^t g(X_r,r)\,\mathrm{d}B_r\,,\end{aligned} \tag{19}$$

where the integral is understood in Ito's sense. Existence and uniqueness of strong solutions of this SDE is given –for instance– in [34, Theorem 5.2.1].

Lemma II.4. *Given Lipschitz functions $g, s, v : \mathbb{R}\times\mathbb{R}_{\ge 0}\to\mathbb{R}$, with v and s bounded, let (X_t, Z_t) be the process defined above. Assume $\mathbb{E}\{g(X_t,t)^2\} = 1$ for all $t \ge 0$. Further consider the state evolution iteration of Eq. (11), whereby $\widehat{g}_k$ is defined recursively via*

$$\widehat{g}_k(x) \equiv \frac{g(x, k\delta)}{\mathbb{E}\{g(X_{k-1}^{\delta}, k\delta)^2\}^{1/2}}\,. \tag{20}$$

Then, there exists a coupling of $(X_k^{\delta})_{k\ge 0}$ and $(X_t)_{t\ge 0}$ such that

$$\max_{k\le\lfloor\overline{q}/\delta\rfloor} \mathbb{E}\big(|X_k^{\delta} - X_{k\delta}|^2\big)\Big\} \le C\delta\,, \tag{21}$$

$$\mathbb{E}\big(|Z^{\delta} - Z_{\overline{q}}|^2\big) \le C\sqrt{\delta}\,, \tag{22}$$

$$\begin{aligned}&\delta\sum_{k=1}^{\lfloor\overline{q}/\delta\rfloor-1} \mathbb{E}\{(U_k^{\delta})^2\}\,\mathbb{E}\{\widehat{g}_k(X_{k-1}^{\delta})\}\,\mathbb{E}\{\widehat{g}_k(X_{k-1}^{\delta})^2\} \\ &= \int_0^{\overline{q}} \mathbb{E}\{g(X_t,t)\}\,\mathrm{d}t + O(\delta^{1/4})\,.\end{aligned} \tag{23}$$

(Here C is a constant depending only on the bounds on g, v, s, and on $\overline{q}$. Further the $O(\delta^{1/4})$ error is bounded as $|O(\delta^{1/4})| \le C\delta^{1/4}$ for the same constant.)

Proof: Throughout this proof, we will write $t_k = k\delta$ and denote by C a generic constant that depends on the bounds on g, s, v, and can change from line to line. Note that, by construction, $\widehat{q}_j = 1$ for all j, and therefore $(U_j^{\delta})_{j\ge 0} \sim_{iid} \mathsf{N}(0,1)$. Hence we can construct the discrete and continuous processes on the same space by letting $\sqrt{\delta}U_j^{\delta} = B_{t_{j+1}} - B_{t_j}$.

We then decompose the difference between the two processes as

$$\begin{aligned}X_{k\delta} - X_k^{\delta} = &\sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta} \big[v(X_t,t) - v(X_j^{\delta}, t_{j+1})\big]\,\mathrm{d}t \\ &+ \sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta} \big[s(X_t,t) - s(X_j^{\delta}, t_{j+1})\big]\,\mathrm{d}B_t\,.\end{aligned}$$

By taking the second moment, and using the fact that X_t is measurable on $(B_s)_{s\le t}$ and X_j^{δ} is measurable on

$(B_s)_{s\le t_j}$, we get

$$\begin{aligned}
&\mathbb{E}\big\{\big[X_{k\delta}-X_k^\delta\big]^2\big\}\\
&\le 2k\sum_{j=0}^{k-1}\delta\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{\big[v(X_t,t)-v(X_j^\delta,t_{j+1})\big]^2\big\}\,\mathrm{d}t\\
&\quad+2\sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{\big[s(X_t,t)-s(X_j^\delta,t_{j+1})\big]^2\big\}\,\mathrm{d}t\,.
\end{aligned}$$

Next notice that by the boundedness of s, v, we have $\mathbb{E}\{|X_t-X_s|^2\}\le C|t-s|$. Let $\Delta_k\equiv\mathbb{E}(|X_{t_k}-X_k^\delta|^2)$. Assuming without loss of generality $\delta<1$,

$$\begin{aligned}
&\mathbb{E}\big\{\big[v(X_t,t)-v(X_j^\delta,t_{j+1})\big]^2\big\}\\
&\quad\le C\mathbb{E}(|X_t-X_{t_j}|^2)+C\mathbb{E}(|X_{t_j}-X_j^\delta|^2)\\
&\qquad+C|t-t_{j+1}|^2\le C\Delta_k+C\delta\,.
\end{aligned}$$

The same bound holds for $\mathbb{E}\{[s(X_t,t)-s(X_j^\delta,t_{j+1})]^2\}$. Substituting above, we get

$$\Delta_k\le C(\overline{q}+1)\delta\sum_{j=0}^{k-1}(\Delta_j+\delta).$$

This implies bound $\mathbb{E}(|X_{t_k}-X_k^\delta|^2)\le C\delta$ as stated in (21).

In order to prove Eq. (22), note that

$$\begin{aligned}
&\mathbb{E}\{g(X_{k-1}^\delta,t_k)^2\}\le\\
&\quad\le\mathbb{E}\{[g(X_{t_{k-1}},t_k)+C|X_{k-1}^\delta-X_{t_{k-1}}|]^2\}\\
&\quad\le 1+C\sqrt{\delta}\,.
\end{aligned}$$

Hence

$$\mathbb{E}\big\{[\widehat{g}_k(X_{k-1}^\delta)-g(X_{k-1}^\delta,t_k)]^2\big\}\le C\sqrt{\delta}\,.$$

Let $K=\lfloor\overline{q}/\delta\rfloor$, and write

$$Z_{K\delta}-Z^\delta=\sum_{j=0}^{K-1}\int_{t_j}^{t_j+\delta}\big[g(X_t,t)-\widehat{g}_{j+1}(X_j^\delta)\big]\,\mathrm{d}B_t\,.$$

Therefore

$$\begin{aligned}
&\mathbb{E}\big(|Z_{K\delta}-Z^\delta|^2\big)\\
&=\sum_{j=0}^{K-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{[g(X_t,t)-\widehat{g}_{j+1}(X_j^\delta)]^2\big\}\,\mathrm{d}t\\
&\le 2\sum_{j=0}^{K-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{[g(X_t,t)-g(X_j^\delta,t_{j+1})]^2\big\}\,\mathrm{d}t\\
&\quad+2\sum_{j=0}^{K-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{[\widehat{g}_{j+1}(X_j^\delta)-g(X_j^\delta,t_{j+1})]^2\big\}\,\mathrm{d}t\\
&\le C\delta\sum_{j=0}^{K-1}\int_{t_j}^{t_j+\delta}(\Delta_j+\delta)\,\mathrm{d}t+C(\overline{q}+1)\sqrt{\delta}\\
&\le C(\overline{q}+1)\sqrt{\delta}\,.
\end{aligned}$$

The bound of Eq. (22) follows since

$$\mathbb{E}\big(|Z_{q_*}-Z_{K\delta}|^2\big)=\int_{K\delta}^{q_*}\mathbb{E}\big\{g(X_t,t)^2\big\}\mathrm{d}t\le\delta\,.$$

Finally, Eq. (23) follows by the same estimates. ∎

We now collect the main findings of this section in a theorem. This characterizes the values of the objective function achievable by the above algorithm.

Theorem 3. *Let $g,s,v:\mathbb{R}\times\mathbb{R}_{\ge0}\to\mathbb{R}$ be Lipschitz continuous, with v and s bounded, and define the process (X_t,Z_t) using the SDE (18) with initial condition $X_0=Z_0=0$. Assume $\mathbb{E}\{g(X_t,t)^2\}=1$ for all $t\ge0$. Further assume $\partial_x g(x,t)\partial_x s(x,t)\partial_x v(x,t)$ to exist and be Lipschitz continuous.*

Define the incremental AMP iteration $(\boldsymbol{u}^k)_{k\ge0}$, and let $\boldsymbol{z}$ be given by Eq. (13). Finally, let $\psi:\mathbb{R}\to\mathbb{R}$ be a pseudo-Lipschitz function. Then, for any $\varepsilon>0$ there exist $\delta_*(\varepsilon)>0$, and for any $\delta\ge\delta_*(\varepsilon)$ there exist $M_*(\varepsilon,\delta)<\infty$ such that, if $\delta\le\delta_*(\varepsilon)$ and $M\ge M_*(\varepsilon,\delta)$, we have

$$\begin{aligned}
&\left|\operatorname*{p-lim}_{n\to\infty}\frac{1}{2n}\langle\boldsymbol{z},\boldsymbol{A}\boldsymbol{z}\rangle-\int_0^{\overline{q}}\mathbb{E}\{g(X_t,t)\}\,\mathrm{d}t\right|\le\varepsilon\,,\\
&\left|\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi(z_i)-\mathbb{E}\{\psi(Z_{\overline{q}})\}\right|\le\varepsilon\,.
\end{aligned}$$

(Further the above limits in probability are non-random quantities.)

Proof: This follows immediately from Lemma II.3 and Lemma II.4. ∎

III. Proof of the Main Theorem

A. Choosing the nonlinearities

In view of Theorem 3, we need to choose the coefficients g,s,v in the SDE (18) as to solve the following stochastic optimal control problem:

$$\begin{aligned}
\text{maximize}\quad&\int_0^{\overline{q}}\mathbb{E}\{g(X_t,t)\}\,\mathrm{d}t\,,\\
\text{subject to}\quad&\mathbb{P}(Z_{\overline{q}}\in[-1,1])=1\,,&(24)\\
&\mathrm{d}X_t=v(X_t,t)\,\mathrm{d}t+s(X_t,t)\,\mathrm{d}B_t\,,&(25)\\
&g,s,v\in\mathrm{Lip}(\mathbb{R}\times[0,1])\,.
\end{aligned}$$

By Theorem 3, the value of this problem is the asymptotic optimal value achieved by the IAMP algoritm, for problem (1). Notice that related (but not identical) optimal control problems were studied, among others, in [4], [23], to construct useful representations for the Parisi formula. Here we will not attempt to solve directly this problem, and instead we will compare it with

the structure of the Parisi formula. This will motivate a guess for the three functions g, s, v, which enables us to prove Theorem 2 (after taking $\beta \to \infty$). Note that it follows *a posteriori* that this guess is an optimizer of the above stochastic optimal control problem (again, for large β).

Throughout this section we set $\beta > \beta_0$ as per Assumption 1. We also set $\overline{q} = q_* = q_*(\beta)$ and $\mu = \mu_\beta$ the unique minimizer of the Parisi functional. We also fix Φ to be the solution of the PDE (2) with $\mu = \mu_\beta$.

There is a natural SDE associated with the Parisi's variational principle, that was first introduced in physics [28], and recently studied in the probability theory literature [4], [23]:

$$\mathrm{d}X_t = \beta^2 \mu(t) \partial_x \Phi(t, X_t)\, \mathrm{d}t + \beta\, \mathrm{d}B_t\,. \tag{26}$$

Unless otherwise stated, it is understood that we set the initial condition to $X_0 = 0$. Motivated by this, we set the coefficients g, s, v as follows

$$\begin{aligned} &v(x,t) = \beta^2 \mu(t) \partial_x \Phi(t,x)\,, \quad s(x,t) = \beta\,,\\ &g(x,t) = \beta \partial_{xx} \Phi(t,x)\,. \end{aligned} \tag{27}$$

We collect below a few useful regularity properties of Φ, which have been proved in the literature.

Lemma III.1. (i) *$\partial_x^j \Phi(t,x)$ exists and is continuous for all $j \ge 1$.*

(ii) *For all $(t,x) \in [0,1] \times \mathbb{R}$,*

$$\begin{aligned} &\big|\partial_x \Phi(t,x)\big| \le 1\,, \quad 0 < \partial_x^2 \Phi(t,x) \le 1\,,\\ &\big|\partial_x^3 \Phi(t,x)\big| \le 4\,. \end{aligned}$$

(iii) *$\partial_t \partial_x^j \Phi(t,x) \in L^\infty([0,1] \times \mathbb{R})$ for all $j \le 0$.*

(iv) *$\partial_x \Phi(t,x)$, $\partial_x^2 \Phi(t,x)$ are Lipschitz continuous on $[0,1] \times \mathbb{R}$.*

Proof: Points (i) and (iii) are Theorem 4 in [23]. Point (ii) is Proposition 2.(ii) in [4]. Finally, point (iv) follows immediately from points (iii), (iv). ∎

This Lemma implies that the choice (27) satisfies the regularity assumptions in Theorem 3. We next have to check the normalization condition, and compute the resulting distribution.

Lemma III.2. *We have*

$$Z_t = \partial_x \Phi(t, X_t)\,. \tag{28}$$

In particular $\mathbb{P}(Z_t \in [-1,1]) = 1$ for all t.

Proof: By Lemma 2 in [4], we have, for any $t_1 < t_2$

$$\partial_x \Phi(t_2, X_{t_2}) - \partial_x \Phi(t_1, X_{t_1}) = \int_{t_1}^{t_2} \beta \partial_{xx} \Phi(t, X_t)\, \mathrm{d}B_t\,, \tag{29}$$

which is exactly Eq. (28). Lemma III.1.(ii) implies $|Z_t| \le 1$ almost surely. ∎

Lemma III.3. *For all $0 \le t \le q_*$, we have*

$$\mathbb{E}\big\{\big(\partial_x \Phi(t, X_t)\big)^2\big\} = t\,, \tag{30}$$

$$\mathbb{E}\big\{\big(\beta \partial_{xx} \Phi(t, X_t)\big)^2\big\} = 1\,. \tag{31}$$

Proof: Equation (30) is Proposition 1 in [13]. For Eq. (31) note that by Eq. (39) in the same paper, we have, for any $t_1 < t_2 \le q_*$

$$\begin{aligned} &\mathbb{E}\{(\partial_x \Phi(t_2, X_{t_2}))^2\} - \mathbb{E}\{(\partial_x \Phi(t_1, X_{t_1}))^2\}\\ &\quad = \int_{t_1}^{t_2} \mathbb{E}\big\{\big(\beta \partial_{xx} \Phi(t, X_t)\big)^2\big\}\, \mathrm{d}t\,, \end{aligned}$$

and therefore the claim follows from Eq. 30. ∎

Lemma III.4. *For any $0 \le t \le q_*$, we have*

$$\mathbb{E}\{\partial_{xx} \Phi(t, X_t)\} = \int_t^1 \mu(s)\, \mathrm{d}s\,.$$

Proof: Consider $t \in [0, q_*]$ a continuity point of μ. Then the proof of Lemma 16 in [23] yields

$$\begin{aligned} \partial_{xx} \Phi(t, X_t) =& 1 - \mu(t) \big(\partial_x \Phi(t, X_t)\big)^2\\ &- \mathbb{E}\left\{ \int_t^1 \big(\partial_x \Phi(s, X_s)\big)^2 \mu(\mathrm{d}s) \right\}, \end{aligned}$$

Taking expectation and using Fubini's alongside Eq. (30), we get

$$\begin{aligned} \mathbb{E}\{\partial_{xx} \Phi(t, X_t)\} &= 1 - \mu(t) t - \int_t^1 s\, \mu(\mathrm{d}s)\\ &= \int_t^1 \mu(s)\, \mathrm{d}s \end{aligned}$$

The claim follows also for t not a continuity point because the right hand side is obviously continuous in t. The left hand side is continuous because $\partial_{xx} \Phi$ is Lipschitz (cf. Lemma III.1) and $\mathbb{E}\{|X_t - X_s|^2\} \le C|t-s|$ because the coefficients of the SDE are bounded Lipschitz. ∎

We summarize the results of this section in the following theorem. Here and below, for $\boldsymbol{x} \in \mathbb{R}^n$, $S \subseteq \mathbb{R}^n$, we let $d(\boldsymbol{x}, S) \equiv \inf\{\|\boldsymbol{x} - \boldsymbol{y}\| : \boldsymbol{y} \in S\}$.

Theorem 4. *Under Assumption 1 let $g, s, v : \mathbb{R} \times \mathbb{R}_{\ge 0} \to \mathbb{R}$ be defined as per Eq. (27), and set $\overline{q} = q_*(\beta)$ for $\beta > \beta_0$. Further let*

$$\mathcal{E}(\beta) \equiv \frac{\beta}{2}[1 - (1 - q_*(\beta))^2] - \frac{\beta}{2} \int_0^1 s^2\, \mu_\beta(\mathrm{d}s)\,.$$

Define the incremental AMP iteration $(\boldsymbol{u}^k)_{k \ge 0}$ via Eqs. (6), (9), (10), with $\widehat{g}_k$ given by Eq. (20), and

let $\boldsymbol{z}$ be given by Eq. (13). Then, for any $\varepsilon > 0$ there exist $\delta_*(\varepsilon) > 0$, and for any $\delta \geq \delta_*(\varepsilon)$ there exist $M_*(\varepsilon,\delta) < \infty$ such that, if $\delta \leq \delta_*(\varepsilon)$ and $M \geq M_*(\varepsilon,\delta)$, we have

$$\left|\operatorname*{p-lim}_{n\to\infty} \frac{1}{2n}\langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{z}\rangle - \mathcal{E}(\beta)\right| \leq \varepsilon\,,$$
$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n} d(\boldsymbol{z}, [-1,1]^n)^2 \leq \varepsilon\,.$$

(Further the above limits in probability are non-random quantities.)

Proof: First notice that $d(\boldsymbol{z},[-1,1]^n)^2 = \sum_{i=1}^n \psi(z_i)$ with $\psi(z_i) = d(z_i,[-1,1])^2$ a pseudo-Lipschitz function. Further, integration by parts yields

$$\mathcal{E}(\beta) = \beta \int_0^{q_*} \int_t^1 \mu(s)\,\mathrm{d}s\,\mathrm{d}t\,.$$

Hence the claims of this theorem follow immediately from Theorem 3 upon checking those assumptions using the lemmas given in this section. ∎

B. Sequential rounding and putting everything together

Theorem 4 constructs a vector $\boldsymbol{z} \in \mathbb{R}^n$. It is not difficult to round this to a vector with entries in $\{+1,-1\}$, as detailed in the next lemma.

Lemma III.5. *There exist an algorithm with complexity $O(n^2)$, and an absolute constant $C > 0$ such that the following happens with probability at least $1 - e^{-n}$. Given $\boldsymbol{A} \sim \mathsf{GOE}(n)$ and a vector $\boldsymbol{x} \in \mathbb{R}^n$ such that $d(\boldsymbol{x},[-1,1]^n)^2 \leq n\,\varepsilon_0$, the algorithm returns a vector $\boldsymbol{\sigma}_* \in \{+1,-1\}^n$ such that*

$$\frac{1}{2n}\langle \boldsymbol{\sigma}_*, \boldsymbol{A}\boldsymbol{\sigma}_*\rangle \geq \frac{1}{2n}\langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle - 20\Big(\sqrt{\varepsilon_0} + \frac{1}{\sqrt{n}}\Big)\,.$$

Proof: Recall the definition of Hamiltonian $H_n(\boldsymbol{x}) \equiv \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle/2$ (which we view as a function on $\mathbb{R}^n$). We also define $\tilde{H}_n(\boldsymbol{x}) = H_n(\boldsymbol{x}) - \sum_{i=1}^n A_{ii}x_i^2/2 = \sum_{i<j\leq n} A_{ij}x_ix_j$.

We construct $\boldsymbol{\sigma}_*$ in two steps. First we let $\tilde{\boldsymbol{z}}$ to be the projection of $\boldsymbol{z}$ onto the hypercube $[-1,+1]^n$ (i.e. $\tilde{\boldsymbol{z}} \in [-1,+1]^n$ is such that $\|\tilde{\boldsymbol{z}} - \boldsymbol{z}\|^2 = d(\tilde{\boldsymbol{z}},[-1,+1]^n)^2 \leq n\,\varepsilon_0$). Note that this can be constructed in $O(n)$ time (simply by projecting each coordinate $\tilde{z}_i$ onto $[-1,+1]$).

Second, note that the function $\tilde{H}_n(\boldsymbol{x})$ is linear in each coordinate of $\boldsymbol{x}$. Namely, for each ℓ $\tilde{H}_n(\boldsymbol{x}) = x_\ell h_{1,\ell}(\boldsymbol{x}_{\sim\ell};\boldsymbol{A}) + h_{0,\ell}(\boldsymbol{x}_{\sim\ell};\boldsymbol{A})$, where $\boldsymbol{x}_{\sim\ell} = (x_i)_{i\in[n]\setminus\ell}$ and $h_{1,\ell}(\boldsymbol{x}_{\sim\ell};\boldsymbol{A}) = \sum_{j\neq\ell} A_{\ell j}x_j$. We then construct a sequence $\tilde{\boldsymbol{z}}(0),\dots\tilde{\boldsymbol{z}}(n)$ as follows. Set $\tilde{\boldsymbol{z}}(0) = \tilde{\boldsymbol{z}}$ and, for each $1 \leq \ell \leq n$:

$$\tilde{\boldsymbol{z}}(\ell)_i = \begin{cases} \tilde{\boldsymbol{x}}(\ell-1)_i & \text{if } i \neq \ell,\\ \operatorname{sign}\big(h_{1,\ell}(\tilde{\boldsymbol{z}}(\ell-1)_{\sim\ell};\boldsymbol{A})\big) & \text{if } i = \ell. \end{cases}$$

Finally we set $\boldsymbol{\sigma}_* = \tilde{\boldsymbol{z}}(n)$. This procedure takes $O(n^2)$ operations.

The lemma then follows straightforwardly from the following three claims:

(*i*) $\tilde{H}_n(\boldsymbol{\sigma}_*) \geq \tilde{H}_n(\tilde{\boldsymbol{z}})$.
(*ii*) $|\tilde{H}_n(\boldsymbol{\sigma}_*) - H_n(\boldsymbol{\sigma}_*)| \leq 20\sqrt{n}$, $|\tilde{H}_n(\boldsymbol{\sigma}_*) - H_n(\boldsymbol{\sigma}_*)| \leq 20\sqrt{n}$ with probability at least $1 - e^{-2n}$.
(*iii*) $|H_n(\boldsymbol{z}) - H_n(\tilde{\boldsymbol{z}})| \leq 20n\sqrt{\varepsilon_0}$ with probability at least $1 - e^{-2n}$.

Claim (*i*) is immediate since $\tilde{H}_n(\tilde{\boldsymbol{z}}(\ell+1)) \geq \tilde{H}_n(\tilde{\boldsymbol{z}}(\ell+1))$ for each ℓ.

Claim (*ii*) holds since, for any $\boldsymbol{x} \in [-1,+1]^n$,

$$|\tilde{H}_n(\boldsymbol{x}) - H_n(\boldsymbol{x})| \leq \frac{1}{2}\sum_{i=1}^n |A_{ii}| \equiv \tau(\boldsymbol{A})\,.$$

Now we have $\mathbb{E}\tau(\boldsymbol{A}) = \sqrt{n/\pi}$, and τ is a Lipschitz function of the Gaussian vector $(A_{ii})_{i\leq n}$. hence the desired bounds follow by Gaussian concentration.

For claim (*iii*), let $\boldsymbol{v} = \boldsymbol{z} - \tilde{\boldsymbol{z}}$ and note that (denoting by $\lambda_{\max}(\boldsymbol{A})$ the maximum eigenvalue of $\boldsymbol{A}$)

$$\begin{aligned}\big|H_n(\boldsymbol{z}) - H_n(\tilde{\boldsymbol{z}})\big| &\leq \frac{1}{2}|\langle \boldsymbol{v}, \boldsymbol{A}\boldsymbol{v}\rangle| + |\langle \boldsymbol{v}, \boldsymbol{A}\tilde{\boldsymbol{z}}\rangle|\\ &\leq \frac{1}{2}\lambda_{\max}(\boldsymbol{A})\|\boldsymbol{v}\|^2 + \lambda_{\max}(\boldsymbol{A})\|\boldsymbol{v}\|\,\|\tilde{\boldsymbol{z}}\|\\ &\leq n\lambda_{\max}(\boldsymbol{A})\Big[\frac{1}{2}\varepsilon_0 + \sqrt{\varepsilon_0}\Big] \leq 2n\lambda_{\max}(\boldsymbol{A})\sqrt{\varepsilon_0}\,.\end{aligned}$$

The desired probability bound follows by concentration of the largest eigenvalue of GOE matrices [2]. ∎

We finally need to show that the quantity $\mathcal{E}(\beta)$ of Theorem 4 converges to the asymptotic optimum value, for large β. This is achieved in the two lemmas below.

Lemma III.6. *Let $\mathcal{E}_0(\beta) \equiv (\beta/2)(1 - \int_0^1 t^2\,\mu_\beta(\mathrm{d}t))$. Then, almost surely,*

$$\mathcal{E}_0(\beta) \leq \lim_{n\to\infty} \frac{1}{2n} \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle \leq \mathcal{E}_0(\beta) + \frac{\log 2}{\beta}\,.$$

Proof: By Gaussian concentration, it is sufficient to consider the expectation $E_n = \mathbb{E}\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} h_n(\boldsymbol{\sigma})/n$ (recall that $H_n(\boldsymbol{\sigma}) = \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle/2$. Recall the definition of partition function $Z_n(\beta) = \sum_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \exp(\beta H_n(\boldsymbol{\sigma}))$, and define the associated Gibbs measure $\nu_\beta(\boldsymbol{\sigma}) = \exp(\beta H_n(\boldsymbol{\sigma}))/Z_n(\beta)$ and free energy

density $F_n(T) \equiv (T/n)\mathbb{E}\log Z_n(\beta = 1/T)$. A standard thermodynamic identity [27] yields $F_n(T) = \mathbb{E}\nu_{1/T}(H_n(\boldsymbol{\sigma})) + TS(\nu_{1/T})$, where $S(q)$ is the Shannon entropy of the probability distribution q. Further $F_n'(T) = S(\nu_{1/T}) \geq$ and $F_n(T) \to E_n$ as $T \to 0$. Hence

$$\mathbb{E}\nu_\beta(H_n(\boldsymbol{\sigma})) \leq E_n \leq F_n(1/\beta) \leq \mathbb{E}\nu_\beta(H_n(\boldsymbol{\sigma})) + \frac{\log 2}{\beta}\,.$$

On the other hand, $\partial_\beta(\beta F_n(\beta)) = \mathbb{E}\nu_\beta(H_n(\boldsymbol{\sigma}))$. Since $\beta F_n(\beta) \to \mathsf{P}_\beta(\mu_\beta)$ by Theorem 1, $F_n(\beta), \mathsf{P}_\beta(\mu_\beta)$ are convex with $\mathsf{P}_\beta(\mu_\beta)$ differentiable [43], it follows that

$$\lim_{n\to\infty} \mathbb{E}\nu_\beta(H_n(\boldsymbol{\sigma})) = \frac{\mathrm{d}}{\mathrm{d}\beta}\mathsf{P}_\beta(\mu_\beta) = \mathcal{E}_0(\beta)\,.$$

(The last equality is proved in [43], with a difference in normalization of β.) ∎

Lemma III.7. *For any $\beta > \beta_0$,*

$$\lim_{\beta\to\infty} \beta^2(1 - q_*(\beta))^2 \leq 1\,.$$

Proof: The PDE (2) can be solved for $t \in (q_*, 1]$ using the Cole-Hopf transformation $\Phi = \log u$. This yields $\Phi(q_*, x) = ((1 - q_*)/2) + \log 2\cosh x$, whence $\partial_x\Phi(q_*, x) = \tanh(x)$ and $\partial_{xx}\Phi(q_*, x) = 1 - \tanh(x)^2$. Substituting in Eqs. (30), (30), we get

$$\mathbb{E}\big\{\tanh(X_{q_*})^2\big\} = q_*\,,$$
$$\beta^2\mathbb{E}\big\{\big(1 - \tanh(X_{q_*})^2\big)^2\big\} = 1\,.$$

Hence

$$\beta^2(1 - q_*)^2 = \beta^2\mathbb{E}\big\{1 - \tanh(X_{q_*})^2\big\}^2$$
$$\leq \beta^2\mathbb{E}\big\{\big(1 - \tanh(X_{q_*})^2\big)^2\big\} = 1\,.$$

∎

The proof our main result, Theorem 2, follows quite easily from the findings of this section.

Proof of Theorem 2: Let $E_* \equiv \lim_{n\to\infty}\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} H_n(\boldsymbol{\sigma})/n$. This limit exists by Corollary I.1, and we further have $E_* \geq 1/2$ (this can be proved by the same thermodynamic argument as in the proof of Lemma III.6, noting that $(1/n)\log_n Z_n(\beta) \to \log 2 + (\beta^2/4)$ for $\beta \leq 1$ [37]). It is therefore sufficient to output $\boldsymbol{\sigma}_*$ such that, with high probability, $H_n(\boldsymbol{\sigma}_*)/n \geq E_* - (\varepsilon/3)$.

Let $\beta = 10/\varepsilon$. By Lemma III.6 and Lemma III.7, we have $\mathcal{E}(\beta) \geq E_* - (\varepsilon/5)$. Applying the algorithm of Theorem 4 thus we obtain, with high probability, a vector $\boldsymbol{x} \in \mathbb{R}^n$ such that $H_n(\boldsymbol{z}) \geq E_* - \varepsilon/4$ and $d(\boldsymbol{x}, [-1, 1]^n)^2 \leq \varepsilon^2/10^6$. The proof is completed by using the rounding procedure of Lemma III.5. ∎

IV. Relation with the TAP equations

In this section we prove that the algorithm described in Section II, when used in conjunction with the specific choice of functions g_k, s, v in Section III actually constructs an approximate solution of the TAP equations (under Assumption 1). As in the previous section, we set $\overline{q} = q_*$, $v(x,t) = \beta^2\mu(t)\partial_x\Phi(t,x)$, $s(x,t) = \beta$, $g(x,t) = \beta\partial_{xx}\Phi(t,x)$, and

$$\widehat{g}_k(x) \equiv \frac{g(x, k\delta)}{\mathbb{E}\{g(X^\delta_{k-1}, k\delta)^2\}^{1/2}}\,. \tag{32}$$

Using these settings, we recall that $\boldsymbol{x}^k$ and $\boldsymbol{z}$ are given by

$$\boldsymbol{x}^k = \boldsymbol{x}^{k-1} + v(\boldsymbol{x}^{k-1}, k\delta)\,\delta + \beta\sqrt{\delta}[\boldsymbol{u}_k]_M\sqrt{\delta}\,, \tag{33}$$
$$\boldsymbol{z} = \sqrt{\delta}\sum_{k=1}^{\lfloor\overline{q}/\delta\rfloor} g_k(\boldsymbol{u}_0, \dots, \boldsymbol{u}_{k-1}) \odot \boldsymbol{u}^k\,. \tag{34}$$

Finally, we will repeatedly use the fact that the PDE (2) can be solved on $(q_*, 1]$ using the Cole-Hopf transformation, which yields $\Phi(q_*, x) = \log 2\cosh(x) + (1 - q_*)/2$.

Lemma IV.1. *Setting $k_* = \lfloor q_*/\delta\rfloor$, we have*

$$\lim_{\delta\to 0}\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{z} - \tanh(\boldsymbol{x}^{k_*})\right\|^2 = 0\,.$$

Proof: By Lemma II.2, we have

$$\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{z} - \tanh(\boldsymbol{x}^{k_*})\right\|^2 = \mathbb{E}\left\{\left[Z^\delta - \partial_x\Phi(q_*, X^\delta_{k_*})\right]^2\right\}\,.$$

On the other hand, using Lemma II.4, we obtain

$$\lim_{\delta\to 0}\mathbb{E}\left\{\left[Z^\delta - \partial_x\Phi(q_*, X^\delta_{k_*})\right]^2\right\}$$
$$= \mathbb{E}\left\{\left[Z_{q_*} - \partial_x\Phi(q_*, X_{q_*})\right]^2\right\}$$
$$= \mathbb{E}\left\{\left[\int_0^{q_*}\beta\partial_{xx}\Phi(t, X_t)\,\mathrm{d}B_t - \partial_x\Phi(q_*, X_{q_*})\right]^2\right\}$$
$$= 0\,.$$

where the last identity follows from Eq. (29). ∎

Lemma IV.2. *Setting $k_* = \lfloor q_*/\delta\rfloor$, let*

$$\boldsymbol{E}_{n,M,\delta} \equiv \beta\boldsymbol{A}\boldsymbol{z} - \boldsymbol{x}^{k_*} - \beta^2(1 - q_*)\tanh(\boldsymbol{x}^{k_*})\,.$$

Then,we have

$$\lim_{\delta\to 0}\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{E}_{n,M,\delta}\right\|^2 = 0\,.$$

Proof: Throughout the proof, we will write $\boldsymbol{f}_k \equiv f_k(\boldsymbol{u}_0, \dots, \boldsymbol{u}_k)$. By the basic iteration (6), we have

$$\begin{aligned}\boldsymbol{A}\boldsymbol{z} &= \sqrt{\delta}\sum_{k=1}^{k_*} \boldsymbol{A}\boldsymbol{f}_k \\ &= \sqrt{\delta}\sum_{k=1}^{k_*} \boldsymbol{u}^{k+1} + \sqrt{\delta}\sum_{k=1}^{k_*}\sum_{\ell=1}^{k} \mathsf{b}_{k\ell}\boldsymbol{f}_{\ell-1}\,.\end{aligned}$$

Using Eqs. (16) and (17), together with the fact that $\|\boldsymbol{f}_k\|^2/n$, $\|\boldsymbol{u}^k\|^2/n$ are bounded by Lemma II.2, we get

$$\begin{aligned}&\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{E}_{n,M,\delta}\right\|^2 = \\ &\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\Big\|\beta\sqrt{\delta}\sum_{k=1}^{k_*}\boldsymbol{u}^{k+1} - \beta^2(1-q_*)\tanh(\boldsymbol{x}^{k_*}) \\ &\quad +\beta\sqrt{\delta}\sum_{k=1}^{k_*}\mathbb{E}\{\widehat{g}_k(X^\delta_{k-1})\}\boldsymbol{f}_{k-1} - \boldsymbol{x}^{k_*}\Big\|^2 \\ &= \mathbb{E}\left\{\left[\beta\sqrt{\delta}\sum_{k=1}^{k_*}U^\delta_{k+1} - \beta^2(1-q_*)\tanh(X^\delta_{k_*})\right.\right. \\ &\left.\left.+\beta\sqrt{\delta}\sum_{k=1}^{k_*}\mathbb{E}\{\widehat{g}_k(X^\delta_{k-1})\}\widehat{g}_{k-1}(X^\delta_{k-2})U^\delta_{k-1} - X^\delta_{k_*}\right]^2\right\}. \qquad (35)\end{aligned}$$

Next, using again Lemma II.4, we have $\sqrt{\delta}\sum_{k=1}^{k_*}U^\delta_{k+1} \xrightarrow{L_2} B_{q_*}$, $X^\delta_{k_*}\xrightarrow{L_2} X_{q_*}$ and

$$\begin{aligned}&\sqrt{\delta}\sum_{k=1}^{k_*}\mathbb{E}\{\widehat{g}_k(X^\delta_{k-1})\}\widehat{g}_{k-1}(X^\delta_{k-2})U^\delta_{k-1} \xrightarrow{L_2} \\ &\int_0^{q_*}\mathbb{E}\{g(X_t,t)\}\,g(X_t,t)\mathrm{d}B_t \\ &= \beta^2\int_0^{q_*}\mathbb{E}\{\partial_{xx}(t,X_t)\}\,\partial_{xx}\Phi(t,X_t)\mathrm{d}B_t \\ &= \beta^2\int_0^{q_*}\int_t^1\mu(s)\,\mathrm{d}s\,\partial_{xx}\Phi(t,X_t)\mathrm{d}B_t\,,\end{aligned}$$

where in the last step we used Lemma III.4. By Fubini's theorem

$$\begin{aligned}&\beta^2\int_0^{q_*}\int_t^1\mu(s)\,\mathrm{d}s\,\partial_{xx}\Phi(t,X_t)\mathrm{d}B_t \\ &= \beta^2\int_0^{q_*}\mu(s)\int_0^s\partial_{xx}\Phi(t,X_t)\mathrm{d}B_t\,\mathrm{d}s \\ &\quad+\beta^2\int_{q_*}^1\mu(s)\int_0^{q_*}\partial_{xx}\Phi(t,X_t)\mathrm{d}B_t\,\mathrm{d}s \\ &= \beta\int_0^{q_*}\mu(s)\,\partial_x\Phi(X_s,s)\,\mathrm{d}s + \beta(1-q_*)\partial_x\Phi(X_{q_*},q_*)\,,\end{aligned}$$

where in the last step we used once more Eq. (29). Substituting these limits in Eq. (35), we get

$$\begin{aligned}&\lim_{\delta\to 0}\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{E}_{n,M,\delta}\right\|^2 = \\ &= \mathbb{E}\left\{\left[\beta B_{q_*} + \beta^2\int_0^{q_*}\mu(s)\,\partial_x\Phi(X_s,s)\,\mathrm{d}s\right.\right. \\ &\qquad+\beta^2(1-q_*)\partial_x\Phi(X_{q_*},q_*) - X_{q_*} \\ &\qquad\left.\left.-\beta^2(1-q_*)\tanh(X_{q_*})\right]^2\right\} \\ &= \mathbb{E}\left\{\left[\beta^2(1-q_*)\partial_x\Phi(X_{q_*},q_*) - X_{q_*}\right.\right. \\ &\qquad\left.\left.-\beta^2(1-q_*)\tanh(X_{q_*})\right]^2\right\} = 0\,.\end{aligned}$$

Where we used the fact that X_t solves te SDE (26), and $\Phi(q_*,x) = \log 2\cosh(x) + (1-q_*)/2$. ∎

We can therefore state our result about constructing solutions to the TAP equations.

Theorem 5 (Constructing solutions to the TAP equations)**.** *Under Assumption 1 let $g, s, v : \mathbb{R}\times\mathbb{R}_{\ge 0}\to\mathbb{R}$ be defined as per Eq. (27), and set $\overline{q} = q_*(\beta)$ for $\beta > \beta_0$. Define the incremental AMP iteration $(\boldsymbol{u}^k)_{k\ge 0}$ via Eqs. (6), (9), (10), with $\widehat{g}_k$ given by Eq. (20), and let $\boldsymbol{z}$ be given by Eq. (13). (The same iteration is given explicitly in Eqs. (33), (34).)*

Set $k_* = \lfloor q_*/\delta\rfloor$. Then, for any $\varepsilon > 0$ there exist $\delta_*(\varepsilon) > 0$, and for any $\delta \le \delta_*(\varepsilon)$ there exist $M_*(\varepsilon,\delta) < \infty$ such that, if $\delta \le \delta_*(\varepsilon)$ and $M \ge M_*(\varepsilon,\delta)$, we have, with high probability

$$\frac{1}{n}\left\|\beta\boldsymbol{A}\tanh(\boldsymbol{x}^{k_*}) - \boldsymbol{x}^{k_*} - \beta^2(1-q_*)\tanh(\boldsymbol{x}^{k_*})\right\| \le \varepsilon\,.$$

Proof: The theorem follows immediately from Lemma IV.1 and Lemma IV.2, using the fact that, with high probability, $\boldsymbol{A}$ has operator norm bounded by $2+\varepsilon$ [2]. ∎

V. Universality

In this section we use the universality results of [7] to generalize Theorem 2 to other random matrix distributions. Namely, we will work under the following assumption:

Assumption 2. *The matrix $\boldsymbol{A} = \boldsymbol{A}(n)$ is symmetric with $A_{ii} = 0$ and $(A_{ij})_{1\le i<j\le n}$ a collection of independent random variables, satisfying $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/n$. Further, the entries are subgaussian, with common subgaussian parameter C_*/n. (Namely, $\mathbb{E}\{\exp(\lambda A_{ij})\} \le \exp(C_*\lambda^2/2n)$ for all $i < j \le n$.)*

Using [7, Theorem 4], and proceeding exactly as for Proposition II.1, we obtain the following.

Proposition V.1. *Consider the AMP iteration (6), with $\boldsymbol{A} = \boldsymbol{A}(n)$ satisfying Assumption 2. Further, assume $f_k : \mathbb{R}^{k+2} \to \mathbb{R}$ to be a fixed polynomial (independent of n). Then for any $k \in \mathbb{N}$, and any pseudo-Lipschitz function $\psi : \mathbb{R}^{k+2} \to \mathbb{R}$, we have*

$$\frac{1}{n}\sum_{i=1}^{n} \psi(u_i^0, \dots, u_i^k; y_i) \xrightarrow{p} \mathbb{E}\psi(U_0, \dots, U_k; Y)\,. \tag{36}$$

Here $(U_j)_{j\ge 1}$ is a centered Gaussian process independent of (U_0, Y) with covariance $\widehat{\boldsymbol{Q}} = (\widehat{Q}_{kj})_{k,j\ge 1}$ determined recursively via

$$\widehat{Q}_{k+1,j+1} = \mathbb{E}\big\{f_k(U_0, \dots, U_k; Y) f_j(U_0, \dots, U_j; Y)\big\}\,. \tag{37}$$

Notice an important difference with respect to Proposition (6): instead of Lipschitz functions, we require the functions f_k to be polynomials. However, this result is strong enough to allow us prove the following generalization of Theorem 2.

Theorem 6. *Let $\boldsymbol{A} = \boldsymbol{A}(n)$, $n \ge 1$ be random matrices satisfying Assumption 2. Under Assumption 1, for any $\varepsilon > 0$ there exists an algorithm that takes as input the matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, and outputs $\boldsymbol{\sigma}_* = \boldsymbol{\sigma}_*(\boldsymbol{A}) \in \{+1, -1\}^n$, such that the following hold: (i) The complexity (floating point operations) of the algorithm is at most $C(\varepsilon)n^2$. (ii) We have $\langle \boldsymbol{\sigma}_*, \boldsymbol{A}\boldsymbol{\sigma}_*\rangle \ge (1-\varepsilon)\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n}\langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle$.*

Proof: Let $\widehat{g}_k(x)$, $v(x,t)$, $s(x,t)$ be defined as in the proof of Theorem 2 for $k \le 1/\delta$. For each $M \in \mathbb{Z}$, and each $k \le 1/\delta$, we construct a polynomial $\hat{p}_{k,M} : \mathbb{R}^{k-1} \to \mathbb{R}$ which approximates the dynamics defined by $\widehat{g}_k(\,\cdot\,)$, $v(\,\cdot\,, k\delta)$, $s(\,\cdot\,, k\delta)$, in a sense that we will make precise below.

We define the IAMP iteration, analogously to (9), (10)

$$f_k(u_0, \dots, u_k) = \hat{p}_{k,M}\left(u_1, \dots, u_{k-1}\right) \cdot u_k\,, \tag{38}$$

We then claim that we can construct these polynomial approximations $\hat{p}_{k,M}$ so that, for any $k \le 1/\delta$, and any pseudo-Lipschitz function $\psi : \mathbb{R}^{k+2} \to \mathbb{R}$, we have

$$\lim_{M\to\infty} \operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(u_i^0, \dots, u_i^k) = \mathbb{E}\psi(U_0^\delta, \dots, U_k^\delta)\,, \tag{39}$$

where the independent random variables $(U_\ell^\delta)_{\ell\ge 0}$ are defined as in Lemma II.2. Given this claim, the rest of the proof of Theorem 2 can be applied verbatimly to this –slightly different– algorithm.

In order to prove the claim (39), we proceed as in the proof of Lemma II.2. Namely, by applying Proposition V.1, we get

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(u_i^0, \dots, u_i^k) = \mathbb{E}\psi(U_0, \dots, U_k^{\delta,M})\,,$$

where $(U_\ell^{\delta,M})_{\ell\ge 0}$ is a centered Gaussian process. Using the same argument as in Lemma II.2, we obtain that the Gaussian random variables $(U_\ell^{\delta,M})_{\ell\ge 0}$ are independent. Further, letting $\widehat{q}_\ell^{\,M} \equiv \mathbb{E}\{(U_\ell^{\delta,M})^2\}$, Proposition V.1 yields the following recursion

$$\widehat{q}_{k+1}^{\,M} = \mathbb{E}\{\hat{p}_{k,M}(U_1^{\delta,M}, \dots, U_{k-1}^{\delta,M})^2\} \cdot \widehat{q}_k^{\,M}\,. \tag{40}$$

The claim (39) follows by showing the we can choose polynomials $(\hat{p}_{\ell,M})_{\ell\ge 0}$ so that $\lim_{M\to\infty}\widehat{q}_\ell^{\,M} = \widehat{q}_\ell$ for each $\ell \le 1/\delta$. This can be done by induction over k. As a preliminary, notice that there is $c_0 = c_0(\delta) > 0$ sufficiently small so that, for the sequence of random variables defined recursively via Eq. (11), we have $2c_0 \le \widehat{q}_k \le 1/(2c_0)$ for all $k \le 1/\delta$ (the existence of such $c_0 > 0$ can also be shown by induction over k using the fact that $\widehat{g}_k, v, s$, are bounded Lipschitz).

The basis of the induction $\lim_{M\to\infty}\widehat{q}_0^{\,M} = \widehat{q}_0$ is trivial. Then assume that the induction claim is true for all $\ell \le k$. Without loss of generality we can consider that, for any $M \ge 1$ we have $c_0 \le \widehat{q}_1^{\,M}, \dots, \widehat{q}_k^{\,M} \le 1/c_0$. Indeed by the induction hypothesis this holds for all M large enough, and we can always renumber the polynomials $\hat{p}_{\ell,M}(\cdots)$ so that it holds for all $M \ge 1$. Then notice that the random variable X_k^δ of Eq. (11) can be written as $X_k^\delta = h_k(U_0, U_1^\delta, \dots, U_{k-1}^\delta)$ for a certain function h_k that is bounded by a polynomial. We then choose the polynomials $\hat{p}_{k,M}(\,\cdot\,)$ so that

$$\mathbb{E}\Big\{\big|h_k(U_0, U_1^{\delta,M}, \dots, U_{k-1}^{\delta,M}) - \hat{p}_{k,M}(U_0, U_1^{\delta,M}, \dots, U_{k-1}^{\delta,M})\big|^2\Big\} \le \frac{1}{M}\,.$$

Such polynomials can be constructed, for instance, by considering the expansion of h_k in the basis of multivariate Hermite polynomials (suitably rescaled as to form an orthonormal basis with in $L^2(\mathbb{R}^{k-1}, \mu_k)$, where μ_k is the joint distribution of $U_0, U_1^{\delta,M}, \dots, U_{k-1}^{\delta,M}$.) The variance bound $c_0 \le \widehat{q}_1^{\,M}, \dots, \widehat{q}_k^{\,M} \le 1/c_0$ is used in controlling the error term.

The induction claim then follows by

$$\begin{aligned}\lim_{M\to\infty} &\mathbb{E}\{\hat{p}_{k,M}(U_1^{\delta,M},\dots,U_{k-1}^{\delta,M})^2\}\\ &= \lim_{M\to\infty} \mathbb{E}\{h_k(U_1^{\delta,M},\dots,U_{k-1}^{\delta,M})^2\}\\ &= \mathbb{E}\{h_k(U_1^{\delta},\dots,U_{k-1}^{\delta})^2\}\,,\end{aligned}$$

where the last equality holds by dominated convergence. ∎

Corollary I.2 follows by applying Theorem 6 with $\boldsymbol{A}$ a suitably centered and normalized adjacency matrix.

Proof of Corollary I.2: Given a graph $G \sim \mathcal{G}(n,p)$, construct the matrix $\boldsymbol{A} = \boldsymbol{A}^{\mathsf{T}} \in \mathbb{R}^{n\times n}$, by setting $A_{ii} = 0$ and, for $i \neq j$:

$$A_{ij} = \begin{cases} -\sqrt{\frac{1-p}{np}} & \text{if } (i,j)\in E,\\ \sqrt{\frac{p}{n(1-p)}} & \text{if } (i,j)\notin E,\end{cases}$$

It is easy to verify that this matrix satisfies Assumption 2. Further, we have

$$\mathsf{CUT}_G(\boldsymbol{\sigma}) = \frac{1}{2}|E_n| - \frac{p}{4}\langle\boldsymbol{\sigma},\mathbf{1}\rangle^2 + \frac{\sqrt{np(1-p)}}{4}\langle\boldsymbol{\sigma},\boldsymbol{A}\boldsymbol{\sigma}\rangle\,.$$

Recall that we know from [17] $\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n}\mathsf{CUT}_G(\boldsymbol{\sigma}) = |E_n|/2 + (n^3p(1-p)/2)^{1/2}\mathsf{P}_* + o(n^{3/2})$. Let $\boldsymbol{\sigma}_1$ denote the output of the algorithm of Theorem 6, on input $\boldsymbol{A}$. Applying this theorem and Lemma V.1, we get

$$\begin{aligned}\text{p-lim}\inf_{n\to\infty}\frac{1}{2n}\langle\boldsymbol{\sigma}_1,\boldsymbol{A}\boldsymbol{\sigma}_1\rangle &\ge (1-\varepsilon)\mathsf{P}_*\,,\\ \text{p-lim}_{n\to\infty}\frac{1}{n}\langle\boldsymbol{\sigma}_1,\mathbf{1}\rangle &= 0\,.\end{aligned}$$

We construct $\boldsymbol{\sigma}_*$ by balancing $\boldsymbol{\sigma}_1$. Namely, if $|\langle\boldsymbol{\sigma}_1,\mathbf{1}\rangle| = \ell$, we obtain $\boldsymbol{\sigma}_*$ by flipping $\lfloor\ell/2\rfloor$ entries of $\boldsymbol{\sigma}_1$ so that $|\langle\boldsymbol{\sigma}_*,\mathbf{1}\rangle| \le 1$. We then have, with high probability

$$\begin{aligned}&\mathsf{CUT}_G(\boldsymbol{\sigma}_*) - \frac{1}{2}|E_n|\\ &\ge -\frac{p}{4} + \frac{1}{4}\sqrt{np(1-p)}\,\langle\boldsymbol{\sigma}_*,\boldsymbol{A}\boldsymbol{\sigma}_*\rangle\\ &\quad - \frac{1}{4}\sqrt{np(1-p)}\,|\langle\boldsymbol{\sigma}_*,\boldsymbol{A}\boldsymbol{\sigma}*\rangle - \langle\boldsymbol{\sigma}_1,\boldsymbol{A}\boldsymbol{\sigma}_1\rangle|\\ &\ge \frac{1}{2}|E_n| + \frac{1}{4}(1-\varepsilon)\sqrt{np(1-p)}\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n}\langle\boldsymbol{\sigma},\boldsymbol{A}\boldsymbol{\sigma}\rangle\\ &\quad - \sqrt{n}\|\boldsymbol{A}\|_{\text{op}}\|\boldsymbol{\sigma}_1\|\,\|\boldsymbol{\sigma}_* - \boldsymbol{\sigma}_1\|\,.\end{aligned}$$

(Here $\|\boldsymbol{A}\|_{\text{op}}$ denotes the operator norm of matrix $\boldsymbol{A}$.) Therefore, since $|\langle\boldsymbol{\sigma}_1,\mathbf{1}\rangle|/n = \ell/n \xrightarrow{p} 1$, and $\|\boldsymbol{A}\|_{\text{op}} \le 2.01$ with high probability [2], we get

$$\begin{aligned}&\mathsf{CUT}_G(\boldsymbol{\sigma}_*) - \frac{|E_n|}{2}\\ &\ge (1-\varepsilon)\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n}\left\{\mathsf{CUT}_G(\boldsymbol{\sigma}) - \frac{|E_n|}{2}\right\} - n\sqrt{\ell}\|\boldsymbol{A}\|_{\text{op}}\\ &\ge (1-\varepsilon)\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n}\left\{\mathsf{CUT}_G(\boldsymbol{\sigma}) - \frac{|E_n|}{2}\right\} - o(n^{3/2})\,,\end{aligned}$$

which completes the proof. ∎

Acknowledgements

I am grateful to Eliran Subag for an inspiring presentation of his work [42] delivered at the workshop 'Advances in Asymptotic Probability' in Stanford, and for a stimulating conversation. This work was partially supported by grants NSF DMS-1613091, CCF-1714305, IIS-1741162 and ONR N00014-18-1-2729.

Appendix

As mentioned in the main text, Proposition II.1 is a consequence of the general analysis of AMP algorithms available in the literature. In particular it can be obtained from a reduction to the setting of [24, Theorem 1]. Let us briefly recall the class of algorithms considered in [24], adapting the notations to the present ones. (we limit ourselves to consider the 'one-block' case in the language of [24]).

Fixing $T \ge 1$ consider a sequence of Lipschitz functions

$$\begin{aligned}F_t :&\mathbb{R}^T\times\mathbb{R}^2\to\mathbb{R}^T\,,\\ &(x_1,\dots,x_T,z_1,z_2)\mapsto F_t(x_0,x_1,\dots,x_T,z_1,z_2)\,.\end{aligned}$$

Given two matrices $\boldsymbol{x}\in\mathbb{R}^{n\times(T+1)}$, $\boldsymbol{z}\in\mathbb{R}^{n\times 2}$, we let $F_t(\boldsymbol{x};\boldsymbol{z})\in\mathbb{R}^{n\times(T+1)}$ be the matrix whose i-th row is given by $F_t(\boldsymbol{x}_i,\boldsymbol{z}_i)$ (where $\boldsymbol{x}_i$ is the i-th row of $\boldsymbol{x}$ and $\boldsymbol{z}_i$ is the i-th row of $\boldsymbol{z}$).

Then [24] analyzes the following AMP iteration, which produces a sequence of iterates $\boldsymbol{x}^t\in\mathbb{R}^{n\times(T+1)}$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}F_t(\boldsymbol{x}^t;\boldsymbol{z}) - F_{t-1}(\boldsymbol{x}^{t-1};\boldsymbol{z})\mathsf{B}_t^{\mathsf{T}}\,. \qquad (41)$$

Here $\mathsf{B}_t\in\mathbb{R}^{T\times T}$ is a matrix with entries defined by

$$(\mathsf{B}_t)_{ij} = \frac{1}{n}\sum_{\ell=1}^n (D_{\boldsymbol{x}}F_t(\boldsymbol{x}_\ell^t;\boldsymbol{z}_\ell))_{ij} = \frac{1}{n}\sum_{\ell=1}^n \frac{\partial F_t}{\partial x_{\ell,j}^t}(\boldsymbol{x}_\ell^t;\boldsymbol{z}_\ell)\,.$$

Under the assumption that $\boldsymbol{x}^0,\boldsymbol{z}$ are independent of $\boldsymbol{A}$, and $\hat{p}_{\boldsymbol{x}^0,\boldsymbol{z}} \equiv n^{-1}\sum_{i=1}^n \delta_{\boldsymbol{x}_i^0,\boldsymbol{z}_i}$ converges in W_2, [24, Theorem 1] determines the asymptotic empirical distribution of $\boldsymbol{x}^t,\boldsymbol{z}$.

Proposition II.1 can be recast as a special case of this setting. First notice that we can always choose an n-independent T such that the time horizon k in Eq. (7) satisfies $k \le T$. We then consider the iteration (41) with initialization $\boldsymbol{x}^0 = \boldsymbol{0}$, data vectors $\boldsymbol{z} = (\boldsymbol{u}^0, \boldsymbol{y})$, and update functions given by

$$F_t(x_1, x_2, \dots, x_T, z_1, z_2)_\ell = f_{\ell-1}(z_1, x_1, \dots, x_{\ell-1}; z_2)\,.$$

With this setting, the vector $(x^t_{i,\ell})_{i\le n} \in \mathbb{R}^n$ coincides with $\boldsymbol{u}^\ell$ as given in Eq. (6), for all $t \ge \ell$. The recursion of Eq. (8) follows from the analogous recursion in [24, Theorem 1].

In this appendix we provide a simplified version of the algorithm of Theorem 2, for the reader's convenience. In this presentation we simplify certain technical details that have been introduced in the main text to simplify the proof. In the pseudo-code below $\odot$ denotes entrywise multiplication between vectors. Further, when a scalar function is applied to a vector, it is understood to be applied componentwise. In particular, note that $\|\partial_{xx}\Phi(k\delta, \boldsymbol{x}^k)\|$ is the ℓ_2 norm of the vector whose i-th component is $\partial_{xx}\Phi(k\delta, x^k_i)$.

Algorithm 1: IAMP algorithm to optimize SK Hamiltonian

Data: Matrix $\boldsymbol{A} \sim \mathsf{GOE}(n)$, parameters δ, $\beta > 0$
Result: Near optimum $\boldsymbol{\sigma}_* \in \{+1-1\}^n$ of the SK Hamiltonian
Compute minimizer μ_β of the Parisi functional $\mathsf{P}_\beta(\mu)$ (cf. Eq. (3));
Compute solution Φ PDE (2), with $\mu = \mu_\beta$;
Compute $q_*(\beta) = \sup\{q :\ q \in \mathrm{supp}(\mu_\beta)\}$ (Edwards-Anderson parameter);
Initialize $\boldsymbol{u}^{-1} = \boldsymbol{0}$, $\boldsymbol{u}^0 \sim \mathsf{N}(0, \boldsymbol{I}_n)$, $\boldsymbol{g}^{-1} = \boldsymbol{1}$, $\boldsymbol{g}^{-2} = \boldsymbol{0}$, $\mathsf{b}_0 = 0$;
for $k \leftarrow 0$ **to** $\lfloor q_*/\delta \rfloor$ **do**
  $\boldsymbol{u}^{k+1} = \boldsymbol{A}(\boldsymbol{g}^{k-1} \odot \boldsymbol{u}^k) - \mathsf{b}_k \boldsymbol{g}^{k-2} \odot \boldsymbol{u}^{k-1}$;
  $\boldsymbol{x}^k = \boldsymbol{x}^{k-1} + \beta^2 \mu(k\delta)\, \partial_x \Phi(k\delta, \boldsymbol{x}^{k-1})\, \delta + \beta\sqrt{\delta}\boldsymbol{u}^k$;
  $\boldsymbol{g}^k = \sqrt{n}\partial_{xx}\Phi(k\delta, \boldsymbol{x}^k)/\|\partial_{xx}\Phi(k\delta, \boldsymbol{x}^k)\|$;
  $\mathsf{b}_{k+1} = \sum_{i=1}^n g^k_i/n$;
end
Compute $\boldsymbol{z} = \sqrt{\delta}\sum_{k=1}^{\lfloor q_*/\delta \rfloor} \boldsymbol{g}^{k-1} \odot \boldsymbol{u}^k$;
Round $\boldsymbol{z}$ to $\boldsymbol{\sigma}_* \in \{-1, +1\}^n$;
return $\boldsymbol{\sigma}_*$

Notice that this pseudo-code does not describe how to minimize the Parisi functional and to solve the PDE (2). As discussed in the introduction, we believe this can be done efficiently because of the strong convexity and continuity of $\mu \mapsto \mathsf{P}_\beta(\mu)$. Indeed highly accurate numerical solutions (albeit with no rigorous analysis) were developed already in [15], [35], [40].

Further, the pseudo-code does not specify the rounding procedure, which is given below.

Algorithm 2: Round

Data: Matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, vector $\boldsymbol{z} \in \mathbb{R}^n$
Result: Integer solution $\boldsymbol{\sigma}_* \in \{+1-1\}^n$
for $i \leftarrow 1$ **to** n **do**
  Set $\tilde{z}_i \leftarrow \min(\max(z_i, -1), +1)$;
end
for $i \leftarrow 1$ **to** n **do**
  Compute $h_i = \sum_{j\neq i} A_{ij}\tilde{z}_j$;
  Set $\tilde{z}_i \leftarrow \mathrm{sign}(h_i)$;
end
return $\boldsymbol{\sigma}_* = \tilde{\boldsymbol{z}}$.

References

[1] Louigi Addario-Berry and Pascal Maillard. The algorithmic hardness threshold for continuous random energy models. arXiv:1810.05129, 2018. 4

[2] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Cambridge University Press, 2009. 10, 12, 14

[3] Sanjeev Arora, Eli Berger, Hazan Elad, Guy Kindler, and Muli Safra. On non-approximability for quadratic programs. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 206–215. IEEE, 2005. 1, 3

[4] Antonio Auffinger and Wei-Kuo Chen. The Parisi formula has a unique minimizer. *Communications in Mathematical Physics*, 335(3):1429–1444, 2015. 2, 8, 9

[5] Antonio Auffinger and Wei-Kuo Chen. Parisi formula for the ground state energy in the mixed p-spin model. *The Annals of Probability*, 45(6b):4617–4631, 2017. 2

[6] Antonio Auffinger, Wei-Kuo Chen, and Qiang Zeng. The SK model is Full-step Replica Symmetry Breaking at zero temperature. arXiv:1703.06872, 2017. 2

[7] Mohsen Bayati, Marc Lelarge, and Andrea Montanari. Universality in polytope phase transitions and message passing algorithms. *The Annals of Applied Probability*, 25(2):753–822, 2015. 3, 4, 12, 13

[8] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. on Inform. Theory*, 57:764–785, 2011. 4

[9] Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. arXiv:1708.03950, 2017. 4, 5

[10] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014. 4

[11] Jean-Philippe Bouchaud, Leticia F Cugliandolo, Jorge Kurchan, and Marc Mézard. Out of equilibrium dynamics in spin-glasses and other glassy systems. *Spin glasses and random fields*, pages 161–223, 1998. 3

[12] Philippe Carmona and Yueyun Hu. Universality in sherrington–kirkpatrick's spin glass model. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, volume 42, pages 215–222. Elsevier, 2006. 2

[13] Wei-Kuo Chen. Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound. *The Annals of Probability*, 45(6A):3929–3966, 2017. 9

[14] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy. arXiv:1812.05066, 2018. 4

[15] Andrea Crisanti and Tommaso Rizzo. Analysis of the ∞-replica symmetry breaking solution of the Sherrington-Kirkpatrick model. *Physical Review E*, 65(4):046137, 2002. 2, 3, 15

[16] Leticia F. Cugliandolo and Jorge Kurchan. On the out-of-equilibrium relaxation of the sherrington-kirkpatrick model. *Journal of Physics A: Mathematical and General*, 27(17):5749, 1994. 3

[17] Amir Dembo, Andrea Montanari, and Subhabrata Sen. Extremal cuts of sparse random graphs. *The Annals of Probability*, 45(2):1190–1217, 2017. 2, 3, 14

[18] David L. Donoho, Arian Maleki, and Andrea Montanari. Message Passing Algorithms for Compressed Sensing. *Proceedings of the National Academy of Sciences*, 106:18914–18919, 2009. 4

[19] Andreas Engel and Christian Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, 2001. 1

[20] David Gamarnik. Computing the partition function of the sherrington-kirkpatrick model is hard on average. *arXiv:1810.05907*, 2018. 3

[21] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376. ACM, 2014. 4

[22] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003. 2

[23] Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the parisi functional. *Proceedings of the American Mathematical Society*, 144(7):3135–3150, 2016. 1, 2, 3, 8, 9

[24] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013. 4, 5, 14, 15

[25] Yoshiyuki Kabashima. A CDMA multiuser detection algorithm on the basis of belief propagation. *J. Phys. A*, 36:11111–11121, 2003. 4

[26] Dmitriy Kunisky and Afonso S Bandeira. A tight degree 4 sum-of-squares lower bound for the sherrington-kirkpatrick hamiltonian. *arXiv:1907.11686*, 2019. 3

[27] Marc Mézard and Andrea Montanari. *Information, Physics and Computation*. Oxford, 2009. 1, 11

[28] Marc Mézard, Giorgio Parisi, and Miguel A. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, 1987. 1, 2, 3, 4, 9

[29] Marc Mézard, Giorgio Parisi, and Riccardo Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002. 1

[30] Marc Mézard and Miguel Angel Virasoro. The microstructure of ultrametricity. *Journal de Physique*, 46(8):1293–1307, 1985. 4

[31] Andrea Montanari. Local algorithms for two problems on locally tree-like graphs. https://simons.berkeley.edu/talks/andrea-montanari-02-22-2016, February 22, 2016. Talk given at Simons Institute for the Theory of Computing. 3

[32] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 814–827. ACM, 2016. 3

[33] Hidetoshi Nishimori. *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Oxford University Press, 2001. 1

[34] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003. 7

[35] Reinhold Oppermann, Manuel J Schmidt, and David Sherrington. Double Criticality of the Sherrington-Kirkpatrick Model at $T = 0$. *Physical Review Letters*, 98(12):127201, 2007. 2, 15

[36] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Annals of Mathematics*, pages 383–393, 2013. 1

[37] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013. 1, 11

[38] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979. 1

[39] Manuel J Schmidt. Replica symmetry breaking at low temperatures. Ph.F. Thesis, 2008. 3

[40] Manuel J Schmidt and Reinhold Oppermann. Method for replica symmetry breaking at and near T= 0 with application to the Sherrington-Kirkpatrick model. *Physical Review E*, 77(6):061104, 2008. 2, 15

[41] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975. 1

[42] Eliran Subag. Following the ground-states of full-RSB spherical spin glasses. arXiv:1812.04588, 2018. 4, 14

[43] Michel Talagrand. Parisi measures. *Journal of Functional Analysis*, 231(2):269–286, 2006. 11

[44] Michel Talagrand. The Parisi formula. *Annals of Mathematics*, pages 221–263, 2006. 1

[45] Michel Talagrand. *Mean Field Models for Spin Glasses: Volume I*. Springer-Verlag, Berlin, 2010. 1, 3

[46] Fabio Lucio Toninelli. About the almeida-thouless transition line in the sherrington-kirkpatrick mean-field spin glass model. *EPL (Europhysics Letters)*, 60(5):764, 2002. 2

[47] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008. 4

[48] Peter G Wolynes and Vassiliy Lubchenko. *Structural glasses and supercooled liquids: Theory, experiment, and applications*. John Wiley & Sons, 2012. 1